TEMPERATURE, COOLING RATE AND HYDRATION DEPENDENCE OF PROTEIN CONFORMATIONAL ENSEMBLES AND RADIATION DAMAGE IN BIOMOLECULAR X-RAY CRYSTALLOGRAPHY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Hakan Atakisi

August 2021

TEMPERATURE, COOLING RATE AND HYDRATION DEPENDENCE OF PROTEIN CONFORMATIONAL ENSEMBLES AND RADIATION DAMAGE IN BIOMOLECULAR X-RAY CRYSTALLOGRAPHY

Hakan Atakisi, Ph.D.

Cornell University 2021

X-ray crystallography is the gold standard technique for the structural determination of biological macromolecules such as proteins, nucleic acids, viruses and their complexes. Much of our information in structural biology, in particular about proteins, come from X-ray crystallography studies.

This dissertation covers different aspects of protein X-ray crystallography which is an inherently multifaceted tool with many adjustable experimental knobs. We present temperature, cooling rate and relative humidity dependence of protein structures which, when put together, offers a much more complete picture of protein structure, dynamics and function. We also quantify radiation damage to protein samples at cryogenic temperatures by establishing 'local' Fourier-space relationship between radiation damage and dose which is well supported by our simple physics-based model, yet very consistent with both diffraction and imaging experiments over more than two orders of magnitude in resolution, explaining the underlying mechanism.

BIOGRAPHICAL SKETCH

Hakan Atakisi was born in 1990 and spent his childhood in Istanbul of Turkey. He got attracted to various science contests in high school which motivated him to get his Bachelor of Science degree in Physics from Bilkent University (Ankara, Turkey) in 2013. He joined the Physics Department at Cornell University in the same year as a graduate student. He obtained his Master of Science degree in Physics from Cornell University in 2016 and he is currently working towards obtaining his Ph.D. degree under the supervision of Professor Robert E. Thorne.

Dedicated to my family.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude for my Ph.D. advisor Professor Robert E. Thorne for providing me with academic guidance with his unique understanding of physics concepts from both engineering and theoretical perspective. He has not only guided me in my research, without whom this dissertation would not have been possible, but also been insanely patient with me in my difficult times during this long journey.

I also wish to thank Professor Veit Elser and Professor Sol Gruner for taking time to serve in my committee.

I must also thank my lab mate David Moreau, one of the kindest people I have ever met in my life, for being a great company during sleepless nights of data collection at CHESS.

I am deeply indebted to my undergraduate advisor Professor Ozgur Oktel for his extremely friendly guidance during my undergraduate research and my graduate school applications.

I also would like to take this opportunity to express my deepest appreciation to my friends Daniyar Saparov, Azer Eyvazov, Hasan Sezer, Abdurrahman Gumus, Omer Bilgen, Jorge Arriaza, Melik Turker, and my special thanks to Arslan and Ghaffur families–always felt like family–for all their invaluable friendship and sharing the happiest moments with me in Ithaca.

Finally, of course, I cannot overemphasize my dear family's endless support and encouragement during my Ph.D.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The vast majority of three dimensional structures of biological macromolecules such as proteins, nucleic acids and viruses have been determined by X-ray crystallography. Now being over 100 years old, since the discovery of the diffraction of X-rays from single crystals, it is likely to continue to be the most favored structure determination technique in near future too. To date, biomacromolecular X-ray crystallography (MX) has played a central role in elucidation of enzyme mechanisms and greatly helped in our understanding of protein-ligand interactions enabling structure based drug design.

Less than a few decades ago the protein X-ray data collection used to be performed at room temperature and it did not take long for crystallographers to realize that X-ray rapidly damages the protein at room temperature (or physiological temperatures) limiting the amount and the quality of data to be collected from a given crystal sample. In order to reduce the X-ray radiation damage, plunging samples into a liquid nitrogen and collecting data at very low temperatures typically ~100 K has become a standard protocol since 1990s. Convenience of flash-cooling and storing crystal samples combined with the subsequent advances in X-ray sources, beamline optics and detectors, automation, analysis software and computing brought us to the present era of mail-in, high-throughput synchrotron data collection and semi-automated data processing, model building, and validation.

