Technische Universität München

Lehrstuhl für Bioverfahrenstechnik

A Crystallographic and Molecular Dynamics Simulation Study of Protein Crystallization

Johannes Hermann

Vollständiger Abdruck der von der Fakultät für Maschinenwesen der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:		Prof. Dr.-Ing. Harald Klein
Prüfer:	1.	Prof. Dr.-Ing. Dirk Weuster-Botz
	2.	Prof. Dr. Martin Zacharias

Die Dissertation wurde am 11.12.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Maschinenwesen am 19.03.2021 angenommen.

Acknowlegements

The presented doctoral thesis originated during my work at the Institute of Biochemical Engineering at the Technical University of Munich (TUM) under supervision of Prof. Dr.-Ing. Dirk Weuster-Botz. My gratitude goes to Prof. Weuster-Botz who trusted me in developing a new theoretical field at his institute. Though our diverse expertise, his unrestricted confidence and support was a key for a successful study.

Furthermore, I thank Prof. Dr. Martin Zacharias (Chair of Theoretical Biophysics (T38), TUM) for not only guiding my attention to the field of Theoretical Biophysics at the very end of my Physics Master's study but also for continuing support during my doctoral study. Thanks for inviting me to the Winter School of his institute and thanks to Sonja Ortner for great organization thereof.

I am also thankful to late Dr. Dariusch Hekmat, Phillip Grob, and Daniel Bischoff for fruitful interdisciplinary discussions on experimental and theoretical aspects of the project. I thank my students Christina Ditter, Ruben Rossbach, and Ramla Cherif for supporting me during my study. Further thanks goes to the staff of the department: thank you Gabriele Herbrik, Ellen Truxius, Marlene Schocher, Patrick Meins, Markus Amann, and Norbert Werth for organisatorial and technical help.

I am further thankful to Dr. Robert Janowski (Institute of Structural Biology, Helmholtz-Zentrum München) and Dr. Sabine Schneider (Chair of Biochemistry, TUM) for X-ray diffraction measurement of protein crystals and Dr. Tobias Schrader (Jülich Centre for Neutron Science) and Dr. Andreas Ostermann (Heinz Maier-Leibnitz Centre, TUM) for help in neutron diffraction measurements.

I am thankful for all my colleagues during my time at the Institute of Biochemical Engineering. Thanks for all the laughs, talks and critical professional discussions. A special thanks goes to Ingmar Polte and Christoph Mähler for answering all my question on laboratory and biological backgrounds and to Korbinian Liebl (Chair of Theoretical Biophysics (T38), TUM) for discussing any theoretical findings. I gratefully acknowledge Korbinian Liebl, Christoph Mähler, Sabine Wagner and Rolf Frey for proof-reading.

I owe my deep gratitude to my partner Lena for the support in the last years. Mein größter Dank geht an meine Familie und insbesondere meine Eltern. Auf Eure Unterstüzung kann ich immer zählen.

Table of Contents

Chapter 1

Introduction

Protein-protein interactions drive various functions in biology and human live. Interactions range from transient protein-protein interaction networks e. g. catalyzing metabolic enzymatic steps [Durek and Walther, 2008], and initiating RNA splicing, transcription and translation [Staley and Woolford Jr, 2009] to stable protein-protein complexes e. g. assembling in a quaternary structure [Janin et al., 2008]. A special form of stable interaction is protein crystallization where proteins bind to each other in a periodic, symmetric order to form a crystal. Protein crystallization in living cells is normally a pathological phenomenona and therefore disfavored by evolution [Doye et al., 2004]. Crystallization in living cells leads to disorders or diseases like eye cataract [Pande et al., 2001; Siezen et al., 1985], homozygous hemoglobin C disease causing a form of anaemia [Lessin et al., 1969], and so-called 'Charcot-Lyden crystals' leading to bronchial asthma [Charcot and Robin, 1853; Von Leyden, 1872]. However, protein crystallization may also be put into use: protein crystallization is an inevitable step in structure determination by diffraction methods. X-rays require highly ordered structures like crystals to provide information in their diffraction patterns. The first X-ray crystal structure was solved by scientists Max Perutz and John Kendrew in the 1950s [Kendrew et al., 1958; Perutz et al., 1960] for which they were awarded with the Nobel Prize in Chemistry in 1962. In addition, crystallization has been used for purification purpose long before its application in crystallography. Osborne [1892] extended the work of Ritthausen [1880; 1881] to prepare pure specimens

of globulins by crystallization. Nowadays, technical protein crystallization became into focus as an attractive and effective purification method compared to conventional, costly chromatographic methods [Hekmat, 2015; Smejkal et al., 2013]. Furthermore, protein crystallization is well suited as a delivery tool of biopharmaceuticals [Basu et al., 2004]. Protein-protein interactions rely on molecular docking mechanisms and require specific interactions depending on the associated function [De Las Rivas and Fontanillo, 2010]. Protein crystallization and the associated molecular mechanisms are of high importance as crystallization may be used for technical purposes but may also cause diseases.

Chapter 2

Motivation and Objective

Thermodynamically, protein crystallization is a phase transition from a solvated into a bound, crystalline state which occurs in two subsequent steps: nucleation and crystal growth. For both, different nucleation [Becker and Döring, 1935; Gibbs, 1878; Ten Wolde and Frenkel, 1997; Volmer and Weber, 1926] and crystal growth theories [Burton et al., 1951] have been developed. In experiment, protein crystallization requires highly elaborate and costly work: conventionally, crystallization of proteins is enabled or improved by manipulating the proteins environment in a large crystallization conditions screen [Luft et al., 2011]. A different approach is to manipulate the protein itself by protein engineering. One of the earlierst works of protein engineering for the purpose of crystallization can be found in the work of Lawson et al. [1991]. Protein engineering was sucessfully applied in further studies [Charron et al., 2002; Dale et al., 2003; D'Arcy et al., 1999]. These studies lead to a currently prominent protein engineering strategy: the surface entropy reduction (SER) strategy [Derewenda, 2004]. SER is empirically motivated and relies on semi-rational substitution of amino acids on a protein's surface.

Controlling protein crystallization conflicts with competing theories, semi-rational engineering strategies, and high experimental effort. The molecular mechanisms enabling protein crystallization on an atomic level are not visible during a crystallization experiment. In order to resolve these issues, the aim of this study was to unravel mechanistic interactions and develop an *in silico* model for the crystallization process.

Assumptions on molecular mechanisms of any kind have been made based on atomic models created by crystallography. For example, neutron and X-ray diffraction have been used to identify re-organization of hydrogen bonds, active site residues, and solvent upon ligand binding in enzymes [Koruza et al., 2019]. Successful co-enzyme engineering has been performed purely based on crystal X-ray structures of enzymes within the same family [Mähler et al., 2019]. Though crystallography may provide valuable static properties, it is blind to dynamic properties and thermodynamic quantities. The constructed atomistic model resembles a static snapshot found in a protein crystal.

Molecular dynamics (MD) simulations provide a possibility to investigate interactions dynamically and to calculate thermodynamic quantities. The structures determined with crystallographic methods serve as input parameters. MD simulations have been termed 'computational microscope' [Dror et al., 2012; Lee et al., 2009] as they provide information on an atomic level for processes not accessible by experiment. Processes ranging from DNA unwinding [Liebl and Zacharias, 2020] to gamma secretase substrate recognition [Hitzenberger and Zacharias, 2019] were successfully unraveled with MD simulations. For biological protein-protein interactions, even automated MD based programs for prediciting binding sites have been developed [Schindler et al., 2015]. However, up-to date no atomistic MD study exists for the process of protein crystallization. Protein crystallization happens on the time scale from seconds to days. The average net frequency of one successfull binding event in protein crystallization has been estimated to 0.1 s^{-1} [Yau and Vekilov, 2000]. Protein crystals can grow up to 1.0 mm^3 [Piccoli et al., 2007]. Usually, MD simulation set-ups and methods capture time scales of a few µs and length scales of a few hundred nm.

In this thesis, protein crystallization is studied for the example protein alcohol dehydrogenase from *Lactobacillus brevis* (*Lb*ADH). *Lb*ADH is an *R*-specific, nicotinamide adenine dinucleotide phosphate [NADP(H)]-dependent enzyme. It is known for its high stability at elevated temperatures, high stereoselectivity, tolerance to organic solvents and broad substrate range making it industrially relevant [Hummel, 1997; Niefind et al., 2003; Schlieben et al., 2005]. In previous experimental work, wild type (WT) *Lb*ADH and mutants were crystallized and the crystallization behavior was monitored [Grob et al., 2020; Nowotny et al., 2019].

In the first part of the thesis, a detailed crystallographic study of *Lb*ADH WT and mutant crystals with neutron and X-ray diffraction is performed to determine static crystal properties and to provide valid structure models:

- Room temperature neutron and cryocooled X-ray diffraction are performed on the same crystal to elucidate the influence of cryocooling.
- The symmetries found in *Lb*ADH crystals are analyzed concerning their influence on crystal packing.
- Further structure models of *Lb*ADH WT and mutant crystals are generated and compared concerning their structure and crystal packing.
- The physical origins of different symmetries in *Lb*ADH crystals are discussed.

In the second part, the structure models obtained in the first part are parameterized for atomistic MD simulations in order to explore dynamic and thermodynamic properties of the crystallization process:

- Structure model systems, set-ups, and methods for MD simulations are developed.
- Free energy changes are calculated to thermodynamically quantify a protein's crystallization ability *in silico*.
- The free energy changes are traced back to mechanistic interactions which are unraveled in atomic detail.
- With the new findings, established crystal growth and nucleation theories, previous experimental crystallization behavior, and implications for crystal engineering strategies are discussed.

Chapter 3

Protein Crystallization

3.1 Thermodynamics

In this study, the crystallization process of water-solvable proteins is considered: a thermodynamic system consisting of proteins in aqueous polar solution undergoes a phase transition. A system is referred to as a multitude of particles within defined borders to its environment or reservoir. The particles give rise to measurable thermodynamic quantities. The borders separating the system from the surrounding environment may be impermeable or permeable. Accordingly, thermodynamic quantities can either be constant state variables or changeable due to coupling with an external reservoir. For example, in crystallization experiments the borders of the system usually allow temperature coupling (e. g. in microbatch crystallization experiments). In the initial state of the system, the proteins are solvated in an aqueous solution, whereas in the final state, the proteins are bound to each other in crystalline order forming a crystal in solution. Both equilibrium states are connected by a thermodynamic process.

The free energy is an important thermodynamic quantity of a system. At thermodynamic equilibrium, the free energy of a system reaches its minimum. At constant pressure and temperature the Gibbs free energy G can be

formulated

$$G = H - TS, \tag{3.1}$$

where H denotes the enthalpy, T the temperature and S the entropy of a system. A thermodynamic process as discussed in the following can be evaluated according to its associated change of the free energy ΔG between the initial and final state. For the process of protein crystallization, changes in entropy can be decomposed into protein $\Delta S_{\text{protein}}$ and solvent $\Delta S_{\text{solvent}}$ contribution. The free energy change results in

$$\Delta G = \Delta H - T \left(\Delta S_{\text{protein}} + \Delta S_{\text{solvent}} \right). \tag{3.2}$$

A process can occur spontaneously if it induces a negative free energy change ($\Delta G < 0$), i. e. the system evolves in a lower free energy state. The free energy change for the process of protein crystallization has been estimated in the range of -10 to $-100\,\text{kJ mol}^{-1}$ [Derewenda and Vekilov, 2006]. During crystallization, the protein molecules are transferred from solution into the crystal lattice. The newly formed bonds in the crystal induce a negative enthalpy gain ($\Delta H < 0$) promoting the crystallization process. The entropy of the protein is decreased ($\Delta S_{\text{protein}} < 0$) as the incorporation reduces the proteins' degrees of freedom. The ordered arrangement results in a loss of translational and rotational degrees of freedom. As the protein's entropy enters with a negative sign in equation (3.2), the arrangement results in a positive contribution to the free energy difference ΔG. The situation is different for the entropy of the solvent. In the solvated phase, solvent molecules like water molecules are ordered at the surface of protein. When the proteins approach each other in order to form a crystal, the water molecules are released at the surface areas which are now crystal contacts. The solvent molecules which are then freely moving in the bulk solvent gain entropy ($\Delta S_{\text{solvent}} > 0$) by gaining degrees of rotational and translational freedom. The solvent entropy change promotes the crystallization process as it contributes to a negative free energy change ΔG.

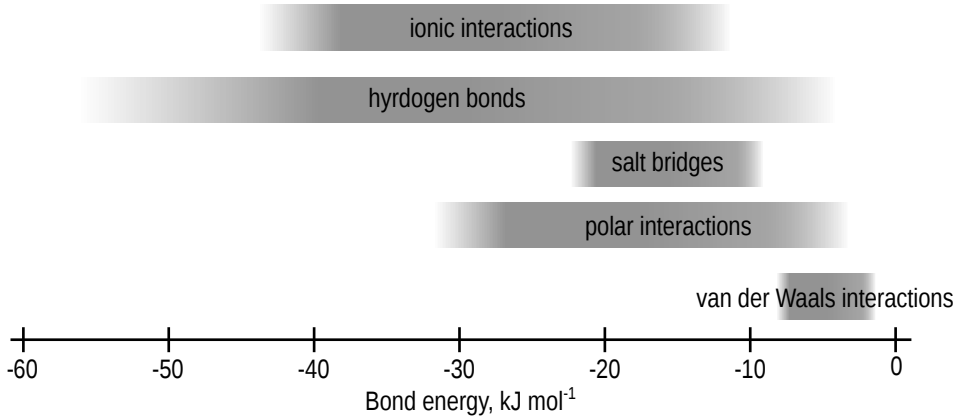

Fig. 3.1 Typical ranges for bond energies of molecular interactions. The typical interaction energies for ionic interactions, H-bonds, salt bridges, polar interactions, and van der Waals interactions are displayed.

3.2 Intermolecular Protein-Protein Interactions

The crystallization process relies on the change of non-covalent intermolecular interactions for both solvent and protein. Non-covalent intermolecular interactions include ionic interactions, hydrogen bonds (H-bonds), salt bridges, polar interactions, van der Waals interactions, and hydrophobic effects. In general, multiple effects contribute to one binding site and it is rare to find pure interactions of precisely one or the other character [Rupp, 2009]. The typical energy ranges of intermolecular interactions elucidated in the following are displayed in Fig. 3.1.

Ionic interactions. Ionic interactions are electrostatic interactions between charged residues. Typical examples are interactions between negatively charged acidic residue and positively charged basic residue. These interactions are typical for crystal contacts. For these long range electrostatic interactions, no exact distance criterion can be given but moderate interactions may exceed distance of 4 Å [Karshikoff and Jelesarov, 2008]. Ionic interactions are in the range of -12.5 to -40 kJ mol^{-1} [Rupp, 2009].

Hydrogen bonds (H-bonds). In order to be classified as H-bond, a hydrogen donor and hydrogen acceptor group needs to participate in the interaction. H-bonds can occur when a hydrogen atom is covalently polar bound, i. e.

the binding partner is more electronegative than hydrogen. This group is termed the hydrogen donor group. The now partially positively charged hydrogen atom is 'accepted' by any atom with free electrons. Examples for hydrogen donors are amide or imide groups. Good hydrogen acceptors are acidic residues and oxygen atoms of the polypeptide backbone. H-bond donor-acceptor distances for moderate interactions are in the range of 2.5 up to 3.5 Å which results in a hydrogen atom and acceptor distance of 1.5 to 2.5 Å. [Jeffrey and Jeffrey, 1997]. The electrostatic attractions between these polar groups are in the range of -6 to -29 kJ mol^{-1} [Rupp, 2009]. Others report the interaction energy of moderate H-bonds to be between -16.5 to -63 kJ mol^{-1} [Jeffrey and Jeffrey, 1997; Karshikoff and Jelesarov, 2008].

Salt bridges. Salt bridges are a combination of ionic interaction and H-bonding. An interaction is classified as salt bridge if the centroids of the functional group of the binding partners have a maximal distance of 4 Å [Kumar and Nussinov, 1999, 2002]. In another definition, a distance smaller than 4 Å between nitrogen of the positively charged basic residue and oxygen of the negatively charged acidic residue (e. g. amino group of lysine interacting with carboxyl group of aspartic acid) is used as salt bridge classification [Barlow and Thornton, 1983]. These definitions are often applied in studies on X-ray structures where no hydrogen atoms are present. In studies with hydrogen atoms (e. g. molecular dynamics (MD) simulations), a salt bridge defined as a couple of functional groups connected by H-bonds may be most appropriate [Karshikoff and Jelesarov, 2008]. The interaction energy of a salt bridge are reported as -12.5 to - 21 kJ mol^{-1} [Anderson et al., 1990].

Polar interactions. Polar interactions are dipole-dipole interactions between partners carrying opposing dipole moments (the same as H-bonds, but not involving hydrogen atoms). The strength of the interaction depends on the difference in the dipole moments and can vary in the range of -4 to -33.5 kJ mol^{-1} [Rupp, 2009].

Van der Waals interactions. Compared to polar interactions, van der Waals interactions are weak. Van der Waals interactions result from a quantum mechanical treatment where electron wave functions are considered. Each

Van der Waal interaction is weak, in the range of -2 to -6 kJ mol^{-1} but as they occur between every atom their sum can be significantly higher [Rupp, 2009].

Hydrophobic interactions. Hydrophobic interactions are indirect forces (e. g. in contrast to polar electrostatic interactions). As already stated for the crystallization process, water molecules around hydrophobic residues have to satisfy their H-bonds themselves, i. e. an ordered solvation layer around hydrophobic clusters is formed. When two hydophobic clusters are brought together, the formerly ordered water molecules are released in the solvent, which results in an solvent entropy gain. This stabilizes the crystal (c. f. equation (3.2)). It is important to emphasize that this is not a real force but rather an indirect thermodynmically stabilizing effect. Hence, no concrete range as for the other direct specific interactions can be given.

3.3 Crystal Nucleation and Growth Theories

The process of protein crystallization can be formally divided into a crystal nucleation and growth process. Nucleation is the initiation of the formation of a new thermodynamic phase. Crystal growth completes the phase transition. Crystal nucleation occurs in a thermodynamic environment termed supersaturation. When reaching supersaturation in crystallization solution, there are different pathways suggested in literature for a nucleus to form and grow.

3.3.1 Supersaturation

Supersaturation is a meta-stable state when a solution contains more solute than its solubility limit allows under normal physical conditions. Both nucleation and crystal growth can only happen in a supersaturated state. This non-equilibrium state is left into an equilibrium state by lowering its free energy. This can either be carried out by precipitation or by phase transition, i. e. creation of a nucleus and crystal growth [McPherson, 2004]. A supersaturated solution can be reached by varying parameters like salt concentration, temperature, or pH [McPherson and Gavira, 2014].

3.3.2 Nucleation

One Step Nucleation (Classical Nucleation Theory)

Based on Gibb's early formulation of the relation between work and stability of a nucleus of a new phase [Gibbs, 1878], Volmer and Weber [1926], and Becker and Döring [1935] formulated a thermodynamic concept of nucleation which is now called classical nucleation theory (CNT). Due to Brownian motion crystal building blocks encounter each other. There is constant attachment and detachment, where building blocks stay attached due to intermolecular interactions to form clusters. The main assumption in CNT is that small clusters have the same properties as the bulk material. This means that the clusters can only grow to nuclei if the building blocks are attached in the same way as in the desired phase (c. f. Fig. 3.2 **a**). The cluster formation is thermodynamically determined by enthalpic gain due to newly formed intermolecular bonds, and an entropic loss due to emergence of a cluster-liquid interface surface tension (positional entropy loss). For an ideal model system (cluster is a spherical object), the energetic gain scales with volume, where as the loss scales with surface of the cluster. Initially the surface loss dominates. At a certain cluster size, the free energy landscape reaches a local unstable maximum. The cluster has equal probability to grow or to dissolve. The nucleus has reached its critical size. Addition of one more building block will doom the nucleus to grow. As the phase transition is hampered by only one energy barrier, CNT is seen as one-step nucleation mechanism. The above formulation of CNT can be applied to any kind of phase transition. In case of crystallization, CNT implies that the crystal building blocks are attached to each other via the same interactions found in a fully grown crystal. Hence, in CNT the nucleus formation is driven by oriented attachments between building blocks. The building blocks bind to each other via the same protein-protein interactions sites which are later seen as crystal contacts. The building blocks can only stay attached, i. e. the cluster can only grow to a critical nucleus, if the building blocks are orderly attached, i. e. exhibit the same properties as the phase in which it will evolve. In the case of protein crystallization: the building blocks are bound according to crystal symmetry.

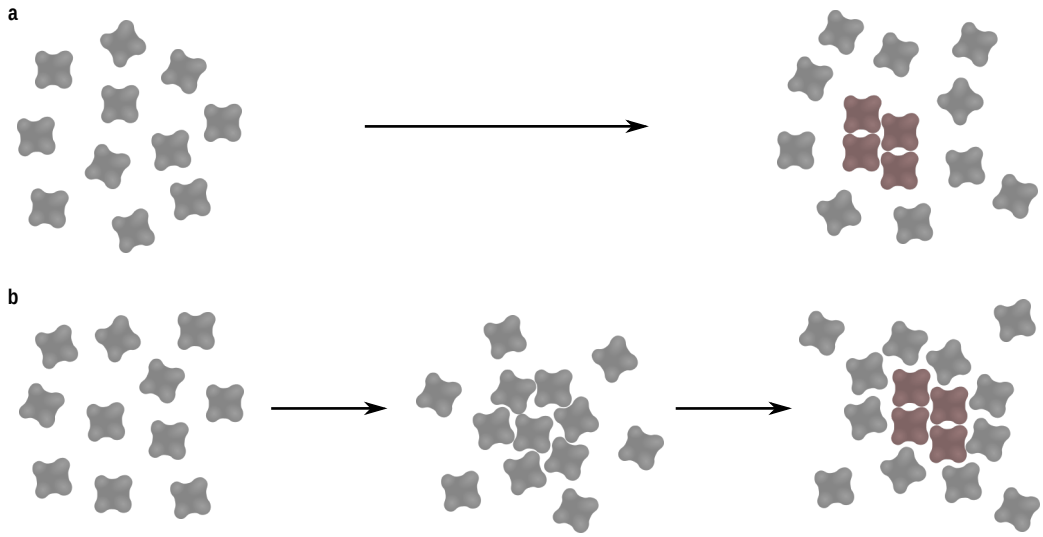

Fig. 3.2 One step (CNT) and two step nucleation. **a** In CNT, the crystalline nucleus forms in one step. The nucleus formation is hampered by only one energy barrier. **b** In the two step nucleation theory, a dense cluster originates before a crystalline nucleus is formed within this cluster. Nucleation is hampered by two energy barriers as it evolves in two steps.

Two Step Nucleation

Based on numerical calculations, ten Wolde and Frenkel [1997] suggested an alternative nucleation path. Numerically, they showed a thermodynamically favored nucleation pathway involving a dense liquid cluster before transformation into a crystalline nucleus. This dense cluster exhibits a different structure (c. f. Fig. 3.2 **b**). As the nucleation pathway involves another stage, it is referred to as 'two step' nucleation pathway. Two step nucleation involves two energy barriers. Hence, two steps can be rate defining.
Studies have described the nucleation time beeing dependent on formation of the metastable phase [Kashchiev et al., 2005]. Whereas newer publications found the dense clusters (meta stable phase) in which nuclei are built have much shorter time scale than nucleus formation in those clusters. Dense clusters (meta stable phase) evolve instantenously [Vekilov, 2010].
In any case, the nucleation process involves the initial formation of a cluster with a different structure, which then transforms into a stable crystalline state. The transformation into the crystalline nucleus follows the same principles as described for the classical pathway: orderly formation and attachment of crystal building blocks.

Direct Experimental Observations

Indirect observations rely e. g. on kinetic considerations. Here, four key studies for nucleation pathway search are described which rely on direct observation of the nucleation.

Experimentally, the one step nucleation pathway was first observed in the year 2000 on a molecular level with atomic force microscopy (AFM) for the nucleation of apoferritin [Yau and Vekilov, 2000]. Near-critical-size clusters (pre-critical nuclei) contained 20 to 50 molecules depending on the level of supersaturation. The number of molecules decreased with the level of supersaturation. The cluster exhibited the same structure as in appoferitin crystals.

In 2010, the first direct observation of a two step nulcation process was published. AFM showed the two-dimensional nucleation of S-layer protein SbpA. The nucleation was shown to occure by phase transition from amorphous clusters into crystalline clusters on supported lipid bilayers [Chung et al., 2010]. However, the two step nature may be related to conformal transformation of the monomers while phase transition and the liquid like nature of the supporting substrate [Sleutel and Van Driessche, 2018].

Another AFM study revealed the two-dimensional nucleation of glucose isomerase on a solid substrate according to one step nucleation. In this study the critical size of the nucleus could be estimated to 20 molecules [Sleutel et al., 2014].

Further AFM studies exists. However, the problem with AFM is, that it is limited to systems that crystallize on substrates. The studies are blind to the actions in fluids. The nucleation is influenced by the supporting substrate. Furthermore, the nucleation on substrate is naturally a heterogenous nucleation [Sleutel and Van Driessche, 2018].

To overcome these shortcomings, a recent study used time resolved cryo-transmission electron microscopy (TEM). Samples of glucose isomerase were plunge frozen at various intervals. From bright-field TEM images thereof crystallographic parameters could be calculated which corresponded to the macroscopic crystal parameters. No evidence for amorphous pre-critical clusters were observed [Van Driessche et al., 2018].

Homogenous and Heterogenous Nucleation

The models described above were developed for homogenous nucleation. Homogenous nucleation describes the process of spontaneous formation of a nucleus in the bulk of the solution. In heterogenous nucleation the nuclei is formed with the help of a 'seed'. This can be a foreign particle, the crystallization container or an already formed crystal. Hetereogenous nucleation reduces the energy barrier to form a nucleus: volume to surface ratio is shifted towards volume if a seed is involved. Hence, the surface entropy loss is effectively reduced and nucleation can happen at a lower supersaturation level [McPherson, 1999].

3.3.3 Crystal Growth

Crystal growth by two-dimensional nucleation and by spiral dislocations was a model developed mainly by Burton, Cabrera, and Frank around 1950 [Burton et al., 1951]. So far, it has been identified as the main growth mechanisms in protein crystallization [Malkin et al., 1995; McPherson et al., 2003, 2001; Vekilov and Chernov, 2003; Vekilov and Rosenberger, 1996]. The growth originates from steps at crystal surfaces which are produced by a two-dimensional nucleation process or by spiral growth advancing from dislocations.

The nucleation of new step edges or nucleation islands on the crystal surface is followed by tangential growth and extension of step edges at the surface. At the latest, when the new surface layer is completed, i. e. the surface is flat, new step edges or islands have to be initiated by two-dimensional surface nucleation. At high supersaturation level or small crystal size, two-dimensional growth islands originate randomly all over a crystal surface. With increasing crystal size and decreasing supersaturation level, two-dimensional surface nucleation becomes restricted to facet edges and, finally, to facet corners [Malkin et al., 1995; Vekilov and Rosenberger, 1996]. Two-dimensional nucleation and the addition of these building blocks is from a physical view point similar to the nucleation process of classical nucleation theory [Durbin and Feher, 1986]. The underlying molecular mechanims follow CNT and the growth is often modelled according to CNT [Schmit and Dill, 2012]. This

growth method is facilitated by productive binding of building blocks in crystal orientation and by a favor of association over dissociation of building blocks. This need of orientated binding determines the rate of crystal growth [Asthagiri et al., 2000].

In spiral growth mode which is energetically favored at low supersaturations the step edges evolve from screw dislocations [Burton et al., 1951]. This dislocation serve as growth hillocks. Due to the spiral nature this growth hillocks never disappear.

It was found that different crystal faces of the same crystal may show different preferred growth mechanism. Generally, it was found that at high supersaturation level and small crystal size, two-dimensional nucleation is favored, whereas at low supersaturation level and increasing crystal size, crystals predominatly grow by spiral growth [Malkin et al., 1995; Vekilov and Rosenberger, 1996].

Independent on the exact growth mechanism, the critical step in the growth of crystals is the directed incorporation of building blocks into the growing crystal in crystalline order. Crystal building blocks which are not orderly bound would lead to the growth to slow down or would lead to an amorphous phase.

3.4 Protein Engineering for Crystallization

One of the earlierst works of protein engineering for the purpose of crystallization can be found in the work of Lawson and co-workers in 1991 [Lawson et al., 1991]. While horse spleen ferritin and rat liver L-chain ferritin could be crystallized, crystallization experiments of human H-chain ferritin did not yield X-ray diffracting crystals. Lawson et al. found L-ferritins exhibit a metal ion (Cd^{2+}) mediated crystal contacts via an amino acid not present in human H-ferritin. By point mutation they successfully enabled this crystal contact also in human H-ferritin and therefore were able to obtain crystals of point mutated human H-chain ferritin. This approach to not only vary the crystallization conditions but to use the protein as a variable in crystallization experiments, was investigated in further studies [Dale et al., 2003; D'Arcy

et al., 1999]. Charron et al. directly targeted interactions at crystal contacts of aspartyl-tRNA synthetase from *T. thermophilus* and monitored improved or decreased crystal quality [Charron et al., 2002].

A particular focus was layed on the mutation of lysines and glutamines on the protein's surface. Human RhoGDI GTPase was studied in the Derewenda group to investigate lysine to alanine mutation [Longenecker et al., 2001] and glutamic acid to alanin or aspartic acid [Mateja et al., 2002]. They found both strategies successful for producing crystallizable protein variants. The strategy to mutate lysines and glutamates was further developed to the so called surface entropy reduction (SER) strategy and the associated surface entropy reduction prediction server (SERp server) [Cooper et al., 2007; Derewenda, 2004, 2010; Derewenda and Vekilov, 2006; Goldschmidt et al., 2007]. SERp server is a program for the design of crystallizable protein variants. Incorporating large flexible residues at a protein's surface in an ordered crystal, results in a loss of entropy. Examples are lysines and glutamates. They both have a large solvent exposed area and their side chains are very flexible. Incorporation in a crystal contact results in loss of flexibility. Replacing large hydrophilic residues by alanies, results in a local reduction of conformational entropy and reduces the entropic effort of ordering side chains at a newly formed crystal contact. Furthermore, this generates homogeneous surface patches for contacts. In solution, water molecules are generally ordered at crystal surface, both at hydrophilic and hydrophobic patches. Water molecules may be firmly attached and ordered at polar surface or ordered by themselves at hydrophobic patches. Upon crystallization, water molecules are released from the surface where crystal contacts are formed. Upon release, the previously ordered water molecules gain entropy as they are now free in solution. This is thought to be a driving force. Small residues allow better ordering of water molecules. Large flexible side chains prevent water molecules from ordering.

SERp server [Goldschmidt et al., 2007] combines three steps for predicting mutations which might enhance crystallization: (1) prediction of the surface entropy, (2) prediction of the secondary structure, and (3) analysis of the conservation of residues. It suggests mutations on the surface which may facilitate crystallization. SERp server is based purely on the sequence of

the protein and only predicts the secondary structure as it is designed for proteins which do not crystallize.

SER substitutes large flexible charged residues by small uncharged residues. This reduces the effort of incorporating large flexible residues at the crystal contact. And it fabricates hydrophobic patches with small residues for better ordering of water molecules which results in a greater entropy gain upon release of these water molecules. SER reduces the 'entropy shield' of a protein.

3.5 Chemical Composition of Protein-Protein Interfaces

In general, the amino acid composition of a protein surface is rich in polar and charged residues (asparagine, glutamine, glutamic acid, lysine) and depleted in hydrophobic residues (isoleucine, leucine, methionine, valine) [Bahadur et al., 2004; Conte et al., 1999]. Protein-protein interfaces can be divided into biological protein-protein interfaces (interfaces tailored by nature) and interfaces developed in the crystallization process (crystal contacts).

Biological interfaces. It has been found that lysines and glutamines are least favored amino acids at biological protein-protein interactions. Conte et al. [1999] analyzed 75 protein-protein complexes and compared the general surface attribution of amino acids vs. the interface surface attribution. They found that compared to the average protein surface, in biological interfaces much more aromatic residues (histidine, tyrosine, phenylalanine, and tryptophan) with 21% vs. 8% and more aliphatic residues (leucine, isoleucine, valine and methionine) with 17% vs. 11% occur. The contrary was observed for charged residues like aspartic acid, glutamic acid and lysine. In particular, in the study of oligomeric proteins, lysine and glutamic acid were the most abundant amino acids at protein surfaces but underrepresented in biological interfaces (14.9% vs. 5.4% for lysine and 10.3% vs. 4.1% for glutamic acid). A similar depletion was found for aspartic acid (7.8% vs. 4.8%). An exception

was found for arginin which was found to be depleted at protein surfaces but to make a similar attribution in biological interfaces.

In a later study, with a considerable lager data set, the generally preferential occurrence of large hydrophobic and uncharged polar residues and the preferential suppression of charged residues at the interfaces was confirmed [Bordner and Abagyan, 2005]. While they also found depleted amino acids aspartic acid, glutamic acid, and lysine in the same order, the order for most prominent residues (phenylalanine, cysteine, leucine, methionine, isoleucine, tyrosine, tryptophan, valine, histidine, alanine) differed a little bit probably due to the different composition of the data set (homodimers, heterodimers and multimers) [Bordner and Abagyan, 2005].

The generally enrichment of aromatic and alipatic residues and depletion of lysine and glutamic acid at biological interfaces was reported in several more studies and is well established [Bahadur et al., 2004; Chakrabarti and Janin, 2002; Jones and Thornton, 1996; Ofran and Rost, 2003].

Crystal contact interfaces. Analysis of crystal contact interfaces resulted in no significant difference in composition compared to the average surface of a protein [Bahadur and Zacharias, 2008; Bahadur et al., 2004]. Polar and charged residues are generally abundant and hydrophobic residues are depleted.

Chapter 4

Protein Crystallography

4.1 Diffraction Theory

Crystallography relies on the scattering, i. e. diffraction of incident waves. In X-ray diffraction, electromagnetic waves scatter on electrons, or in particular on the electron density around the atoms; in neutron diffraction, neutrons scatter on nuclei. For simplicity of reading, in the following the general theoretical explanations of diffraction theory and model reconstruction is tailored for X-ray diffraction. However, exchanging the terms 'X-ray' for 'neutron beam' and 'electron density' or 'electron distribution' for 'nuclei distribution' or 'nuclei density' yields an neutron diffraction equivalent.

The structural motifs within an unit cell (UC) are repeated periodically throughout the crystal which is why the crystal can be treated as 3D diffraction grating. This results in directions of positive and negative interference according to Bragg's law. The incident wave is scattered by atoms in lattice planes separated by interplanar distance d. Constructive interference occurs when the waves scattered at different planes stay in phase: i. e. the path length difference of the scattered waves is an integer multiple n of the wavelength λ. The path length difference can be calculated as $2d \sin(\theta_G)$ where θ_G is the glancing angle (see Fig. 4.1.) Therefore, the condition for constructive

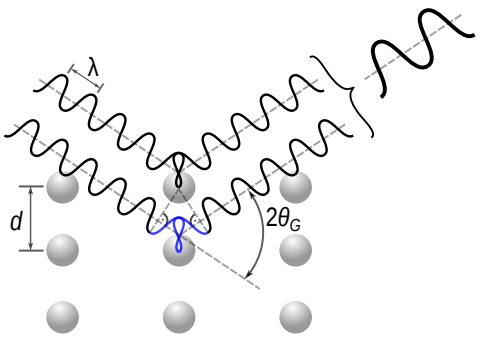

Fig. 4.1 Bragg's law for constructive interference. Two incident waves with wavelength λ scatter on atoms (represented by spheres) in lattice planes separated by a distance d. If the path length difference (colored blue) of $2d\sin(\theta_G)$ after scattering results in an integer n phase shift ($n\,\lambda$) the scattered waves interfere constructively.

interference, known as Bragg's law, yields

$$2\,d\,\sin(\theta_G) = n\,\lambda\,. \tag{4.1}$$

The planes in the crystal on which the incident waves diffract are referred to as crystallographic planes and noted by Miller indices (h,k,l). Diffraction spots (monitored on a detector plate; see Fig. 4.2) occur where scattered waves satisfy Bragg's law for a crystallographic plane (h,k,l). Hence, the direction of the positive interference, i. e. position of the diffraction spots, depends on the translative crystallographic UC (size and shape) and the wave length of the X-rays. The position of the atoms in the cell define the intensity of the diffraction spots [Drenth, 2007; Wlodawer et al., 2008, 2013].

Mathematically, the X-rays perform a discrete Fourier transform (FT) of the electron distribution $\rho(\mathbf{r})$ (or equivalent nuclei distribution in neutron diffraction) from real space (x,y,z) into reciprocal space (h,k,l) where reciprocal lattice points define the diffracting plane in the crystal. The function in reciprocal space is called structure factor $\mathbf{F}(hkl)$. The electron density can be reconstructed by an inverse Fourier transform (IFT) of the structure factor $\mathbf{F}(hkl)$:

$$\rho(x,y,z) = \sum_{h,k,l} \mathbf{F}(h,k,l)e^{-2\pi i(hx+ky+lz)}. \tag{4.2}$$

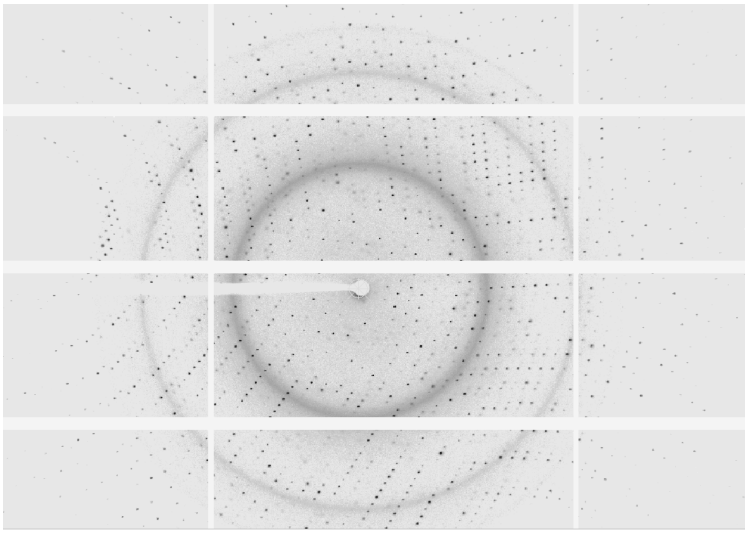

Fig. 4.2 Diffraction pattern of an alcohol dehydrogenase from *Lactobacillus brevis* (*Lb*ADH) crystal. Diffraction spots are collected on an image plate detector. During data reduction, the diffraction spots are identified and labeled with a scattering vector (or Miller indices) (h, k, l).