Despite the enormous progress that has been made in the field in the last decades there are still several challenges and concerns associated with the current MX data collection protocol that need to be addressed. First, in spite of the overall similarity of room (or biological) temperature and cryogenic structures, unfortunately, cryocooling may locally modify main chain and side chain conformations including highly flexible

active sites of proteins [1–3], thus making them less likely to be biologically relevant. Conformational differences between room and cryogenic temperature structures may also occur due to the use of penetrating cryoprotectants such as glycerol and PEGs during cryocooling of samples to prevent ice formation [4]. Another relevant parameter for cryocooling for samples is the cooling rate at which the crystal sample is cooled. In theory, higher temperature states may be kinetically trapped by increasing cooling rate if the evolution of those states are slower than or comparable to the cooling rate. Therefore, systematic temperature and cooling rate dependence of protein conformation is of interest.

Second, as radiation damage manifests even at cryogenic temperatures (with the use of small crystals and intense micro beams), it is essential to quantify and model spatio-temporal evolution of protein diffraction intensity in order to be able to improve the accuracy of structure factors and structural models.

Third, protein crystals, 20–80% water in content, are fragile in the sense that they are in a delicate thermodynamic equilibrium with their mother liquor, so they are prone to unintentional dehydration during routine sample handling. Dehydration, both intentional and accidental, has been shown to induce conformational changes on the protein structure [5–7], if not a complete loss of diffraction. Therefore, it deserves special attention.

The overarching theme in this dissertation is improving and getting the most information out of MX experiments by focusing on i) temperature, cooling rate and relative humidity dependence of protein structures, and ii) better understanding and modeling of X-ray radiation damage to crystal samples. Here is a high level summary of each chapter.

The second chapter presents measurements and analyses of the temperature and cooling rate dependent structure of the enzyme urease obtained from the organism Klebsiella aerogenes (KAU). Temperature dependent measurements over the full range from 300 K to 100 K follow the evolution of structural elements near the active site, primarily the mobile flap (consisting of ~25 residues) flanking the active site, that undergo substantial change between room temperature and $T = 100$ K. Variable temperature data without soaking crystals in penetrating cryoprotectants showed a systematic evolution of the conformation of the flexible flap. Another set of variable temperature data by soaking samples in 25% glycerol–typical cryoprotectant concentration used in MX–revealed that glycerol significantly disrupts the trend of evolution of the flap conformation as a function of temperature. We also determined $T = 100$ K structure vs. cooling rate for rates spanning a factor of $\sim 10^5$ to show that the room temperature conformation of the active site flap can be more accurately captured using faster cooling rates of ~10,000 K/s. Finally, we determined the structures of the native crystals soaked in the solutions of the most commonly used denaturants, urea and guanidine, prepared at various concentrations to get insights about differences and similarities in their denaturation mechanism.

The third chapter of the dissertation adds another dimension to MX experiments, relative humidity (r.h.), another thermodynamic variable. The modulation of main-chain and side-chain conformational heterogeneity and solvent structure in monoclinic lysozyme crystals by dehydration (related to water activity) and temperature are examined. Decreasing the relative humidity (from ~99 to ~11%) and decreasing the temperature to $T = 100$ K both lead to contraction of the unit cell, to an increased area of crystal contacts and to remodeling of primarily contact and solvent-exposed residues. Both lead to the depopulation of some minor side-chain conformers and to the generation of new conformations. Dehydration from 99 to 93% r.h. and cooling from 298 to 100 K

result in a comparable number of remodeled residues, with dehydration-induced remodeling somewhat more likely to arise from contact interactions. These results illuminate the qualitative and quantitative similarities between structural perturbations induced by modest dehydration, which routinely occurs in samples prepared for 298 and 100 K data collection, and cryocooling. Differences between these perturbations in terms of energy landscapes and occupancies, and implications for variable-temperature crystallography between 180 and 298 K, are discussed.