The structure factor $\mathbf{F}(hkl)$ is a discrete complex function and can be separated into its amplitude $|\mathbf{F}(hkl)| = F(hkl)$ and phase $\varphi(hkl)$:

$$\mathbf{F}(hkl) = F(hkl)e^{i\varphi(hkl)}. \tag{4.3}$$

The amplitude $F(hkl)$ corresponds to the intensity of the diffraction spot $I(hkl)$ and is calculated as $F(hkl) = \sqrt{I(hkl)}$. However, the phase information $\varphi(hkl)$ can not be measured in the diffraction experiment. For reconstruction of electron density from the structure factors via IFT (equation (4.2)) the phases have to be obtained [Drenth, 2007; Wlodawer et al., 2008, 2013].

4.2 Processing of Crystal Diffraction Images

4.2.1 Data Reduction

In a first step, the diffraction images (see Fig. 4.2) measured in a diffraction experiment at an X-ray source have to be evaluated. This is called data reduction. The data is corrected for background noise, the diffraction spots are detected, and the associated scattering vectors (h, k, l) are determined.

The diffraction spots are analyzed and UC geometry and space group (SG) are calculated. Depending on the obtained UC and SG, the reflections are reduced to a list of reflections (h, k, l) and their intensity $I(hkl)$ including an error estimate thereof [Drenth, 2007; Wlodawer et al., 2008, 2013].

4.2.2 Model Construction

The second step starts with a first initial phase reconstruction. In molecular replacement (MR) the 'phase problem' (loss of phase information in the diffraction experiment) is targeted. For an initial phase estimate different methods exists. The most common method is MR based on Patterson methods. The Patterson function is the FT of the square of the structure factor amplitudes $[F(hkl)]^2$ [Patterson, 1935]. Hence, the Patterson function is calculated based on the available experimentally observed data, i. e. the structure factor amplitudes $F(hkl)$ but not the structure factor itself $\mathbf{F}(hkl)$. Therefore, the Patterson function does not yield the atomic positions, but it was shown to result in a map of interatomic vectors [Patterson, 1935]. A structural homologue is used as search model and fitted in the calculated Patterson map. Based on the fitted structure model homologue, an initial phase estimate can be calculated.

Before the refinement process starts, a certain amount of reflections (normally 5%) is labeled as test set and not used during reciprocal space refinement. This set of reflection is referred to as *free* set and is used for cross validation (see below). These reflections are excluded from refinement. The rest of the reflection set is termed *working* set and used during reciprocal refinement.

During refinement the atomic positions and atomic displacement parameters known as B-factor are changed. The B-factor models atomic fluctuations and static disorder. Depending on the experimental data quality and resolution the B factor is modeled isotropic, anisotropic or by anisotropic TLS (Translation/Libration/Screw) refinement parameters.

The refinement process consists of iterative repetition of two refinement protocols: (1) Global refinement in reciprocal space where the observed and calculated structure factors are minimized against a likelihood target function. (2) Local refinement in real space where the atomic positions and

B-factors are manually changed while viewing calculated electron density maps. Both refinement methods are alternated. In both refinements atomic geometrical weights are put on atomic positions (geometrical restraints) by prior knowledge of stereochemical geometry [Drenth, 2007; Wlodawer et al., 2008, 2013].

Reciprocal Space Refinement

In order to assess how well the current model reflects the measurement, residuals or so-called crystallographic R-factors are calculated. The R-factors evaluate how well the observed structure factor amplitude $|\mathbf{F}(hkl)| = F_\mathrm{o}(hkl)$ and the structure factor amplitude calculated based on the current model $F_\mathrm{c}(hkl)$ coincide. The first R-factor is calculated based on all the reflections used in refinement and is referred to as R_work

$$R_\mathrm{work} = \frac{\sum\limits_{\{h,k,l\}\notin\text{`free'}} |F_\mathrm{o}(hkl) - F_\mathrm{c}(hkl)|}{\sum\limits_{\{h,k,l\}\notin\text{`free'}} F_\mathrm{o}(hkl)}. \tag{4.4}$$

Equivalently a residual R_free is calculated based purely on the *free* set of reflections

$$R_\mathrm{free} = \frac{\sum\limits_{\{h,k,l\}\in\text{`free'}} |F_\mathrm{o}(hkl) - F_\mathrm{c}(hkl)|}{\sum\limits_{\{h,k,l\}\in\text{`free'}} F_\mathrm{o}(hkl)}. \tag{4.5}$$

The lower the R-factors are, the better the constructed model reflects the measured quantities. Monitoring the progression of R_free and R_work evaluates the success of the refinement steps, i. e. if the latest modifications to the model did lead a better or worse agreement with the measured structure factor amplitudes. It is particular important to monitor R_free as the free set is not used during refinement: if only R_work goes down while R_free stays the same or goes up, it is a clear sign of over-fitting and over-interpretation [Drenth, 2007; Wlodawer et al., 2008, 2013].

Real Space Refinement

In local real space refinement the model is viewed against electron density maps. The first electron density map is calculated from structure factor amplitude ($2F_o - F_c$). This could be best understood as a sum of the 'observed' map (calculated from F_o) and difference map (calculated from ($F_o - F_c$)). It therefore shows the electron density and missing/spurious features simultaneously. The second electron density map, which is viewed in real space refinement, is the difference map ($F_o - F_c$). It is displayed at two contour levels: a positive contour level to indicate that the model contains something which is not observed and a negative contour level to indicate that there might be something missing in the model.

However, one has always to keep in mind that the phases needed to calculate the electron density map with the structure factors via IFT are calculated according to the modeled electron density (encoded in the modeled atom positions). The calculated electron density is not an *observed* quantity but depends on the already positioned atoms, i. e. on itself. Changing one atom position influences in a certain degree the whole electron density as it influences the phase information which is calculated based on the atom positions [Drenth, 2007; Wlodawer et al., 2008, 2013].

4.3 Peculiarities of Neutron and X-ray Diffraction

4.3.1 X-ray and Neutron Sources

Nowadays, the most common methods to gain diffraction data are synchrotron sources. In-house X-ray sources are only used for testing purposes. In this study, the diffraction quality of crystals were initially tested at the Catalysis Resarch Center (CRC), Technical University of Munich (TUM), in a diffractometer with a molybdenum source and a copper anode. The powerful X-ray radiation at synchroton enables measuring a whole data set within minutes whereas it takes days using an in-house source. At synchrotrons charged particles, e. g. electrons, are accelerated in a closed loop. Magnetic fields bend the particles to run in a closed fixed loop. The X-ray

radiation is a result of the so-called *bremsstrahlung*: charged particles which are accelerated radially emit electromagnetic radiation. Diffraction data was collected at the Synchrotron Radiation Facility (ESRF) in France, at the Swiss Light Source (SLS) in Switzerland, and data for initial test structures was obtained at Diamond Light source in England (the source is specified in the corresponding protein data bank (PDB) entry).

In principle, two techniques for generating neutrons for diffraction experiments are used. Neutrons are either produced in nuclear reactors or in spallation sources. For this study, the neutron diffractometer BioDiff [Ostermann and Schrader, 2015] at the nuclear reactor FRMII in Garching was used. The continously produced neutrons in nuclear reactions have energies of several MeV and therefore have to be slowed down first. These slow neutrons are also called cold neutrons [Niimura and Podjarny, 2011].

4.3.2 Wave-Length Relation

Neutron diffraction and X-ray diffraction have different energy-wavelength-relations. Neutrons are baryon particles with mass, whereas X-rays, i. e. electromagnetic waves or photons, are boson particles without mass. For a neutron with a mass m, the relationship between the energy E, velocity v and *de Broglie* wavelength λ is written as

$$E = \frac{1}{2}mv^2 = \frac{h^2}{2m\lambda^2}, \tag{4.6}$$

where h is the Planck's constant. For X-rays the energy E of a photon is related to the wavelength λ as

$$E = \frac{hc}{\lambda}, \tag{4.7}$$

where c is the speed of light. Equation 4.6 shows that the wavelength of neutrons can be controlled via their velocity. Hence, the wavelength can be adjusted to the structure under analysis. For crystallography, a wavelength of the order of 1 Å is chosen. A wavelength of 1 Å corresponds to an energy of 82 meV for neutrons and 12.4 keV for X-rays.

4.3.3 Cryoprotection

The energy of X-rays may cause radiation damage as the covalent bonding energies in the protein are in the range of eV. Covalent bonds in the protein may break, if the crystal is not prepared properly, e. g. with liquid-helium or liquid-nitrogen freezing. Due to the weaker energy of the neutron beam, neutron diffraction can be performed at room temperature [Blakeley et al., 2008; Niimura and Podjarny, 2011]. Before cryocooling the crystals, the crystals are soaked in mother liquor supplemented with cryoprotectant like glycol. The cryocooling has to be performed in cooling times of 0.1 to 1 s [Kriminski et al., 2003; Teng and Moffat, 1998]: protein crystals typically contain between 20 and 90% (w/w) water [Kriminski et al., 2002]. When flash-frozen with cooling times of 0.1 to 1 s, the water molecules do not undergo a phase transition from liquid into hexagonal ice phase but into an amorphous phase, i. e. glassy or microcrystalline phase [Kriminski et al., 2002]. This prevents a considerably high volume change of water and the associated distortion and eventually cracking of the crystal.

4.3.4 Diffraction Physics

Another peculiarity between neutron and X-ray diffraction was already stated in brief: neutrons are scattered at the nucleus, whereas X-rays or, in the particle picture, photons are scattered at the electron density around an atom or chemical bond [Blakeley et al., 2008; Drenth, 2007; Niimura and Podjarny, 2011; Wlodawer et al., 2008, 2013]. Hence, hydrogen atoms can hardly be seen with X-ray diffraction. Neighboring elements in the periodic table differ only by one electron and are hard to distinguish with x-ray diffraction. Neutrons interact with the nuclei of the atoms. The scattering power of a nucleus is quantified in the so-called scattering length of the specific element. Neutron diffraction not only shows hydrogen atoms but also distinguishes between different isotopes. This means hydrogen bonding, protonation states, and hydration of the protein, including the orientation of water molecules, can be determined [Niimura and Podjarny, 2011].

A common technique is to deuterate the protein crystal before neutron diffraction experiments. Neutron diffraction then reveals the exchange

ratio of hydrogen atoms with deuterium atoms (H/D ratio). This ratio provides information on whether the hydrogen atoms are easily accessible. Furthermore, deuteration reduces the high background noise in the diffraction pattern produced by incoherent scattering at hydrogen atoms [Niimura and Podjarny, 2011].

4.3.5 Scattering Lengths

The term 'scattering length' of an atom describes the probability or 'strength' of an atom to scatter incident neutron or photon. It is proportional to the scattering cross-section. For neutron diffraction the scattering length of an atom varies by element and isotop whereas the scattering of X-rays increases with atomic number [Blakeley et al., 2008; Drenth, 2007; Niimura and Podjarny, 2011; Wlodawer et al., 2008, 2013].

An important characteristic of neutron diffraction is that elements do not only have positive but also negative scattering lengths. The scattering lengths of atoms in close groups add up. Hence, in some cases it may occur the scattering in specific groups cancel each other out, and therefore the groups cannot be seen. This is for example the case for methylene groups. Furthermore, sulfur-sulfur bonds can hardly be dissolved as the scattering length of sulfur is relatively small. On the other hand for example nitrogen has a rather large scattering length and can be identified easily [Niimura and Podjarny, 2011].

4.3.6 Refinement Possibilities

Besides refining neutron and X-ray diffraction data separately there is the possibility for *joint refinement* in which both data sets are combined. For joint refinement, neutron and X-ray diffraction experiments should be conducted on the same crystal and at the same temperature. X-ray diffraction data provides information on heavy atoms, whereas neutron data may yield information on hydrogen atoms, protonation states and isotopes. However, if the neutron data is of sufficient quality, there is no need for joint X-ray data and the neutron structure alone may yield insightful results [Blakeley et al., 2008; Niimura and Podjarny, 2011].

4.4 Crystallographic Terms and Relations

Asymmetric unit (AU). In some literature confusingly called unit cell. An asymmetric unit (AU) is the smallest portion needed to build up a crystal by symmetry operations. These operations are defined in the symmetry group (or point group or crystallographic group) [Lattman and Loll, 2008]. The structure and information of an AU is published in the PDB.

Space group (SG). In crystallography also called crystallographic groups. The SG defines the symmetry operations under which the AU is invariant within the crystal. The operations include translation, reflection, rotation, crew axis and glide plane symmetry operations. A protein crystal comprises of chiral elements. Therefore, the symmetry operations in a protein crystal are restricted to rotation, screw axes and glide planes in order not to change the hand of objects. This narrows down the number of possible SGs from 230 to 65 [Lattman and Loll, 2008].
Examining the crystal structures in the PDB, it can be seen that in protein crystals some SGs are favored over others. The most frequently occurring SGs for protein crystals are the orthorhombic group $P2_12_12_1$ followed by the monoclinic group $P12_11$ (http://www.rcsb.org/pdb).

Unit cell (UC). Also called translative cell. A UC is the smallest unit needed to build up a crystal by translation. A UC contains one or more copies of an AU [Lattman and Loll, 2008].

Biological unit. Also called functional unit, biological assembly. The biological unit is the quaternary structure that represents the physiologically functioning form of a macromolecule [Krissinel and Henrick, 2007]. In general, the biological unit does not coincide with the AU. There are several programs that attempt to predict the biological unit from the crystallographic information. A popular and frequently used program is PISA (Protein Interfaces, Surfaces and Assemblies) [Krissinel and Henrick, 2007].

Relation between the AU and the biological unit. The contacts or interaction in the quaternary structure of the protein are termed biological relevant interactions or significant interactions. When a protein crystallizes it is assumed that these biological assemblies are unchanged in the crystal structure. The additionally formed contacts are called superficial crystal contacts or artifacts of crystal packing [Krissinel, 2015].

When the biological unit consists of more than one protein, this macromolecular complex may exhibit 'crystallographic' symmetry, which is superimposed with the symmetry in the crystal. The crystal symmetry is determined according to the smallest unit, which is needed to reconstruct the whole crystal, i. e. the AU. Hence, from looking at the crystal structure alone, it cannot necessarily be distinguished if the contacts occur due to the crystallization process or already existed before in the quaternary structure of the protein.

Mosaicity. The mosaicity α of a crystal or the mosaic spread, is a measure of the imperfection of a crystal [Zachariasen et al., 1945]. A crystal is considered to consist of an ensemble of small perfect crystals, so-called crystallites, which are not perfectly aligned [Camattari et al., 2015]. The relative angle between the crystallites is thought to be Gaussian distributed [Darwin, 1922]. This broadens the diffraction spot as the Bragg condition is met for different diffraction angles. The mosaicity is defined as the full width at half maximum (FWHM) of the diffraction peaks. Due to this definition, not only the intrinsic mosaicity of the crystal, but also the 'mosaicity' due to beam divergence and wave length distribution is measured. Both contribute to the broadening of the diffraction peak. In X-ray sources of high brilliance e. g. synchrotron these effects are neglegtibe, and the measured mosaicity resembles the intrinsic mosaicity of the crystal. However, in in-house X-ray sources or neutron sources the beam divergence and wave length distribution may dominate.

Chapter 5

Thermodynamic and Computational Concepts

5.1 MD Simulations

5.1.1 Classical Atomistic Model

The most principal choice for studying a system in numerical calculations concerns the level of detail. Ideally, the level of detail reflects the question which are aimed to be answered. The smallest units simulated in molecular dynamics (MD) simulations are atoms and the interaction between them is described in a classical potential. Molecular mechanics potentials provide an atom-centered description for all interactions - in contrast to a quantum mechanics approach, where electron coordinates are the principal variables. The approximation, where only the atoms are considered is called Born-Oppenheimer approximation. In general, an atom centered potential V is composed of a potential for bonded V_{bonded} and non-bonded $V_{\text{non-bonded}}$ interactions:

$$V = V_{\text{bonded}} + V_{\text{non-bonded}} \tag{5.1}$$

$$V_{\text{bonded}} = V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihedral}} \tag{5.2}$$

$$V_{\text{non-bonded}} = V_{\text{electrostatic}} + V_{\text{van der Waals}} \tag{5.3}$$

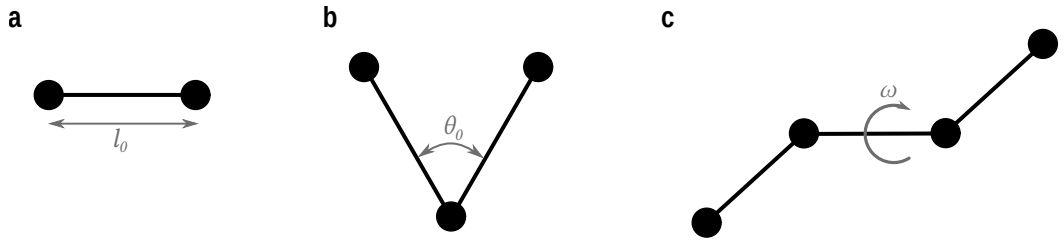

Fig. 5.1 Scheme of intermolecular interactions in an MD potential for bonded interactions. Atoms are represented by black spheres and covalent bonds are indicated as black lines. In an MD potential bonded interactions between atoms are parametrized according to their **a** bond length, **b** angle and **c** dihedral angle.

The potentials for MD simulations differ in their functional form and the way their parameters are derived. Commonly, bonded V_{bonded} and non-bonded $V_{\text{non-bonded}}$ potentials take the general form [Cornell et al., 1995]:

$$\begin{aligned} V = & \sum_{\text{bonds}} k_b(l - l_0)^2 + \sum_{\text{angles}} k_a(\theta - \theta_0)^2 \\ & + \sum_{\text{dihedrals}} \sum_{n} \frac{1}{2} k_d[1 + \cos(n\omega - \zeta)] \\ & + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}. \end{aligned} \tag{5.4}$$

The first three terms describe the potential V_{bonded} for **bonded interactions**, i. e. potential between covalently bound atoms (c. f. Fig. 5.1): Bond length l (first term) and angles θ (second term) are modeled with harmonic potentials with equilibrium bond length l_0 and angle θ_0 and force constants k_b and k_a, respectively. The dihedral angle ω (third term) is modeled with a cosine function with phase ζ, multiplicity n and scaling factor k_d.

The last term in equation (5.4) summarizes the potential $V_{\text{non-bonded}}$ for the **non-bonded interactions** where r_{ij} refers to the distance between atom i and j. It is described by a Lennard-Jones and a Coulomb term. The Lennard-Jones potential with parameters r_0 and ϵ_{ij} describes short range repulsive interactions resembling Pauli-exclusion principle and attractive long range interactions resembling attractive van der Waals and dipol-dipol interactions. The Coulomb term describes electrostatic interactions for a point-charge model with charge q_i for atom i and charge q_j for atom j and vacuum

dielectric constant ϵ_0. All parameters are specific for atom types and groups of atom types. I. e. there are several sets of parameters for one atom type (element) depending on the surroundings of the atom. A potential of the presented form (5.4) together with all its parameters is called MD force field. The time evolution of all particles is calculated based on the force field and is described in the following sections.

5.1.2 Hamiltonian and Newtonian Mechanics

In classical mechanics, a system can be characterized by its classical Hamiltonian $\mathcal{H}$. A Hamiltonian $\mathcal{H}$ is a 'characteristic function' which depends on a set of canonical coordinates $\boldsymbol{r} = (\boldsymbol{q}, \boldsymbol{p})$, where $\boldsymbol{q} = (q_1, q_2, ..., q_{3N})$ are the coordinates and $\boldsymbol{p} = (p_1, p_2, ..., p_{3N})$ the momentum of all N particles in the system. In classical mechanical systems the Hamiltonian $\mathcal{H}$ is the energy of a system as a function of the phase space. The time evolution of a system is calculated from its Hamiltonian $\mathcal{H}$ by employing the Hamilton's equations

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}} \tag{5.5}$$

and

$$\frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}t} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}}. \tag{5.6}$$

In case of N particles with mass m in a non relativistic potential V the Hamiltonian is the sum of kinetic and potential energy

$$\mathcal{H} = \frac{\boldsymbol{p}^2}{2m} + V(\boldsymbol{q}). \tag{5.7}$$

With the force field potential description in (5.4), the Hamiltonian equations (5.5) and (5.6) for particle i are then calculated to

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial V}{\partial q_i} \tag{5.8}$$

and

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = \frac{p_i}{m}. \tag{5.9}$$

This results in Newtons equation of motion

$$m\frac{\mathrm{d}^2 q_i}{\mathrm{d}t^2} = F_i = -\frac{\partial V}{\partial q_i}, \tag{5.10}$$

where F_i is the force on particle i.

5.1.3 Numerical Time Integration

To compute the time evolution of a system defined by Hamiltonian $\mathcal{H}$, Newtons equations of motion (5.10) for all N particles in the system have to be solved. The integration is performed numerically in small steps as only for particle numbers $N \leq 2$ the set of N equations can be solved analytically. Different algorithms exist. Leap-frog integrator is a frequently used algorithm to integrate ordinary differential equations and is applied in all MD simulations in this thesis. The positions q_i and velocities $\dot{q}_i = v_i$ are updated at interleaved time points (therefore the name *leapfrog*). The algorithm is similar to velocity-Verlet method (based on the same principles) and sometimes these names are used interchangable. In leap-frog algorithm, the positions and velocities at the next time point $(t + \Delta t)$ are calculated as

$$q_i(t + \Delta t) = q_i(t) + \Delta t\, v_i(t) + \frac{\Delta t^2}{2\,m} F_i(t) \tag{5.11}$$

and

$$v_i(t + \Delta t) = v_i(t) + \frac{\Delta t}{2\,m}(F_i(t) + F_i(t + \Delta t)). \tag{5.12}$$

The Leap-frog algorithm is a symplectic integrator (conservation of phase space) and is therefor suitable for calculating the long-term evolution of chaotic Hamiltonian systems such as molecular dynamics Hamiltonian systems [Frenkel and Smit, 2001].

5.1.4 Constant Temperature and Pressure Simulations

Integration of Newtons equation of motion (5.10) results in constant energy trajectories: simulating N particles in volume V with conserved total energy E in a microcanocial ensemble NVE. However, in reality often not the energy but the temperature T is constant. Several algorithms (thermostats) exist for generating a canonical ensemble NVT. The most simple method is to re-scale all velocities v_i according to

$$v_i^{\text{new}} = \left(\frac{T_0}{T}\right)^{1/2} v_i^{\text{old}}, \tag{5.13}$$

with T_0 reference temperature and the actual temperature T. In its simple scheme, this allows no fluctuations and produces no true constant temperature ensemble. A more sophisicated version including a stochastic term allows generation of a true canonical distribution [Bussi et al., 2007]. In another frequently used thermostat the velocity is adjusted with a coupling constant or relaxation time τ_T:

$$v_i^{\text{new}} = v_i^{\text{old}} \left(1 + \frac{\Delta t}{\tau_T}\left(\frac{T_0}{T} - 1\right)\right)^{1/2}. \tag{5.14}$$

This coupling sheme is called Berendsen thermostat [Berendsen et al., 1984].

Depending on the biological system, a simulation under constant pressure p instead of constant volume may be more adequate. This can be achieved by a Berendsen barostat (similar to Berendsen thermostat), where the particle distances and hence also the simulation box volume is adjusted [Tuckerman, 2010]. This creates an isothermal-isobaric ensemble NPT.

5.1.5 Computational Approximations

Periodic Boundary Conditions

An important concept used in MD simulations are periodic boundary conditions (PBCs). In physiological conditions a molecule is surrounded by large layer of solvent molecules, e. g. water molecules and ions. In order to reduce

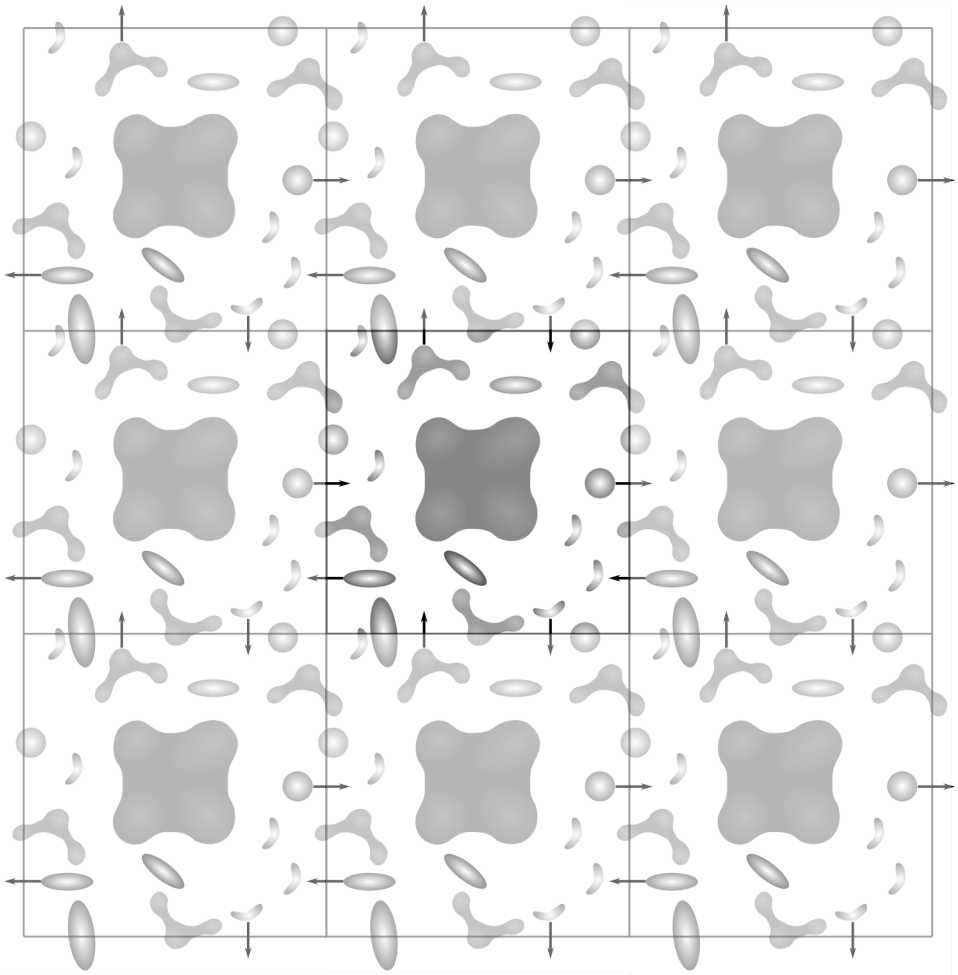

Fig. 5.2 Concept of periodic boundary conditions (PBCs) visualized in two dimensions. The central cell is periodically replicated infinitely in all directions (here: exemplary eight copies of the central cell are displayed). Particles leaving the cell into one direction enter the cell from the opposite direction. The central cell interacts with all its periodic images.

the computational effort but still simulate a large number of surrounding molecules PBC can be used. Furthermore, simulating within a finte box leads to artifacts as particles tend to stick at boundary surfaces [Frenkel et al., 1997]. The simulated system is repeated infinitely in all space directions. The simulation box can be viewed as a unit cell (UC) in a crystal lattice. A particle leaving the simulation box on one side re-enters the simulation box on the opposite side. The periodicity implies that all particles interact with each other (c. f. Fig. 5.2).

The concept of PBC introduced above reduces artifacts but leads to an infinite system. In order to still be able to numerically calculate interactions,

approximations are introduced. In particular, the non-bonded interactions which are computed pairwise for all particles result in a computational effort which scales quadratically with system size.

Cut off Radii for Lennard-Jones Potential

In order to efficiently calculate the Lennard-Jones potential, the van der Waals interactions are truncated at a cut-off radius. It has been shown that the short range Lennard-Jones interactions ($\frac{1}{r^6}$) decrease rapidly and can be truncated a cut-off radius of 9 to 12 Å. Usually the potential is shifted to zero at the cut-off radius to avoid discontinuities [Tuckerman, 2010].

Particle Mesh Ewald for Coulomb Potential

As electrostatic interactions are of long range nature, a simple cut-off as for the van der Waals interactions is not possible. For the treatment of Coulomb interactions, the Ewald summation method is applied: therefore, the interactions are methodologically separated in short range and long range interactions by introducing Gaussian-distributed charges to screen and compensate the point charges. The resulting short-range electrostatic contribution is then calculated with a cut-off approximation as described for the Lennard-Jones potential. The Gaussian-distributed screening charges interact periodically (as the simulation box is infinitely repeated in all directions over PBC) and can be efficiently calculated in Fourier space. The numerical effort can be further improved by introducing a mesh (or grid) on which the charges are interpolated. This allows the application of Fast Fourier Transformations. The method is referred to as particle mesh Ewald (PME) [Tuckerman, 2010].

5.1.6 Hamiltonian Replica Exchange

Hamiltonian replica exchange (H-RE) is a method frequently used to enhance sampling. H-RE can be applied when systems with slightly different Hamiltonians $\mathcal{H}_i$ are simulated in parallel (e. g. for alchemical systems as described later in section 5.3.2). At certain time intervals the configurations of two

systems $\mathcal{H}_1$ and $\mathcal{H}_2$ are allowed to exchange with a probability according to the Metroplis criterion

$$p = \begin{cases} e^{-\Delta U / k_B T}, \Delta U \geq 0 \\ 1, \Delta U < 0 \end{cases} \tag{5.15}$$

Therefore, the internal energy U of the configuration in the respective other system is calculated and the difference $\Delta U = \mathcal{H}_1 - \mathcal{H}_2$ is evaluated. An exchange is always performed when an exchange results in a lower overall energy. If an exchange results in an overall unfavorable energy change, the exchange is only accepted according to a Boltzmann weighted probability with Boltzmann constant k_B [Tuckerman, 2010].

5.2 Calculation of Thermodynamic Quantities

In MD simulations a sequence of points in phase space, called trajectory, is calculated. The phase space consists of all possible micro states available for a system. An ensemble contains only those micro states (points in phase space) that are consistent with given macroscopic properties (e. g. temperature, energy, volume). Hence, all points or micro sates in an ensemble are a subset of all possible phase space points or micro states. At thermodynamic equilibrium macroscopic observables do not change in time [Tuckerman, 2010].

In order to characterize an ensemble, a partition function is calculated. The partition function can be seen as a measure of the number of micro states in phase space which are accessible by an given ensemble. At thermodynamic equilibrium the partition function Q in a canonical ensemble is calculated as

$$Q = \frac{1}{h^{3N}} \int \int \exp\left(-\frac{1}{k_B T} \mathcal{H}(\boldsymbol{q}, \boldsymbol{p})\right) d\boldsymbol{q} d\boldsymbol{p}, \tag{5.16}$$

with Planck's constant h and Boltzmann factor k_B. The integral is performed over the complete phase space, i. e. all canonical coordinates $\boldsymbol{r} = (\boldsymbol{q}, \boldsymbol{p})$. With the partition function the expectation value, i. e. ensemble average, of any

observable X can be calculated

$$\langle X \rangle = \frac{1}{h^{3N}} \int Q(\boldsymbol{r}) X(\boldsymbol{r}) d\boldsymbol{r}. \tag{5.17}$$

A central assumption in MD simulations is the applicability of the ergodic hypothesis. It states that the time spent by a system in some region of the phase space resembles the ensemble probability of these microstates. Hence, the average of an observable over a statistical ensemble (5.17) can be calculated by a time average

$$\langle X \rangle = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau X(t) dt. \tag{5.18}$$

Thus, macroscopic properties can be computed as time average. However, this requires sufficiently long simulation times, commonly referred to as *convergence*.

5.3 Free Energy Calculation

5.3.1 Free Energy

The free energy of a system is a thermodynamical quantity, whereby systems evolve such that it is minimized. Depending on the thermodynamical ensemble different free energies are applicable. In the canonical ensemble, or NVT-ensemble (constant number of particles N, constant volume V, and constant temperature T), the Helmholtz free energy F is defined as

$$F = U - TS, \tag{5.19}$$

where U is the internal energy of the system and S the entropy. In the NPT-ensemble (constant number of particles N, constant pressure p, and constant temperature T) the Gibbs free energy G is calculated as

$$G = H - TS, \tag{5.20}$$

with the enthalpy H of the system. The enthalpy is defined as

$$H = U + pV. \tag{5.21}$$

In thermodynamic equilibrium the free energies G and H, respectively, depending on the ensemble, are at a minimum. In most biological systems which are simulated no major volume change is expected and the quantities do not depend on pressure or volume variation. Therefore, the differentiation between Gibbs G and Helmholtz F free energy is purely of theoretical nature. In statistical mechanics, the Helmholtz free energy F is calculated from the partition function Q of a system via

$$F = -\frac{1}{\beta} \ln Q, \tag{5.22}$$

where $\beta = 1/k_B T$ is calculated from the inverse Boltzmann factor k_B and temperature T. The (absolute) free Helmholtz energy, i. e. equation (5.22), can analytically only be calculated for very small systems.

It is not possible to measure the absolute free energy in a computer simulation as the free energy is directly related to the volume in phase space which is accessible to a system (c. f. Equation (5.22) and (5.16)). The analytic solution is only possible for easy systems, e. g. the ideal gas. However, it is possible to calculate the free energy difference between two states. There are two main methodological categories for free energy calculations (FECs): thermodynamic integration (TI) [Kirkwood, 1935] and free energy perturbation (FEP) [Zwanzig, 1954]. Both methods are alchemical methods and thus require construction of specific Hamiltonian systems, which is discussed in the following.

5.3.2 Hamiltonians for Alchemical Methods

Let $\mathcal{H}_A$ denote the Hamiltonian of state 'A' of a biological system, e. g. a protein wild type (WT), and $\mathcal{H}_B$ the Hamiltonian of state 'B' of a biological system, e. g. a protein variant. A combined Hamiltonian describing both

systems is constructed as

$$\mathcal{H}(\lambda) = (1 - \lambda)\mathcal{H}_{\mathrm{A}} + \lambda\mathcal{H}_{\mathrm{B}}, \text{ with } \lambda \in [0, 1], \tag{5.23}$$

where λ is a *coupling* parameter. For $\lambda = 0$, $\mathcal{H}$ describes state A and for $\lambda = 1$, $\mathcal{H}$ represents state B. The system described by $\mathcal{H}(\lambda)$ for all values of $\lambda \in]0, 1[$ does not necessary has to represent a biological system: e. g. in the case where state A and B are different variants of the protein, $\mathcal{H}(\lambda)$ with $\lambda \in]0, 1[$ represents an *alchemical* system.

Furthermore, N discrete systems described by $\mathcal{H}_{\mathrm{i}}$ between system A ($\mathcal{H}_{\mathrm{A}}$) and system B ($\mathcal{H}_{\mathrm{B}}$) can be constructed. Therefore, alchemical systems $\mathcal{H}(\lambda_{\mathrm{i}}) \equiv \mathcal{H}_{\mathrm{i}}$ with discrete values $\lambda_{\mathrm{i}} \in [0, 1]$ are designed.

5.3.3 Thermodynamic Integration

The free energy is not an average over phase space and cannot be directly determined from the simulation trajectories. However, the derivatives of the free energy can be measured. This method goes back to Kirkwood [1935]. In essence, the reversible work along a selected thermodynamic path is calculated. The energy difference between two states does not depend on the selected path. Thus, the advantage of a simulation is that there is no limitation to a physical thermodynamical path, but an arbitrary path linking initial and final state can be constructed.

By employing the partition function (5.16) and the Helmholtz free energy (5.22) one can show that the change of the free energy F with the alchemical coupling parameter λ can be related to the Hamiltonian $\mathcal{H}$ as

$$\frac{\partial F}{\partial \lambda} = -\frac{1}{\beta Q}\frac{\partial Q}{\partial \lambda} = \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle, \tag{5.24}$$

where $\langle\rangle$ denotes the ensemble average. An alchemical Hamiltonian $\mathcal{H}$ with a coupling parameter λ (c. f. equation (5.23)) as constructed in section 5.3.2 will be used. The Hamiltonian of system A is transformed to the Hamiltonian of system B via the coupling parameter λ. The average of the associated force on the system is integrated along λ:

$$\Delta F_{AB} = F(\lambda = 1) - F(\lambda = 0) = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda. \tag{5.25}$$

The important difference to free energy equation (5.22) is that in equation (5.25) the free energy difference is calculated via an ensemble average. It is important to note, that Equation (5.25) is still exact and no approximation was made to derive the energy difference. In order to calculate the integral numerically, a numerical integration scheme according to

$$\Delta F_{AB} = \sum_{i=1}^{K} W_i \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda_i}, \tag{5.26}$$

is applied, with W_i weighting factors depending on the numerical integration scheme used [Paliwal and Shirts, 2011]. Equation 5.26 can be evaluated numerically with K equilibrium simulations. The alchemical Hamiltonian $\mathcal{H}(\lambda)$ is discretized in K systems $\mathcal{H}_{1...\mathrm{K}}$. From the simulation of K alchemical systems $\mathcal{H}_\mathrm{i}$ the average $\langle \frac{\partial \mathcal{H}_\mathrm{i}}{\partial \lambda} \rangle_{\lambda_i}$ can be calculated.