And the last chapter describes our attempt to determine the underlying 'local' Fourier-space relationship between radiation damage and dose, which is key to probing and understanding radiation damage by X-rays and energetic particles in both diffraction and imaging experiments. We have shown that crystallographic data have been strongly perturbed by the effects of spatially nonuniform crystal irradiation and diffraction during data collection in MX experiments. We have established a relationship between 'local' Fourier-space radiation damage and dose by fitting our model to experimental intensity curves in three different studies. Reanalysis of the data revealed a strong resolution dependence of radiation-induced diffraction decays implying that the previously proposed 20–30 MGy dose limit for protein crystallography should be replaced by a resolution-dependent dose limit that, for atomic resolution data sets, will be much smaller. In an attempt to explain the resolution-dependent dose limit, we considered a simple physics-based model in which damage events occurring at random locations within a sample cause energy deposition and blurring of the electron density within a small volume. Our simple model turned out to be very consistent with both diffraction and imaging experiments over more than two orders of magnitude in resolution. The results suggest that the physics underlying basic experimental trends in radiation damage at $T \simeq 100$ K is straightforward and universal.

CHAPTER 2

EVOLUTION AND TRAPPING OF CONFORMATIONAL SUBSTATES IN UREASE USING VARIABLE TEMPERATURE AND VARIABLE COOLING RATE CRYSTALLOGRAPHY

This chapter is part of a manuscript in preparation with contributions of Matthew A. Warkentin, Jesse B. Hopkins and P. Andrew Karplus under the supervision of Robert E. Thorne. Robert E. Thorne, Matthew A. Warkentin, Jesse B. Hopkins and P. Andrew Karplus conceived of the presented idea. Matthew A. Warkentin and Jesse B. Hopkins collected the data. Hakan Atakisi derived the models and analysed the data. Robert E. Thorne and Hakan Atakisi wrote the manuscript. All authors discussed the results and contributed to the final manuscript. James Fraser provided useful input for the data analysis.

2.1 Introduction

2.1.1 Temperature in crystallography

The widespread adoption of cryo-crystallographic methods in the 1990s revolutionized protein crystallography and structural biology. Cooling protein crystals to $T = 100$ K reduced radiation damage by a factor of 50 or more [8–11] allowing complete data sets to be obtained from a single crystal. Reduced thermal motions at 100 K often improved resolution and allowed better identification of solvent atoms [12, 13]. "Looping" and cooling crystals was easier than mounting them in X-ray capillaries, and cold crystals could be stored for weeks and then shipped for measurement. Initial concerns as to whether biomolecular structures obtained at cryogenic temperature reflected the

biologically relevant form diminished as comparisons with available room-temperature structures seldom showed significant differences at the resolutions then routinely achievable. Subsequent advances in X-ray sources, beamline optics and detectors, and automation, and in analysis software and computing brought us to the present era of mail-in, high-throughput synchrotron data collection and semi-automated data processing, model building, and validation. Not surprisingly, cryogenic temperature data collection has dominated for the last 20 years. For most proteins in the Protein Data Bank (PDB), no room temperature structure is available.

Despite the overall similarity of room (or biological) temperature and $T = 100$ K structures, cooling has many effects on crystal and biomolecular structure. Unit cells and protein volume typically contract by 2–4% and 1–2%, respectively [14]. Core packing and the area of crystal contacts tend to increase. Crystal mosaicity, a measure of lattice disorder, significantly increases and crystals are less likely to be isomorphous [15, 16] [17, 18]. Primary conformations of side chains and of larger structures such as mobile loops and flaps may change substantially. Improved methods for identifying and modeling low-density features in electron density maps show that minority side chain conformations – including catalytically important minority conformations in active sites deduced from room temperature NMR or crystallography – may depopulate or be altered on cooling to cryogenic temperatures [19–21] [1, 22].

Conformational differences between cryogenic temperature and room or biological temperature structures may reflect the protein's underlying energy landscape and how that evolves as temperature modulates hydrophobic and other interactions. Conformational differences of surface residues and also of interior residues may arise from changes (generally, increases) in packing interactions between molecules in the crystal lattice on cooling, as occur during, e.g., crystal dehydration. Conformational differences

may also occur if solute/cryoprotectant concentrations differ between crystals. Glycerol, PEGs, and alcohols at concentrations of 20–30% typical in cryocrystallography affect protein stability, can modulate more labile regions e.g., near active sites, and can obscure or masquerade as weakly bound ligands [4].