5.3.4 Free Energy Perturbation

A second class of FEC are perturbation methods. One of them is FEP which was originally introduced by Zwanzig in 1954 [Zwanzig, 1954]. The difference between the free energy of system A (F_A) and the free energy of system B (F_B) is calculated by employing free energy equation (5.22) for two states A and B defined by Hamiltonian $\mathcal{H}_\mathrm{A}$ and $\mathcal{H}_\mathrm{B}$ and taking the difference

$$\Delta F_\mathrm{AB} = F_\mathrm{B} - F_\mathrm{A} = -\frac{1}{\beta} \ln \frac{Q_\mathrm{B}}{Q_\mathrm{A}}. \tag{5.27}$$

With partition function (5.16) and the expectation value of an observable (5.17) this results in

$$\Delta F_{AB} = -k_B T \ln \left\langle \exp\left(-\frac{\mathcal{H}_\mathrm{B} - \mathcal{H}_\mathrm{A}}{k_\mathrm{B} T} \right) \right\rangle_\mathrm{A}, \tag{5.28}$$

where $\langle\rangle$ denotes the ensemble average of state A. The ensemble is generated for one system and evaluated in both. The prerequisite for FEP is that the ensembles of both systems A and B overlap. Depending on the systems under investigation, state A, described by $\mathcal{H}_\mathrm{A}$, and state B, described by $\mathcal{H}_\mathrm{B}$, are physically so far apart, that the ensembles of state A and B generated via MD simulations exhibit no phase space overlap in a finite simulation time.

As the free energy is a state variable, a path between the free energy of system A (F_A) and system B (F_A) is constructed and the free energy difference along that path is summed up. N systems between system A ($\mathcal{H}_\mathrm{A}$) and system B ($\mathcal{H}_\mathrm{B}$) are introduced by constructing alchemical systems $\mathcal{H}(\lambda_\mathrm{i}) \equiv \mathcal{H}_\mathrm{i}$ with discrete values $\lambda_\mathrm{i} \in [0, 1]$. The corresponding free energy difference ΔF_{ij} between system i and system j is calculated through

$$\Delta F_{ij} = -k_B T \ln\left\langle \exp\left(-\frac{\mathcal{H}_\mathrm{j} - \mathcal{H}_\mathrm{i}}{k_\mathrm{B} T}\right)\right\rangle_\mathrm{i}. \tag{5.29}$$

Finally, the free energy difference $\Delta F_{A\to B}$ evaluates as the summation of free energy differences ΔF_{ij} between system i and system j, where the initial system $\mathcal{H}(\lambda_{\mathrm{i}=0} = 0)$ and the final system $\mathcal{H}(\lambda_{\mathrm{j}=\mathrm{N}} = 1)$ represents system A and B, respectively.

In practice, the systems $\mathcal{H}_\mathrm{i}$ and $\mathcal{H}_\mathrm{j}$ are chosen such that the ensemble of system i generated with MD simulations within a finite simulation time contains also configurations of state j, i. e. sampling of one state, also samples the other state. A resulting free energy difference ΔF_{ij} smaller than $k_\mathrm{B}T$ indicates such an overlap.

By construction, system A and B (and therefore system i and j) are interchangeable. Hence, the free energy difference can be calculated in forward and backward direction. The average of these forward and backward simulations is called double end FEP.

However, the FEP approach is assymmetric: in practice, the energy difference calculated via FEP depends on the direction of the transformation [Klimovich et al., 2015]. This is due to an undersampling of high energy states [Pohorille et al., 2010].

5.3.5 Bennett Acceptance Ratio

Another perturbation method was formulated by Bennett [1976]. The formulation known as Bennett acceptance ratio (BAR) states an equation for the free energy difference ΔG_{AB} between two systems A and B described by Hamiltonians $\mathcal{H}_{\mathrm{A}}$ and $\mathcal{H}_{\mathrm{B}}$ as

$$\exp\left(\frac{1}{k_B T}\left(\Delta G_{\mathrm{AB}} - C\right) + \ln(n_B/n_A)\right) = \frac{\left\langle f\left(\frac{1}{k_B T}\left(\mathcal{H}_{\mathrm{A}} - \mathcal{H}_{\mathrm{B}} + C\right)\right)\right\rangle_B}{\left\langle f\left(\frac{1}{k_B T}\left(\mathcal{H}_{\mathrm{B}} - \mathcal{H}_{\mathrm{A}} - C\right)\right)\right\rangle_A}, \quad (5.30)$$

with Fermi function $f(x) = 1/\big(1 + \exp(x)\big)$ and n_B and n_A the number of sampled configurations. Following Bennett [1976], equation 5.30 is numerically solved by finding variable C to satisfy the self consistency equation

$$\left\langle f\left(\frac{1}{k_B T}\left(\mathcal{H}_{\mathrm{A}} - \mathcal{H}_{\mathrm{B}} + C\right)\right)\right\rangle_B = \left\langle f\left(\frac{1}{k_B T}\left(\mathcal{H}_{\mathrm{B}} - \mathcal{H}_{\mathrm{A}} - C\right)\right)\right\rangle_A. \quad (5.31)$$

The free energy difference then calculates to

$$\Delta G_{\mathrm{AB}} = C - k_B T \ln(n_B/n_A). \quad (5.32)$$

As for FEP described before, the systems for which the free energy difference is calculated need to exhibit enough phase space overlap. This is again solved by calculating intermediate free energy changes ΔG_{ij} between all N alchemical systems $\mathcal{H}_{\mathrm{i}}$'s. Formally, all 'A's and 'B's in equations (5.30) to (5.32) are substituted by 'i's and 'j's indicating alchemical intermediate states described by Hamiltonians $\mathcal{H}_{\mathrm{i}}$ and $\mathcal{H}_{\mathrm{j}}$. The full free energy change ΔG_{AB} is the sum of all intermediate free energy changes ΔG_{ij}.

In contrast to FEP, BAR eliminates the bias of the transformation and inherently combines forward and backward transformation to minimize the statistical error and improve convergence.

5.3.6 Thermodynamic Cycle

In section 5.3 methods for calculating the free energy difference between two thermodynamic states were introduced. In order to quantify the effect of a mutation, a thermodynamic cycle between thermodynamic states can be constructed. This can be formulated in terms of protein crystallization for calculating the energy difference between crystallizing WT vs. mutant protein. A thermodynamic path from the unbound solvated crystal building blocks to the bound crystal state for WT and mutant can be constructed (see Fig. 5.3). Each corner represents a thermodynamic state. The states are connected by thermodynamic paths or 'legs' of a thermodynamic cycle: the horizontal paths represent the crystallization process of WT and mutant with associated free energy differences $\Delta G^{\mathrm{WT}}_{\mathrm{crystallization}}$and $\Delta G^{\mathrm{Mut}}_{\mathrm{crystallization}}$. The free energy difference

$$\Delta\Delta G \equiv \Delta\Delta G^{\mathrm{WT}\rightarrow\mathrm{Mut}}_{\mathrm{crystallization}} = \Delta G^{\mathrm{Mut}}_{\mathrm{crystallization}} - \Delta G^{\mathrm{WT}}_{\mathrm{crystallization}} \tag{5.33}$$

defines which process has a lower free energy minimum, which process is more likely to happen, and which end state (crystal) is more stable. In the definition of equation (5.33) a negative $\Delta\Delta G$ indicates a thermodynamic favor of mutant crystallization over WT crystallization. However, the individual free energy $\Delta G^{\mathrm{WT}}_{\mathrm{crystallization}}$and $\Delta G^{\mathrm{Mut}}_{\mathrm{crystallization}}$are not accessible in simulations as this involves the lengthy non-equilibrium process of crystallization. In a closed thermodynamic cycle the individual free energy differences sum up to zero

$$\Delta G^{\mathrm{Mut}}_{\mathrm{crystallization}} - \Delta G^{\mathrm{WT}}_{\mathrm{crystallization}} - \Delta G^{\mathrm{WT}\rightarrow\mathrm{Mut}}_{\mathrm{crystal}} + \Delta G^{\mathrm{WT}\rightarrow\mathrm{Mut}}_{\mathrm{solvated}} = 0. \tag{5.34}$$

The free energy changes upon mutation in both state $\Delta G^{\mathrm{WT}\rightarrow\mathrm{Mut}}_{\mathrm{crystal}}$ and $\Delta G^{\mathrm{WT}\rightarrow\mathrm{Mut}}_{\mathrm{solvated}}$ (vertical paths in Fig. 5.3) can be calculated by alchemically transforming the residues (transforming the force field parameters of the residues in the MD simulations) and monitoring the free energy change $\Delta G^{\mathrm{WT}\rightarrow\mathrm{Mut}}_{\mathrm{crystal}}$and $\Delta G^{\mathrm{WT}\rightarrow\mathrm{Mut}}_{\mathrm{solvated}}$. Equation (5.33) and (5.34) result in a formulation of $\Delta\Delta G$ which can be calculated by MD free energy simulations:

$$\Delta\Delta G = \Delta G^{\mathrm{WT}\rightarrow\mathrm{Mut}}_{\mathrm{crystal}} - \Delta G^{\mathrm{WT}\rightarrow\mathrm{Mut}}_{\mathrm{solvated}}. \tag{5.35}$$

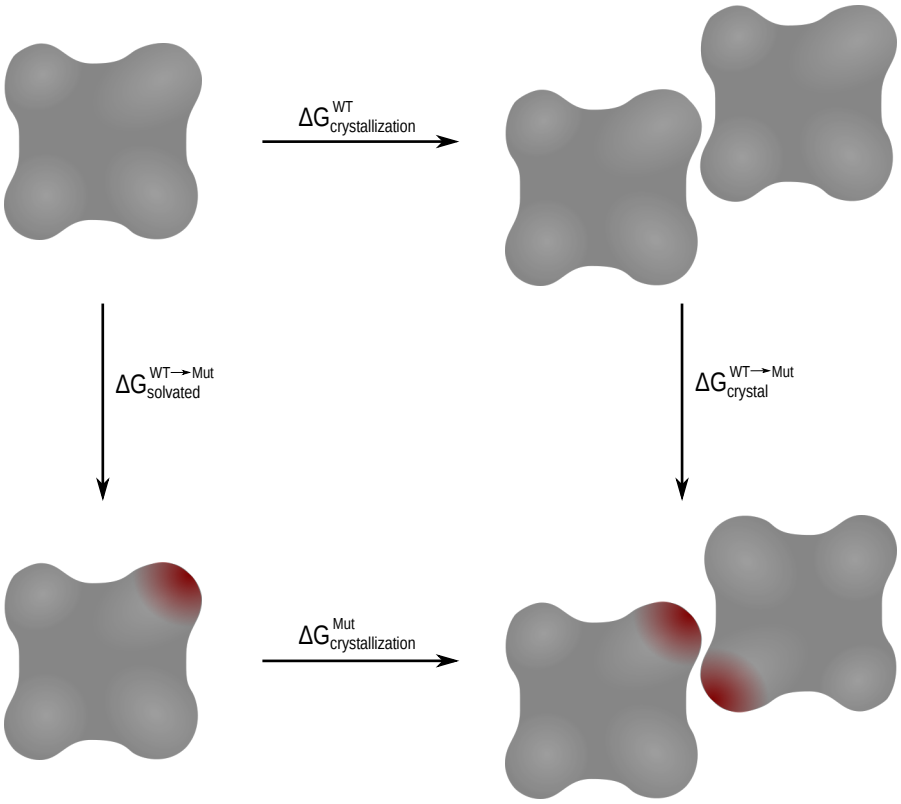

Fig. 5.3 Thermodynamic cycle for free energy difference $\Delta\Delta G$ calculation. The thermodynamic cycle is constructed for crystallization of WT and mutant protein. The thermodynamic states (corners of the cycle) are connected by thermodynamic paths or 'legs'. A negative free energy difference between mutant (Mut) $\Delta G^{\text{Mut}}_{\text{crystallization}}$ and WT $\Delta G^{\text{WT}}_{\text{crystallization}}$ crystallization ($\Delta\Delta G = \Delta G^{\text{Mut}}_{\text{crystallization}} - \Delta G^{\text{WT}}_{\text{crystallization}}$) denotes a thermodynamic favor of mutant over WT crystallization. As both values are not accessible by simulation, the numerically equivalent free energy difference $\Delta\Delta G = \Delta G^{\text{WT}\rightarrow\text{Mut}}_{\text{crystal}} - \Delta G^{\text{WT}\rightarrow\text{Mut}}_{\text{solvated}}$ is calculated. Both individual values are accessible in free energy simulations in which the solvated unbound WT is alchemically transformed into solvated mutant protein ($\Delta G^{\text{WT}\rightarrow\text{Mut}}_{\text{solvated}}$) and the crystal bound WT into crystal bound mutant protein ($\Delta G^{\text{WT}\rightarrow\text{Mut}}_{\text{crystal}}$).

5.4 Charge Corrections in Free Energy Calculations

As already introduced, PBC can be applied to mimick a biological system. While this provides many advantages as described in section 5.1.5, it introduces artifacts for electrostatic interactions due to their long range nature. Electrostatic interactions exceed far beyond the simulation box. The effective treatment of these interactions with lattice sum methods (section 5.1.5) introduces artifacts [Chen et al., 2018; Reif and Hünenberger, 2011; Reif and Oostenbrink, 2014; Rocklin et al., 2013]:

1. The simulated complex interacts with its periodically replicated images and with a neutralizing background charge which compensates non-zero charges in the simulation box.
2. The solvent around the complex under consideration is perturbed by the periodic images and therefore induces an undersolvation.
3. In an ideal non-periodic system the electrostatic potential is set to zero at infinity. This is not possible in a periodic system. When using lattice sum electrostatics, the zero is set by averaging over the computational box.

This spurious behavior results in artifacts when calculating free energy changes in charged systems with alchemical transformation methods. In systems without net-charge change, the spurious behavior is not dominant in the calculated free energy change as it occurs in both end states of the alchemical transformation. However, it is not reasonable to restrict oneself to investigate mutations without charge change. Therefore, corrections schemes are of need. In the following, two methodologically different correction schemes are presented to correct the effective approximate electrostatic interaction calculation for simulations with finite size and PBC. The first correction scheme is applied during the simulation whereas the second correction scheme is based on post-simulation numerical calculations.

5.4.1 Instantaneous Correction

The instantaneous correction scheme described here, adresses the net-charge of the system as main source for inaccuracies [Chen et al., 2018; Lin et al., 2014]. Keeping the net charge zero for both end points (no charge change) minimizes the finite size effects. Chen et al. [2018] proposed a co-alchemical ion approach. Here, the net charge of the system is kept zero by transforming an ion, i. e. the co-alchemical ion, simultaneously during the alchemical transformation. The ion can be in principal transformed into anything which allows a charge compensation for the alchemical transformation of the protein. Here, the dummy co-alchemical ion approach was chosen, where the ion is decharged but Lennard-Jones parameters are kept. The ion is transformed in a dummy particle with mass and Lennard-Jones parameters. In contrast to annealing the ion, this ensures no conflicts with singularities during annealing. The 'direction' of the charge change, e. g. $\Delta q = +e$ or $\Delta q = -e$, depends on the alchemical transformation of amino acids (mutation) in the protein. Therefore, the co-alchemical ion is either transformed (de-charged) into a dummy particle with no charge. Or a dummy particle with no charge is transformed (charged) into a co-alchemical ion. Chen et al. [2018] include the charging or de-charging free energy of the co-alchemical ion into the free energy change ΔG of one thermodynamic leg (c. f. section 5.3.6). This charging or de-charging free energy of a co-alchemical ion cancels as it occurs in both thermodynamic legs (c. f. equation (5.35)). Here a co-alchemical ion is simulated in the system with its appropriate charge, but it is excluded in the free energy calculation. In the MD engine GROMACS (GROningen Machine for Chemical Simulations), perturbated atoms are explicitly defined with a 'B-state'. The co-alchemical ion is treated as regular atom with varying charge over all discrete alchemical Hamiltonians $\mathcal{H}_\mathrm{i}$, but not with a 'B-state'. Hence, the charging or de-charging free energy of the co-alchemical ions do not enter in the calculated free energy change. This further reduces potential errors entering due to two charging free energy calculations. For charge changes of $\Delta q = \pm 1e$, a monovalent chloride ion (Cl^{-1}) was used as a co-alchemical ion; for charge changes of $\Delta q = \pm 2e$ a divalent magnesium ion (Mg^{+2}) was used.

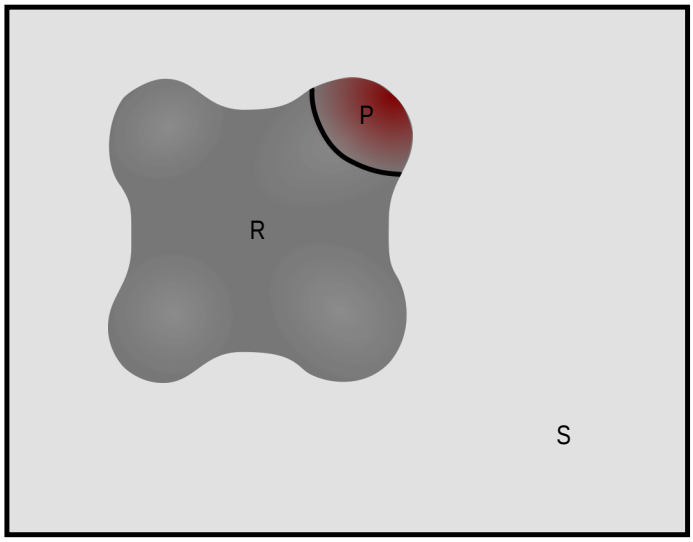

Fig. 5.4 Scheme for post simulation charge correction. The system is formally divided into perturbated (P), i. e. alchemcially transformed, atoms, and remaining (R) atoms which are surounded by solvent (S) atoms.

5.4.2 Numerical Post Simulation Correction

The presented numerical post simulation charge change correction was originally proposed by Reif et al. [2011; 2014] and is similar to analytic corrections proposed by Rockling et al. [2013]. Following Reif et al. the raw free energies ΔG_{raw} can be corrected by a free energy correction ΔG_{cor}

$$\Delta G = \Delta G_{\text{raw}} + \Delta G_{\text{cor}}. \tag{5.36}$$

The correction term ΔG_{cor} consists of individual correction terms correcting for spurious direct interactions ΔG_{dir}, artifacted undersolvation or solvent polarization ΔG_{pol}, and incorrect calculation of zero solvent potential ΔG_{dsm}

$$\Delta G_{\text{cor}} = \Delta G_{\text{pol}} + \Delta G_{\text{dir}} + \Delta G_{\text{dsm}}. \tag{5.37}$$

For the calculation of the individual correction terms, the whole alchemical system is formally divided into perturbed (P) atoms, remaining (R) atoms, and surrounding solvent (S) atoms (c. f. Fig. 5.4). Perturbated (P) atoms summarize all atoms which are alchemically transformed whereas remaining (R) atoms denote all atoms which are unchanged in both states.

Solvent Polarization

In periodic lattice sum schemes like PME the solvent around the charged complex under consideration is not only interacting with the complex itself but also with its periodic images which leads to undersolvation or underhydration [Kastenholz and Hünenberger, 2006; Reif and Hünenberger, 2011]. This is further influenced by a potential inaccurate dielectric permittivity of the solvent model used in the MD simulation.

The correction is calculated as the difference between charging free energy of the perturbed atoms under macroscopic conditions $\Delta G^{\text{macro}}_{\text{chg,P(S)}}$ (fully Coulomb interactions; no PBC) and charging free energy calculated under PBC with PME (lattice sum) treatment of electrostatic interactions $\Delta G^{\text{sim}}_{\text{chg,P(S)}}$

$$\Delta G_{\text{pol}} = \Delta G^{\text{macro}}_{\text{chg,P(S)}} - \Delta G^{\text{sim}}_{\text{chg,P(S)}}. \tag{5.38}$$

Each free energy change can be computed from electrostatic potentials Φ which are calculated from continuum-electrostatic calculations (Poisson-Boltzmann (PB)) at the given charge state defined by λ

$$\Delta G^{\text{env}}_{\text{chg,P(S)}} = \sum_{i=1}^{N_{\text{P}}} \Delta q_i \int_0^1 d\lambda \, \Phi(\boldsymbol{r_i}; \lambda; \text{env}; \epsilon_{\text{sol}}) - \Phi(\boldsymbol{r_i}; \lambda; \text{env}; \epsilon_0), \tag{5.39}$$

where env $\in$ (macro, sim), $i = 1, ..., N_{\text{P}}$ are the charge perturbed atoms, and $\Delta q_i = q_i^{\lambda=1} - q_i^{\lambda=0}$ the total partial charge change of atom i. The solvent dielectric permittivity ϵ_{sol} has to be adjusted to the environment: for calculating ΔG^{sim}, ϵ_{sol} is set to the dielectric perimittivity of the water model used. For TIP3P water model [Jorgensen et al., 1983] ϵ_{sol} is set to 82.0 [Van der Spoel et al., 1998]. For calculating $\Delta G^{\text{macro}}_{\text{chg,P(S)}}$, ϵ_{sol} is set to 78.4 accounting for (real) water. The electrostatic potential shows linear dependence on the charge state defined by λ. The integration in equation (5.39) can therefore be

performed by applying the trapezoidal rule

$$\Delta G^{\text{env}}_{\text{chg,P(S)}} = \sum_{i=1}^{N_{\text{P}}} \frac{1}{2}\Delta q_i \left[\Phi(\boldsymbol{r_i}; \lambda = 0; \text{env}; \epsilon_{\text{sol}}) + \Phi(\boldsymbol{r_i}; \lambda = 1; \text{env}; \epsilon_{\text{sol}})\right] - \frac{1}{2}\Delta q_i \left[\Phi(\boldsymbol{r_i}; \lambda = 0; \text{env}; \epsilon_0) + \Phi(\boldsymbol{r_i}; \lambda = 1; \text{env}; \epsilon_0)\right]. \tag{5.40}$$

The potentials are computed with the program *dGslv_ pbsolv* included in the simulation package GROMOS (GROningen MOlecular Simulation) [Christen et al., 2005]. This PB solver was recently edited to be used with other file formats than GROMOS files [Öhlknecht et al., 2020]. *dGslv_ pbsolv* now allows to take 'pqr'-files which can be generated with GROMACS.

Direct Non-solvent Interactions

When using PME to treat the electrostatic interactions, the perturbed atoms interact with the periodic images. This is similar to the spurious interaction of the solvent with the periodic images of the perturbed atoms. The correction term for direct self interactions can be calculated from two electrostatic free energies as

$$\Delta G_{\text{dir}} = \Delta G^{\text{macro}}_{\text{chg,P(R)}} - \Delta G^{\text{sim}}_{\text{chg,P(R)}}. \tag{5.41}$$

This calculates the difference between macroscopic ($\Delta G^{\text{macro}}_{\text{chg,P(R)}}$) and simulation ($\Delta G^{\text{sim}}_{\text{chg,P(R)}}$) environment for charging of perturbed atoms (P) due to the remaining atoms (R). Therefore, the Hamiltonian of the system $\mathcal{H}$ (only including electrostatic interactions in vacuum) is formally separated into a Hamiltonian for perturbed atoms $\mathcal{H}^{\text{P}}$ and a Hamiltonian for both groups, perturbed and remaining atoms, $\mathcal{H}^{\text{PR}}$ (c. f. Fig. 5.4). The system is constructed such that for $\lambda = 0$ the perturbed atoms (P) are uncharged and therefore electrostatically non-interacting, but only the remaining atoms (R) are. For $\lambda = 1$ all atoms, perturbed (P) and remaining (R), are interacting. This allows to separate the alchemical Hamiltonian as

$$\mathcal{H} = (1 - \lambda)\mathcal{H}^{\text{R}} + \lambda\mathcal{H}^{\text{PR}}. \tag{5.42}$$

As demonstrated in section 5.3.3 the free energy can be calculated by the partial derivative of $\mathcal{H}$ with respect to λ and integrating over λ (see equation (5.25)). Therefore, the partial derivative is calculated to

$$\frac{\partial \mathcal{H}}{\partial \lambda} = \mathcal{H}^{\mathrm{PR}} - \mathcal{H}^{\mathrm{R}} = E_{\mathrm{ele}}^{\mathrm{PR}} - E_{\mathrm{ele}}^{\mathrm{R}}. \tag{5.43}$$

By construction of $\mathcal{H}$, $E_{\mathrm{ele}}^{\mathrm{PR}}$ and $E_{\mathrm{ele}}^{\mathrm{R}}$ are the electrostatic interactions of the atoms PR and R, respectively. Employing free energy calculation equation (5.22), yields the free energy correction

$$\Delta G_{\mathrm{chg,P(R)}}^{\mathrm{env}} = \int_0^1 \left(\left\langle E_{\mathrm{ele}}^{\mathrm{PR}} \right\rangle - \left\langle E_{\mathrm{ele}}^{\mathrm{R}} \right\rangle \right) d\lambda. \tag{5.44}$$

Potential from Discrete Solvent Molecules

In macroscopic systems the zero of the solvent generated potential is set at infinity. This is not possible in periodic systems. To resemble this situation in a simulation with finite box size, the zero has to be set 'outside' the simulation box. To correct for the incorrect summation implied by lattice summation method PME a correction term is calculated to

$$\Delta G_{\mathrm{dsm}} = -N_{\mathrm{A}}(6\epsilon_0)^{-1}\gamma_{\mathrm{S}}\,\Delta q\,N_{\mathrm{S}}V_{\mathrm{B}}^{-1}, \tag{5.45}$$

where γ_{S} quadrupole moment trace of the solvent, N_{A} Avogadro constant, ϵ_0 vacuum dielectric permittivity, N_{S} number of solvent molecules, V_{B} box volume, and Δq net-charge change in the system. The γ_{S} quadrupole moment trace of TIP3P water model is set to $\gamma_{\mathrm{S}} = 0.007641\ e\,\mathrm{nm}^2$ [Jorgensen et al., 1983].

Practical Guidelines

The free energy correction for the potential from solvent molecules ΔG_{dsm} can be calculated based on simulation systems details. The correction for discrete solvent molecules ΔG_{dsm} does not depend on time. Hence, for ΔG_{dsm} one calculation for one system is required. The correction for spurious solvation

ΔG_{pol} and direct interactions ΔG_{dir} are time dependent, i. e. configuration dependent. However, Reif et al. [2014] and Oehlknecht et al. [2020] showed that, while the individual terms ΔG_{pol} and ΔG_{dir} are configuration dependent (time dependent), their sum ($\Delta G_{\text{pol}} + \Delta G_{\text{dir}}$) calculated for the same configuration does not vary much. Hence, the total correction ΔG_{cor} can be calculated from minimum set configurations when for both calculation of spurious undersolvation ΔG_{pol} and direct interactions ΔG_{dir} the same configuration is used.

Chapter 6

Methods

6.1 Crystallographic Details

6.1.1 Neutron Crystal Structure

Before conducting neutron diffraction experiments on crystals of alcohol dehydrogenase from *Lactobacillus brevis* (*Lb*ADH), these crystals had to be prepared: in order to achieve hydrogen/deuterium exchange of *Lb*ADH crystals, the reservoir solution in the microbatch crystallization plates (MRC UnderOil Crystallization Plate, SWISSCI, Neuheim, Switzerland) was replaced four times with D_2O precipitation buffer (1 mM Tris-HCl pD 7.5, 50 mM $MgCl_2$, 273 mM polyethylene glycol (PEG) 550 MME) over 14 days at equal intervals which was handled by Phillip Grob (Institute of Biochemical Engineering, Technical University of Munich (TUM), Garching, Germany). In previous diffraction experiments, six *Lb*ADH crystals with volumes of up to 0.8 mm^3 were tested under cryocooled conditions at T = 100 K. The crystals were cryoprotected with reservoir solution containing 50% (v/w) deuterated glycerol before cooling in a cryostream. The crystals diffracted up to 3.6 Å resolution. In a second experiment, another five crystals with volumes of up to 0.7 mm^3 were tested at room temperature. All of the latter crystals showed diffraction to higher resolution compared with the cryocooled conditions. As room temperature diffraction data explores the crystal content of a crys-

tal with the least possible disturbance (and cryoprotection even decreased the diffraction behavior) diffraction experiments for data collection were performed at room temperature. Prior to the room temperature diffraction experiments, the crystals were mounted in home-made Teflon cryoloops (thread diameter 0.1 mm) on stainless-steel pins (catalogue No. MD-7-410, 18 mm; Molecular Dimensions, Suffolk, England), which were glued into the goniometer base (type GB-B3S; MiTeGen, Ithaca, New York, USA) and inserted into a quartz capillary (catalogue No. Z567361-5EA; Sigma–Aldrich, St Louis, USA). A small drop of reservoir solution was placed at the other end of the capillary to avoid drying of the crystal. The capillary was sealed with epoxy glue. Vacuum grease around the copper shaft of the goniometer base helped to prevent gas emanating from the epoxy glue from entering the inner space of the capillary. The preparation procedure was performed at room temperature by Dr. Tobias Schrader (Jülich Centre for Neutron Science). Room-temperature neutron data were obtained with the BIODIFF instrument [Ostermann and Schrader, 2015] at the Heinz Maier-Leibnitz Centre (MLZ), Garching, Germany.

For the neutron data set presented in this study, a crystal with approximate dimensions of $1.4 \times 1.2 \times 0.4$ mm was used. A neutron data set was collected at a wavelength of 2.7 Å at room temperature within 16 days. Together with Dr. Tobias Schrader a strategy was designed for collecting data in the previously determined space group (SG) of the crystal *I*222. Data collection was performed over a total angular range of 82.5°, with a rotation range of 0.3° per frame and an exposure time of 1.5 h. The data were integrated in SG *I*222, with unit cell (UC) parameters a = 56.5 Å, b = 84.6 Å, c = 115.4 Å, at a resolution of 2.15 Å. Data reduction was performed using HKL-2000 v.705b [Otwinowski and Minor, 1997]. PHENIX [Adams et al., 2010, 2011] software suite was used for structure solution and refinement of the neutron structure. The published ligand-free structure of *Lb*ADH collected at 283 K (PDB entry 1nxq [Niefind et al., 2003]) served as a search model for molecular replacement with Phaser [McCoy et al., 2007]. phenix.refine [Afonine et al., 2012] was used for refinement. Manual model building was performed using Coot [Emsley et al., 2010] employing nuclear maps ($2F_o$ - F_c, F_o - F_c). The final structure was validated with PDB-REDO [Joosten et al., 2014]. Structure solution and refinement statistics are given in Appendix A in Table A.1.

6.1.2 X-ray Crystal Structures

For X-ray data collection of the crystal previously measured in neutron diffraction experiment, the epoxy glue between the copper shaft of the goniometer and the quartz capillary was removed with a cutter and the crystal was detached from the Teflon cryoloop. The ethylene glycol concentration was slowly increased stepwise to 30% (w/v) in 30 min which allowed this large *Lb*ADH crystal to be cooled while retaining its diffraction ability (d_{min} = 1.48 Å). The treatment was handled by Dr. Sabine Schneider (Chair of Biochemistry, TUM, Garching, Germany).

All other crystals could be cryoprotected within one step, by soaking the crystal for two seconds in the crystallization buffer, which was supplemented with 25 - 30% (v/v) ethylene glycol. Afterwards the crystal was mounted on nylon fiber loops and flash-cooled to 100 K in liquid nitrogen. Diffraction data were collected at 100 K at Swiss Light Source (SLS) on beamlines PXI X06SA and SLS PXIII X06DA, and at Synchrotron Radiation Facility (ESRF). Crystal preparation and data collection was performed by Dr. Sabine Schneider or Dr. Robert Janowski (Institute of Structural Biology, Helmholtz-Zentrum München, Neuherberg, Germany).

The diffraction data reduction was indexed and integrated with XDS [Kabsch, 2010] and scaled with SCALE [Evans, 2006]. Model building and refinement was performed in software suite CCP4 [Collaborative et al., 1994]. For WT *Lb*ADH previously published PDB entry 1nxq [Niefind et al., 2003]) served as a search model. For all subsequent crystal structures *in silico* mutated wild type (WT) served as search model in molecular replacement with Phaser [McCoy et al., 2007]. Refinement and manual model building was performed with REFMAC [Murshudov et al., 1997] and COOT [Emsley et al., 2010] (employing electron density maps ($2F_o$ - F_c, F_o - F_c)). Diffraction data, structure solution and refinement statistics for all published structures are given Appendix A in Tables A.1, A.2, and A.3.

6.2 Structural Analysis Details

6.2.1 Root Mean Square Deviation (RMSD) Calculation

In order to investigate the proteins structural similarity, root-mean-square deviation (RMSD) calculations were used. In an RMSD calculation, structures are superimposed and the average distance between atomic positions are calculated. Frequently, proteins are aligned on the backbone C_α atoms of another. The RMSD is calculated for the structural displacement of the C_α atoms, so-called C_α RMSD. Given two sets of n atoms with atomic positions $\mathbf{v} = (\mathbf{v}_1, ..., \mathbf{v}_n)$ and $\mathbf{w} = (\mathbf{w}_1, ..., \mathbf{w}_n)$ the RMSD calculates to

$$\begin{aligned} \text{RMSD} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} ||\mathbf{v}_i - \mathbf{w}_i||} \\ &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left[(v_{i,x} - w_{i,x})^2 + (v_{i,y} - w_{i,y})^2 + (v_{i,z} - w_{i,z})^2\right]} \end{aligned} \tag{6.1}$$

Monomer or tetrameter C_α RMSD. The alignment of the monomers or tetrameters before the RMSD calculation is performed with a least-square fit. The structures are translated and rotated to minimize a scoring function, i. e. the least-square. Here, the method of Kabsch [1978] implemented in VMD (visual molecular dynamics) [Humphrey et al., 1996] was used.

UC C_α RMSD. The structural comparison of two crystals can be performed based on their UCs. The crystallographic UC is the smallest translative unit of a crystal. The UC is not invariant under rotation, but only under translation. Therefore, to compare the content of a crystallographic UC, the superposition has to be performed without rotation. This equals a center of geometry (COG) alignment of C_α atoms.

Naturally, the RMSD depends on the number of compared atomic positions, i. e. in the case of a C_α RMSD number of residues. Hence, the monomer, tetramer, and UC RMSDs can be compared only within each other but not to each other.

6.2.2 Resolution Limit in RMSD Calculations

A prerequisite for structural comparison via RMSD calculation is the reliability of the structure models. A commonly used measure is the resolution limit. Models can only be as accurate as the diffraction data from which they are constructed. I. e. models may differ simply due to inaccuracies associated with resolution limits in the diffraction experiment. It has been shown that RMSDs calculations depend on resolution limits [Carugo, 2003]:

1. The RMSD of two structure models of the identical protein may differ simply because different resolution limits in the diffraction experiments resulted in slightly different models. The RMSD depends on the gap of the resolutions $\Delta d_{min} = d_{min,\ structure\ 1} - d_{min,\ structure\ 2}$.
2. Two structure models of the identical protein with data sets with the same resolution limits, may exhibit a finite RMSD. The RMSD depends on the resolution of the structures $d_{min} = d_{min,\ structure\ 1} = d_{min,\ structure\ 2}$.

Carguo et al. [2003] studied the resolution influence on the RMSD calculation for structures with resolution limits of d_{min} = 1.6 to 2.9 Å and constructed empirical formulas for both of the above cases. With the definition of the normalized RMSD [Carugo and Pongor, 2001] the results can be extrapolated to *Lb*ADH structure models: e. g. structure models of identical *Lb*ADH variants resolved with a resolution of d_{min} = 1.8 Å may exhibit C_α RMSD differences of: 0.20 Å for the monomer, 0.29 Å for the tetramer, and 0.33 Å for the UC. Furthermore, structure models of identical *Lb*ADH variants resolved with a different resolutions of Δd_{min} = 0.3 Å (e. g. $d_{min,\ structure\ 1}$ = 1.6 Å and $d_{min,\ structure\ 2}$ = 1.9 Å) may exhibit C_α RMSD differences of: 0.42 Å for the monomer, 0.62 Å for the tetramer, and 0.72 Å for the UC.

X-ray Structures

The generated X-ray structure models in this study exhibited high resolution limits of d_{min} = 1.09 to 1.80 Å. Hence, the formulas cannot be applied directly, but represent a good estimate [Carugo, 2003] for interpreting the calculated RMSDs: defined conservatively, structures are considered to be identical or undistinguishable within the resolution with the following criterions:

- Monomer C_α RMSD < 0.20 Å
- Tetramer C_α RMSD < 0.29 Å
- UC C_α RMSD < 0.33 Å

Neutron and X-ray Structure of Wild Type *Lb*ADH

The resolution limit obtained in the neutron diffraction experiment was considerably lower. In case of the identical crystal, a resolution limit of d_{min} = 2.15 Å and d_{min} = 1.48 Å was detected in neutron and X-ray diffraction experiments, respectively. It is worth noting that for X-ray diffraction, the detector size and not the crystal diffraction ability was limiting the resolution. The lower resolution limit in neutron diffraction in case of this perfectly diffracting crystal was exclusively due to the naturally considerably weaker flux of the neutron beam at BIODIFF compared to X-ray beam in synchrotrons. Applying the correlation found by Carguo et al. [2003] for Δd_{min} = 0.67 Å yields RMSDs criterions:

- Monomer C_α RMSD < 0.59 Å
- Tetramer C_α RMSD < 0.86 Å
- UC C_α RMSD < 1.00 Å

As the resolution limit of the considered X-ray structure is higher as the considered resolution range in the study of Carguo et al. [2003], these limits should be considered as upper limits.

6.2.3 Crystal Mosaicity

The mosaicity α of a crystal is a measure for the imperfections of a crystal. I. e. how well the crystallites within a crystal are arranged with respect to each other (see section 4.4). Hence, the mosaicity α of a crystal reflects how well the properties concerning the arrangement within the whole crystal can be deduced from a crystallographic UC alone. All X-ray diffraction data exhibited a low mosaicity of α = 0.09 to 0.55° indicating a well ordered crystal (see Table 7.1). The mosaicity in the neutron data was considerably higher (α = 0.99°). However, the reported mosaicity in the neutron diffraction

data, does not resemble the intrinsic mosaicity of the crystal, as in the neutron diffraction experiment the beam divergence (0.6 - 0.7° full width at half maximum (FWHM)) and the wavelength distribution ($\Delta\lambda/\lambda \approx 2.5\%$) dominate the intensity peak widening. Only in the case of diffraction sources of high brilliance like synchrotrons, the detected mosaicity in the diffraction data allows a direct correlation to the intrinsic mosaicity of the crystal.

6.3 Simulation Details

6.3.1 System Construction for Atomistic MD Simulations

Crystal Bound State

For *Lb*ADH protein variants for which a crystal structure existed (T102E, D54F, Q126H, Q126K, K32A, Q207D) crystal systems for both, WT and mutant, were constructed, so that the mutation could be investigated in both crystal systems. For the variants H39A, K45A, and D54A no diffraction data could be obtained. For these variants the mutation was performed only in the WT crystal system. Two representations for the crystal bound state were used: (a) One crystallographic UC with periodic boundary condition (PBC). This bound state set-up was constructed for mutation Q126H in both WT and mutant crystal packing geometries, and additionally for all three WT crystals (*I*222 crystal determined with neutron diffraction, *I*222 and $P2_122_1$ crystals determined with X-ray diffraction) (b) The isolated crystal contact, i. e. reduced to the monomers of the engineered crystal contact (for all mutations). The composition of the UC with PBC set-up (bound version (a)) and the reduced set-up (bound version (b)) are reported in Appendix B Tables B.3 and B.4. The simulation box sizes for the UC set-up (bound version (a)) correspond to the respective UC box vectors (see Table 7.1). For the reduced set-up (bound version (b)) the monomers of the isolated crystal contact were placed in a simulation box leaving at least 30 Å to the next periodic image. This resulted in box sizes ranging from $97.440 \times 142.460 \times 81.86$ Å^3 (in cases where the crystal contact could be reduced to two monomers) to $114.01 \times 125.49 \times 100.88$ Å^3 (in cases where the crystal contact was reduced

to three monomers). The box dimensions for all set-ups are reported in Appendix B Tables B.3 and B.4.

Unbound Solvated State

The solvated unbound state is represented (a) by one tetramer (for mutation Q126H and Q207D) and (b) by one or two monomers (depending on the position of the mutation) placed in a cubic box with PBC leaving at least 30 Å to the next periodic image (for all mutations). This results in a simulation box sizes of $88.00 \times 98.00 \times 93.00$ Å^3 for unbound version (a) and $90.56 \times 90.06 \times 82.02$ Å^3 and $93.72 \times 114.41 \times 76.30$ Å^3 for unbound version (b) in case of one or two monomers, respectively. The details for all unbound set-ups are reported in Appendix B Table B.5.

Preparation and Parametrization

GROMACS (GROningen Machine for Chemical Simulations) tool pdb2gmx was used to protonate the protein at pH 7.0 (resulting in a net charge of -5 e for WT *Lb*ADH) and to parametrize the protein in CHARMM36 [Vanommeslaeghe et al., 2010] with the refined parameters CHARMM36m [Huang et al., 2017]. All simulation boxes were solvated with TIP3P water model [Jorgensen et al., 1983] and neutralized with 25.5 mM $MgCl_2$.

6.3.2 MD System Equilibration

The molecular dynamics (MD) simulations were carried out with the MD engine GROMACS v.2018.6 [Abraham et al., 2015; Berendsen et al., 1995]. The simulation boxes described above were equilibrated and prepared according to the following protocol:

EM, Annealing, NVT

First, energy minimization (EM) via steepest descent with convergence criterion of 8500 steps or 100 kJ mol^{-1} nm^{-1} was performed, followed by 0.9 ns of annealing with a final temperature of 293 K and 1 ns of constant

volume simulation (NVT). A leap-frog integrator (MD) with time step of 1 fs was used and the temperature was controlled with velocity rescaling [Bussi et al., 2007]. Position restraints on backbone atoms with a force constant of 1000 kJ mol^{-1} nm^{-2} were applied. The bond lengths of hydrogen atoms were constrained with Linear Constraint Solver [Hess et al., 1997]. A highest order in the expansion of the constraint coupling matrix of 12 was used. The electrostatic interactions were calculated with particle mesh Ewald (PME) method [Darden et al., 1993; Essmann et al., 1995] using an PME order of 6 and a fourierspacing of 0.12 nm. For the van der Waals interaction a cut-off distance of 1.2 nm was used with a smooth potential switch starting from 1.0 nm. For neighbor search the verlet scheme [Páll and Hess, 2013] was used.

Determining the Box Size / Solvent Ratio

Special care had to be taken when determining the right box size for the simulations. The simulations were chosen to be carried out in the NVT ensemble. Therefore, the box volume (or respectively the number of solvent molecules) had to be determined to yield a physiological environment. The NVT ensemble was chosen because of three main reasons: (1) Constant volume simulations represented a realistic, physiological scenario for a crystal as no major volume changes were expected. (2) In the simulation of the crystal state represented by one UC with PBC (version (a) of the bound state) the proteins protrude over the simulation box boundaries in all directions. In constant pressure (NPT) pre-simulations, it was found that this introduces artifacts due to the coordinate re-scaling in barostats. Crystal contacts tended to be lost and simulation instabilities occurred in pre-simulations as pressure coupling is not designed for 'crowded' simulation boxes when PBC are applied for 'physiological' reasons in contrast to 'normal' use of extending the solvent area and avoiding finite size boundary artifacts. (3) Due to the application of position restraints, the pressure calculation in the simulated system is not well defined: the equilibrium atom positions in the simulation and the measured structure may differ slightly due to minor different equilibrium parameters in force field and due to different temperature of structure determination and simulation (the simulations were carried out at room temperature and the structure was collected at

cryocooled temperature). The restraining force on each atom was small, but when restraining 2008 C_α atoms, the forces accumulate. This introduces an 'internal' pressure of the protein as the pressure is calculated from the virial where also the restraining forces are included. Hence, to avoid artifacted pressure calculation due to position restraints, short unconstrained isothermal-isobaric ensemble (NPT) test simulations (2 to 10 ns) with isotropic Berendsen barostat set to 1.01 bar and a coupling constant of $\tau = 1.0$ were conducted to adjust the box size or number of solvent molecules.

Equilibration Run

The simulation boxes of the previously NVT equilibrated structures were adjusted according to the test simulations and re-equilibrated with a short NVT run. The time step was increased to 2 fs. The position restraints were decreased to C_α atoms and a force constant of 750 kJ mol^{-1} nm^{-2} and the highest order in the expansion of the constraint coupling matrix was decreased to 4. This equilibration run was carried out until the backbone RMSD was constant. Typically, this took 200 ns for a UC simulation (version (a) of the bound state), 75 ns for the isolated crystal contact (version (b) of the bound state), 50 ns for an unbound tetramer (version (a) of the unbound state), and 25 ns for an unbound monomer (version (b) of the unbound state).

Starting Structure Extraction

In all set-ups the last frame was extracted as a starting structure for the free energy perturbation (FEP) simulations and crystal state simulations. Additionally, for free energy calculation (FEC) for variant Q126H the equilibration run of the UC set-up (version (a) of the bound state) was elongated for another 35 ns so that three random starting structures (structures at time points: 200 ns, 220 ns, 235 ns) could be extracted from the trajectory, i. e. system phase space).

6.3.3 Crystal State Simulations

Production Run

For all three determined WT structures, the crystal states (version (a) of the bound state) were simulated. The extracted structures were used as starting structures. Each of the three crystal sates were simulated for 100 ns with the same parameters as for the equilibration run.

Analysis

The generated trajectories were analyzes with python scripts. For hydrogen bond (H-bond) determination a distance smaller than 2.5 Å was used. Error estimate was performed by block averaging over five blocks. For water counts (hydration shell) at the crystal contacts, spheres were constructed between the interfacing interaction areas and a count was performed over all frames.

6.3.4 MD Free Energy Calculations

Alchemical Setup

In order to generate a hybrid protein structures and topologies for FEP simulation from the equilibrated starting structures the GROMACS compatible programm PMX [Gapsys and de Groot, 2017; Gapsys et al., 2015] was used. For the UC bound state (version (a) of the bound) 60 equally spaced alchemical transition windows were constructed. For all the rest of the setups 30 transition windows were employed.

EM, Annealing, NVT, Equilibration Run

For each of the windows EM, Annealing, NVT and equilibration run were conducted as described above with a reduced annealing time of 0.3 ns and a equilibration run time of 5 ns. Soft core potentials were used for the Lennard Jones and Coulomb interactions with soft core parameter alpha of 0.3, a power of 6 for the radial term, and a soft core sigma for particles of 0.25.

Beginning from the equilibration run, hamiltonian replica exchange (H-RE) with an exchange attempt between neighboring transition windows every 1000 steps resulting in transition probabilities of 25 to 40% was performed.

Production Run

The production run of 20 ns was performed with the same parameters as for the equilibration run and the free energy difference of every window to its two neighboring window was calculated during the simulation every 100 steps, i. e. every 0.2 ps.

Analysis

The free energy difference was estimated via the Bennett acceptance ratio (BAR) [Bennett, 1976] implemented in GROMACS and divided by the number of mutations in the respective setup. Error estimate was calculated with GROMACS BAR (block averaging over five blocks).

Charge Change Corrections

As the electrostatic interactions were treated with PME, a periodic lattice-sum method, the finite box with PBCs introduces finite-size effects which impacted the FEC in cases where the mutation involves a charge change. Two methodological different charge correction schemes were applied.

Co-alchemical counterion. Firstly, the co-alchemical counter ion approach was applied which corrects the dominant error: the net charge change [Chen et al., 2018]. The net charge is kept constant by charging or de-charging a divalent magnesium ion (Mg^{2+}) or a chlorid ion (Cl^{-}) during the alchemical transformation while keeping their Van der Waals parameters. For net charge changes of $\Delta q = 2e$ (isolated crystal contact of D54F, K32A, Q207D) Mg^{2+} and for changes of $\Delta q = e$ (isolated crystal contact of Q126H, Q126K, T102E, K45A, H39A) Cl^{-} was used as alchemical counter ion, respectively. The charging free energies of alchemical counter ions were not included in the free energy calculations (see section 5.4.1).

Numerical post simulation charge correction. Secondly, a post simulation charge change correction orginally proposed by Reif et al. [Reif and Hünenberger, 2011; Reif and Oostenbrink, 2014] was applied (see section 5.4.2). The associated GROMOS (GROningen MOlecular Simulation) Poisson-Boltzmann (PB) solver was recently edited to be used with other file formats than GROMOS files [Öhlknecht et al., 2020]. This approach was applied for mutations Q126K and K32A which showed the smallest free energy change $\Delta\Delta G$. PB solver grid spacings were tested for Q126K bound transformation crystal contact 1. Grid spacings of 0.016 nm, 0.018 nm, 0.020 nm, 0.022 nm, 0.024 nm were used. The results of the grid spacing testing (Appendix C Table C.1) yielded that a spacing of 0.022 nm can be used without loss of information. This spacing was used for all further PB solver calculations. Five snapshots for each end states were extracted in equidistant time intervals from the trajectory for analysis. Free energy correction for spurious direct interactions ΔG_{dir} and spurious solvent polarization ΔG_{pol} were calculated based on the same snapshots as suggested by Öhlknecht et al. [2020].

6.4 Hardware Details

The calculations were carried out on local workstations and high performance supercomputers. The supercomputers 'SuperMuc' and 'SuperMuc-NG' were used which were operated by the Gauss Centre for Supercomputing/Leibnitz Supercomputing Centre, Garching, Germany. 'SuperMuc' was used until the end of 2018 when it retired for its successor 'SuperMuc-NG'. 'SuperMuc' had a peak performance of up to 3.19 PetaFlops/s and main memory of 340 TB. 'SuperMuc-NG' offered a peak performance of 26.9 PetaFlop/s with a main memory of 719 TB. Computational time was provided by several computing grants consisting of 'summer of simulation' projects and a regular project under grant pr62ci. A total of 26.1 MioCpuH were provided. One local workstation consisted of two Intel Xeon Broadwell-EP Series Processor E5-2680, 2.40 GHz, 14-Core Socket 2011-3, 64GB DDR4 memory, and four Geforce GTX 1080. The other workstation was built up by one Intel Core i9 9900K 8x 3.60 GHz, 32 GB DDR4 memory, and one GeForce RTX 2080.

Chapter 7

Protein Crystal Structures[1]

The model protein alcohol dehydrogenase from *Lactobacillus brevis* (*Lb*ADH) has already been crystallized in previous studies [Niefind et al., 2003; Schlieben et al., 2005]. *Lb*ADH wild type (WT) and mutant structures were determined (protein data bank (PDB)-IDs 1zk4, 1zjy, 1zjz, 1zk0, 1zk1, 1zk2, 1zk3 [Schlieben et al., 2005], and 1nxq [Niefind et al., 2003]). However, *Lb*ADH investigated in this study, was genetically modified compared to the studies of Schlieben et al. [2005] and Niefind et al. [2003]: a hexahistidine (His_6) purification tag and a glycine-serine-glycine (GSG) linker was fused at the N-terminus of the 'original' *Lb*ADH. In this study, this version is considered as 'WT *Lb*ADH'. Genetically, the two versions of *Lb*ADH differ and it has been argued that flexible purification tags and linkers may influence the structure and particularly the crystal arrangement [Carson et al., 2007]. Hence, new structure models for *Lb*ADH used in this study have to be generated and all structural comparisons in this study are based on His_6-tagged *Lb*ADH and mutants thereof.

Up-to date, X-ray diffraction is the most common method for protein structure determination [Wlodawer et al., 2008]. The protein crystals have to be cryo-protected in order to not get destroyed by the powerful X-ray beams. It has been reported that cryocooling reorders side chains [Halle, 2004] or that the crystal undergoes a re-packing [Juers and Matthews, 2001]; while others

[1]The following chapter has been published in parts in Hermann et al. 2018.

suggested that cryocooling captures representative conformations [Rader and Agard, 1997]. Juers and Matthews [2001] further describe a reversible lattice re-packing which induces substantial changes of the crystal contacts by flash-freezing. The re-arrangement changed which amino acids mediate the crystal packing. Furthermore, small changes of the crystallographic unit cell (UC) parameters may accompany the re-packing. UC contractions of 2 to 7% were reported upon cryocooling [Fraser et al., 2011; Juers and Matthews, 2001; Young et al., 1994]. Besides the ambiguous studies concerning crystal packing and side chain conformation, there has been consent that the backbone structure and fold of the protein is not affected [Fraser et al., 2011; Halle, 2004].

Neutron crystallography provides a method to collect diffraction data at room temperature and to circumvent potential artifacts due to cryocooling. Due to lower energy of the incident diffraction beam, the diffraction experiments can be conducted without cryoprotecting the crystal. However, neutron diffraction has major disadvantages compared to X-ray diffraction: in neutron diffraction, crystals of large volumes (over 0.5 mm^3) are needed due to a lower beam flux [Niimura and Podjarny, 2011]. This constitutes a major constraint on all experimental crystallization work. Furthermore, crystals require elaborate treatment before measurement: deuterium / hydrogen exchange has to be performed to reduce background noise in the diffraction experiment [Niimura and Podjarny, 2011]. Lastly, data in X-ray diffraction experiments of cryocooled crystals may be collected in a few minutes, while data collection in room temperature neutron diffraction takes days. As additionally considerably fewer neutron than X-ray sources exist, the availability of neutron crystallography is diminutive compared to X-ray crystallography. Hence, it is desirable to use structure models reconstructed from X-ray instead of neutron diffraction experiments.

For investigating protein crystallization with molecular dynamics (MD) simulations, the exact knowledge of side chain conformations is not decisive. MD simulations are intended to explore their dynamic behavior. However, cryocooling should not influence the general crystal packing, i. e. change the crystal contacts. In this chapter, the crystal properties of *Lb*ADH are determined. It is explored if X-ray structure models may be used as input structures in MD simulations: room temperature neutron and cryocooled

X-ray diffraction is conducted on the same *Lb*ADH WT crystal (section 7.1). The crystal structure models are reconstructed. The determined space groups (SGs) and their influence on the crystal packings are analyzed (section 7.2). Further WT structure models are generated and the models are structurally aligned. Root-mean-square deviation (RMSD) of UCs, tetramers, and monomers are computed. The crystallographic analysis is extended to crystals of *Lb*ADH mutants (section 7.3). Lastly, the findings are summarized and the influence of cryocooling and mutants on the free energy landscape of SGs in *Lb*ADH crystals and crystal packing is discussed (section 7.4).

7.1 Cryocooling of an *Lb*ADH Wild Type Crystal

The structures measured from protein crystals may be affected by cryocooling the crystal before measurement via X-ray diffraction. Therefore, room temperature neutron and low temperature X-ray diffraction data were collected from the same crystal. After neutron diffraction data was collected, the crystal was flash frozen as described in 6.1.2. The X-ray data indicated a reduction in the lengths of the a, b and c axes of the UC by 0.8%, 1.5% and 0.9%, respectively, resulting in a whole UC volume reduction of 3.3% (see Table 7.1). Moreover, this was accompanied by a phase transition within the orthorhombic crystal systems from the SG $I222$ in neutron diffraction to SG $P2_122_1$ in X-ray diffraction.

7.2 Space Groups and Crystal Packing

The collected diffraction data of WT *Lb*ADH allowed *in silico* reconstruction of the crystals. *Lb*ADH biologically assembles itself as homotetramer. Two of these biological units, i. e. crystal building blocks, build up a crystallographic UC. As will be shown later all structures of *Lb*ADH and variants thereof diffracted in the orthorhombic SGs found in the previous room temperature neutron and cryocooled X-ray diffraction experiment: SGs $P2_122_1$ and $I222$ (Hermann-Mauguin / international short symbol notation). The UC lattice vectors in an orthorhombic Bravais lattice are mutually orthogonal but have three different lengths. The Bravais lattice in $P2_122_1$ is primitive P with

lattice points at the cell corners. The lattice in *I*222 is body centered *I* (from the German term '*innenzentriert*') with lattice points on the cell corners with one additional point at the center of the cell. In *Lb*ADH crystals with *I*222 symmetry the asymmetric unit (AU) contains one monomer. Eight symmetry operations (with twofold rotation along all axes) on the AU build up the crystallographic UC. In *Lb*ADH $P2_122_1$ crystals, the AU consists of two monomers. Four symmetry operations (with twofold screw rotary translation along the x and z axes) are applied on the AU to generate a UC.

Both crystal systems result in the same crystal packing displayed in Fig. 7.1 **a** and **b**. The crystal contacts indicated in Fig. 7.1 occur in both crystal systems. The difference between both crystal systems is, that in SG *I*222, the tetramers are aligned parallel (translational invariant), while the tetramers in SG $P2_122_1$ are rotated with respect to each other around the y-axis. Fig. 7.2 shows the projection of both UCs into the x-y-plane. The vectors **w** and **v** indicate the orientation of the tetramers. The vectors point from the center of geometry (COG) of the backbone C_α atoms of the bottom two monomers, to the COG of the backbone C_α atoms of top two monomers. The angle γ for rotation of the tetramers with respect to each other is then defined as the angle between the two COG vectors **w** and **v**:

$$\gamma = \angle(\mathbf{w}, \mathbf{v}). \tag{7.1}$$

The crystal contacts at the 'corners' of the tetramer were termed 'corner contacts'. Because the tetramers are rotated with respect to each other in crystals with $P2_122_1$, two different corner contacts exist: corner contact 1 and 2 (c. f. Fig. 7.1). The corner contacts in crystals with *I*222 symmetry are identical. The crystal contact in y direction of the crystallographic UC between the 'edges' of the tetramer is termed 'edge contact' and the crystal contact in x-direction is termed 'side contact' (see Fig. 7.1).

7.3 Structural Comparison

7.3.1 *Lb*ADH Wild Type Crystals

As described, the WT crystal, which was first measured via neutron diffraction, underwent a phase transition from SG *I*222 to $P2_122_1$ upon cryocooling and a small volume reduction of 3.3%. This is on the lower limit of the described UC volume changes between room temperature and cryo-temperature in literature [Fraser et al., 2011; Juers and Matthews, 2001; Young et al., 1994]. In order to quantify the structural changes within the UC, geometrical analysis is performed. Structural comparison of the crystals based on the UC alone is possible due to the low and comparable moscaicity α for the measured crystals (see section 6.2.3 and Table 7.1). The monomer and tetramer exhibited a C_α RMSD of 0.13 Å and 0.28 Å, respectively (c. f. Table 7.1). These values indicate an identical or undistinguishable structure of the C_α atoms within the resolution limit for both monomer and tetramer (c. f. section 6.2.2). As stated, the difference between both crystal packings is the orientation of the tetramers. The rotation of the tetramers in WT $P2_122_1$ crystal (Fig. 7.2 **b**) calculates to 5.2°. UC C_α RMSD between room temperature *I*222 neutron structure and cryocooled $P2_122_1$ X-ray structure calculated to 1.16 Å. This RMSD value was slightly higher as an RMSD value which could be explained with the resolution limits in the respective diffraction (c. f. section 6.2.2). It indicated an in principal similar structure and stems from the slight rotation of the tetramers, i. e. phase transition, during cryocooling.

In addition, to the WT 'X-ray $P2_122_1$', and 'neutron *I*222' crystal, another WT crystal which diffracted in SG *I*222 in a cryocooled X-ray diffraction experiment could be found. The WT '*I*222 UC' from the room temperature neutron diffraction experiment can be translational aligned onto the WT '*I*222 UC' of the cryocooled X-ray diffraction experiment. The UC C_α RMSD for WT *I*222 neutron and X-ray measured crystals calculated to 0.55 Å. Considering the resolution difference of $\Delta\ d_{min} = 0.8$ Å the two structures are undistinguishable within the resolution limits.

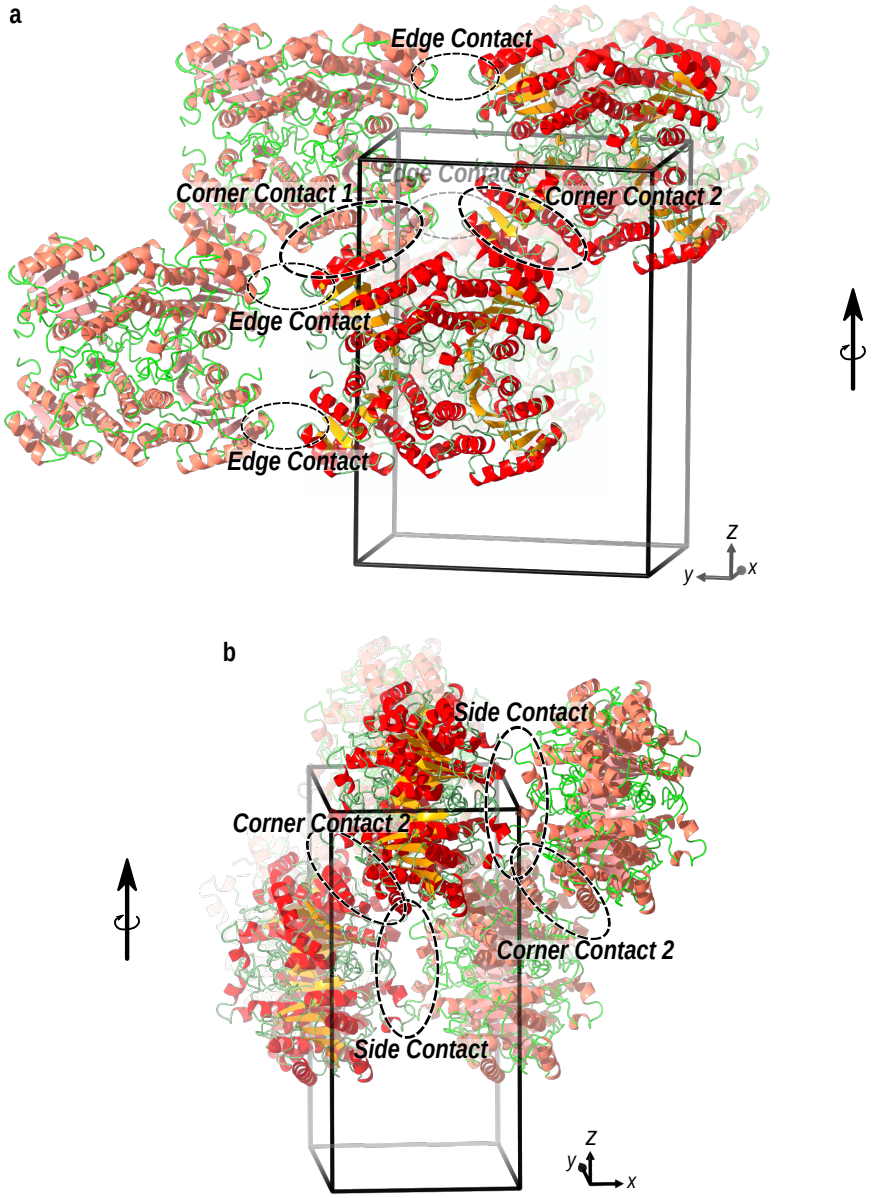

Fig. 7.1 Crystal packing and crystallographic UC of *Lb*ADH. UC is drawn in black. The two tetramers building the UC are colored thick, adjacent UCs are colored light. **a** The view in the y-z-plane shows the 'edge contact' and the 'corner contacts'. In SG $P2_122_1$ 'corner contact 1' and 'corner contact 2' structurally slightly differ as the two tetramers are rotated with respect to each other around the y-axis (In SG $I222$ (not displayed here) the two tetramers are translationally invariant and therefore the 'corner contacts' coincide; c. f. Fig. 7.2). **b** The view in the x-z-plane shows the third crystal contact: the 'side contact' between two tetramers.

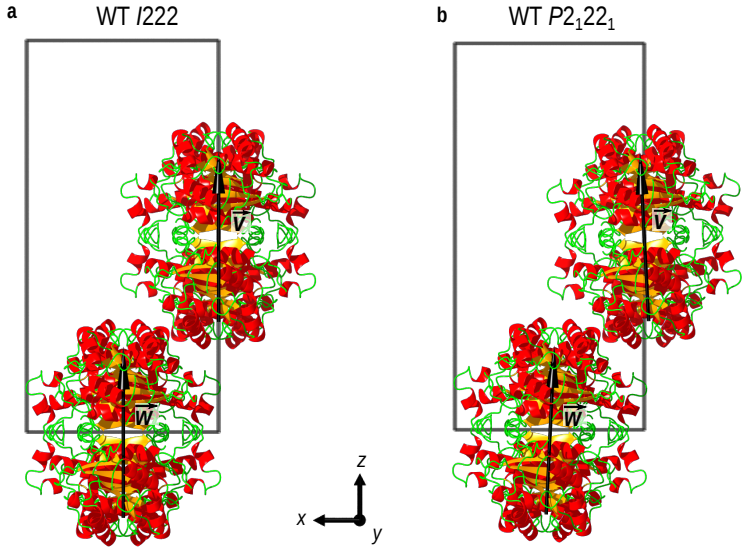

Fig. 7.2 Crystallographic UCs of *Lb*ADH *I*222 and $P2_122_1$ crystals. The UCs are projected onto the x-z-plane. Vectors **w** and **v** point from the COG of the backbone C_α atoms of bottom two monomers to the COG of the backbone C_α atoms of top two monomers. **a** In crystals exhibiting *I*222 symmetry, both vectors are parallel. The tetramers are translational invariant. **b** In crystals exhibiting $P2_122_1$ symmetry, the vectors still lie in the x-z-plane but include an angle $\gamma = \angle(\mathbf{w}, \mathbf{v}) \neq 0$ ($\gamma = 5.2°$ for WT $P2_122_1$ crystal). The tetramers are rotated against each other around the y-axis.

Table 7.1 Structural properties of *Lb*ADH WT and variant crystals. For all measured crystals SG, resolution limit d_{min}, mosaicity α, rotation of tetramers with respect to each other γ, UC paramters, UC volume V_{UC}, UC volume ratio with regard to UC volume of cryocooled WT X-ray $P2_122_1$ crystal $r = (V_{UC} - V_{UC,WT,P2_122_1})/V_{UC,WT,P2_122_1}$ are reported. Further, UC, monomer (M), and tetramer (T) C_α RMSD after alingement on WT $I222$ and $P2_122_1$ cryocooled crystal, respectively, are given.

protein variant	SG	d_{min}, Å	α, °	γ, °	UC parameters, Å^3	V_{UC}, Å^3	r, %	RMSD after alignment on WT, Å					
								$I222$ crystal			$P2_122_1$ crystal		
								UC	M	T	UC	M	T
WT	$P2_122_1$	1.48	0.156	5.21	$56.03 \times 83.31 \times 114.38$	533.91	-	0.97	0.16	0.22	-	-	-
D54H	$P2_122_1$	1.09	0.156	4.52	$55.65 \times 81.02 \times 114.10$	514.45	-3.64	0.91	0.15	0.24	0.48	0.12	0.17
Q207Y	$P2_122_1$	1.40	0.165	5.33	$56.29 \times 83.35 \times 114.38$	536.64	0.51	1.00	0.19	0.26	0.17	0.14	0.14
D54F	$P2_122_1$	1.41	0.145	4.99	$55.76 \times 81.09 \times 113.13$	511.53	-4.19	1.08	0.17	0.25	0.40	0.16	0.19
T102E	$P2_122_1$	1.44	0.168	4.56	$55.62 \times 81.15 \times 115.56$	521.59	-2.31	0.96	0.17	0.23	0.46	0.13	0.17
K32A_Q126K	$P2_122_1$	1.40	0.102	4.80	$56.28 \times 82.03 \times 114.59$	529.02	-0.92	0.92	0.19	0.26	0.20	0.14	0.16
WT	$I222$	1.33	0.551	-	$55.80 \times 82.88 \times 115.01$	531.89	-0.38	-	-	-	0.97	0.16	0.22
K32A	$I222$	1.55	0.211	-	$55.58 \times 81.78 \times 114.89$	522.21	-2.19	0.31	0.12	0.13	0.96	0.15	0.23
Q126K	$I222$	1.21	0.349	-	$55.77 \times 84.23 \times 113.56$	533.45	-0.09	0.52	0.13	0.17	1.09	0.17	0.26
Q126H	$I222$	1.22	0.267	-	$56.05 \times 80.57 \times 113.47$	512.42	-4.02	0.73	0.20	0.23	1.06	0.19	0.26
T102E_Q126K	$I222$	1.80	0.086	-	$55.95 \times 84.50 \times 114.75$	542.51	1.61	0.44	0.15	0.15	1.10	0.15	0.20
Q207D	$I222$	1.20	0.131	-	$55.93 \times 80.69 \times 115.27$	520.24	-2.56	0.61	0.22	0.26	1.05	0.17	0.25
WT Neutron Diffraction	$I222$	2.15	0.994	-	$56.51 \times 84.57 \times 115.42$	551.59	3.31	0.55	0.24	0.29	1.16	0.23	0.28

7.3.2 *Lb*ADH Variant Crystals

*Lb*ADH variant crystals were found to diffract in both SGs $P2_122_1$ (D54H, Q207Y, D54F, T102E, K32A_Q126K) and $I222$ (K32A, Q126K, Q126H, T102E_Q126K, Q207D). No preference for either of the SGs was found. The structural details including resolution limit d_{min} and mosaicity α of the crystal are reported in Table 7.1. The low mosaicity (α = 0.10 to 0.55°) allowed comparision of the crystals based on one crystallographic UC (see section 6.2.3). In case of SG $P2_122_1$, the angle γ between the tetramers were calculated. UC C_α RMSDs, tetramer and monomer C_α RMSDs were calculated for alignment on both, $I222$ and $P2_122_1$, WT X-ray structures. Monomer and tetramer deviated on average by 0.17 ± 0.03 Å and 0.22 ± 0.05 Å, respectively. This practically means the same structure was built by the proteins' C_α atoms within the resolution limits of the diffraction experiments (c. f. section 6.2.2). Hence, there are no differences in the crystal building blocks (tetramers) between *Lb*ADH variants.

Next, it was analyzed how the crystal building blocks were arranged in the UCs of different *Lb*ADH variant crystals. The average tetramer torsion angle in $P2_122_1$ crystals was calculated to $\gamma = 4.92 \pm 0.31°$. Torsion angles lay close together (c. f. Table 7.1) and suggested a common minimum in free energy landscape for rotation of the tetramers. The low average UC C_α RMSDs of 0.35 ± 0.13 Å supported this suggestion. The structural arrangement was highly similar. The calculated average RMSD lay slightly above a value which would be explainable simply by different resolution limits of the structure models (c. f. section 6.2.2). The same observation was made for $I222$ crystals: a low average UC C_α RMSDs of 0.53 ± 0.14 Å was calculated indicating a highly similar arrangement. While the structural arrangement within the SGs were highly similar, UC C_α RMSDs between different SGs ($I222$ and $P2_122_1$) clearly indicated the slightly different arrangement of the tetramers (see Table 7.1). The average UC C_α RMSDs between different SGs calculated to 1.01 ± 0.07 Å and resembled a slight difference while still indicating the same general arrangement, i. e. crystal packing (c. f. section 6.2.2).

7.4 Conclusion

7.4.1 Crystal Packing

The diffraction experiments allowed to reconstruct and analyze the proteins structures, and crystals itself. Though SG and UC parameters slightly differed (see Table 7.1) all *Lb*ADH mutants and WT crystals exhibited the same crystal packing independent on mutation and diffraction experiment (room temperature neutron or cryocooled X-ray diffraction). This means that the overall arrangement, in particular the crystal contacts are the same in all *Lb*ADH crystals. The crystal contacts are composed of the same amino acids. The small tetramer C_α RMSD of all *Lb*ADH mutant structures showed identical or undistinguishable crystal building blocks within the resolution limit of the generated structure models. UC C_α RMSDs within the same SG exhibited RMSDs values slightly above RMSDs values explainable by the diffraction limits. The two SGs differ by the rotation of the tetramers in the UC against each other. All *Lb*ADH variant crystals exhibit highly similar but slightly altered arrangement of the crystal building blocks (tetramers) within the UC. In particular, cryocooling did not change the crystal structure, as the monomers, tetramers and UC RMSDs of the room temperature neutron diffraction structure and low temperature X-ray diffraction structures of WT *Lb*ADH exhibited low values indicating undistinguishable structures within the resolution limit. No major re-arrangement with change of crystal contacts as e. g. found in a study by Juers and Matthews [2001] was observed.

7.4.2 Space Groups

There are two different physical explanations for the occurrence of two SGs in *Lb*ADH WT and variant crystals:
(1) The free energy landscape has two minima for crystallization of *Lb*ADH. *Lb*ADH crystallizes in either $I222$ or $P2_122_1$. Cryofreezing captures the respective 'crystal state'. The phase transition observed for the crystal analyzed with room temperature neutron diffraction and afterwards with cryo-temperature X-ray diffraction, may have been induced by crystal handling between the

experiments. For example, the mother liquor was exchanged with a mixture of cryoprotectant and mother liquor. The crystal was exposed to physical stress when it was removed from the loop after neutron diffraction.
(2) *Lb*ADH crystallizes at room temperature in *I*222. Cryocooling altered the free energy landscape as suggested in [Juers and Matthews, 2001]. This alteration created a second minimum for the arrangement of the tetramers (crystal building blocks) enabling a slight re-arrangement in SG $P2_122_1$. Such a phase transition within an orthorhombic crystal system has previously been described by Pinard et al. [2016]. They observed a phase transition from $P2_122_1$ to *I*222 without alteration of the overall arrangement of monomers following cryoannealing. Cryoannealing is described as re-equilibration of the cryocooled crystal to room temperature prior to repeated cryocooling [Yeh and Hol, 1998]. In their study, cryoannealing increased the symmetry [Pinard et al., 2016]. This might be consistent with the presented case in this study, where a higher symmetry was detected at room temperature compared to cryocooled conditions for the same crystal.

Independent on the physical origins of the two different SGs, the arrangement within both crystal systems is stable: the neutron experiment took in total 16 days and was performed in two separate diffraction experiments with a time gap of one month between each other. This means that phase transition from *I*222 to $P2_122_1$ did not happen spontaneously over time. Hence, grown in *I*222, the crystal was stable.

No *Lb*ADH mutants were reported in this study which diffracted in both SGs. However, as WT *Lb*ADH crystals were found in both SGs it is unlikely that other variants of *Lb*ADH can crystallize only in either one of the SGs. Hence, it is likely that the two minima are independent on mutation of *Lb*ADH. The fact that for *Lb*ADH mutants not both SGs could be reported, may be due to fewer analyzed mutant crystals compared to WT crystals.

Interestingly, all *Lb*ADH crystals in the study of Schlieben et al. [2005] and Niefind et al. [2003] exhibited *I*222 symmetry. Only *Lb*ADH mutant G37D (PDB-ID 1zk3) was found in another symmetry: SG P1 [Schlieben et al., 2005]. However, crystal reconstruction indicated severe problems with the SG, UC and AU determination. SG P1 should not be considered further. Hence, in total seven *Lb*ADH variants were reported in SG *I*222. The *Lb*ADH

investigated in this study had an His_6 tag and GSG linker fused at its N-terminus in contrast to the *Lb*ADH in the study of Schlieben et al. [2005] and Niefind et al. [2003]. It can be hypothesized that the flexibility introduced by the His_6 tag and GSG linker enabled the assembly in two different SGs.

7.4.3 X-ray Crystallography Structure Models

It can be summarized that X-ray diffraction is well suited as structure determination method to provide input structures for MD simulations. No re-arrangement was found upon cryocooling according to the high structural similarity between the measured crystals. Diffraction data from room temperature and cryocooled conditions resulted in structure models exhibiting the same crystal packing and same crystal contacts. *Lb*ADH crystals occurred in two different SGs exhibiting only a marginal difference in the crystal structure. All *Lb*ADH WT and mutant crystals showed the same general crystal packing, i. e. same crystal contacts. Fast cryocooled X-ray diffraction can be favored over lengthy room temperature neutron diffraction.

Chapter 8

Modeling Protein Crystallization[2]

Crystallographic methods like neutron and X-ray diffraction allow the generation of protein structure models. These structure models resemble static snapshots of a protein's configuration found in a crystal. In molecular dynamics (MD) simulations the protein structures determined with crystallographic methods are parameterized with a force field and their time evolution is computed. The simulated trajectories reveal interactions on an atomic level and following the ergodic hypothesis thermodynamic quantities can be calculated. Although MD simulations explore dynamic properties they are limited to a small timescale (depending on the MD method and set-up this ranges from ns to a few µs). Therefore, it is desirable to get structure models as close to biological reality as possible. Further insights into the biology of the system can be used to guide the dynamics within known boundaries: constraints and restraints can be applied as 'boundaries', which facilitate the sampling of the system's conformational phase space.