More information about protein conformational ensembles and interactions can in principle be obtained by examining crystallographic structure versus temperature – not just at 300 K and 100 K but at multiple temperatures between room/biological temperature and ∼180 K where temperature-dependent structural evolution typically ceases – and versus cooling rate, especially when crystals are prepared in the same way with the same solute concentrations within their internal solvent cavities. Temperature-dependent occupancies of side chain rotamers and of larger conformational entities can provide quantitative information about the underlying local energy landscapes. Temperature-dependent data is essential for verifying that observed differences between 300 K and 100 K structures are due to temperature and not differences in, e.g., crystal growth or post-growth handling (e.g., dehydration). Temperature dependent data can be particularly critical in distinguishing the contributions of static disorder – arising from, e.g., frustrated crystal packing interactions, chemical heterogeneity, and other effects – from those of the dynamic disorder relevant to function; static disorder is typically constant or increases while dynamic disorder typically decreases (depending on the roughness of the local energy landscape) with decreasing temperature.

Despite this potential, only a handful of studies have examined protein structure versus temperature. Crystal structures of myoglobin determined between 220 and 300 K showed evolution of conformational substates that were interpreted in terms of barrier heights in the underlying energy landscape [23]. Examination of B-factors in ribonuclease A showed a transition to liquid-like motions around 220 K (subsequently identified

as the protein-solvent glass or dynamical transition), and this transition coincided with the onset of substrate binding [24, 25]. More recent studies have illustrated the potential of variable temperature crystallography to reveal information about protein conformational ensembles including within active sites and how they relate to the protein function [22, 26], and to provide insight into the nature of the protein-solvent transition.

Relatively few studies have examined the temperature dependence of protein structure, especially at temperatures between ~240 K and ~180 K. In this temperature range, ice forms rapidly, and use of very large cryoprotectant concentrations up to 75% have been observed [25].

Measurements of structure versus cooling rate can add kinetic information to this largely equilibrium picture. Ultrafast cooling to cryogenic temperature can in principle capture the room temperature structure. In practice, relaxations occur on a broad range of time scales depending on local and global barriers, attempt frequencies, and diffusivities – which depend on the detailed nature of the motions involved – that control the kinetics of the relaxation. Small motions e.g., side chain rotamers, should relax very quickly and have small barriers, while flaps, lids, loops and other structures involving many residues should relax more slowly [27–29] [30, 31]. As a result, at a given cooling rate different regions of the molecule (and crystal) will fall out of equilibrium at different temperatures, resulting in a heterogenous and nonequilibrium low temperature state.

To date, the only crystallographic studies using cooling rate as a variable compared $T = 100$ K structures obtained using conventional plunge cooling, gives cooling rates of 300–1000 K/s for 100–300 μm crystals – and using programmed slow cooling in an N_2 gas stream at rates of ~1 K/s [32, 33]. Very slow cooling allows relaxation of solvent distributions (driven by changes in internal solvent and solvent cavity volumes) and of crystal packing that are incomplete at more typical cooling rates, and often yields

smaller $T = 100$ K crystal mosaicity.

Here we report high resolution crystallographic measurements and analysis of the temperature and cooling rate dependent structure of the enzyme urease from Klebsiella aerogenes (KAU). As reviewed in Section 2.1.2, a key aspect of this remarkable enzyme is a large and highly mobile flap that covers its active site. Measurements using uncryoprotected crystals at several temperatures spanning the full range from 340 K or 300 K to 100 K, with a particular emphasis on temperatures near and above the protein-solvent glass transition, allow us to follow the evolution of the flap and other structural elements near the active site that exhibit substantial conformation differences in $T = 300$ K and 100 K structures. Complete temperature series collected from cryoprotectant free crystals and from crystals soaked in 25% w/v glycerol are used to explore possible effects of cryoprotection on the mobile flap. We also determined $T = 100$ K structures vs. cooling rate for rates spanning a factor of 10^5. These measurements show that the room temperature conformation of the flap can be more accurately captured by cooling at rates of $\sim 10,000$ K/s. Implications of these results for the timescales of important motions, for the mechanism of enzymatic action and for broader protein crystallography are discussed. Finally, we determined the structures of native crystals soaked in solutions of urea and guanidine having concentrations ranging from 0.1 M to 4 M.