Following the experimental investigation of crystallization behavior of alcohol dehydrogenase from *Lactobacillus brevis* (*Lb*ADH) in Grob et al. [2020] the term 'enhanced/*reduced* crystallizability' is defined as 'correlating observations of a higher/*lower* number of crystals (equivalent to a higher/*lower* nucleation rate), lower/*longer* induction time, lower/*longer* time span until crystallization equilibrium, and crystallization at reduced/*increased* concentrations of protein

[2]The following chapter has been published in parts in Hermann et al. 2021.

and crystallization agent'. For the investigation of enhanced crystallizing mutants of *Lb*ADH with MD simulations, the mutants T102E, D54F, Q126H, Q126K, K32A which were experimentally analyzed in detail [Grob et al., 2020; Nowotny et al., 2019] were chosen. As negative control mutants K45A, H39A, D54A and Q207D were studied. K45A and H39A showed lowered success rates in Nowotny et al. [2019]. Based on the work of Nowotny et al. [2019] and Grob et al. [2020] the order of mutants with increased crystallizabilty follows T102E ≈ D54F > Q126H > Q126K > K32A > WT. For mutants with reduced crystallizability, Phillip Grob (Institute of Biochemical Engineering, Technical University of Munich (TUM), Garching, Germany) provided the crystallization order as WT > Q207D > H39A, K45A, D54A where the latter three mutants did not crystallize under the tested conditions. The investigated protein variants are listed in Table 8.1. Though the crystallization behavior was experimentally examined, the underlying thermodynamics and mechanics are largely unknown.

Table 8.1 Investigated protein variants of *Lb*ADH. The *Lb*ADH variants are listed according to their crystallization behavior in descending order. Furthermore, the targeted contact (see Fig. 7.1) and the availability of a crystal structure is noted.

	targeted contact	structure model
T102E	corner contact	yes
D54F	edge contact	yes
Q126H	corner contact	yes
Q126K	corner contact	yes
K32A	edge contact	yes
Q207D	side contact	yes
K45A	corner contact	no
H39A	corner contact	no
D54A	edge contact	no

In order to investigate the crystallization process by means of MD simulations, the free energy difference $\Delta\Delta G$ of crystallizing a wild type (WT) vs. mutant protein is calculated. Free energy calculation (FEC) is performed with free energy perturbation (FEP) calculation of an alchemical transformation. The free energy difference is calculated according to the thermodynamic path constructed in section 5.3.6. Crystal state simulations (i. e. end states in the crystallization process) allow the investigation of altered interactions. The free energy difference $\Delta\Delta G$ defines which crystallization process (WT or

mutant) is more likely to happen and which crystal is energetically more stable. The advantage of alchemical free energy simulations is that they inherently correctly sample the Boltzmann distribution of microstates. As all atom MD simulations are used, entropic and enthalpic effects for both solvent and protein are inherently included.

This chapter aims to calculate dynamic and thermodynamic properties to model protein crystallization *in silico*: system, set-ups and methods for a stable protein crystal simulation are developed (section 8.1). The set-up is tested on *Lb*ADH WT crystals. Methods to perform reliable FEC calculations via FEP are implemented. Free energy differences $\Delta\Delta G$ for the crystallization process of WT vs. mutant *Lb*ADH are calculated (section 8.2). Mechanistic interactions for all *Lb*ADH variants dictating the crystallization behavior are unraveled (section 8.3). Lastly, the findings are summarized and discussed under the perspective of nucleation and crystal growth theories, previous experimental results, and engineering strategies (section 8.4).

8.1 System, Set-up and Method Development

8.1.1 Biological and Crystal Contact Interfaces

In preliminary simulations it was found, that crystal contacts were too weak to be investigated without additional restraints. If a WT *Lb*ADH crystal contact was simulated by extracting the interacting monomers of the respective contact (version (b) of the bound state described in 6.3.1) or if the crystal contact was represented by two tetramers (crystal building blocks) the interactions were lost within a few ns and the crystal contact drifted apart. In contrast, simulation of one tetramer in solution yielded an intact tetramer over a period of 200 ns. This further stresses the already mentioned difference between crystal contact interfaces and biological interfaces as found in an oligomer. The biological assembly did not break up whereas the crystal contact was highly fragile. For the investigation of crystal contacts and crystallization in general with MD simulations further restrictions had to be made: it was chosen to restrain atom positions on positions measured in X-ray diffraction experiments.

Two different restraints were tested: (a) strong position restraints (restraints on all backbone atoms with 1000 kJ mol^{-1} nm^{-2}) and (b) weak position restraints (restraints on all C_α atoms with 750 kJ mol^{-1} nm^{-2}). The influence of the restraints was tested by calculating the free energy change for alchemically transforming the crystal bound state. This corresponds to the free energy change of one leg of the thermodynamic cycle $\Delta G^{\mathrm{WT}\rightarrow\mathrm{Mut}}_{\mathrm{crystal}}$ (c. f. Fig. 5.3). *Lb*ADH mutation Q126H was chosen and simulated in the reduced crystal contact set-up where the interacting monomers are extracted. The free energy difference ΔG for mutating Q126H into WT *Lb*ADH calculated to 79.73 ± 0.39 kJ mol^{-1} for strong position restraints and to 79.75 ± 0.62 kJ mol^{-1} for weak position restraints. The free energy differences are within the error of each other. Inspection of the interaction patterns yielded intact interactions in both cases. Weak restraints were chosen in all following simulations.

8.1.2 Simulation of *Lb*ADH Wild Type Crystals

Three *Lb*ADH WT protein crystal structures were obtained in the X-ray diffraction experiments (c. f. chapter 7):

- '*I*222 neutron diffraction crystal': A crystal was reconstructed from neutron diffraction data of a crystal diffracting in *I*222.
- '$P2_122_1$ X-ray diffraction crystal': Cryocooling of the '*I*222 neutron diffraction crystal' yielded a crystal which diffracted in space group (SG) $P2_122_1$ in X-ray diffraction.
- '*I*222 X-ray diffraction crystal': A third crystal could be reconstructed from X-ray data of a crystal diffracting in *I*222.

In section 7.3, it was shown that these crystals exhibited same crystal packing and high structural similarity based on their backbone C_α root-mean-square deviations (RMSDs). In order to verify that cryocooled X-ray diffraction experiments were adequate to determine protein crystals, all three crystals were simulated in crystal state simulations. Selected interactions at the crystal corner and edge contacts (for definition of the contacts see Fig. 7.1) were analyzed. In the following, the interactions are quantified. The role of these interactions concerning the crystallization behavior of WT *Lb*ADH will be elucidated in section 8.3.

Interaction of Glutamine Q126

The first interaction to be quantified was a hydrogen bond (H-bond) between glutamine at position 126 (Q126) and serine at position 40 (S40) found at the corner contact. In *Lb*ADH WT the amide oxygen of Q126 is hydrogen acceptor for the hydroxy hydrogen of S40 (see Fig. 8.1). This interaction was present in 9.6 ± 3.7% of the simulation time in the '*I*222 neutron crystal' and in 8.2 ± 2.9% in the '*I*222 X-ray diffraction crystal'. In the '$P2_122_1$ X-ray diffraction crystal' two structurally slightly different corner contacts existed (c. f. Fig. 7.1). In corner contact 1 and 2, this interaction was present in 0.8 ± 0.4% and 16.8 ± 8.1% of the simulation time, respectively, resulting in an average of 8.8 ± 3.9%. The crystal contact and MD simulations are displayed in Fig. 8.1 for structures generated from X-ray diffraction experiments of WT *I*222 and $P2_122_1$ crystals. Hence, all three crystals revealed the same interaction pattern as observed in MD simulations.

Interaction of Lysine K45

Another important interaction could be found in the salt bridge between lysine at position 45 (K45) and glutamine at position 66 (Q66) at the corner contact. This interaction was found in 48.1 ± 4.7% of the simulation time in the '*I*222 neutron crystal' and in 49.8 ± 9.0% in the '*I*222 X-ray diffraction crystal'. In the '$P2_122_1$ X-ray diffraction crystal' this salt bridge occurred in 64.0 ± 8.5% of the simulation time at contact 1 and 39.5 ± 8.0% at contact 2 resulting in an average of 51.8 ± 5.2%. Again, the interaction times in all three WT crystals are highly similar. The interaction of K45 and Q66 is displayed in Fig. 8.2 for structures generated from X-ray diffraction experiments of WT *I*222 and $P2_122_1$ crystals.

Interaction of Histidine H39

An amino acid responsible for a decisive interaction can be found in histidine at position 39 (H39) at the corner contact. H39 interacts with arginine at position 127 (R127) in a key-lock manner. The center of geometry (COG) distance between the imidazole ring of H39 and guanino group of R127

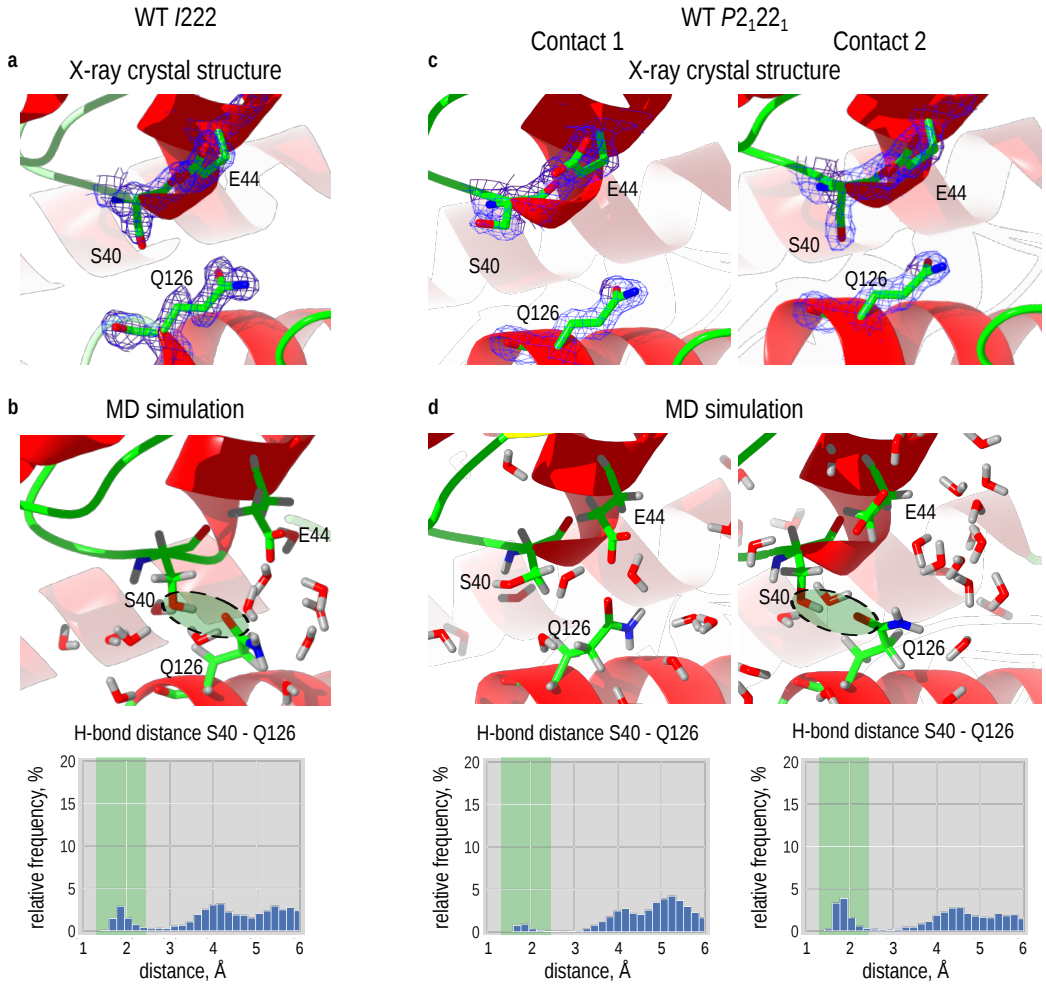

Fig. 8.1 X-ray structure and MD simulation analysis of the corner crystal contact around Q126 in WT *Lb*ADH *I*222 and $P2_122_1$ crystals. **a** X-ray structure of corner contact in WT *I*222 crystal with measured electron density map contoured in blue for displayed residues. **b** MD simulation and analysis of H-bond distance between hydroxy hydrogen of serine (S40) and amide oxygen of glutamine (Q126) in WT *I*222 crystal. **c** X-ray structure of corner contact 1 and 2 in WT $P2_122_1$ crystal with measured electron density map contoured in blue for displayed residues. **d** MD simulation and analysis of H-bonds between S40 and Q126 in WT $P2_122_1$ crystal. H-bond distances (< 2.5 Å) are dyed in green. In both crystals on average a similar H-bond distribution occurred. Electron density maps are calculated with structure factor amplitude $2F_o - F_c$.

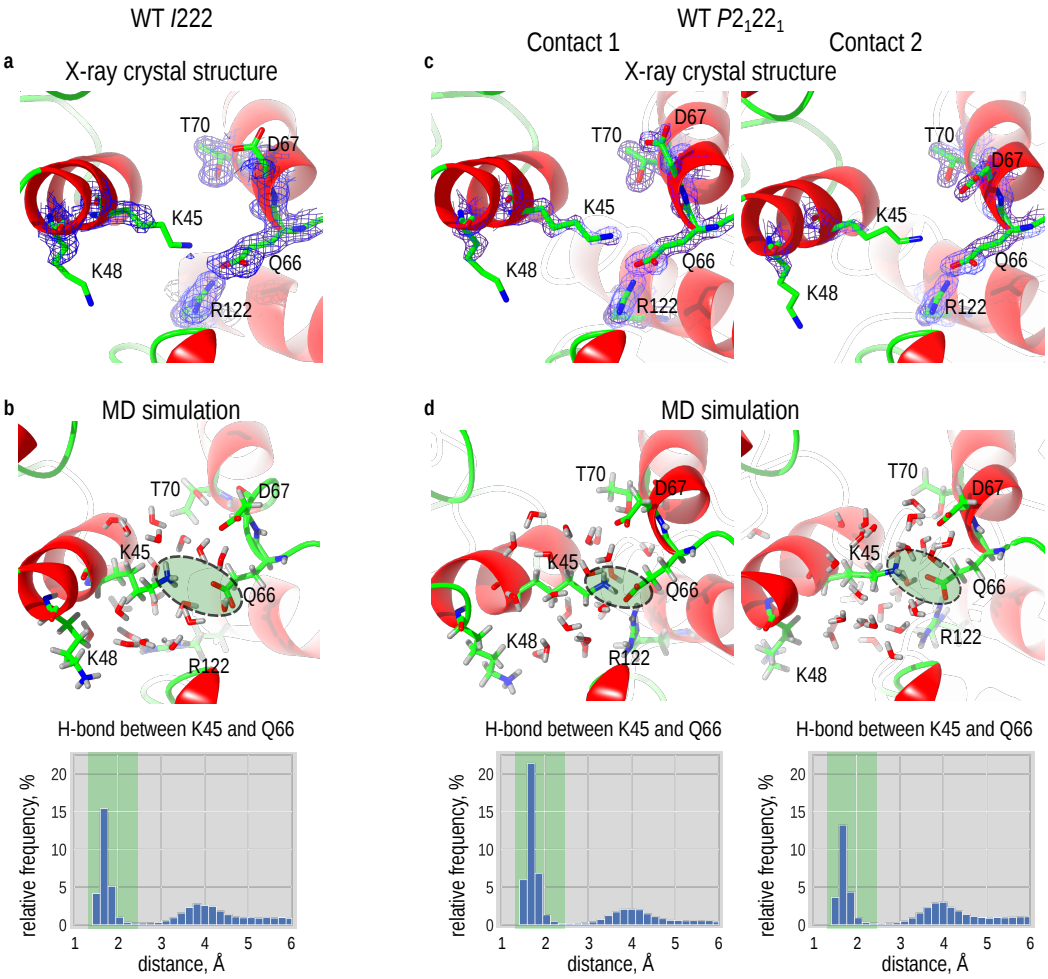

Fig. 8.2 X-ray structure and MD simulation analysis of corner contact around K45 in WT *Lb*ADH *I*222 and $P2_122_1$ crystals. **a** X-ray structure of crystal corner contact around K45 in WT *I*222 crystal. Relevant residues are displayed with measured electron density contoured in blue. **b** Snapshot from MD simulation trajectory including nearby water molecules and analysis of the minimum distance between the positive amino group of K45 and negative carboxy group of Q66 (circled in green). Distance values in H-bond distance (< 2.5 Å) are dyed in green. **c** X-ray structure of crystal corner contact 1 and 2 around K45 in WT $P2_122_1$ crystal. Relevant residues are displayed with measured electron density contoured in blue. **d** MD simulation and analysis of the salt bridge between K45 and Q66. Distance values in H-bond distance (< 2.5 Å) are dyed in green. A similar H-bond pattern was observed on average in all crystals. Electron density maps are calculated with structure factor amplitude $2F_o - F_c$.

allows no additional solvent in-between. The average distance in the '*I*222 neutron diffraction crystal' calculated to 3.8 ± 0.2 Å. An average distance of 3.7 ±0.3 Å was found for '*I*222 X-ray diffraction crystal' and '$P2_122_1$ X-ray diffraction crystal'. The key-lock interaction is present in all three WT crystals. The crystal contact and MD simulations are displayed in Fig. 8.3 for structures generated from X-ray diffraction experiments of WT *I*222 and $P2_122_1$ crystals.

Interaction at the Edge Contact

In *Lb*ADH WT crystals the edge contact is composed of symmetrically interfacing aspartic acids at position 54 (D54) and threonines at position 52 (T52). D54 and T52 are opposite of symmetry related D54 and T52 of another monomer (see Fig. 8.4). The charged functional groups especially of D54 initiate a water mediated H-bonding between the monomers. Therefore, the water shell at this wet contact [Janin, 1999] was analyzed. The number of water molecules within a 3.3 Å shell around the mean positions of C_α atoms of oppositely lying D54s and T52s were counted. This resulted in a water count of 4.0 ± 2.0 in the '*I*222 neutron diffraction crystal', 5.0 ± 2.3 in the '$P2_122_1$ X-ray diffraction crystal', and 4.5 ± 2.2 in '*I*222 X-ray diffraction crystal'. All counts were within the error of each other and revealed a similar interaction pattern at this 'wet contact'.

Hence, it could be verified that based on highly similar structures retrieved from neutron and X-ray diffraction experiments (see section 7.3) MD simulations reproduce highly similar interactions. For all subsequent simulations, X-ray diffraction data was used. In case of WT *Lb*ADH, the '$P2_122_1$ X-ray diffraction crystal' served as input structure.

8.1.3 System and Method Robustness

The robustness of the simulated system was investigated by evaluating if the constructed system, set-up and method yielded reproducible free energy change results. Three independent FEC simulations for the bound state of mutant Q126H were set-up. The bound state was represented by a fully periodically interacting unit cell (UC) (see Methods 6.3.1). After equilibration

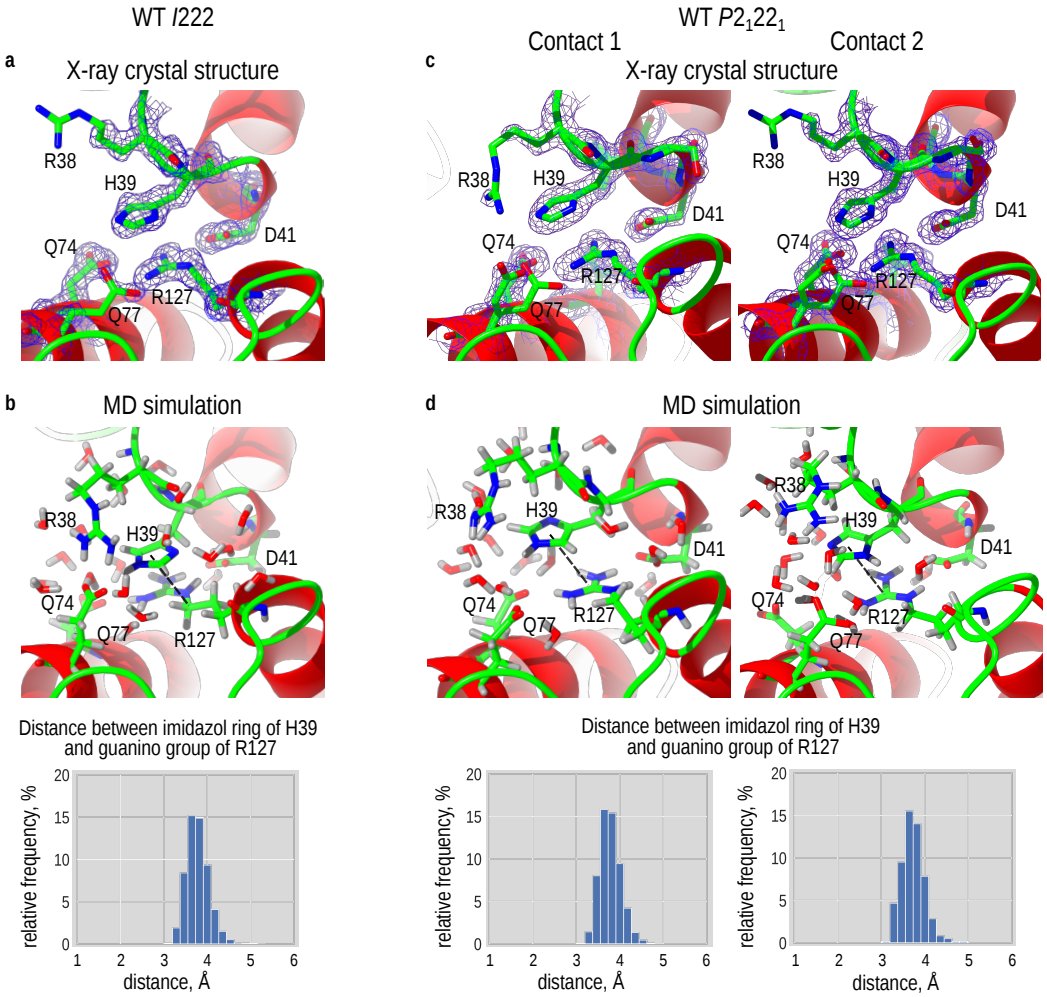

Fig. 8.3 X-ray structure and MD simulation analysis of corner contact H39 in WT *Lb*ADH *I*222 and $P2_122_1$ crystals. a X-ray structure of crystal corner contact in WT *I*222 crystal. Relevant nearby residues are displayed with measured electron density contoured in blue. **b** Snapshot from MD simulation trajectory including nearby water molecules and analysis of the COG distances between imidazole ring of H39 and guanino group of R127. **c** X-ray structure of crystal corner contact in WT $P2_122_1$ crystal. Relevant nearby residues are displayed with measured electron density contoured in blue. **d** MD simulation analysis of the COG distances between imidazole ring of H39 and guanino group of R127. An average binding distance is 3.7 Å was calculated for all crystals. Electron density maps are calculated with structure factor amplitude $2F_o - F_c$.

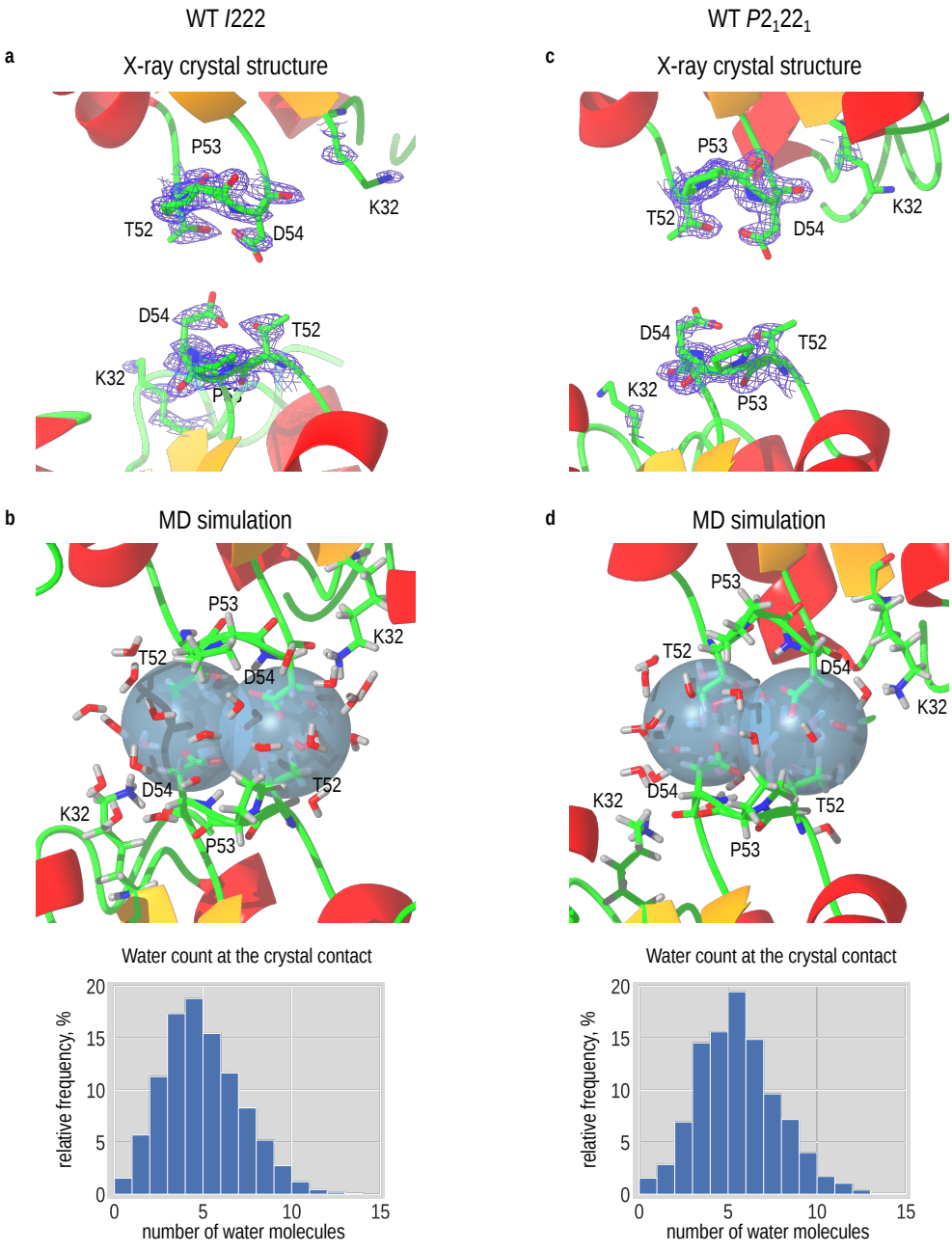

Fig. 8.4 X-ray structure and MD simulation analysis of contact edge around amino acid position 54 in WT *Lb*ADH *I*222 and $P2_122_1$ crystals. a X-ray structure of the crystal contact in *I*222 crystal including electron density for displayed relevant residues. **b** Snapshot from the MD simulation trajectory. Spheres are constructed around the midpoint of oppositely lying C_α atoms of residues D54 and T52. The water molecules inside the spheres were counted to 4.5 ± 2.2. **c** X-ray structure of the edge contact in $P2_122_1$ crystal. Electron density is contoured blue for all displayed residues. **d** MD simulation analysis yielded 5.0 ± 2.3 water molecules at the crystal contact. Electron density maps are calculated with structure factor amplitude $2F_o - F_c$.

the simulation was prolonged to extract three independent configurations as starting structure from the system phase space. In this set-up, the free energy change in each simulation was calculated for eight mutations in a UC at once. The values were in close agreement within the calculated errors (see Table 8.2). The low standard deviation of 0.58 kJ mol^{-1} for eight mutations at once demonstrated the robustness of method and set-up.

Table 8.2 Three independent FEC simulations of the bound UC set-up with periodic boundary condition (PBC) for mutation Q126H in the crystal packing of Q126H. The starting frames were taken by prolonging the simulation after equilibration for 0 ns (simulation 1), 15 ns (simulation 2), and 35 ns (simulation 3). The values are for a whole UC simulation, i. e. eight mutations at once. All values in kJ mol^{-1}.

simulation 1	simulation 2	simulation 3
637.66±0.61	637.94±0.84	638.77±0.48

8.1.4 System Reduction

Direct translation of the bound crystal state into the model system results in the simulation of a crystallographic UC with PBC. This system set-up is not suited for the application of charge correction scheme (see section 5.4). Therefore, the reduction of the system to an appropriate set-up was evaluated. The bound crystal state of a mutant with no charge change (Q126H) was simulated in (a) a fully periodically interacting UC set-up and (b) a reduced set-up where only the interacting monomers of the respective contact are placed in a simulation box (see Appendix B Table B.4). Furthermore, the unbound state biologically consisted of a solvated tetramer. In order to save computer power, the unbound solvated state of Q126H was simulated as (a) tetramer and (b) two monomers of the tetramer. For both cases the free energy change was calculated in WT and mutant crystal system to check if the system is reducible without loss of information. Each free energy calculation was started from two different starting structures where one structure was equilibrated as WT and and the other as mutant. The low standard deviation of all calculations (below 0.44 kJ mol $^{-1}$) showed that each system could be reduced enabling investigation of any mutation (see Table 8.3).

Table 8.3 Individual free energy change upon mutation ΔG for different systems (mutant crystal system, WT crystal system, and unbound state) and different set-ups (bound UC set-up with PBC, reduced set-up isolated crystal contact, solvated unbound tetramer, solvated unbound dimer). The values are calculated for mutation Q126H. For each system and set-up two simulations were performed where one equilibration was performed with WT parameters (EQ WT) and the other one with mutant parameters (EQ Mut). All values in kJ mol $^{-1}$.

		EQ WT	EQ Mut	mean(ΔG)	SD(ΔG)
Mutant Crystal	UC	79.77±0.05	79.67±0.13	79.85±0.18	0.17
	reduced	79.75±0.62	80.07±0.39		
WT crystal	UC	82.49 ±0.06	82.21±0.13	82.71±0.17	0.44
	reduced	83.15±0.36	83.01±0.58		
unbound	Tetramer	83.92 ±0.15	83.55±0.17	83.56±0.14	0.25
	reduced	83.40±0.27	83.37±0.42		

8.1.5 Free Energy Charge Change Corrections

The calculated free energy differences have to be corrected for charge changes during the alchemical transformation. Therefore, two different methodolodigcal different correction schemes were tested. Firstly, the co-alchemical counterion approach was applied [Chen et al., 2018]. A counterion is alchemically transformed alongside the mutation to keep a zero net charge during the alchemical transformation (see section 5.4.1). Secondly, a numerical post simulation correction proposed by Reif et al. [2011; 2014] was used (see section 5.4.2). This correction schemes were compared for two mutations (*Lb*ADH WT into mutants Q126K and K32A) in WT crystal system which showed the lowest free energy changes $\Delta\Delta G$ (see section 8.2). In the following part of this section, the free energy changes calculated via the co-alchemical counterion approach are referred to as $\Delta G_{\text{co-ion}}$. In the numerical correction scheme, the raw free energy change ΔG_{raw} which is calculated in a simulation with net charge change is corrected with three different terms: a correction for spurious polarization ΔG_{pol}, direct interactions ΔG_{dir} and incorrect solvent potential summation ΔG_{dsm}. The numerical corrected free energy change was defined as

$$\Delta G_{\text{num}} = \Delta G_{\text{raw}} + \Delta G_{\text{pol}} + \Delta G_{\text{dir}} + \Delta G_{\text{dsm}} \tag{8.1}$$

if the raw free energy change ΔG_{raw} is corrected with all three correction terms. Furthermore, the numerical corrected free energy energy change was defined as

$$\Delta G'_{\text{num}} = \Delta G_{\text{raw}} + \Delta G_{\text{pol}} + \Delta G_{\text{dir}} \tag{8.2}$$

if the raw free energy change ΔG_{raw} is only corrected by the polarization and direct interaction term. It was discussed that $\Delta G'_{\text{num}}$ corrects similar to the co-alchemical approach [Öhlknecht et al., 2020].

Charge Correction for K32A

The numerical charge correction for mutation K32A was calculated for the set-up without a divalent magnesium ion (Mg^{2+}) between interfacing D54 (see section 8.2). In Table 8.4 the raw free energy changes and the individual correction terms are reported. The direct ΔG_{dir} and polarization ΔG_{pol} correction were calculated based on same snapshots taken in discrete intervals from the trajectory. As it has been pointed out by Reif et al. [2014] and Öhlknecht et al. [2020] the values for ΔG_{dir} and ΔG_{pol} for each snapshot may differ greatly, but the sum $\Delta G_{\text{dir}} + \Delta G_{\text{pol}}$ for each snapshot may exhibit a significantly lower standard deviation. This behavior could be confirmed. The sum was calculated to $\Delta G_{\text{dir}} + \Delta G_{\text{pol}} = -0.41 \pm 3.27\,\text{kJ mol}^{-1}$ for the bound crystal state and $\Delta G_{\text{dir}} + \Delta G_{\text{pol}} = 1.45 \pm 2.95\,\text{kJ mol}^{-1}$ for the unbound solvated state. The values for ΔG_{dir} and ΔG_{pol} for each snapshot for the bound and unbound state are reported in the Appendix C in Table C.5 and C.6, respectively.

Table 8.4 Raw free energy change ΔG_{raw} and correction terms ΔG_{dir}, ΔG_{pol}, and ΔG_{dsm} for mutating *Lb*ADH WT into K32A. The free energy corrections are reported for bound crystal and solvated unbound (unb.) state.

	ΔG_{raw}, kJ mol^{-1}	ΔG_{dir}, kJ mol^{-1}	ΔG_{pol}, kJ mol^{-1}	ΔG_{dsm}, kJ mol^{-1}
bound	200.40 ± 0.22	-54.97	54.56	71.52
unb.	200.19 ± 0.31	14.72	-13.28	72.21

From the individual corrections terms and raw energy change values, the corrected free energy change was calculated according to equations (8.1) and

(8.2). The free energy change calculated with the numerical schemes as well as with the counter ion approach are reported in Table 8.5.

Table 8.5 Corrected free energy changes for mutating *Lb*ADH WT into K32A for different charge correction schemes. The free energy changes are reported for bound crystal and solvated unbound (unb.) state.

	$\Delta G'_{\text{num}}$, kJ mol^{-1}	ΔG_{num}, kJ mol^{-1}	$\Delta G_{\text{co-ion}}$, kJ mol^{-1}
bound	199.99 ± 3.27	271.50 ± 3.27	200.23 ± 0.38
unb.	203.14 ± 2.96	275.35 ± 2.96	199.83 ± 0.29

For both states, the corrections $\Delta G'_{\text{num}}$ and $\Delta G_{\text{co-ion}}$ lay in the same range but the standard deviation in $\Delta G'_{\text{num}}$ was too high for further valid interpretation. The final free energy change for crystallizing *Lb*ADH WT vs. mutant K32A was calculated to $\Delta\Delta G_{\text{num}} = -3.85 \pm 4.40\,\text{kJ mol}^{-1}$ for the numerical correction scheme and $\Delta\Delta G_{\text{co-ion}} = 0.39 \pm 0.48\,\text{kJ mol}^{-1}$ for co-alchemical ion approach. Though the numerical correction scheme exhibited a lower value (as might be expected from experiment) the error was too large for valid interpretation.

Charge Correction for Q126K

The behavior of the correction schemes noted for mutating *Lb*ADH WT into K32A, could also be confirmed in case of mutating WT into Q126K. The correction terms and raw free energy values are reported in Table 8.6. The direct and polar correction summed up to $\Delta G_{\text{dir}} + \Delta G_{\text{pol}} = -1.13 \pm 0.65\,\text{kJ mol}^{-1}$ for the bound crystal state and to $\Delta G_{\text{dir}} + \Delta G_{\text{pol}} = -2.14 \pm 1.14\,\text{kJ mol}^{-1}$ for the unbound solvated state (the individual values for the snapshots are given in Appendix C Table C.2 for bound state contact 1, C.3 for bound state contact 2, and C.4 for unbound solvated state).

Table 8.6 Raw free energy change ΔG_{raw} and correction terms ΔG_{dir}, ΔG_{pol}, and ΔG_{dsm} for mutating *Lb*ADH WT into Q126K. The free energy corrections are provided for bound crystal and solvated unbound (unb.) state.

	ΔG_{raw}, kJ mol^{-1}	ΔG_{dir}, kJ mol^{-1}	ΔG_{pol}, kJ mol^{-1}	ΔG_{dsm}, kJ mol^{-1}
bound	21.13 ± 0.61	17.59	-18.73	-69.28
unb.	18.90 ± 0.46	-26.21	24.07	-70.20

The corrected free energy changes $\Delta G'_{\text{num}}$ and $\Delta G_{\text{co-ion}}$ for both states lay in the same range but again $\Delta G'_{\text{num}}$ exhibited a larger error (c. f. Table 8.7). The

Table 8.7 Corrected free energy changes for mutating *Lb*ADH WT into mutant Q126K for different charge correction schemes. The free energy changes are reported for bound crystal and solvated unbound (unb.) state.

	$\Delta G'_{\text{num}}$, kJ mol^{-1}	ΔG_{num}, kJ mol^{-1}	$\Delta G_{\text{co-ion}}$, kJ mol^{-1}
bound	20.00 ± 0.90	-49.28 ± 0.90	20.72 ± 0.33
unb.	16.76 ± 1.23	-53.44 ±1.23	19.57 ± 0.27

corrected free energy changes for crystallizing *Lb*ADH WT vs. mutant Q126K was calculated to $\Delta\Delta G_{\text{num}} = 4.16 \pm 1.52\,\text{kJ mol}^{-1}$ for the numerical correction scheme and $\Delta\Delta G_{\text{co-ion}} = 1.15 \pm 0.43\,\text{kJ mol}^{-1}$ for co-alchemical ion approach. Both values exhibited the same sign, and the free energy changes were not in the error within each other. However, the numerical scheme exhibited again a much larger error. In all further simulations the free energy changes $\Delta\Delta G$ was calculated with the co-alchemical ion approach.