2.1.2 Urease structure and function

Urease, which catalyzes the hydrolysis of urea to the final products ammonia and carbamate, is one of the most proficient known enzymes, providing a rate enhancement of $\sim 10^{14}$ over the uncatalyzed reaction [34]. Historically, urease was the first enzyme to be crystallized, first used to demonstrate the proteins are enzymes [35], and first shown

to have nickel atoms essential to biological activity in its active site [36]. But despite a century of study, the reasons for its proficiency are incompletely understood. Major scientific and industrial efforts have focused on inhibition of urease activity: urease is a virulence factor enabling many antibiotic-resistant human pathogens, and urea-based synthetic nitrogen fertilizers – widely used around the globe – are rapidly hydrolyzed by abundant urease present in the soil, causing economic losses and environmental pollution [37].

Ureases from diverse organisms (plants, fungi, algae, procaryotes and eukaryotes) have high amino acid sequence homology, but vary in the number of subunits assembled

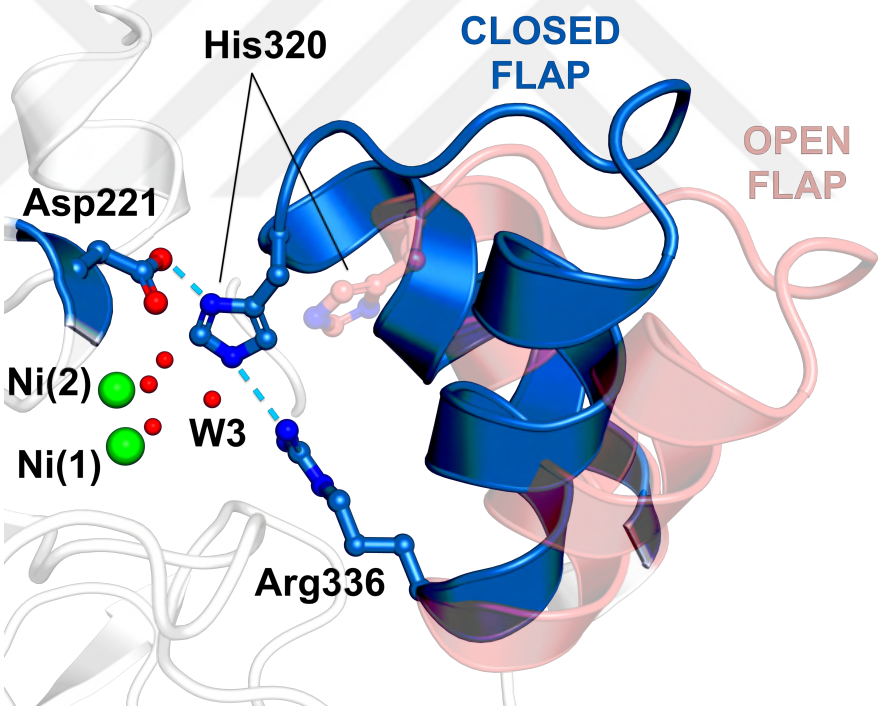

Figure 2.1: Open and closed flap conformations of KAU urease flanking the active site. Tetrahedrally ordered network of four water molecules ligates the bi-nickel site. Highly flexible flap comprising ~25 residues (310–335 of the α-subunit) is shifted up to ca. 7 Å between the two states. Catalytically crucial His320 protruding into the nickel site is shown in both flap states (solid blue in the closed flap and transparent pink in the open flap) along with two other catalytically important residues, Asp221 and Arg336, which are thought take a role in the stabilization of His320 by forming hydrogen bonds with it.

in the biologically active structure. Most crystallographic studies have focused on urease from Sporosarcina pasteurii (SPU) (formerly known as Bacillus pasteurii urease (BPU)) and Klebsiella aerogenes (KAU) as their crystals diffract to high resolution. In all structures, the active site contains a bi-nickel center close to the surface which is flanked by a mobile flap having a helix-loop-helix motif and totaling ~25 residues (in KAU, residues 310–335 in the α-domain). As seen in Figure 2.1, the flap moves between two distinct states – "open" and "closed", thereby acting as a gate for the entrance of substrate and the release of products. In moving between these states the flap residues shifted by up to ~7 Å, depending on where they are located relative to the flap's "hinge".