8.2 Free Energy Changes

In order to theoretically study the influence of the mutation the free energy difference for WT vs. mutant crystallization $\Delta\Delta G$ was calculated via the thermodynamic cycle described in Fig. 5.3. For mutants for which an X-ray structure was available (T102E, D54F, Q126H, Q126K, Q207D; c. f. Table 8.1), the free energy change was calculated in both WT and mutant crystal structure geometries in full and reduced set-ups as described in methods (section 6.3.1). The free energy difference $\Delta\Delta G$ is displayed in Fig. 8.5 (the values are reported in Appendix Table B.1). Free energy calculations in WT crystal structure explored the influence of the mutation as if the mutant crystallized in WT crystal structure; and vice versa. The mutation additionally induced a slight re-arrangement, i. e. coordinate transformation. The associated free energy shift represents the gap between the values for both crystal systems. The real free energy change including a coordinate transformation lies between those two values. The theoretical calculations yielded negative free energy differences for all mutants with experimentally observed increased crystallizability in both crystal packings (WT and mutant) except for mutants Q126K and K32A in WT crystal packing. No defined sign could be assigned within the calculated free energy error for these two mutants

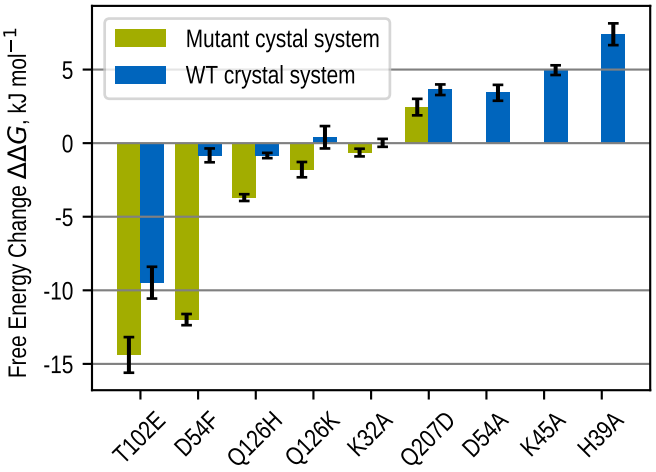

Fig. 8.5 Free energy difference calculation. Free energy change $\Delta\Delta G$ for *Lb*ADH WT vs. mutant crystallization. Green and blue bars indicate calculation in mutant and WT crystal structure, respectively.

in WT crystal packing. For all mutants with decreased crystallizability a positive free energy change was calculated. In case of mutant Q207D, where the free energy change could be calculated in both crystal systems, both values were positive. As the real free energy values lies in between both values of mutant and WT crystal packing calculations it can be deduced that the free energy simulations reproduce the experimental order: T102E ≈ D54F > Q126H > Q126K > K32A > WT > Q207D > D54A, K45A, H39A (c. f. Fig. 8.5).

8.3 Mechanistic Explanations

8.3.1 Transforming a Hydrophilic into a Hydrophobic Contact

A particular interesting contact is found around amino acid position 54 at the edge contact (see Fig. 8.6). In WT *Lb*ADH, two symmetry related aspartic acids (D54) are facing each other with a nearest distance of the carboxylate groups oxygens of 5.3 Å (the other carboxylate group oxygen participates in H-bonding with the hydroxyl group of threonine (T52)) (see Fig. 8.6 **a**). MD simulations showed that the two negatively charged carboxylate groups interact with each other via dynamic water mediated H-bonds. On

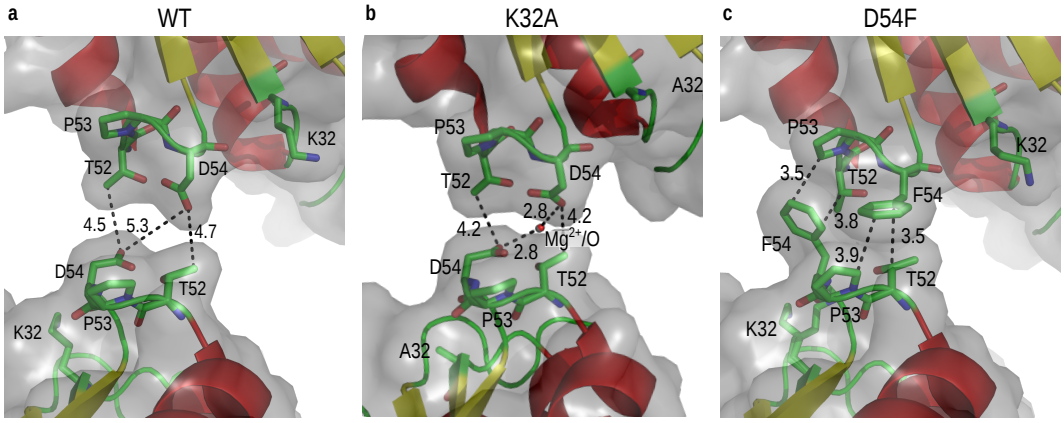

Fig. 8.6 X-ray crystal structure of the edge contact in *Lb*ADH WT, K32A, and D54F crystals. The proteins surface is displayed transparent. **a** In WT crystals the polar side chains leave enough space to enable water mediated interactions across the crystal contact interfaces (wet contact). **b** In K32A crystals the crystal contacts are in closer proximity allowing additional coordination of an Mg^{2+} ion. **c** In D54F crystals, nonpolar chains are in direct contact with no space for water molecules (dry contact). All distances are given in Å.

average 4.5 ± 2.2 water molecules were calculated at the WT crystal contact mediating the attractive interaction of oppositely lying residues (see Fig. 8.4). This is considered as 'wet contact': charged side chains interact with each other via water molecules [Janin, 1999]. In order to verify this attractive interaction, this wet interaction was destroyed by creating mutant D54A. With alanine at position 54 (A54), a positive $\Delta\Delta G$ value was calculated in accordance with no crystallization of D54A in experiments. Mutant K32A also influenced the edge contact. In mutant K32A, it was observed in the measured X-ray structure that to some extent (60%) an interaction is not only mediated by water but also by a divalent magnesium ion (Mg^{2+}). The distance between D54s in K32A was slightly smaller compared to WT and the negatively charged carboxylate groups were able to coordinate Mg^{2+} (see Fig. 8.6 **b**). Besides the attractive interaction mediated by Mg^{2+}, additionally 2.7 ± 1.5 water molecules were counted (see Fig. 8.7 **a** and **b**). The free energy change was calculated with and without alchemically inserting an Mg^{2+} ion at the binding position (excluding its solvation free energy) and weighted according the occupancy measured in the X-ray structure experiments (the individual values are reported in Appendix B Table B.2). Presumably the positively charged, flexible nearby lysine at position 32 (K32) disturbed the crystal contact in WT and substitution of K32 by small alanine (A32) enabled

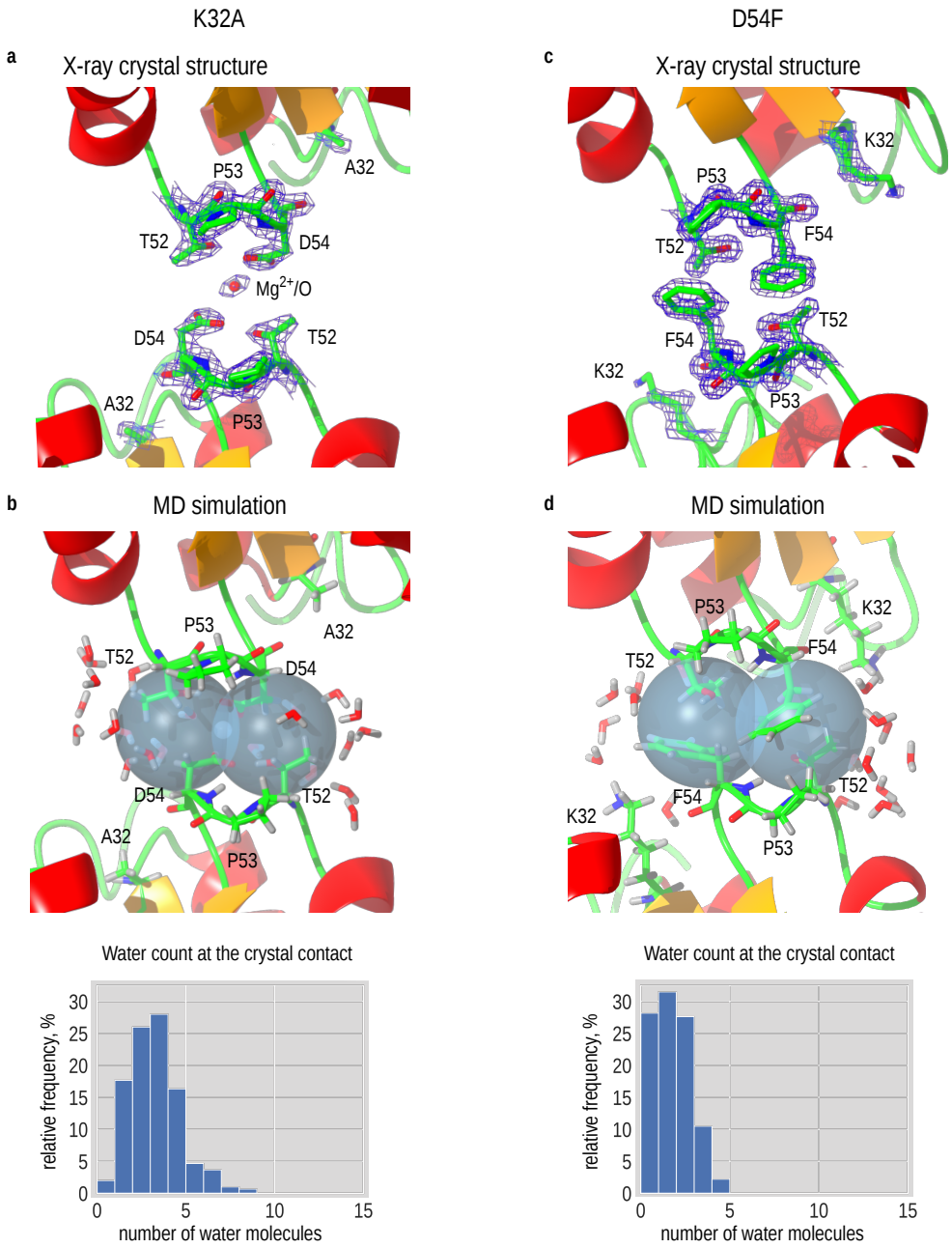

Fig. 8.7 X-ray structure and MD simulation analysis of edge contact around amino acid position 54 in *Lb*ADH mutants K32A and D54F crystals. **a** X-ray structure of the edge contact in K32A crystal including electron density for displayed relevant residues. **b** Snapshot from the MD simulation trajectory. Spheres are constructed around the midpoint of oppositely lying C_α atoms of residues D54 and T52. The water molecules inside the spheres were counted to 2.7 ± 1.5. **c** X-ray structure of the edge contact in D54F crystal. Electron density is contoured blue for all displayed residues. **d** MD simulation analysis yielded 1.3 ± 1.1 water molecules at the crystal contact. Electron density maps are calculated with structure factor amplitude $2F_o - F_c$.

Mg^{2+} binding. The situation was completely changed in mutant D54F (see Fig. 8.6 **c**). The introduced phenylalanines at position 54 (F54s) increased the crystal contact area. The opposing side chains were close together (3.5 Å to 3.9 Å) and did not leave any space for water molecules. Only 1.3 ± 1.1 water molecules were observed (see Fig. 8.7 **b**). The chemical composition at the newly formed interface (threonine (T52), proline (P53), and phenylalanine (F54)) consisted of nonpolar bound atoms. A hydrophobic crystal contact, i. e. a 'dry' contact, was created: water molecules are not present within but only around the contact interface [Janin, 1999].

8.3.2 Fine Tuning a Hydrogen Bonding Interaction

An example for H-bond tuning at a crystal contact could be found at the corner contact at amino acid position 126. In MD simulations of *Lb*ADH WT an H-bond was detected between the hydroxy hydrogen of serine (S40) and the amide oxygen of glutamine (Q126). However, this bond was highly unstable. Assuming a maximum H-bond distance of 2.5 Å, this conformation existed on average in 9.6 ± 3.7% of the time (see Fig. 8.1). In mutant Q126K which exhibited a slightly negative free energy change and experimentally slightly improved crystallization behavior, an H-bond / salt bridge was detected between ammonium group of lysine (K126) and carboxylate group of glutamate (E44) in 66.0±7.7% of the time (see Fig. 8.8 **a** and **b**). The situation was even better in mutant Q126H which exhibited an even more negative free energy change and a better experimental crystallization behavior: an H-bond between imidazole nitrogen of histidine (H126) and hydroxy hydrogen of S40 was present in 74.3 ± 7.3% of the time (see Fig. 8.8 **c** and **d**). At this position, the crystallizability was optimized by fine tuning H-bonds.

8.3.3 Destroying Interactions

Both mutants H39A and K45A (targeting the corner contact) did not yield any crystals in the experiments and MD free energy simulations calculated a considerably positive $\Delta\Delta G$ value compared to WT crystallization. It was found that H39 and K45 mediated attractive interactions in WT *Lb*ADH: X-ray structure analysis and MD simulation of WT *Lb*ADH crystal revealed

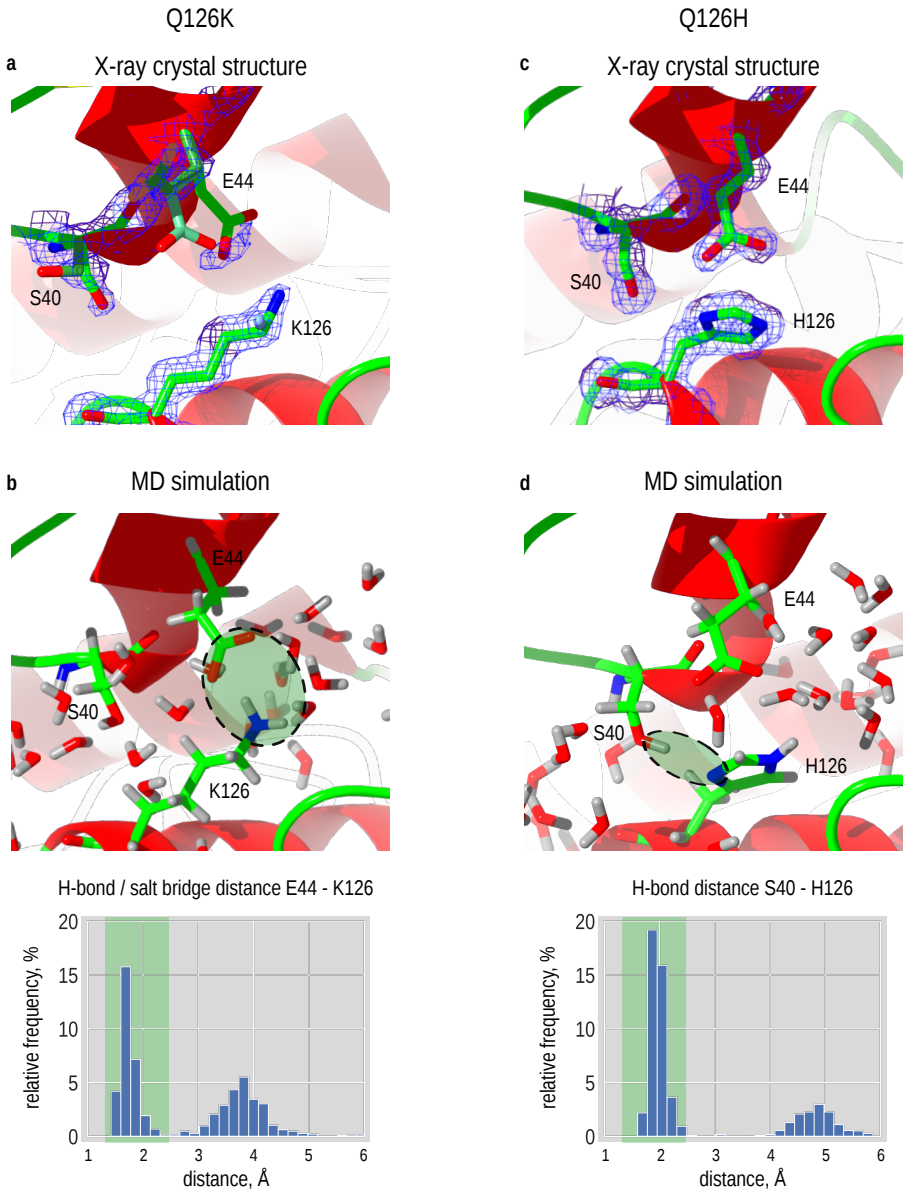

Fig. 8.8 X-ray structure and MD simulation analysis of corner crystal contact around amino acid position 126 in *Lb*ADH mutant Q126K and Q126H crystals. a X-ray crystal structure of Q126K with electron density contoured in blue for displayed residues. Different side chain conformations are displayed in different shades. **b** MD simulation revealed a H-bond / salt bridge between positively charged ammonium group of lysine (K126) and negatively charged carboxylate group of glutamate (E44) in 66.0 ± 7.7% of the time. **c** X-ray crystal structure of Q126H with electron density contoured in blue for displayed residues. **d** Of all variants with mutation at position 126, the most stable H-bond was found in Q126H between imidazole nitrogen of histidine (H126) and hydroxy hyrogen of serin (S40), which was present in 74.3 ± 7.3% of the time. Note: As Q126K and Q126H diffracted in space group *I*222 corner contact '1' and '2' coincide. Electron density maps are calculated with structure factor amplitude $2F_o - F_c$.

that histidine (H39) fills the space with nearby residues arginine (R38) and aspartic acid (D41) to the opposite side with glutamines (Q74 and Q77) and arginine (R127) with no place for water molecules in a 'key/lock'-manner (see Fig. 8.3). The mean distance of 3.7 ± 0.3 Å between the imidazole ring of H39 and guanino group of R127 for both contacts resulted in an attractive binding between those two groups [Heyda et al., 2010]. This interaction was presumably destroyed by substituting histidine by alanine in mutant H39A. In case of non-crystallizable mutant K45A it was found, that an attractive interaction originated from lysine K45 in WT *Lb*ADH. An interaction of the positively charged amino group of K45 with the negatively charged carboxy group of glutamine (Q66) was observed in MD simulation of *Lb*ADH WT crystal (see Fig. 8.2). An H-bond existed on average in 51.8 ± 5.2% of the simulation time inducing a salt bridge between basic K45 and acidic Q66. Mutating lysine into alanine eliminated this attractive interaction.

Mutant Q207D targeted the side contact. A positive free energy difference $\Delta\Delta G$ was calculated which correlated with the experimentally observed reduced crystallizability. In WT crystals, symmetry related glutamines (Q207) are facing each other and were able to interact directly with each other. MD simulations showed that in 48.2 ± 8.8% of the simulation time oxygen atoms and hydrogen atoms of the amine groups are in H-bond distance to each other (see Fig. 8.9 **a** and **b**). Mutating glutamine into aspartic acid destroyed this interaction (see Fig. 8.9 **c**).

8.3.4 Introducing an Ionic Interaction

At the corner contact in WT *Lb*ADH an unpolar threonine (T102) is opposite of lysine (K48). The distance and chemical composition allowed no interaction. T102E has been identified as one of the best performers [Grob et al., 2020] and correlated with the free energy difference calculations. X-ray structure analysis and MD simulation of T102E revealed that the negatively charged carboxy group of glutamic acid (E102) interacted with positively charged ammonium group of lysine (K48). This salt bridge was found to be highly stable. E102 and K48 were in H-bond interaction distance in 92.0 ± 5.1% and 91.5 ± 7.3% of the simulation time at contact 1 and 2, respectively.

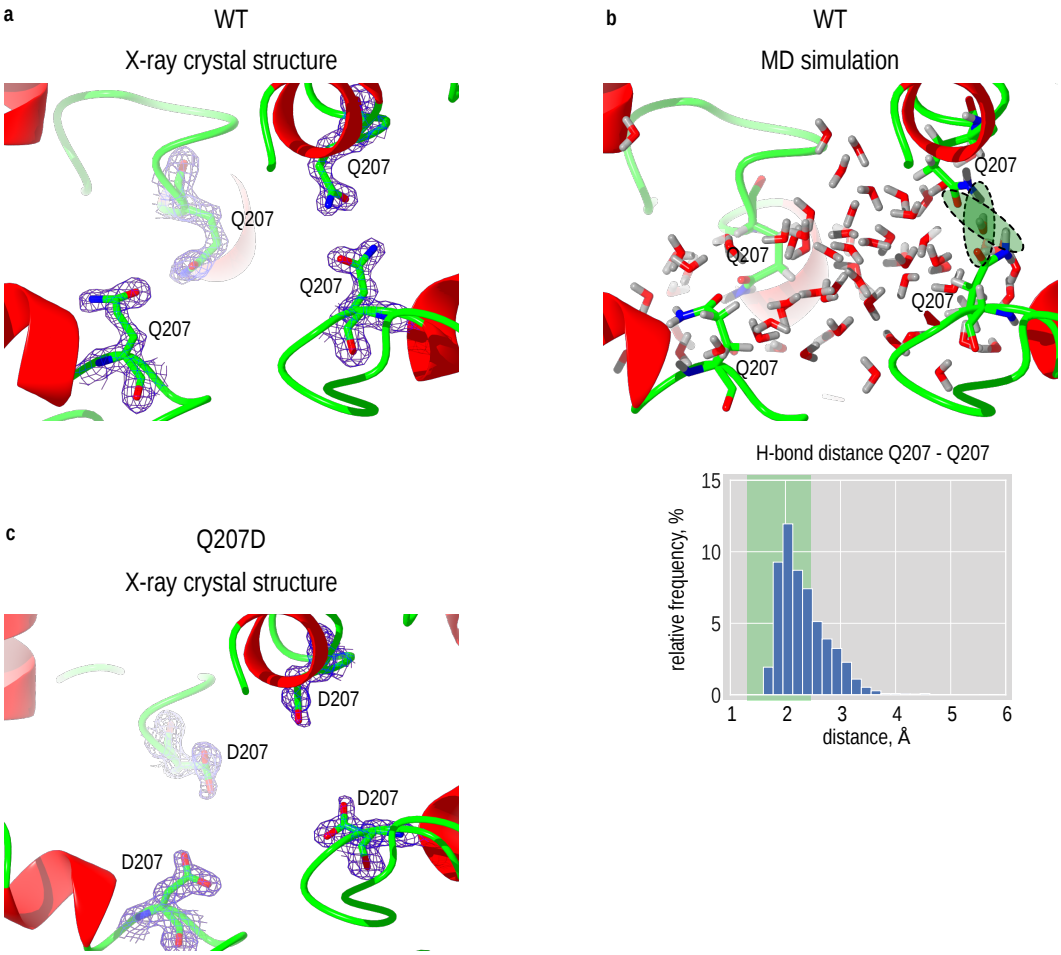

Fig. 8.9 Crystal structure and MD simulation analysis of side contact in *Lb*ADH WT and Q207D crystals. **a** X-ray structure of WT crystal with electron density contoured in blue for displayed residues. **b** MD simulation and analysis of H-bond interactions between interfacing Q207 residues showed an interaction in 48.2 ± 8.8% of the simulation time. H-bond interactions (distance < 2.5 Å) are displayed in green. **c** X-ray structure of Q207D crystal variant with electron density contoured in blue for displayed residues. D207 residues are too far apart from each other to participate in an interaction. Electron density maps are calculated with structure factor amplitude $2F_o - F_c$.

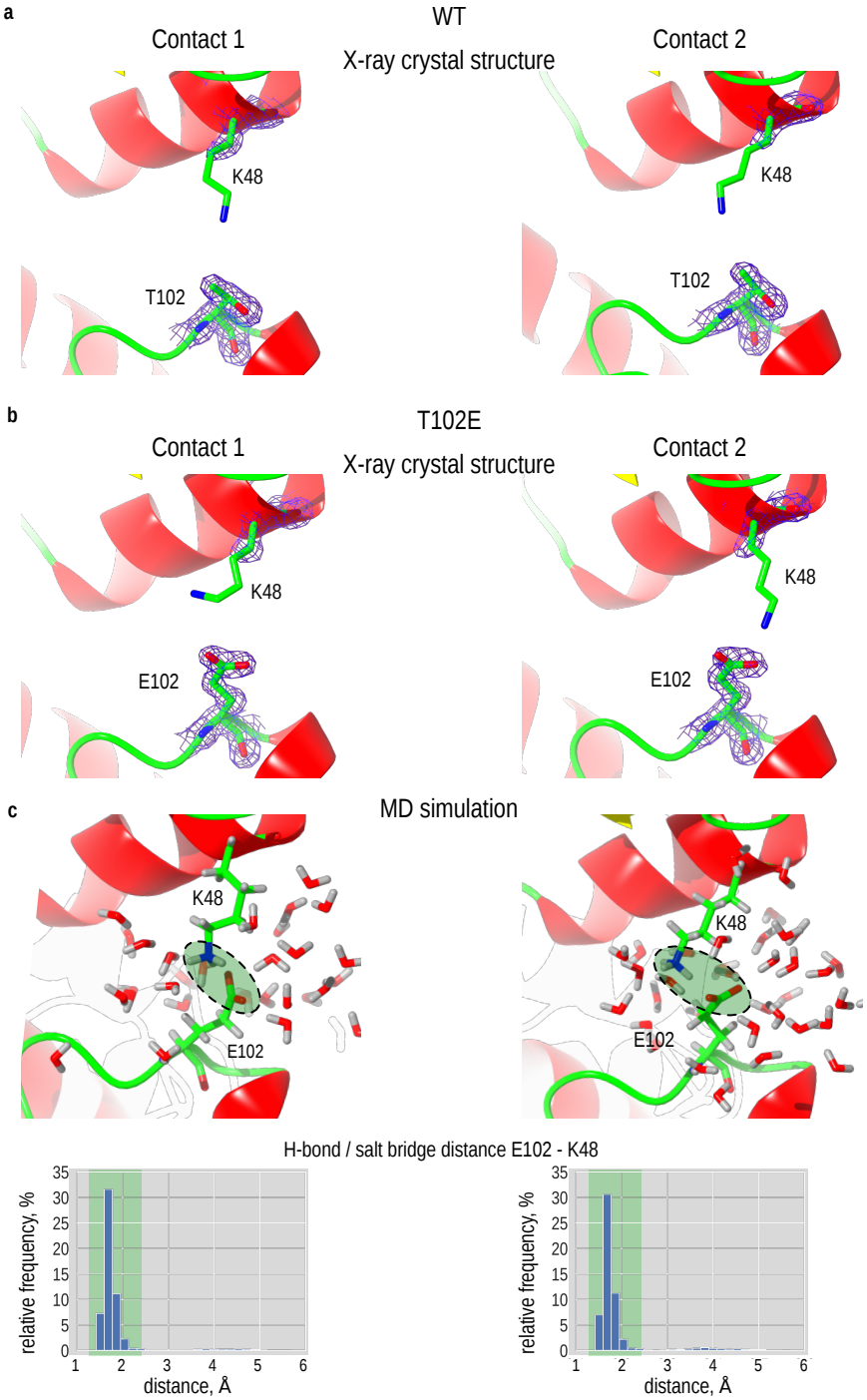

Fig. 8.10 Crystal structure and MD simulation analysis of corner contact around amino acid position 102 in *Lb*ADH WT and T102E crystals. **a** K48 is too far away from opposing T102 to interact. **b** X-ray structure of T102E crystal with measured electron density contoured in blue for displayed residues. **c** MD simulation and analysis of contact 1 and 2 revealed a highly stable H-bond between K48 and E102 in 92.0 ± 5.1% and 91.5 ± 7.3% of the time, respectively. Electron density maps are calculated with structure factor amplitude $2F_o - F_c$.

8.4 Conclusion

8.4.1 System, Set-up and Method

The protein crystal structure was not stable in unrestrained MD simulations due to the weak nature of crystal contacts in contrast to the tetramer itself (crystal building block) which stayed intact. Therefore, the backbone C_α atoms were restrained on their measured positions. Position restraints may introduce artifacts when investigating protein properties in solution but restraining on a structure model retrieved from a crystal. However, when investigating protein crystal properties it is in accordance with biology to restrain on positions measured in diffraction experiments of a crystal. While side chains may be altered during crystal treatment before measurement, the backbone structure is not affected. Therefore, restraining C_α atoms on crystal positions narrows down the phase space sampling by introducing valid boundaries for the proteins' configurations.

In this restrained set-up, interactions in three different WT *Lb*ADH crystals were identical present in MD simulations independent on the temperature of the diffraction experiment (cryocooled X-ray and room temperature neutron diffraction) and the SG of the crystal ($I222$ and $P2_122_1$). Possible alterations in the crystal packing introduced by crystal treatment before measurement did not influence the interactions at the crystal contact which were revealed via MD simulations. Hence, crystal structure models retrieved from cryocooled X-ray diffraction (which is much more practicable than room temperature neutron diffraction) were suitable.

It could be demonstrated that with the developed set-up, FEC simulations yielded consistent results when choosing different starting structures from the systems phase space. As the mutations involve charge changes, FEC results are artifacted due to spurious interactions in alchemical transformations. The system could be reduced to be suited for charge correction schemes.

Although the numerical post correction scheme introduced by Reif et al. [2011; 2014] represents a strict theoretical treatment of spurious interactions, it was found not to be suitable in the case of FECs of crystal systems. Large standard deviations and long calculation time of the Poisson-Boltzmann (PB) solver

made the correction scheme impractical for crystal systems. The correction scheme has been primarily used in small model systems [Öhlknecht et al., 2020; Reif and Oostenbrink, 2014]. The instantaneous correction scheme by keeping a zero net-charge [Chen et al., 2018] required no additional computational time and yielded reasonable results.

8.4.2 Correlation to Experimental Data

The detailed atomistic MD simulations on the basis of X-ray structure analysis precisely revealed interactions at the crystal contacts leading to the calculated free energy difference characterizing the crystallization behavior. A good agreement of theoretically calculated free energy differences $\Delta\Delta G$ and experimentally observed crystallization behavior was determined. Although, a higher free energy difference gap between Q207D and D54A might have been expected, as crystals could be obtained for Q207D but not for D54A. Further calculations and screening conditions might be needed to investigate D54A. However, $\Delta\Delta G$ could be correlated to crystallization behavior and the altered interactions could be traced back and quantified. Free energy difference calculation relies on the end states of the crystal building blocks (solvated unbound and crystal bound state) and is independent on the pathway through the free energy landscape itself. Hence, the crystallization behavior of *Lb*ADH mutants is not dependent on the energy profile of the process itself but can be altered by changing the energy levels of the end states. Thermodynamically, the stability of the crystals was increased or decreased.

8.4.3 Nucleation and Crystal Growth Theories

The results of this study can be viewed under the perspective of classical and two step nucleation, and crystal growth. In classical nucleation theory (CNT) the crystal building blocks are attached to each other in the same orientation as found in a fully grown crystal [Becker and Döring, 1935; Gibbs, 1878; Volmer and Weber, 1926]. I. e. the building blocks bind to each other in the nucleation process via the same protein-protein interaction sites which are later seen as crystal contacts. The cluster formation is thermostatically determined by

enthalpic gain due to the formation of new intermolecular bonds, and an entropic loss due to formation of a cluster-liquid interface surface tension. At a critical cluster size, the free energy landscape reaches a local unstable maximum. Addition of one more building block will doom the nucleus to grow. In CNT the nucleus formation is driven by oriented attachments of the building blocks. Enhancing or decreasing specific stabilizing interactions at the crystal contact may therefore facilitate or hamper growth of a critical nucleus.

In the case of two step nucleation, crystal building blocks form a metastable dense cluster exhibiting an unordered structure before a crystalline ordered nucleus emerges within this cluster [Ten Wolde and Frenkel, 1997]. Crystalline nucleus formation within this cluster relies - as for CNT - on oriented attachment. As this involves two steps (cluster formation and nucleus formation), two energy barriers can be rate defining. Studies have described the nucleation time being dependent on formation of the metastable phase [Kashchiev et al., 2005]. Whereas newer publications found the formation of dense clusters has a much shorter time scale compared to nucleus formation in those clusters. Dense clusters were found to evolve instantaneously [Vekilov, 2010]. These ambiguous findings can be clarified, with the results of this study, if *Lb*ADH crystallized in a two step nucleation: the two-step-nucleation-rate was not dependent on the formation of the unordered metastable phase (first step) but on the formation of a crystalline nucleus (second step). The altered interactions leading to altered crystallizability were demonstrated to modify specific interactions which influence the second step. No hints for altered interactions leading to random agglomeration for the first step were found.

Crystal growth by two-dimensional nucleation and by spiral dislocations have been identified as the main growth mechanisms for crystals [Malkin et al., 1995; McPherson et al., 2003, 2001]. In crystal growth by two-dimensional nucleation, nucleation of new step edges or nucleation islands on the crystal surface is followed by tangential growth and extension of step edges at the surface [Burton et al., 1951]. At the latest, when the new surface layer is completed, i. e. the surface is flat, new step edges or islands are initiated by two-dimensional surface nucleation. At high supersaturation level or small crystal size, two-dimensional growth islands originate randomly all over a crystal surface. With increasing crystal size and decreasing supersaturation level,

two-dimensional surface nucleation becomes restricted to facet edges and, finally, to facet corners [Malkin et al., 1995; Vekilov and Rosenberger, 1996]. In spiral growth mode which is energetically favored at low supersaturations the step edges evolve from screw dislocations [Burton et al., 1951]. These dislocations serve as growth hillocks. Due to the spiral nature, this growth hillocks never disappear. Independent on the exact growth mechanism, a critical important step in both growth methods is the directed incorporation of crystal building blocks into the growing crystal in crystal symmetry order. Crystal building blocks which are not orderly bound would lead the growth to slow down or would lead to an amorphous phase.

8.4.4 Scalability and Transferability of Crystallization

An interesting fact was found by Grob et al. [2020] where the scalability and transferability of improved crystallization was demonstrated: enhanced crystallizing *Lb*ADH variants exhibited the same 'crystallizabiltiy order' independent of crystallization with purified protein in static µL experiments or non-purified protein (dialyzed cell lysate) in stirred 5-mL-crystallizers. Further, the increased crystallizability in the static µL experiments was correlated to enhanced nucleation.

The nucleation process in static µL experiments with purified protein represents homogenous nucleation whereas the nucleation in stirred 5-mL-crystallizers in dialyzed cell lysate resembles heterogeneous nucleation. In heterogenouse nucleation, the nucleation barrier is greatly reduced and nucleation is greatly facilitated [McPherson, 1999]. Hence, the crystallization differences resulting from different nucleation behavior between *Lb*ADH variants are minimized in a heterogenous crystallization experiment. In heterogenous environment, the main difference is traceable to differences in crystal growth behavior. As Grob and co-workers reported the same 'crystallizability order' for enhanced nucleation in homogenous nucleation environment as well as increased crystallization in heterogenous nucleation environment, this indicates that homogenous nucleation and crystal growth are correlated. This can be explained with the results from this study: not the attachment process itself which is clearly different in both crystallization experiments but the end states dictate the overall crystallization ability. In

theory, specific interactions play a decisive role in both nucleation and crystal growth methods. The here presented results suggest, that nucleation and crystal growth are governed by direct interactions.

8.4.5 Engineering Strategies

The results of this study may also impact protein engineering strategies. Surface entropy reduction (SER) to facilitate crystallization relies on the substitution of large charged flexible amino acids by small alanines [Cooper et al., 2007]. Large flexible charged residues (lysines, glutamines and glutamic acids) are replaced by small uncharged alanines to reduce a protein's 'entropy shield' and generate hydrophobic patches at the protein's surface [Cooper et al., 2007; Derewenda, 2004, 2010; Derewenda and Vekilov, 2006; Goldschmidt et al., 2007]. Incorporating and ordering large flexible side chains at crystal contacts requires more entropic effort compared to small side chains. Furthermore, the entropic gain due to release of ordered water molecules during crystallization is improved. Water molecules are more easily ordered at surface patches with small side chains while flexible side chains prevent ordering. This is also viewed and justified empirically, as aspartic acid, glutamic acid, and lysine are most the depleted amino acids in biological protein-protein interfaces [Bordner and Abagyan, 2005; Conte et al., 1999]. Surfaces of water-solvable proteins are generally rich in polar and charged residues, and depleted in hydrophobic residues. Hydrophobic residues are primarily found inside water solvable proteins. Compared to their average occurrence on a protein surface, large hydrophobic and uncharged polar residues are abundant whereas charged residues are depleted at biological interfaces [Bahadur et al., 2004; Bordner and Abagyan, 2005; Chakrabarti and Janin, 2002; Jones and Thornton, 1996; Ofran and Rost, 2003]. In a study with homodimers, heterodimers and multimers Bordner et al. [2005] found the ten most abundant amino acids to be phenylalanine, cysteine, leucine, methionine, isoleucine, tyrosine, tryptophan, valine, and histidine. In the presented study, no preference of dry (hydrophobic) or wet (electrostatic) crystal contacts could be identified. In fact, the two best performers D54F and T102E rely on hydrophobic and electrostatic interactions, respectively. Mutating lysine into alanine in mutant K45A actually destroyed a crystal con-

tact and produced non-crystallizable protein. Introducing lysine at position 126 in mutant Q126K enhanced the crystallization compared to WT *Lb*ADH. Hence, lysines at crystal contacts can also be responsible for productive interactions and can generally not be considered to hamper crystallization. The re-creation of biological contacts in SER is justified by referring to studies investigating biological interfaces [Derewenda and Vekilov, 2006]. However, protein-protein contact interfaces in biological assemblies originate differently and serve different purposes. Bordner et al. [2005] investigated quaternary structure assemblies. The protein assembles itself in its quaternary structure upon production and stays in this assembly as it is decisive for its function. The agglomeration of hydrophobic clusters is an immediate event after protein production. It further reduces the hydrophobicity at the surface and improves the solubility. Introducing hydrophobic patches as suggested by SER accordingly reduces solubility as noted by the SER 'inventors' themselves [Derewenda, 2004]. Hence, agglomeration may be facilitated but crystallization requires specific binding in crystal order. This productive binding should only be present when a supersaturated environment is created, and not randomly according to hydrophobic events. Hydrophobic interactions or effects are unspecific interactions. It has been shown in two-dimensional simulations that unspecific interactions facilitate the general attachment of crystal building blocks but that crystals require an ordered attachment, which is facilitated by specific interactions alone [Whitelam, 2010]. It has been confirmed in this study, that specific targeted interactions dictate a protein's crystallizability. SER has proven to sucessfully produce crystallizable variants of non-crystallizable WT proteins [Derewenda, 2010]. SER might be the weapon of choice for proteins which do not crystallize. However, if a protein already crystallizes, specific, targeted modification of interactions at crystal contacts should be considered.