Adjacent to the bi-nickel center, four water molecules form a tetrahedral cluster, and these are displaced upon substrate binding. Mutation studies indicate that the key residues in either substrate binding and/or the catalytic step are located in the immediate vicinity of the bi-nickel site, and no allosteric modulators have so far been reported. Side chains and main chain oxygen atoms of the key active site residues stabilize the substrate and accelerate the reaction primarily by forming hydrogen bonds with the substrate and with each other. One crucial residue, His320 (in KAU, His323 in SPU) is located on the mobile flap and shifts by ~5 Å when the flap closes, bringing it into proximity with the bi-nickel site where it can interact with the substrate. Mutation studies show that His320 plays a crucial role in the catalytic step but is not involved in substrate binding [38, 39]. In the closed state only, His320 forms hydrogen bonds with the sidechains of Asp221 (located on the non-flap, rigid part of the protein) and Arg336 (located right on the "hinge"); mutations of these residues reduce k_{cat} by orders of magnitude with modest or no effect on K_m [39]. These findings suggest that catalysis occurs when the flap is closed so that His320 protrudes into the substrate binding site, with additional substrate stabilization provided by Asp221 and Arg336.

While other models for the enzymatic mechanism of urease have been discussed [40, 41], the current model, supported by inhibition and kinetic studies [42, 43] [44, 45], is as follows: (1) With the flap in its open conformation, urea enters the active site and binds to the more electrophilic nickel Ni(1), replacing three active site water molecules (W_1, W_2 and W_3). (2) Urea binds to Ni(2) through one of its amino nitrogen atoms. The flap fluctuates to its closed conformation, reducing the active site volume, and allowing hydrogen bonds to form between urea and the key flap/active site residues. These facilitate the orientation and the stabilization of the substrate. (3) Nickel bridging water, W_B, attacks the carbon atom of urea to yield a tetrahedral transition state. (4) His320 acts as a catalytic base, stabilizing urea's distal nitrogen, thereby facilitating $C - NH_3^+$ bond breakage and release of ammonia. On product release, the flap opens.

The conformational state of the flap in each step is thus crucial to accurately portraying the enzymatic mechanism. Since the flap is highly mobile, its conformation might be expected to be more sensitive to temperature, pH, solvent composition, and other factors than more rigid parts of the enzyme. Consistent with the historical trend of data collection temperatures in protein crystallography, almost all PDB entries for urease deposited before ~2000 were collected at room temperature, while almost all those deposited after were collected at 100 K. The earlier entries are mostly for KAU urease, whereas SPU urease is most common in later entries (with additional entries for HPU and JBU urease).

Table 2.1 and 2.2 list high resolution KAU and SPU PDB entries that are described in published manuscripts, indicating their data collection temperature, flap state, and the pH of solutions used in crystallization. Some PDB entries, e.g., those with different inhibitors or different mutant variants yielding the same overall structure and flap conformation, were considered redundant and are excluded from the table. Despite having