Chapter 9

Summary

Protein-protein interactions are central in life as they are involved in various processes ranging from metabolic reactions to cell signals. Metabolic enzymatic steps are catalyzed by a complex protein-protein interaction network [Durek and Walther, 2008]. RNA splicing, transcription and translation are initiated by interaction networks [Staley and Woolford Jr, 2009]. While interaction networks are considered to be transient, also stable interactions are found: proteins assemble themselves to form a permanent quaternary structure fundamental for their function [Janin et al., 2008]. A special form of stable protein-protein interactions are protein crystals, where proteins assemble themselves periodically in a crystal lattice. Protein crystallization is useful in technical applications. Preparative protein crystallization resembles an attractive purification method in comparison to conventional, costly chromatographic methods [Hekmat et al., 2015; Smejkal et al., 2013]. However, protein crystallization is normally disfavored by evolution [Doye et al., 2004] as it leads to pathological phenomena (e. g. eye cataract [Pande et al., 2001; Siezen et al., 1985]). Protein-protein interaction interfaces found in protein crystals and biological complexes differ from each other [Bahadur and Zacharias, 2008]. It is desirable to gain control over the crystallization ability of a protein in order to prevent diseases or utilize protein crystallization in technical applications. High empirical effort is required to change a protein's crystallization behavior as the mechanisms and interactions driving protein crystallization on an atomic level are largely unknown. Semi-rational

engineering strategies to improve a protein's crystallization ability are often based on empirical data of biological protein-protein interfaces aiming to rebuild biological contacts.

The aim of this study was to model protein crystallization *in silico* in order to be able to unravel mechanistic interactions, and to calculate thermodynamic quantities as a measure of a protein's crystallization ability. Therefore, structure models of crystals were obtained by means of neutron and X-ray diffraction for subsequent investigation with molecular dynamics (MD) simulations (c. f. Fig. 9.1). The study was conducted on alcohol dehydrogenase from *Lactobacillus brevis* (*Lb*ADH) and mutants thereof which were produced, crystallized and characterized concerning their crystallizability in previous works [Grob et al., 2020; Nowotny et al., 2019].

In the first part of this thesis, crystallographic analysis of *Lb*ADH wild type (WT) and mutant crystals was conducted (see Fig. 9.1 **a** and **b**). *Lb*ADH assembled itself in a quaternary structure of a homotetramer. These tetramers were the natural crystal building blocks and two tetramers built up a crystallographic unit cell (UC). Room temperature neutron diffraction of *Lb*ADH WT yielded space group (SG) $I222$. After cryocooling, this crystal diffracted in SG $P2_122_1$ in an X-ray diffraction experiment. Cryocooling induced a phase transition within the orthorombic SGs from $I222$ to $P2_122_1$. Both crystal systems yielded the same crystal packing. This means that the crystal contacts, i. e. interaction interfaces between crystal building blocks, were composed of the same amino acids. No major re-arrangement upon cryocooling as reported in a study of Juers and Metthews [2001] was found, but a representative capturing of conformations could be confirmed [Rader and Agard, 1997]. The crystal systems differed by rotation of the tetramer against each other. While in SG $I222$ the two tetramers in the UC were translational invariant, in SG $P2_122_1$ the tetramers were slightly rotated against each other on average by $\gamma = 4.92 \pm 0.31°$ around the y-axis. Additional to the 'WT X-ray $P2_122_1$ crystal' and the 'WT neutron $I222$ crystal', a 'WT X-ray $I222$ crystal' was found. All other *Lb*ADH mutant crystals were found in either one of the two SGs $I222$ and $P2_122_1$ with no indication of a preferred SG. Structural analysis of all crystals yielded highly similar backbone C_α atom positions for the monomers and tetramers. The root-mean-square deviations (RMSDs) calculated to 0.17 ± 0.03 Å and 0.22 ± 0.05 Å, respectively. Following the RMSD

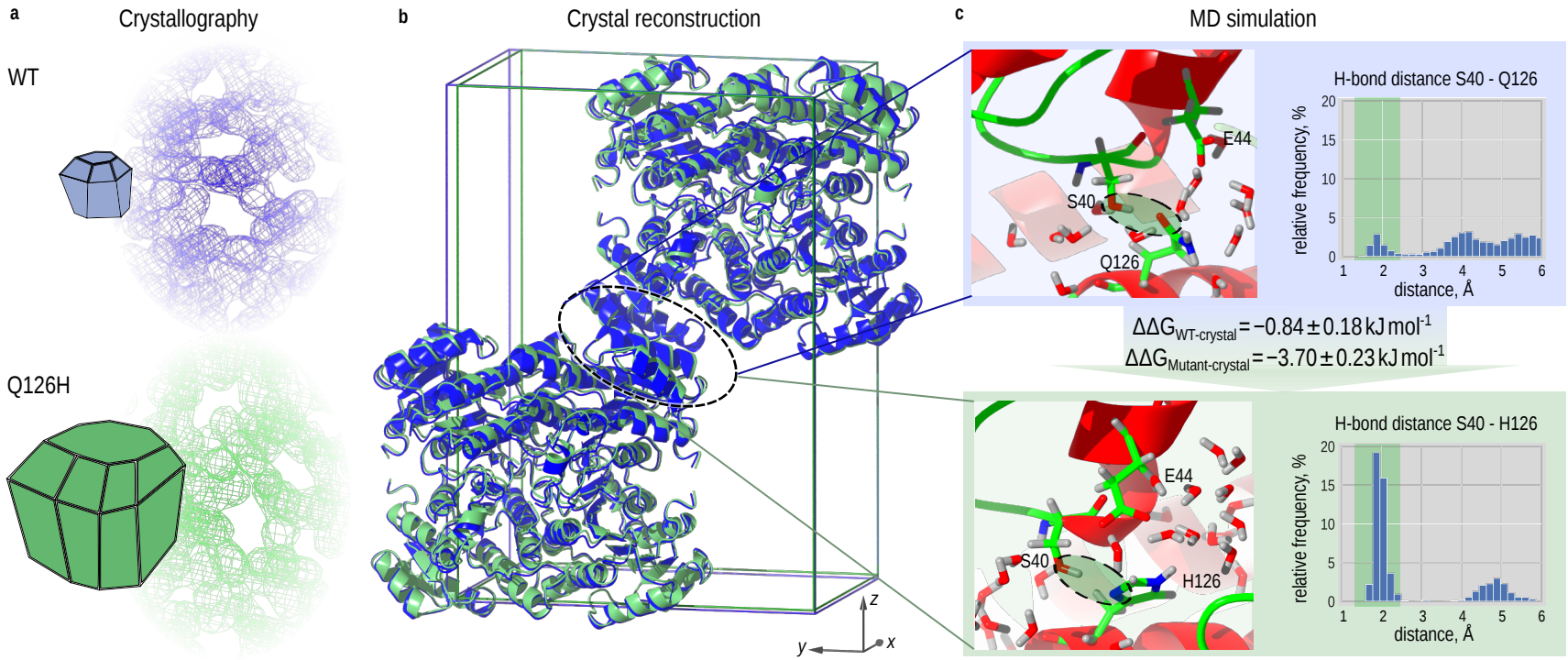

Fig. 9.1 Altered crystallizability investigated with crystallographic and MD methods exemplarily for *Lb*ADH WT and mutant Q126H. *Lb*ADH WT (blue) and mutant Q126H (green) showed different crystallizability (e. g. different crystal size) in previous experimental work. **a** The crystals were crystallographically analyzed and a structure model encoded in electron density (blue and green mesh) was iteratively reconstructed from the diffraction data. **b** The *in silico* reconstructed crystal UCs exhibited high structural similarity though different SGs were found: in WT $P2_122_1$ crystal the two tetramers in the UC were slightly rotated against each other by 5.2° around the *y*-axis, while they were translational invariant in Q126H *I*222 crystal. **c** Parametrization of the structure models allowed MD simulations where a favor of mutant over WT crystallization was calculated with free energy differences of -0.84 ± 0.18 and -3.70 ± 0.23 kJ mol^{-1} (WT and Q126H crystal system). The mutation increased the hydrogen bonding at the crystal contact (9.6 ± 3.7% between S40 and Q126H in WT vs. 74.3 ± 7.3% between S40 and H126 in Q126H).

interpretation of Carugo et al. [2003; 2001], this indicated undistinguishable structures, i. e. identical crystal building blocks, within the resolution limits in the associated diffraction experiments. RMSD calculations for the whole UC after translational alignment yielded values slightly above values explainable with the diffraction limits. Hence, highly similar structures with the same crystal packing. From a geometrical point of view, all X-ray diffraction data of cryocooled crystals yielded the same crystal packing and could be used to build valid structure crystal models. Though, cryocooling induced a phase transition within the orthorombic SGs, it did not change the overall crystal packing. Two energy minima in the crystal arrangement of *Lb*ADH were present. As in X-ray diffraction both SGs were observed for *Lb*ADH WT crystals, the free energy landscape of SGs in *Lb*ADH crystals is likely to be independent on the *Lb*ADH variant. Physically, two explanations for the occurrence of two SGs could be found: (1) *Lb*ADH crystallized in both SGs $I222$ and $P2_122_1$, or (2) *Lb*ADH always crystallized in SG $I222$ and the free energy landscape was altered upon cryocooling enabling the tetramers to rotate in SG $P2_122_1$ in some cases. In previous studies of Schlieben et al. [2005] and Niefind et al. [2003] all seven published *Lb*ADH WT and mutant crystals exhibited $I222$ symmetry. The *Lb*ADH investigated in this study has an additional hexahistidine (His_6) purification tag and a glycine-serine-glycine (GSG) linker genetically fused at the N-terminus. Hence, it might be possible that the His_6 tag and the GSG linker introduced flexibility in *Lb*ADH crystal building blocks which enabled two different SGs in this study.

In the second part of this thesis, the structural models were parameterized and MD simulations were performed (see Fig. 9.1 **b** and **c**). It was found that the contacts within the tetramer (crystal building blocks) were stable in contrast to the crystal contacts, i. e. the interaction interfaces between the tetramers. The protein-protein interactions at the crystal contacts were too weak for simulation in an unrestrained set-up. Therefore, the C_α atoms were weakly restrained on the positions which were measured in the previous diffraction experiments. Thereafter, three *Lb*ADH WT crystals were simulated: two $I222$ crystals reconstructed from room temperature neutron and cryocooled X-ray diffraction and a $P2_122_1$ crystal reconstructed from cryocooled X-ray diffraction data. All three crystals exhibited the same interaction patterns at the crystal contacts. Interaction patterns in MD simulations were not artifacted

by cryocooling and independent on the SGs. Free energy calculations (FECs) were conducted to explore the free energy difference $\Delta\Delta G$ between crystallizing *Lb*ADH WT vs. mutant. The free energy difference $\Delta\Delta G$ was calculated with free energy perturbation (FEP) by alchemically transforming *Lb*ADH WT into mutant. Some mutations involved charge changes and the electrostatic interactions were treated with particle mesh Ewald (PME). This induced spurious effects influencing alchemical FEC in systems with net charge change and required correction schemes to be applied. In order for these correction schemes to work, the simulated system needed a special form: it could be shown that the crystal system can be effectively reduced to simulating only the monomers involved in the crystal contact. It was found that an instantaneous correction scheme, where the net-charge is kept constant by simultaneously transforming a co-alchemical ion [Chen et al., 2018], was superior to a post simulation numerical correction scheme [Reif and Hünenberger, 2011; Reif and Oostenbrink, 2014].

The successful validation of system and set-up provided the basis for the main purpose of this study: the free energy differences $\Delta\Delta G$ for crystallizing *Lb*ADH WT vs. mutant were calculated as a potential *in silico* measure of protein crystallization. Additionally, interactions on an atomic level were identified which caused the free energy change and dictated the crystallization behavior. For all simulated *Lb*ADH variants, the calculated free energy change $\Delta\Delta G$ correlated to the experimentally observed crystallization behavior. For all *Lb*ADH mutants the changed crystallizability compared to WT *Lb*ADH could be traced back to specific interactions at the crystal contacts. Negative free energy changes of $-14.39 \pm 1.21\,\mathrm{kJ\,mol^{-1}}$ and $-11.99 \pm 0.38\ \mathrm{kJ\,mol^{-1}}$ for best performers T102E and D54F could be calculated, respectively. Non-crystallizable mutants K45A and H39A yielded considerably positive free energy changes of $4.96 \pm 0.33\ \mathrm{kJ\,mol^{-1}}$ and $7.40 \pm 0.74\ \mathrm{kJ\,mol^{-1}}$, respectively. It could be concluded that the crystallizability was increased or decreased by stabilizing or destabilizing the crystal.

Thermodynamically, the process of protein crystallization resembles a phase transition from a solvated into a crystal state. The phase transition is carried out in two subsequent processes: nucleation and crystal growth.
The thermodynamical findings in this thesis are consistent with classical nucleation theory (CNT) [Becker and Döring, 1935; Gibbs, 1878; Volmer and

Weber, 1926] and two step nucleation theory [Ten Wolde and Frenkel, 1997] and allow to clarify an ambiguity in two step nucleation theory. In CNT, molecules or crystal building blocks are bound in crystal order to form a crystalline nucleus. Hence, the nucleus can only reach a critical size if the building blocks bind to each other at the same interfaces which are later identified in the crystal as crystal contacts. In two step nucleation theory, the crystal building blocks form a dense unordered cluster before a crystalline nucleus emerges within this cluster. It was proposed, that the nucleation time is determined by the cluster formation time [Kashchiev et al., 2005] or by the nucleus formation time within this cluster [Vekilov, 2010]. In this study, a correlation of the nucleation to free energy changes induced by specific interactions was found. For example, mutant Q126H crystallization exhibited a negative free energy change of $-3.70 \pm 0.23\,\mathrm{kJ\,mol^{-1}}$ compared to WT crystallization. This could be traced back to an increased hydrogen bonding at the crystal contact ($9.6 \pm 3.7\%$ in WT vs. $74.3 \pm 7.3\%$ in Q126H, see also Fig. 9.1 **b** and **c**). All interactions were unambiguously traceable to specific interactions which energetically stabilized the crystal contacts. No hints for unspecific interactions which would lead to random agglomeration were found. Hence, this is within the picture of CNT and it strongly indicates that the nucleation time is dictated by nucleus formation within the cluster, if a two step nucleation process is present. The importance of the identified energetically stabilizing interactions at crystal contacts can also be found in crystal growth theories: crystals grow by two-dimensional nucleation and spiral dislocations. Evidently, the nucleation process on a crystal surface is determined by proteins which are bound in crystal order. And spiral dislocations can only serve as crystal hillocks if the proteins are orderly attached. Hence, a negative free energy change and associated stabilizing interactions may facilitate both, crystal nucleation and crystal growth.

With the previous theoretical considerations, the scalability and transferability of improved crystallization which was found by Grob et al. [2020] can be explained. Grob and co-workers ranked the crystallizability of *Lb*ADH WT and mutants in µL crystallization experiments with purified protein and could correlate the crystallizability to nucleation ability of the protein. They found that the crystallizability order did not change when non-purifed protein (cell lysate) is crystallized in stirred mL crystallizers.

The crystallization of purified protein in µL-scale is initiated by homogenous nucleation whereas the nucleation of non-purified protein in stirred mL-scale is of heterogenous nature. In heterogenous environment the energy barrier for nucleation is greatly reduced [McPherson, 1999]. Nevertheless, an astonishing clear correlation between both crystallization experiments was found by Grob et al. [2020]. The enhanced nucleation found in µL-scale cannot explain the enhanced crystallizability in stirred mL-scale crystallization. The preservation of the crystallizability order therefore indicates that the calculated free energy and the associated interactions found in this study impact both: crystal (homogeneous) nucleation and crystal growth. They are both determined by specific energetically stabilizing interactions at crystal contacts. This may also impact crystallization modeling as it further legitimates modeling crystal growth according to homogenous classical nucleation as e. g. in Schmit and Dill [2012].

A lot of effort has been put into the *in silico* control of protein crystallization. A prominent protein engineering strategy is surface entropy reduction (SER) where large flexible amino acids on the protein surface like lysines, glumatines and glutamic acids are mutated into small alanines to create hydrophobic patches [Derewenda, 2004]. In theory, this reduces the entropy loss of the protein by avoiding incorporating large flexible amino acids and maximizes the solvent entropy gain by the release of solvent molecules from newly generated hydrophobic alanine patches at the surface. SER aims to rebuild biological protein-protein interfaces which are abundant of hydrophobic residues and depleted of charged residues [Derewenda and Vekilov, 2006]. However, in this study the two best performers relied on electrostatic interactions (T102E) and hydrophobic interactions (D54F), respectively. Additionally, it was shown that specific electrostatic interactions like hydrogen bonds significantly influenced the crystallizability. For example, lysine at position 45 in *Lb*ADH WT mediated a stable salt bridge in 51.8 ± 5.2% of the simulation time. Replacing lysine by alanine in mutant K45A resulted in non-crystallizable proteins. The discrepancy of these results to the suggestions in SER can be traced back to the different nature of biological and crystal protein-protein interfaces [Bahadur and Zacharias, 2008]. For example, biological assemblies like tetrameters (where the interfaces are abundant in hydrophobic amino acids and depleted in charged amino acids) evolve after protein production and along side

protein folding. In protein folding, hydrophobic residues are arranged inside the protein and hydrophilic residues are exposed at the protein surface in a spontaneous process. Protein crystallization is intended when a protein solution is forced into a supersaturated environment. Random spontaneous crystallization would lead to sever pathological phenomena. SER has proven to sucessfully produce crystallizable variants of non-crystallizable WT proteins [Derewenda, 2010]. When an already crystallizable protein is engineered in order to modify its crystallization behavior, specific interactions should be targeted. This is in contrast to SER where only unspecific hydrophobic interactions are introduced. The importance of energetically stabilizing direct interactions has been demonstrated in this study. Furthermore, the success of the protein engineering effort can be evaluated *in silico* by free energy change calculations. This free energy guided design may control protein crystallization for technical application as well as for preventing pathological phenomena.

Chapter 10

Outlook

In the presented study it was shown that protein crystallization is guided by specific interactions at crystal contacts and that a protein's crystallizability can be *in silico* calculated with molecular dynamics (MD) free energy simulations.

This can be used in future work to guide protein engineering for controlling crystallization. The engineering can be exclusively reduced to mutations targeting the stability of the crystal contact: e. g. introducing opposite charges at crystal contacts and 'prolonging' amino acids (e. g. mutating aspartic acid into glutamic acid) to generate a shorter interaction distance. When selecting suitable mutations based on a wild type (WT) crystal structure, it should be kept in mind that the arrangement at the crystal contact may vary in the sub-Ångstrom scale to adjust itself. A mutation which seem not to result in a perfect bonding distance (e. g. for a newly introduced hydrogen bond) according to the WT crystal structure, may induce a slight re-arrangement for a perfect bonding distance in the mutants crystal structure. Mutations targeting the whole surface of the protein and not considering crystal contacts (as suggested by the surface entropy reduction (SER) strategy) can be neglected. However, the general idea of SER, introduction of hydrophobic patches, can be pursued but should be limited to crystal contact areas. Introducing hydrophobic patches might be particularly constructive at symmetric contacts. At symmetric contacts, two identical amino acids which are symmetry related are interacting with each other. Introducing too many hydrophobic residues at a protein's surface reduces the proteins solubility and leads to agglomeration.

At a symmetric contact the whole (or at least a great part) of a crystal contact is modified with only one mutation. Hence, symmetric contacts are suitable for introducing hydrophobic effects to promote crystallization without potential loss of the protein's solubility. With these practical engineering suggestions a great number of protein variants can be designed. However, experimental investigation of all potential variants is costly and time consuming. The major advantage in future work is that the success of the protein design can be evaluated *in silico*. Free energy guided design of protein variants was found to be a powerful new tool to control protein crystallization. For variants with greatly enhanced or reduced crystallizability, a successful characterization is possible based on the WT protein structure alone (without prior knowledge of the variants structure). The elaborate, costly experimental work of genetic protein modification, protein production, and crystallization experiments can be reduced to a few variants which are pre-selected with free energy MD calculations.

Furthermore, MD simulations might be used to design protein crystal properties important for technical crystallization in down stream processing. Preparative protein crystallization in clarified cell lysate requires subsequent filtration or centrifugation to extract the protein crystals. Crystals are exposed to mechanical stress during solid-liquid separation. Size and shape of crystals were shown to influence the filtration behavior [Radel et al., 2019]. MD simulations may be used not only to control a protein's crystallizability but also to control the crystal properties. Growth in specific directions to influence the crystal shape may be calculated *in silico*. The same applies to the calculation of crystal packing to allow the crystal to endure mechanical stress. The developed tools and concepts in this study, may provide the basis for the design of protein properties for down stream processing.

Glossary

Abreviations

A	Alanine.
AFM	Atomic force microscopy.
AU	Asymmetric unit.
BAR	Bennett acceptance ratio.
CNT	Classical nucleation theory.
COG	Center of geometry.
CPU	Central processing unit.
CRC	Catalysis Resarch Center.
D	Aspartic acid.
E	Glutamic acid.
EM	Energy minimization.
ESRF	Synchrotron radiation facility.
F	Phenylalanine.
FEC	Free energy calculation.
FEP	Free energy perturbation.
FT	Fourier transform.
FWHM	Full width at half maximum.
G	Glycine.
GPU	Graphics processing unit.
GROMACS	Groningen Machine for Chemical Simulations.
GROMOS	Groningen Molecular Simulation.
GSG	Glycine-serine-glycine linker.

H	Histidine.
H-bond	Hydrogen bond.
H-RE	Hamiltonian replica exchange.
His_6	Hexahistidine purification tag.
IFT	Inverse Fourier transform.
JCNS	Juelich Centre for Neutron Science.
K	Lysine.
*Lb*ADH	Alcohol dehydrogenase from *Lactobacillus brevis*.
LRZ	Leibniz-Rechenzentrum.
MD	Molecular dynamics.
MLZ	Heinz Maier-Leibnitz Zentrum.
MR	Molecular replacement.
NADH	Nicotinamide adenine dinucleotide.
NMR	Nuclear magnetic resonance.
P	Proline.
PB	Poisson-Boltzmann.
PBC	Periodic boundary condition.
PDB	Protein data bank.
PEG	Polyethylene glycol.
PISA	Protein Interfaces, Surfaces and Assemblies.
PME	Particle mesh Ewald.
Q	Glutamine.
RMSD	Root-mean-square deviation.
S	Serine.
SER	Surface entropy reduction.
SERp server	Surface entropy reduction prediction server.
SG	Space group.
SLS	Swiss Light Source.
T	Threonine.
TEM	Transmission electron microscopy.
TI	Thermodynamic integration.
TLS	Translation/Libration/Screw refinement parameters.
TUM	Technical University of Munich.
UC	Unit cell.
VMD	Visual molecular dynamics.
WT	Wild type.
Y	Tyrosine.

Greek Symbols

α	Mosaicity of a crystal.
ϵ	Relative dielectric permittivity.
ϵ_0	Vacuum dielectric permittivity.
ϵ_{ij}	Lennard-Jones parameter.
γ	Rotation of tetramers with respect to each other..
γ_S	Quadrupole moment trace of the solvent molecule.
λ	Wavelength of incident wave.
λ	Alchemical coupling parameter.
ω	Dihedral angle between covalently bound atoms in an MD potential.
$\varphi(hkl)$	Phase information in reciprocal space.
$\rho(\mathbf{r})$	Electron density distribution.
θ_G	Glancing angel of incident wave.
θ	Angle between covalently bound atoms in an MD potential.
θ_0	Equilibrium angle between covalently bound atoms in an MD potential.
τ_T	Relaxation time in thermostats.
ζ	Phase for dihedral angle in an MD potential.

Latin Symbols

c	Speed of light.
d	Interplanar distance in crystals.
$\mathrm{d}_{\mathrm{min}}$	Resolution limit in diffraction experiments.
E	Energy.
e	Elementary charge.
F	Helmholtz free energy.
$\mathbf{F}(hkl)$	Structure factor.
$I(hkl)$	Intensity of the diffraction spot.
$F(hkl)$	Structure factor amplitude.
$F_{\mathrm{c}}(hkl)$	Calculated structure factor amplitude.
$F_{\mathrm{o}}(hkl)$	Observed structure factor amplitude.

G	Gibbs free energy.
ΔG	Free energy change in a thermodynamic process.
$\Delta\Delta G$	Change of free energy change in a thermodynamic process.
ΔG_{cor}	Correction term for free energy change.
ΔG_{dir}	Correction term for spurious direct interactions.
ΔG_{dsm}	Correction term for incorrect calculation of zero solvent potential.
ΔG_{pol}	Correction term for spurious solvent polarization.
ΔG_{raw}	Raw uncorrected free energy change.
$\mathcal{H}$	Classical Hamiltonian of a system.
H	Enthalpy.
ΔH	Enthalpy change.
h	Planck's constant.
h, k, l	Miller indices, reciprocal space parameters.
i, j	Particle indices.
k_a	Force constant for angle in an MD potential.
k_{B}	Boltzmann constant.
k_b	Force constant for bond length parameter in an MD potential.
k_d	Scaling factor for dihedral angle in an MD potential.
l	Bond length between covalently bound atoms in an MD potential.
l_0	Equilibrium bond length between covalently bound atoms in an MD potential.
m	Particle mass.
N	Number of particles in the system.
n	Integer number.
N_{A}	Avogadro constant.
NPT	Isothermal–isobaric ensemble.
NVE	Microcanocial ensemble.
NVT	Canonical ensemble.
p	Pressure.
$\boldsymbol{p}$	Momentum vector.
Q	Canonical partition function.
Δq	Charge change of a system.
q	Point charge.
$\boldsymbol{q}$	Coordinate vector.
r	Unit cells volume ratio.

$\boldsymbol{r}$	Canonical coordinates.
r_0	Lennard-Jones parameter.
R_{free}	Residual / reliability factor for test set.
r_{ij}	Distance between two atoms.
R_{work}	Residual / reliability factor for work set.
S	Entropy.
ΔS	Entropy change.
$\Delta S_{\text{protein}}$	Entropy change of the protein.
$\Delta S_{\text{solvent}}$	Entropy change of the solvent.
t	Time.
Δt	Time step.
T	Temperature.
T_0	Reference temperature in thermostats.
U	Internal energy.
V	Potential in molecular dynamics simulation.
V	Volume.
v	Velocity.
$\mathbf{v}, \mathbf{w}$	Vectors.
V_{bonded}	Potential for bonded interaction in MD simulation.
$V_{\text{non-bonded}}$	Potential for non-bonded interaction in MD simulation.
W_i	Weighting factor for numerical integration.
X	Thermodynamic observable.
x, y, z	Coordinate system directions.

List of Figures

List of Tables

Bibliography

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., et al. (2010). Phenix: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):213–221.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Echols, N., Headd, J. J., Hung, L.-W., Jain, S., Kapral, G. J., Kunstleve, R. W. G., et al. (2011). The phenix software for automated determination of macromolecular structures. *Methods*, 55(1):94–106.

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica Section D: Biological Crystallography*, 68(4):352–367.

Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. (1990). ph-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of t4 lysozyme. *Biochemistry*, 29(9):2403–2408.

Asthagiri, D., Lenhoff, A., and Gallagher, D. T. (2000). Role of competitive interactions in growth rate trends of subtilisin s88 crystals. *Journal of crystal growth*, 212(3-4):543–554.

Bahadur, R. and Zacharias, M. (2008). The interface of protein-protein complexes: analysis of contacts and prediction of interactions. *Cellular and Molecular Life Sciences*, 65(7-8):1059–1072.

Bahadur, R. P., Chakrabarti, P., Rodier, F., and Janin, J. (2004). A dissection of specific and non-specific protein–protein interfaces. *Journal of molecular biology*, 336(4):943–955.

Barlow, D. J. and Thornton, J. (1983). Ion-pairs in proteins. *Journal of molecular biology*, 168(4):867–885.

Basu, S. K., Govardhan, C. P., Jung, C. W., and Margolin, A. L. (2004). Protein crystals for the delivery of biopharmaceuticals. *Expert opinion on biological therapy*, 4(3):301–317.

Becker, R. and Döring, W. (1935). Kinetische behandlung der keimbildung in übersättigten dämpfen. *Annalen der Physik*, 416(8):719–752.

Bennett, C. H. (1976). Efficient estimation of free energy differences from monte carlo data. *Journal of Computational Physics*, 22(2):245–268.

Berendsen, H. J., Postma, J. v., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8):3684–3690.

Berendsen, H. J., van der Spoel, D., and van Drunen, R. (1995). Gromacs: a message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56.

Blakeley, M. P., Langan, P., Niimura, N., and Podjarny, A. (2008). Neutron crystallography: opportunities, challenges, and limitations. *Current opinion in structural biology*, 18(5):593–600.

Bordner, A. J. and Abagyan, R. (2005). Statistical analysis and prediction of protein–protein interfaces. *Proteins: Structure, Function, and Bioinformatics*, 60(3):353–366.

Burton, W.-K., Cabrera, N., and Frank, F. (1951). The growth of crystals and the equilibrium structure of their surfaces. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 243(866):299–358.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101.

Camattari, R., Guidi, V., Bellucci, V., and Mazzolari, A. (2015). Thequasi-mosaic'effect in crystals and its applications in modern physics. *Journal of Applied Crystallography*, 48(4):977–989.

Carson, M., Johnson, D. H., McDonald, H., Brouillette, C., and DeLucas, L. J. (2007). His-tag impact on structure. *Acta Crystallographica Section D: Biological Crystallography*, 63(3):295–301.

Carugo, O. (2003). How root-mean-square distance (rmsd) values depend on the resolution of protein structures that are compared. *Journal of applied crystallography*, 36(1):125–128.

Carugo, O. and Pongor, S. (2001). A normalized root-mean-spuare distance for comparing protein three-dimensional structures. *Protein science*, 10(7):1470–1473.

Chakrabarti, P. and Janin, J. (2002). Dissecting protein–protein recognition sites. *Proteins: Structure, Function, and Bioinformatics*, 47(3):334–343.

Charcot, J. M. and Robin, C. (1853). Observation of leukocytosis. *TCRMem. Soc. Biol.*, 4:450–454.

Charron, C., Kern, D., and Giegé, R. (2002). Crystal contacts engineering of aspartyl-trna synthetase from thermus thermophilus: effects on crystallizability. *Acta Crystallographica Section D: Biological Crystallography*, 58(10):1729–1733.

Chen, W., Deng, Y., Russell, E., Wu, Y., Abel, R., and Wang, L. (2018). Accurate calculation of relative binding free energies between ligands with different net charges. *Journal of chemical theory and computation*, 14(12):6346–6358.

Christen, M., Hünenberger, P. H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D. P., Heinz, T. N., Kastenholz, M. A., Kräutler, V., Oostenbrink, C., et al. (2005). The gromos software for biomolecular simulation: Gromos05. *Journal of computational chemistry*, 26(16):1719–1751.

Chung, S., Shin, S.-H., Bertozzi, C. R., and De Yoreo, J. J. (2010). Self-catalyzed growth of s layers via an amorphous-to-crystalline transition limited by folding kinetics. *Proceedings of the National Academy of Sciences*, 107(38):16536–16541.

Collaborative, C. P. et al. (1994). The ccp4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography*, 50(Pt 5):760.

Conte, L. L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. *Journal of molecular biology*, 285(5):2177–2198.

Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M., and Derewenda, Z. (2007). Protein crystallization by surface entropy reduction: optimization of the ser strategy. *Acta Crystallographica Section D: Biological Crystallography*, 63(5):636–645.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197.

Dale, G. E., Oefner, C., and D'Arcy, A. (2003). The protein as a variable in protein crystallization. *Journal of structural biology*, 142(1):88–97.

D'Arcy, A., Stihle, M., Kostrewa, D., and Dale, G. (1999). Crystal engineering: a case study using the 24 kda fragment of the dna gyrase b subunit from escherichia coli. *Acta Crystallographica Section D: Biological Crystallography*, 55(9):1623–1625.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald: An n log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092.

Darwin, C. (1922). Xcii. the reflexion of x-rays from imperfect crystals. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 43(257):800–829.

De Las Rivas, J. and Fontanillo, C. (2010). Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput Biol*, 6(6):e1000807.

Derewenda, Z. S. (2004). Rational protein crystallization by mutational surface engineering. *Structure*, 12(4):529–535.

Derewenda, Z. S. (2010). Application of protein engineering to enhance crystallizability and improve crystal properties. *Acta Crystallographica Section D: Biological Crystallography*, 66(5):604–615.

Derewenda, Z. S. and Vekilov, P. G. (2006). Entropy and surface engineering in protein crystallization. *Acta Crystallographica Section D: Biological Crystallography*, 62(1):116–124.

Doye, J. P., Louis, A. A., and Vendruscolo, M. (2004). Inhibition of protein crystallization by evolutionary negative design. *Physical Biology*, 1(1):P9.

Drenth, J. (2007). *Principles of protein X-ray crystallography*. Springer Science & Business Media.

Dror, R. O., Dirks, R. M., Grossman, J., Xu, H., and Shaw, D. E. (2012). Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics*, 41:429–452.

Durbin, S. and Feher, G. (1986). Crystal growth studies of lysozyme as a model for protein crystallization. *Journal of crystal growth*, 76(3):583–592.

Durek, P. and Walther, D. (2008). The integrated analysis of metabolic and protein interaction networks reveals novel molecular organizing principles. *BMC systems biology*, 2(1):100.

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):486–501.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh ewald method. *The Journal of chemical physics*, 103(19):8577–8593.

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography*, 62(1):72–82.

Fraser, J. S., van den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N., and Alber, T. (2011). Accessing protein conformational ensembles using room-temperature x-ray crystallography. *Proceedings of the National Academy of Sciences*, 108(39):16247–16252.

Frenkel, D. and Smit, B. (2001). *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier.

Frenkel, D., Smit, B., Tobochnik, J., McKay, S. R., and Christian, W. (1997). Understanding molecular simulation. *Computers in Physics*, 11(4):351–354.

Gapsys, V. and de Groot, B. L. (2017). pmx webserver: A user friendly interface for alchemistry. *Journal of Chemical Information and Modeling*, 57(2):109–114.

Gapsys, V., Michielssens, S., Seeliger, D., and de Groot, B. L. (2015). pmx: Automated protein structure and topology generation for alchemical perturbations. *Journal of computational chemistry*, 36(5):348–354.

Gibbs, J. W. (1878). On the equilibrium of heterogeneous systems, part ii. *Trans. Connecticut Acad. Sci*, 3:343–524.

Goldschmidt, L., Cooper, D. R., Derewenda, Z. S., and Eisenberg, D. (2007). Toward rational protein crystallization: A web server for the design of crystallizable protein variants. *Protein Science*, 16(8):1569–1576.

Grob, P., Huber, M., Walla, B., Hermann, J., Janowski, R., Niessing, D., Hekmat, D., and Weuster-Botz, D. (2020). Crystal contact engineering enables efficient capture and purification of an oxidoreductase by technical crystallization. *Biotechnology Journal*, page 2000010.

Halle, B. (2004). Biomolecular cryocrystallography: structural changes during flash-cooling. *Proceedings of the National Academy of Sciences*, 101(14):4793–4798.

Hekmat, D. (2015). Large-scale crystallization of proteins for purification and formulation. *Bioprocess and biosystems engineering*, 38(7):1209–1231.

Hekmat, D., Breitschwerdt, P., and Weuster-Botz, D. (2015). Purification of proteins from solutions containing residual host cell proteins via preparative crystallization. *Biotechnology letters*, 37(9):1791–1801.

Hermann, J., Bischoff, D., Grob, P., Janowski, R., Hekmat, D., Niessing, D., Zacharias, M., and Weuster-Botz, D. (2021). Controlling protein crystallization by free energy guided design of interactions at crystal contacts. *Crystals*, 11(6):588.

Hermann, J., Nowotny, P., Schrader, T. E., Biggel, P., Hekmat, D., and Weuster-Botz, D. (2018). Neutron and x-ray crystal structures of lactobacillus brevis alcohol dehydrogenase reveal new insights into hydrogen-bonding pathways. *Acta Crystallographica Section F: Structural Biology Communications*, 74(12):754–764.

Hess, B., Bekker, H., Berendsen, H. J., Fraaije, J. G., et al. (1997). Lincs: a linear constraint solver for molecular simulations. *Journal of computational chemistry*, 18(12):1463–1472.

Heyda, J., Mason, P. E., and Jungwirth, P. (2010). Attractive interactions between side chains of histidine-histidine and histidine-arginine-based cationic dipeptides in water. *The Journal of Physical Chemistry B*, 114(26):8744–8749.

Hitzenberger, M. and Zacharias, M. (2019). Structural modeling of γ-secretase aβ n complex formation and substrate processing. *ACS chemical neuroscience*, 10(3):1826–1840.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H., and MacKerell, A. (2017). Charmm36m: an improved force field for folded and intrinsically disordered proteins. *nature methods*, 14(1):71–73.

Hummel, W. (1997). New alcohol dehydrogenases for the synthesis of chiral compounds. In *New enzymes for organic synthesis*, pages 145–184. Springer.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38.

Janin, J. (1999). Wet and dry interfaces: the role of solvent in protein–protein and protein–dna recognition. *Structure*, 7(12):R277–R279.

Janin, J., Bahadur, R. P., and Chakrabarti, P. (2008). Protein-protein interaction and quaternary structure. *Quarterly reviews of biophysics*, 41(2):133.

Jeffrey, G. A. and Jeffrey, G. A. (1997). *An introduction to hydrogen bonding*, volume 12. Oxford university press New York.

Jones, S. and Thornton, J. M. (1996). Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences*, 93(1):13–20.

Joosten, R. P., Long, F., Murshudov, G. N., and Perrakis, A. (2014). The pdb_redo server for macromolecular structure model optimization. *IUCrJ*, 1(4):213–220.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935.

Juers, D. H. and Matthews, B. W. (2001). Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions. *Journal of molecular biology*, 311(4):851–862.

Kabsch, W. (1978). A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 34(5):827–828.

Kabsch, W. (2010). Xds. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):125–132.

Karshikoff, A. and Jelesarov, I. (2008). Salt bridges and conformational flexibility: Effect on protein stability. *Biotechnology & Biotechnological Equipment*, 22(1):606–611.

Kashchiev, D., Vekilov, P. G., and Kolomeisky, A. B. (2005). Kinetics of two-step nucleation of crystals. *The Journal of chemical physics*, 122(24):244706.

Kastenholz, M. A. and Hünenberger, P. H. (2006). Computation of methodology-independent ionic solvation free energies from molecular simulations. ii. the hydration free energy of the sodium cation. *The Journal of chemical physics*, 124(22):224501.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R., Wyckoff, H., and Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666.

Kirkwood, J. G. (1935). Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics*, 3(5):300–313.

Klimovich, P. V., Shirts, M. R., and Mobley, D. L. (2015). Guidelines for the analysis of free energy calculations. *Journal of computer-aided molecular design*, 29(5):397–411.

Koruza, K., Mahon, B. P., Blakeley, M. P., Ostermann, A., Schrader, T. E., McKenna, R., Knecht, W., and Fisher, S. Z. (2019). Using neutron crystallography to elucidate the basis of selective inhibition of carbonic anhydrase by saccharin and a derivative. *Journal of structural biology*, 205(2):147–154.

Kriminski, S., Caylor, C., Nonato, M., Finkelstein, K., and Thorne, R. (2002). Flash-cooling and annealing of protein crystals. *Acta Crystallographica Section D: Biological Crystallography*, 58(3):459–471.

Kriminski, S., Kazmierczak, M., and Thorne, R. (2003). Heat transfer from protein crystals: implications for flash-cooling and x-ray beam heating. *Acta Crystallographica Section D: Biological Crystallography*, 59(4):697–708.

Krissinel, E. (2015). Stock-based detection of protein oligomeric states in jspisa. *Nucleic acids research*, 43(W1):W314–9.

Krissinel, E. and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology*, 372(3):774–797.

Kumar, S. and Nussinov, R. (1999). Salt bridge stability in monomeric proteins. *Journal of molecular biology*, 293(5):1241–1255.

Kumar, S. and Nussinov, R. (2002). Relationship between ion pair geometries and electrostatic strengths in proteins. *Biophysical journal*, 83(3):1595–1612.

Lattman, E. E. and Loll, P. J. (2008). *Protein crystallography: A concise guide*. Johns Hopkins University Press, Baltimore.

Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J. M., Livingstone, J. C., Treffry, A., Luzzago, A., Levi, S., Arosio, P., Cesareni, G., et al. (1991). Solving the structure of human h ferritin by genetically engineering intermolecular crystal contacts. *Nature*, 349(6309):541–544.

Lee, E. H., Hsin, J., Sotomayor, M., Comellas, G., and Schulten, K. (2009). Discovery through the computational microscope. *Structure*, 17(10):1295–1306.

Lessin, L. S., Jensen, W. N., and Ponder, E. (1969). Molecular mechanism of hemolytic anemia in homozygous hemoglobin c disease: electron microscopic study by the freeze-etching technique. *The Journal of experimental medicine*, 130(3):443–466.

Liebl, K. and Zacharias, M. (2020). How global dna unwinding causes non-uniform stress distribution and melting of dna. *PloS one*, 15(5):e0232976.

Lin, Y.-L., Aleksandrov, A., Simonson, T., and Roux, B. (2014). An overview of electrostatic free energy computations for solutions and proteins. *Journal of chemical theory and computation*, 10(7):2690–2709.

Longenecker, K. L., Garrard, S. M., Sheffield, P. J., and Derewenda, Z. S. (2001). Protein crystallization by rational mutagenesis of surface residues: Lys to ala mutations promote crystallization of rhogdi. *Acta Crystallographica Section D: Biological Crystallography*, 57(5):679–688.

Luft, J. R., Snell, E. H., and DeTitta, G. T. (2011). Lessons from high-throughput protein crystallization screening: 10 years of practical experience. *Expert opinion on drug discovery*, 6(5):465–480.

Mähler, C., Kratzl, F., Vogel, M., Vinnenberg, S., Weuster-Botz, D., and Castiglione, K. (2019). Loop swapping as a potent approach to increase ene reductase activity with nicotinamide adenine dinucleotide (nadh). *Advanced Synthesis & Catalysis*, 361(11):2505–2513.

Malkin, A., Kuznetsov, Y. G., Land, T., DeYoreo, J., and McPherson, A. (1995). Mechanisms of growth for protein and virus crystals. *Nature structural biology*, 2(11):956–959.

Mateja, A., Devedjiev, Y., Krowarsch, D., Longenecker, K., Dauter, Z., Otlewski, J., and Derewenda, Z. S. (2002). The impact of glu→ ala and glu→ asp mutations on the crystallization properties of rhogdi: the structure of rhogdi at 1.3 å resolution. *Acta Crystallographica Section D: Biological Crystallography*, 58(12):1983–1991.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *Journal of applied crystallography*, 40(4):658–674.

McPherson, A. (1999). *Crystallization of biological macromolecules*.

McPherson, A. (2004). Introduction to protein crystallization. *Methods*, 34(3):254–265.

McPherson, A. and Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallographica Section F: Structural Biology Communications*, 70(1):2–20.

McPherson, A., Kuznetsov, Y. G., Malkin, A., and Plomp, M. (2003). Macromolecular crystal growth as revealed by atomic force microscopy. *Journal of structural biology*, 142(1):32–46.

McPherson, A., Malkin, A., Kuznetsov, Y. G., and Plomp, M. (2001). Atomic force microscopy applications in macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 57(8):1053–1060.

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography*, 53(3):240–255.

Niefind, K., Müller, J., Riebel, B., Hummel, W., and Schomburg, D. (2003). The crystal structure of r-specific alcohol dehydrogenase from lactobacillus

brevis suggests the structural basis of its metal dependency. *Journal of molecular biology*, 327(2):317–328.

Niimura, N. and Podjarny, A. (2011). *Neutron protein crystallography: hydrogen, protons, and hydration in bio-macromolecules*, volume 25. Oxford University Press.

Nowotny, P., Hermann, J., Li, J., Krautenbacher, A., Kloepfer, K., Hekmat, D., and Weuster-Botz, D. (2019). Rational crystal contact engineering of lactobacillus brevis alcohol dehydrogenase to promote technical protein crystallization. *Crystal Growth & Design*, 19(4):2380–2387.

Ofran, Y. and Rost, B. (2003). Analysing six types of protein–protein interfaces. *Journal of molecular biology*, 325(2):377–387.

Öhlknecht, C., Lier, B., Petrov, D., Fuchs, J., and Oostenbrink, C. (2020). Correcting electrostatic artifacts due to net-charge changes in the calculation of ligand binding free energies. *Journal of Computational Chemistry*.

Osborne, T. B. (1892). Crystallized vegetable proteins. *American Journal of Physiology*, 14:662–689.

Ostermann, A. and Schrader, T. (2015). Biodiff: Diffractometer for large unit cells. *Journal of large-scale research facilities JLSRF*, 1:2.

Otwinowski, Z. and Minor, W. (1997). [20] processing of x-ray diffraction data collected in oscillation mode. *Methods in enzymology*, 276:307–326.

Paliwal, H. and Shirts, M. R. (2011). A benchmark test set for alchemical free energy transformations and its use to quantify error in common free energy methods. *Journal of chemical theory and computation*, 7(12):4115–4134.

Páll, S. and Hess, B. (2013). A flexible algorithm for calculating pair interactions on simd architectures. *Computer Physics Communications*, 184(12):2641–2650.

Pande, A., Pande, J., Asherie, N., Lomakin, A., Ogun, O., King, J., and Benedek, G. B. (2001). Crystal cataracts: human genetic cataract caused by protein crystallization. *Proceedings of the National Academy of Sciences*, 98(11):6116–6120.

Patterson, A. L. (1935). A direct method for the determination of the components of interatomic distances in crystals. *Zeitschrift für Kristallographie-Crystalline Materials*, 90(1-6):517–542.

Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. (1960). Structure of hæmoglobin: a three-dimensional fourier synthesis at 5.5-å. resolution, obtained by x-ray analysis. *Nature*, 185(4711):416–422.

Piccoli, P. M., Koetzle, T. F., and Schultz, A. J. (2007). Single crystal neutron diffraction for the inorganic chemist–a practical guide. *Comments on Inorganic Chemistry*, 28(1-2):3–38.

Pinard, M. A., Kurian, J. J., Aggarwal, M., Agbandje-McKenna, M., and McKenna, R. (2016). Cryoannealing-induced space-group transition of

crystals of the carbonic anhydrase psca3. *Acta Crystallographica Section F: Structural Biology Communications*, 72(7):573–577.

Pohorille, A., Jarzynski, C., and Chipot, C. (2010). Good practices in free-energy calculations. *The Journal of Physical Chemistry B*, 114(32):10235–10253.

Radel, B., Funck, M., Nguyen, T., and Nirschl, H. (2019). Determination of filtration and consolidation properties of protein crystal suspensions using analytical photocentrifuges with low volume samples. *Chemical engineering science*, 196:72–81.

Rader, S. D. and Agard, D. A. (1997). Conformational substates in enzyme mechanism: The 120 k structure of α-lytic protease at 1.5 å resolution. *Protein science*, 6(7):1375–1386.

Reif, M. M. and Hünenberger, P. H. (2011). Computation of methodology-independent single-ion solvation properties from molecular simulations. iii. correction terms for the solvation free energies, enthalpies, entropies, heat capacities, volumes, compressibilities, and expansivities of solvated ions. *The Journal of chemical physics*, 134(14):144103.

Reif, M. M. and Oostenbrink, C. (2014). Net charge changes in the calculation of relative ligand-binding free energies via classical atomistic molecular dynamics simulation. *Journal of computational chemistry*, 35(3):227–243.

Ritthausen, H. (1880). Über die eiweisskorper verschiedenen oelsamen . beitraege zur physiologie der samen der culturgewaechse, der nahrungs- und futtermilch. *Pfluegers Arch*, 21:81–104.

Ritthausen, H. (1881). Krystallinisch eiweisskorper aus verschiedenen oelsman. *J. Prakt. Chem.*, 23:481.

Rocklin, G. J., Mobley, D. L., Dill, K. A., and Hünenberger, P. H. (2013). Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: an accurate correction scheme for electrostatic finite-size effects. *The Journal of chemical physics*, 139(18):11B606_1.

Rupp, B. (2009). *Biomolecular crystallography: principles, practice, and application to structural biology*. Garland Science.

Schindler, C. E., de Vries, S. J., and Zacharias, M. (2015). iattract: Simultaneous global and local interface optimization for protein–protein docking refinement. *Proteins: Structure, Function, and Bioinformatics*, 83(2):248–258.

Schlieben, N. H., Niefind, K., Müller, J., Riebel, B., Hummel, W., and Schomburg, D. (2005). Atomic resolution structures of r-specific alcohol dehydrogenase from lactobacillus brevis provide the structural bases of its substrate and cosubstrate specificity. *Journal of molecular biology*, 349(4):801–813.

Schmit, J. D. and Dill, K. (2012). Growth rates of protein crystals. *Journal of the American Chemical Society*, 134(9):3934–3937.

Siezen, R. J., Fisch, M. R., Slingsby, C., and Benedek, G. B. (1985). Opacification of gamma-crystallin solutions from calf lens in relation to cold cataract formation. *Proceedings of the National Academy of Sciences*, 82(6):1701–1705.

Sleutel, M., Lutsko, J., Van Driessche, A. E., Durán-Olivencia, M. A., and Maes, D. (2014). Observing classical nucleation theory at work by monitoring phase transitions with molecular precision. *Nature communications*, 5(1):1–8.

Sleutel, M. and Van Driessche, A. E. (2018). Nucleation of protein crystals–a nanoscopic perspective. *Nanoscale*, 10(26):12256–12267.

Smejkal, B., Agrawal, N. J., Helk, B., Schulz, H., Giffard, M., Mechelke, M., Ortner, F., Heckmeier, P., Trout, B. L., and Hekmat, D. (2013). Fast and scalable purification of a therapeutic full-length antibody based on process crystallization. *Biotechnology and bioengineering*, 110(9):2452–2461.

Staley, J. P. and Woolford Jr, J. L. (2009). Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Current opinion in cell biology*, 21(1):109–118.

Ten Wolde, P. R. and Frenkel, D. (1997). Enhancement of protein crystal nucleation by critical density fluctuations. *Science*, 277(5334):1975–1978.

Teng, T.-Y. and Moffat, K. (1998). Cooling rates during flash cooling. *Journal of applied crystallography*, 31(2):252–257.

Tuckerman, M. (2010). *Statistical mechanics: theory and molecular simulation*. Oxford university press.

Van der Spoel, D., Van Maaren, P. J., and Berendsen, H. J. (1998). A systematic study of water models for molecular simulation: derivation of water models optimized for use with a reaction field. *The Journal of chemical physics*, 108(24):10220–10230.

Van Driessche, A. E., Van Gerven, N., Bomans, P. H., Joosten, R. R., Friedrich, H., Gil-Carton, D., Sommerdijk, N. A., and Sleutel, M. (2018). Molecular nucleation mechanisms and control strategies for crystal polymorph selection. *Nature*, 556(7699):89–94.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., et al. (2010). Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of computational chemistry*, 31(4):671–690.

Vekilov, P. G. (2010). The two-step mechanism of nucleation of crystals in solution. *Nanoscale*, 2:2346–2357.

Vekilov, P. G. and Chernov, A. A. (2003). The physics of protein crystallization. In *Solid State Physics*, volume 57, pages 1–147. Elsevier.

Vekilov, P. G. and Rosenberger, F. (1996). Dependence of lysozyme growth kinetics on step sources and impurities. *Journal of crystal growth*, 158(4):540–551.

Volmer, M. and Weber, A. (1926). Nucleus formation in supersaturated systems. *Zeitschrift für Physikalische Chemie*, 119:277–301.

Von Leyden, E. V. (1872). Zur kenntnis des asthma bronchiale. *Archiv für pathologische Anatomie und Physiologie und für klinische Medizin, Berlin*, 54:324–344; 346–352.

Whitelam, S. (2010). Control of pathways and yields of protein crystallization through the interplay of nonspecific and specific attractions. *Physical review letters*, 105(8):088102.

Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *The FEBS journal*, 275(1):1–21.

Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2013). Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *The FEBS journal*, 280(22):5705–5736.

Yau, S.-T. and Vekilov, P. G. (2000). Quasi-planar nucleus structure in apoferritin crystallization. *Nature*, 406(6795):494–497.

Yeh, J. I. and Hol, W. G. (1998). A flash-annealing technique to improve diffraction limits and lower mosaicity in crystals of glycerol kinase. *Acta Crystallographica Section D: Biological Crystallography*, 54(3):479–480.

Young, A. C., Tilton, R. F., and Dewan, J. C. (1994). Thermal expansion of hen egg-white lysozyme: comparison of the 1· 9 å resolution structures of the tetragonal form of the enzyme at 100 k and 298 k. *Journal of molecular biology*, 235(1):302–317.

Zachariasen, W. H. et al. (1945). Theory of x-ray diffraction in crystals.

Zwanzig, R. W. (1954). High-temperature equation of state by a perturbation method. i. nonpolar gases. *The Journal of Chemical Physics*, 22(8):1420–1426.

Appendix A

Diffraction Data and Refinement Details

Table A.1 Detailed crystallographic data collection and refinement statistics of WT alcohol dehydrogenase from *Lactobacillus brevis* (*Lb*ADH) deposited in protein data bank (PDB) measured with neutron and X-ray diffraction. Values in parentheses are for highest-resolution shell.

Data collection	WT, neutron diffraction	WT, X-ray diffraction
PDB-ID	6h1m	6h07
Beamline	BIODIFF	ESRF ID30A-3
Wavelength, Å	2.67	0.968
Temperature, K	293.15	100
Detector	cylindrical neutron image plate detector	Eiger X4M detector
Crystal-detector distance, mm	199	103
Rotation range per image, °	0.3	0.05
Total rotation range, °	82.5	180
Exposure time per image	1.5 h	0.02 s
space group (SG)	*I*222	$P2_122_1$
unit cell (UC) (a, b ,c), Å	56.5, 84.6 115.4	56.0, 83.3, 114.4
No. of molecules per asymmetric unit	1	2
Refinement		
Resolution range, Å	21.76 - 2.15 (2.23 - 2.15)	47.15 - 1.48 (1.54 - 1.48)
Total No. of reflections	37,191(2,620)	882,139 (82,606)
No. of unique reflections	14,121(995)	89,189 (8,439)
Completeness, %	91.6 (79.0)	99.2 (93.3)
Redundancy	2.6 (2.2)	9.9 (9.8)
$I/\sigma(I)$	5.14 (2.24)	17.40 (1.15)
R_{merge}	0.155 (0.368)	0.0693 (1.664)
R_{meas}	0.192 (0.466)	0.0732 (1.756)
R_{pim}	0.111 (0.282)	0.0232 (0.553)
Overall B-factor from Wilson plot, $Å^2$	14.47	21.43
CC(1/2)	0.995 (0.610)	0.999 (0.643)
No. of reflections	14121 (1201)	88878 (8224)
No. of reflections, test set	1412 (120)	4444 (411)
R_{work} / R_{free}	0.21 (0.27) / 0.25 (0.29)	0.18 (0.34) / 0.20 (0.35)
No. of atoms		
Protein	4061	3780
Ion	2	2
Water	218	391
Total	4281	4173
R.m.s. deviations		
Bond lengths, Å	0.061	0.025
Bond angles, °	2.80	2.16
Ramachandran plot		
Most favored, %	99	98
Additional allowed, %	1	2

Table A.2 Crystallographic data collection and refinement statistics of *Lb*ADH mutants Q126H, Q126K, D54F, Q207D deposited in PDB. Values in parentheses are for highest-resolution shell.

Data collection	Q126H	Q126K	D54F	Q207D
PDB-ID	6y10	6y0z	6y1c	7a2b
Beamline	SLS PXIII X06DA	SLS PXIII X06DA	SLS PXI X06SA	SLS PXIII X06DA
Wavelength	0.966	0.966	1.000	1.000
SG	*I*222	*I*222	$P2_122_1$	*I*222
UC (a, b, c), Å	56.05, 80.57, 113.47	55.77, 84.23, 113.56	55.76, 81.09, 113.13	55.82, 80.52, 115.10
No. of molecules per asymmetric unit	1	1	2	1
Resolution, Å	50-1.22 (1.25-1.22)	50-1.21 (1.28-1.21)	46.83-1.41 (1.44-1.41)	46.83-1.40 (1.40-1.49)
$I/\sigma(I)$	13.5 (1.4)	10.4 (0.8)	15.7 (1.7)	3.04 (3.8)
CC (1/2)	99.9 (70.9)	99.9 (63.8)	99.9 (60.6)	99.9 (96.9)
Completeness, %	99.1 (99.4)	98.9 (97.7)	98.0 (94.5)	99.6 (98.6)
Redundancy	7.0 (6.3)	5.07 (4.4)	13.4 (12.3)	4.0 (3.91)
Refinement				
Resolution, Å	1.22	1.21	1.41	1.40
No. reflections	72057	77296	93011	98425
R_{work} / R_{free}	13.8/16.4	14.9/17.2	12.7/16.2	14.4/17.5
No. atoms				
Protein	2077	2079	4053	2010
Water	407	286	589	318
Other	36	52	81	1
B-factor overall	14.3	19.0	18.0	16.5
R.m.s. deviations				
Bond lengths, Å	0.008	0.010	0.007	0.009
Bond angles, °	1.49	1.59	1.38	1.48
Ramachandran plot				
Most favored, %	96	97	98	98
Additional allowed, %	4	2	2	2

Table A.3 Crystallographic data collection and refinement statistics of *Lb*ADH mutants T102E, K32A_Q126K, T102E_Q126K, K32A deposited in PDB. Values in parentheses are for highest-resolution shell.

Data collection	T102E	K32A_Q126K	T102E_Q126K	K32A
PDB-ID	6y0s	6y1b	6y15	6hlf
Beamline	SLS PXIII X06DA	SLS PXIII X06DA	SLS PXIII X06DA	ESRF ID30A-3
Wavelength	1.000002	1.000009	1.000002	0.968
SG	$P2_122_1$	$P2_122_1$	$I222$	$I222$
UC (a, b, c), Å	55.62, 81.15, 115.56	56.28, 82.03, 114.59	55.95, 84.50, 114.75	55.58, 81.78, 114.89
No. of molecules per asymmetric unit	1	2	1	1
Resolution, Å	50-1.44 (1.48-1.44)	50-1.40 (1.44-1.40)	50-1.80 (1.85-1.80)	47.97-1.55 (1.64-1.55)
$I/\sigma(I)$	16.16 (2.01)	22.1 (2.15)	10.46 (1.61)	17.06 (0.88)
CC (1/2)	99.9 (68.3)	100.0 (76.3)	99.5 (61.8)	99.9 (50.5)
Completeness, %	99.7 (99.9)	98.7 (88.0)	99.2 (99.9)	99.6 (96.5)
Redundancy	12.9 (13.0)	6.2 (3.4)	13.2 (12.5)	9.1 (9.2)
Refinement				
Resolution, Å	1.44	1.40	1.80	1.55
No. reflections	94959	102723	25435	346208
R_{work} / R_{free}	11.9/17.1	10.6/13.8	16.7/20.7	17.2/19.7
No. atoms				
Protein	3939	3952	1967	1867
Water	772	716	324	303
Other	9	13	57	3
B-factor overall	17.0	20.0	18.0	21.68
R.m.s. deviations				
Bond lengths, Å	0.015	0.012	0.018	0.020
Bond angles, °	1.84	1.68	2.01	1.94
Ramachandran plot				
Most favored, %	98	98	98	96
Additional allowed, %	2	2	2	4

Appendix B

MD Simulation Details

B.1 Free Energy Changes

Table B.1 Free energy changes ΔΔG for *Lb*ADH WT vs. mutant crystallization. The free energy changes are calcuated in WT and mutant (Mut.) crystal system. All values in kJ mol $^{-1}$.

Mutation	WT Crystal System	Mut. crystal system
T102E	-9.48 ± 1.08	-14.39 ± 1.21
D54F	-0.83 ± 0.47	-11.99 ± 0.38
Q126H	-0.84 ± 0.18	-3.70 ± 0.23
Q126K	0.40 ± 0.76	-1.80 ± 0.52
K32A	0.02 ± 0.27	-0.64 ± 0.26
Q207D	3.63 ± 0.36	2.45 ± 0.56
D54A	3.42 ± 0.54	-
K45A	4.96 ± 0.33	-
H39A	7.40 ± 0.74	-

Table B.2 Free energy changes ΔΔG for *Lb*ADH WT vs. mutant crystallization for different setups for mutation K32A. All values in kJ mol $^{-1}$.

Mutant crystal	
Mg^{2+} between D54s	-2.92±0.40
alchemical transformed MG^{2+} between D54s	-1.17±0.32
no Mg^{2+} between D54s	0.16±0.45
WT crystal	
Mg^{2+} between D54s	-2.17±0.38
alchemical transformed MG^{2+} between D54s	-0.24±0.32
no Mg^{2+} between D54s	0.39±0.48

B.2 Setup Details

Table B.3 Simulation set-up description of crystal state simulations. Three different *Lb*ADH WT crystals were simulated. The symmetry operators of the respective SG ($P2_122_1$ or $I222$) and UC translations to be applied on the asymmetric unit are displayed. The monomers which are extracted of the respective symmetry mate are given by indicating the chain-IDs (A, B) from the corresponding PBD-entry. Additionally, the simulation box size is noted.

crystal	symmetry operation and extracted monomers of asymmetric unit	simulation box size, Å^3
X-ray $P2_122_1$	**A**+**B** of (x, y, z) + **A**+**B** of $(x + 1/2, -y, -z + 1/2)$ + **A**+**B** of $(-x, y, -z) + (1\ 0\ 0)\}$ + **A**+**B** of $(-x + 1/2, -y, z + 1/2) + (0\ 1\ 0)\}$	$56.03 \times 83.31 \times 114.38$
X-ray $I222$	**A** of (x, y, z) + **A** of $(-x, -y, z)$ + **A** of $\{(x + 1/2, -y + 1/2, -z + 1/2)$ + **A** of $\{(-x + 1/2, y + 1/2, -z + 1/2)$ + **A** of $\{(-x + 1/2, -y + 1/2, z + 1/2)$ + **A** of $\{(x + 1/2, y + 1/2, z + 1/2)$	$55.80 \times 82.88 \times 115.01$
Neutron $I222$	**A** of (x, y, z) + **A** of $(-x, -y, z)$ + **A** of $\{(x + 1/2, -y + 1/2, -z + 1/2)$ + **A** of $\{(-x + 1/2, y + 1/2, -z + 1/2)$ + **A** of $\{(-x + 1/2, -y + 1/2, z + 1/2)$ + **A** of $\{(x + 1/2, y + 1/2, z + 1/2)$	$56.51 \times 84.57 \times 115.42$

Table B.4 Detailed simulation set-up description of free energy calculation (FEC) simulations of the bound state. The bound sate is simulated as the reduced contact or/and as a UC with PBC. For each mutation and crystal system (cryst. sys.), symmetry operators of SG $P2_122_1$ (WT, D54F, T102E) and $I222$ (K32A , Q126H, Q126K, Q207D) and UC translations to be applied on the asymmetric unit are displayed. The monomers which are extracted of the respective symmetry mate are given by indicating the chain-IDs (A, B) from the corresponding structure model. The mutated monomers in the set-up are printed **bold**. Additionally, the simulation box size is noted. Notice: in crystal system $P2_122_1$ two corner contacts (denoted by contact 1 and contact 2) exist.

mut.	cryst. sys.	symmetry operation and extracted monomers of asymmetric unit	simulation box size, Å^3
Q126H	WT	UC with PBC: **A**+**B** of (x, y, z) + **A**+**B** of $(x + 1/2, -y, -z + 1/2)$ + **A**+**B** of $(-x, y, -z) + (1\ 0\ 0)\}$ + **A**+**B** of $(-x + 1/2, -y, z + 1/2) + (0\ 1\ 0)\}$	$56.03 \times 83.31 \times 114.38$
		contact 1: **A**+B of (x, y, z) & A of $(x + 1/2, -y, -z + 1/2)$	$114.01 \times 125.49 \times 100.88$
		contact 2: A+**B** of (x, y, z) & B of $\{(x + 1/2, -y, -z + 1/2) + (-1\ 1\ 0)\}$	$114.01 \times 125.49 \times 100.88$
	mutant	UC with PBC: **A** of (x, y, z) + **A** of $(-x, -y, z)$ + **A** of $\{(x + 1/2, -y + 1/2, -z + 1/2)$ + **A** of $\{(-x + 1/2, y + 1/2, -z + 1/2)$ + **A** of $\{(-x + 1/2, -y + 1/2, z + 1/2)$ + **A** of $\{(x + 1/2, y + 1/2, z + 1/2)$	$56.05 \times 80.57 \times 113.47$
		contact: **A** of (x, y, z) + A of $(-x, -y, z)$ & A of $\{(x + 1/2, -y + 1/2, -z + 1/2) + (-1\ 1\ 0)\}$	$114.01 \times 125.49 \times 100.88$
Q126K	WT	contact 1: **A**+B of (x, y, z) & A of $(x + 1/2, -y, -z + 1/2)$	$114.01 \times 125.49 \times 100.88$
		contact 2: **A**+B of (x, y, z) & B of $\{(x + 1/2, -y, -z + 1/2) + (-1\ 1\ 0)\}$	$114.01 \times 125.49 \times 100.88$
	mutant	contact: **A** of (x, y, z) + A of $(-x, -y, z)$ & A & A of $\{(x + 1/2, -y + 1/2, -z + 1/2) + (-1\ 1\ 0)\}$	$114.01 \times 125.49 \times 100.88$
K32A	WT	contact: **A** of (x, y, z) & **B** of $(x, y, z) + (0\ -1\ 0)$	$97.44 \times 142.46 \times 81.86$
	mutant	contact: **A** of (x, y, z) & **A** of $(-x, -y, z) + (0\ 1\ 0)$	$97.44 \times 142.46 \times 81.86$
D54F	WT	contact: **A** of (x, y, z) & **B** of $(x, y, z) + (0\ -1\ 0)$	$97.44 \times 142.46 \times 81.86$
	mutant	contact: **A** of (x, y, z) & **B** of $(x, y, z) + (0\ 1\ 0)$	$97.44 \times 142.46 \times 81.86$
D54A	WT	contact: **A** of (x, y, z) & **B** of $(x, y, z) + (0\ -1\ 0)$	$97.44 \times 142.46 \times 81.86$

Continued on next page

Table B.4 – *Continued from previous page*

mut.	cryst. sys.	symmetry operation and extracted monomers of asymmetric unit	simulation box size, Å^3
T102E	WT	contact 1: **A**+B of (x, y, z) & B of $(x + 1/2, -y, -z + 1/2) + (-1\ 1\ 0)$	$114.01 \times 125.49 \times 100.88$
		contact 2: A+**B** o (x, y, z) & A of $(x + 1/2, -y, -z + 1/2)$	$114.01 \times 125.49 \times 100.88$
	mutant	contact 1: **A**+B of (x, y, z) & B of $(x + 1/2, -y, -z + 1/2)$	$114.01 \times 125.49 \times 100.88$
		contact 2: A+**B** of (x, y, z) & A of $(x + 1/2, -y, -z + 1/2) + (-1\ 1\ 0)$	$114.01 \times 125.49 \times 100.88$
K45A	WT	contact 1: **A** of (x, y, z) & A+B of $(x + 1/2, -y, -z + 1/2) + (-1\ 0\ 0)$	$114.01 \times 125.49 \times 100.88$
		contact 2: **B** of (x, y, z) & A+B of $\{(x + 1/2, -y, -z + 1/2) + (0\ 1\ 0)\}$	$114.01 \times 125.49 \times 100.88$
H39A	WT	contact 1: **A** of (x, y, z) & A+B of $(x + 1/2, -y, -z + 1/2) + (-1\ 0\ 0)$	$114.01 \times 125.49 \times 100.88$
		contact 2: **B** of (x, y, z) & A+B of $\{(x + 1/2, -y, -z + 1/2) + (0\ 1\ 0)\}$	$114.01 \times 125.49 \times 100.88$
Q207D	WT	contact: **A**+B of (x, y, z) + A+**B** of $\{(-x, y, -z) + (1\ 0\ 0)\}$ & **A**+B of $(-x, y, -z)$ + A+**B** of $\{(x, y, z) + (-1\ 0\ 0)\}$	$141.29 \times 105.26 \times 98.09$
	mutant	contact: **A** of (x, y, z) + **A** of $\{(x, -y, -z) + (0\ 1\ 0)\}$ + A of $\{(-x, y, -z) + (1\ 0\ 0)\}$ + A of $\{(-x, -y, z) + (1\ 1\ 0)\}$ & **A** of $(-x, y, -z)$ + **A** of $\{(-x, -y, z) + (0\ 1\ 0)\}$+ A of $\{(x, y, z) + (-1\ 0\ 0)\}$ + A of $\{(x, -y, -z) + (-1\ 1\ 0)\}$	$141.29 \times 105.26 \times 98.09$

Table B.5 Detailed simulation set-up description of unbound solvated state which is simulated as tetramer or reduced (as described in Methods 6.3.1). Depending on the position of the mutation the tetramer can be reduced to one monomer or two monomers thereof. Mutation Q126H is simulated in both set-ups.

mutation	number of monomers	simulation box, $Å^3$
Q126H	4 (tetramer)	$88.00 \times 98.00 \times 93.00$
	2	$86.71 \times 107.88 \times 69.63$
Q126K	2	$93.88 \times 114.60 \times 76.44$
K32A	1	$90.56 \times 90.06 \times 82.02$
D54F	1	$90.56 \times 90.06 \times 82.02$
D54A	1	$90.56 \times 90.06 \times 82.02$
T102E	2	$93.72 \times 114.41 \times 76.30$
K45A	1	$90.56 \times 90.06 \times 82.02$
H39A	1	$90.56 \times 90.06 \times 82.02$
Q207D	4 (tetramer)	$88.19 \times 98.21 \times 93.20$

Appendix C

Free Energy Charge Correction Details

Table C.1 Q126K bound state contact 1 for different grid spacings in the Poisson-Boltzmann (PB) solver in GROMOS (GROningen MOlecular Simulation).

spacing, nm	0.016	0.018	0.020	0.022	0.024
ΔG_{pol}, kJ mol^{-1}	-19.27	-19.24	-19.22	-19.30	-20.02

Table C.2 Free energy charge correction for Q126K bound crystal state contact 1. Free energy charge correction terms ΔG_{dir}, ΔG_{pol}, and the sum theirof. For start ($\lambda = 0.0$) and end ($\lambda = 1.0$) state of the alchemical transformation, snapshots from the free energy simulation are taken in descrete time intervals of Δt = 5ns.

λ	t, ns	ΔG_{dir}, kJ mol^{-1}	ΔG_{pol}, kJ mol^{-1}	$\Delta G_{\text{pol}} + \Delta G_{\text{dir}}$, kJ mol^{-1}
0.0	15	18.08	-19.22	-1.14
0.0	10	8.86	-9.93	-1.07
0.0	5	28.45	-28.63	-0.18
0.0	0	19.43	-20.32	-0.90
1.0	15	-11.46	8.67	-2.79
1.0	10	-15.75	13.53	-2.22
1.0	5	-0.48	-0.68	-1.16
1.0	0	-1.05	-1.76	-2.82
				-1.53 ± 0.96

Table C.3 Free energy charge correction for Q126K bound crystal state contact 2. Free energy charge correction terms ΔG_{dir}, ΔG_{pol}, and the sum theirof. For start ($\lambda = 0.0$) and end ($\lambda = 1.0$) state of the alchemical transformation, snapshots from the free energy simulation are taken in descrete time intervals of $\Delta t = 5$ns.

λ	t, ns	ΔG_{dir}, kJ mol^{-1}	ΔG_{pol}, kJ mol^{-1}	$\Delta G_{\text{pol}} + \Delta G_{\text{dir}}$, kJ mol^{-1}
0.0	15	53.70	-53.24	0.46
0.0	10	20.33	-21.14	-0.81
0.0	5	25.80	-25.43	0.37
0.0	0	22.89	-24.25	-1.36
1.0	15	44.88	-46.43	-1.55
1.0	10	8.96	-10.37	-1.42
1.0	5	27.52	-29.14	-1.62
1.0	0	31.36	-31.30	0.05
				-0.74 ± 0.89

Table C.4 Free energy charge correction for Q126K unbound solvated state. Free energy charge correction terms ΔG_{dir}, ΔG_{pol}, and the sum theirof. For start ($\lambda = 0.0$) and end ($\lambda = 1.0$) state of the alchemical transformation, snapshots from the free energy simulation are taken in descrete time intervals of $\Delta t = 5$ns.

λ	t, ns	ΔG_{dir}, kJ mol^{-1}	ΔG_{pol}, kJ mol^{-1}	$\Delta G_{\text{pol}} + \Delta G_{\text{dir}}$, kJ mol^{-1}
0.0	15	-62.87	59.25	-3.62
0.0	10	42.93	-43.13	-0.20
0.0	5	-58.86	56.29	-2.57
0.0	0	-10.76	8.14	-2.62
1.0	15	-65.48	63.40	-2.08
1.0	10	15.06	-15.84	-0.78
1.0	5	-36.91	34.73	-2.18
1.0	0	-32.83	29.76	-3.07
				-2.14 ± 1.14

Table C.5 Free energy charge correction for K32A bound crystal state. Free energy charge correction terms ΔG_{dir}, ΔG_{pol}, and the sum theirof. For start (λ = 0.0) and end (λ = 1.0) state of the alchemical transformation, snapshots from the free energy simulation are taken in descrete time intervals of Δt = 5ns.

λ	t, ns	ΔG_{dir}, kJ mol^{-1}	ΔG_{pol}, kJ mol^{-1}	$\Delta G_{\text{pol}} + \Delta G_{\text{dir}}$, kJ mol^{-1}
0.0	15	-82.31	88.99	6.67
0.0	10	-22.91	23.12	0.21
0.0	5	-45.55	43.02	-2.53
0.0	0	-105.21	106.53	1.33
1.0	15	-51.87	49.43	-2.44
1.0	10	-92.42	88.80	-3.63
1.0	5	-38.88	37.56	-1.32
1.0	0	-0.63	-0.94	-1.56
				-0.41 ± 3.27

Table C.6 Free energy charge correction for K32A unbound solvated state. Free energy charge correction terms ΔG_{dir}, ΔG_{pol}, and the sum theirof. For start (λ = 0.0) and end (λ = 1.0) state of the alchemical transformation, snapshots from the free energy simulation are taken in descrete time intervals of Δt = 5ns.

λ	t, ns	ΔG_{dir}, kJ mol^{-1},	ΔG_{pol}, kJ mol^{-1}	$\Delta G_{\text{pol}} + \Delta G_{\text{dir}}$, kJ mol^{-1}
0.0	15	-7.44	6.15	1.29
0.0	10	-45.30	48.85	-3.55
0.0	5	-13.95	12.77	1.18
0.0	0	-27.77	26.15	1.62
1.0	15	-36.92	34.57	2.35
1.0	10	10.24	-17.49	7.25
1.0	5	-3.13	2.12	1.01
1.0	0	6.50	-6.92	0.42
				1.45 ± 2.95