PDB ID	Description	Flap occupancy	Resolution (Å)	Date	Temperature (K)	pH
1KRA	Apo native, no nickels, only one water replacing four waters+nickels, Lys217 not carbamylated	Closed 0.34/0.66	2.30	1995	298	7.5
2KAU	Native, four waters not modeled	0.47/0.53	2.00	1995	298	7.5
1EJR	Asp221Ala, W_3 moves away from other waters by ~0.9 Å	Open 0.64/0.36	2.00	2000	298	7.5
1KRB	His219Ala, Wb and W1 gone	Closed 0.13/0.87	2.50	1995	298	7.5
1EJS	His219Asn mutant	Closed 0.20/0.80	2.00	2000	298	7.5
1EJT	His219GIn mutant	Closed 0.12/0.88	2.00	2000	298	7.5
1KRC	His320Ala mutant, four waters replaced by a single water	Closed 0.47/0.53	2.50	1995	298	7.5
1EJU	His320Asn, W_3 moves away from other waters by ~0.9 Å	Open 0.62/0.38	2.00	2000	298	7.5
1EJV	His320GIn, W_3 moves away from other waters by ~0.9 Å	Open 0.67/0.33	2.40	2000	298	7.5
1EJW	Native at 298 K	Closed 0.50/0.50	1.90	2003	298	7.5
1EJX	Native at 100 K	Open 0.85/0.15	1.60	2003	100	7.5
1FWA	Cys319Ala mutant at pH 7.5	Closed 0.24/0.76	2.00	1997	298	7.5
1FWB	Cys319Ala mutant at pH 6.5	Closed 0.15/0.85	2.00	1997	298	6.5
1FWC	Cys319Ala mutant at pH 8.5	Closed 0.20/0.80	2.00	1997	298	8.5
1FWD	Cys319Ala mutant at pH 9.4	Closed 0.25 / 0.75	2.00	1997	298	9.4
1FWE	Cys319Ala mutant + AHA bound, all four waters replaced by AHA	Open 0.65/0.35	2.00	1997	298	7.5
1FWG	Cys319Ser mutant, W_3 moves away from other waters by ~0.9 Å	Closed 0.47/0.53	2.00	1997	298	7.5
1FWF	Cys319Asp mutant, W_3 moves but not exactly away from other waters	Open 0.74/0.26	2.00	1997	298	7.5
1FWH	Cys319Tyr mutant, W_3 moves away from other waters by ~0.9 Å	Open 0.75/0.25	2.00	1997	298	7.5

Table 2.1: Select high-resolution PDB entries obtained from the organism Klebsiella aerogenes (KAU). If the tetrahedral water network is different from the native structure it is indicated in the description. We re-refined the deposited structures against our two state flap model to obtain flap occupancies. We have excluded some PDB entries that are redundant for our purposes such as different mutant variations yielding the same overall structure and the same flap conformation.

PDB ID	Description	Flap occupancy	Resolution (Å)	Date	Temperature (K)	pH
1S3T	Substrate analog, $B(OH)_3$ (triangular shaped), replaces three waters, W_B remains in place	Open 0.80/0.20	2.10	2004	100	Between 6.3-7.5
2UBP	Native	Open 0.78/0.22	2.00	1999	100	Between 6.3-7.5
3UBP	Transition state analog, DAP (hydrolysis of PPD, tetrahedral shaped) replaces four waters	Closed 0.32/0.68	2.00	1999	100	Between 6.3-7.5 + PPD sol. (pH 8.0)
1UBP	One BME replaces four waters and another BME binds to Cys322	Open 0.90/0.10	1.65	1999	100	6.3-?
4UBP	AHA replaces four waters	Open 0.78/0.22	1.55	2000	100	Between 6.3-7.5 + AHA sol. (pH 8.0)
1IE7	PHO (tetrahedral shaped) replaces four waters	Open 0.79/0.21	1.85	2001	100	6.8
4CEU	Native	Open 0.85/0.15	1.58	2014	100	6.5
4CEX	Floride ion replaces W_B and W_1, W_3 moves away from other waters by ~0.8 Å	Open 0.60/0.40	1.59	2014	100	6.5
5G4H	Cys322 locked by catechol	Open 0.84/0.16	1.50	2016	100	Between 6.3-7.5
5OL4	Transition state analog, MATP (hydrolysis of NBPT, tetrahedral shaped) replaces four waters	Closed 0.17/0.83	1.28	2018	100	Between 6.3-7.5
6H8J	Transition state analog. DAP (hydrolysis of NBPTO, tetrahedral shaped)	Closed 0.38 / 0.62	1.45	2019	100	7.0
6QDY	Finally urea-bound SPU, flouride ion replaces Wb so that catalytic step does not take place	Closed 0.21/0.79	1.42	2019	100	Between 6.3 – 7.5
6RKG	Transition state analog, pH 7.5, DAP (hydrolysis of NBPTO, tetrahedral shaped)	Closed 0.18 / 0.82	1.32	2020	100	7.5
6RP1	Transition state analog, pH 6.5, DAP (hydrolysis of NBPTO, tetrahedral shaped)	Closed 0.43/0.57	1.49	2020	100	6.5

Table 2.2: Select high-resolution PDB entries obtained from the organism Sporosarcina pasteurii (SPU) (formerly known as Bacillus pasteurii urease (BPU)). If the final pH of the crystallization solution was given in the paper, it is listed as a single pH value; otherwise, pH of the protein and the reservoir solution are listed. In the instance of co-crystallization with an inhibitor solution e.g., 3UBP and 4UBP, pH of the inhibitor solution is also listed. We have excluded some PDB entries that are redundant for our purposes such as inhibited structures with many different inhibitors yielding the same overall structure and the same flap conformation.

highly similar active site architectures, the PDB entries for native KAU urease – as well as the great majority of its active site mutant variants – show a closed flap conformation, whereas PDB entries for native SPU – as well as the great majority of its complexes with various inhibitors and ligands – show an open flap conformation. Overall, this difference in flap state strongly correlates with data collection temperature, suggesting that the flap in both KAU and SPU variants undergoes a temperature-dependent conformation change. In unpublished work by Karplus et al. in 2003, KAU urease adopts an open flap conformation at 100 K (PDB code 1EJX), with ten flap residues starting at the catalytically important His320 so disordered that they could not be modelled, and an open conformation at room temperature, (PDB code 1EJW) without any other significant structural difference. These results, as well as the critical role of flap closure in enzymatic mechanism, motivate our selection of urease for study by temperature- and cooling-rate dependent crystallographic data collection. Very recently, the difference in flap state between KAU and SPU structures was interpreted as arising from differences in pH during crystallization, but the evidence for this interpretation is weak (Appendix Section A).

2.2 Materials and methods

2.2.1 Crystallization

Cubic KAU crystals (space group $I2_13$, with one catalytic unit per asymmetric unit) were grown at room temperature using the hanging-drop vapor-diffusion method by equilibrating protein solution (provided by R. Hausner) with a reservoir solution consisting of 100 mM HEPES (pH, 7.5) and 1.6 M Li_2SO_4, as described previously [46]. Crys-

tals used in diffraction studies measured between 250–450 μm in one dimension, had unit cells of 168.9–170.8 Å (depending on data collection temperature), and typically diffracted to ~1.6–1.9 Å. Cubic KAU crystals have a Matthews coefficient of 2.51, a solvent content of 51% v/v, and a maximum solvent cavity size (determined using the room temperature crystallographic structure and the program MAP_CHANNNELS [47]) of ~51 Å.

2.2.2 Variable temperature data collection

X-ray diffraction data were collected on MacCHESS stations A1, F1 and F3 (prior to the CHESS-U upgrade) at the Cornell High-Energy Synchrotron Source (CHESS), over a period of three years. Diffraction images were recorded using ADSC Quantum 210, Quantum 270, and Quantum 4 detectors. Crystals were used as grown, without additional penetrating cryoprotectant beyond the 1.6 M lithium sulfate in the crystallization buffer. High resolution structural data sets were obtained at several temperatures between 340 K and 100 K using crystals cooled (or warmed) in an N_2 gas stream programmed to the final temperature. Ice nucleates rapidly in bulk solution below 240 K, but ice nucleation in internal crystal solvent is strongly suppressed by nanoconfinement within the protein matrix [32, 48]. Ice-free diffraction at reduced temperatures was obtained by carefully removing all external solvent using NVH oil [32]. To evaluate possible effects of cryoprotection on active site structure and specifically the flap conformation, additional data sets were collected at several temperatures between 300 K and 100 K using crystals soaked in 25% w/v glycerol.

2.2.3 Variable cooling rate data collection

Cryoprotectant-free crystals were cooled to ~100 K or below at three different cooling rates. Four crystals of size 250–400 μm were cooled to 100 K using an N_2 gas cryostream, giving cooling rates of order 10^2 K/s. Two smaller crystals, from which excess surface oil was removed, were plunge cooled in liquid N_2 at 77 K. Cold gas normally present above the liquid N_2 surface was removed and replaced with room temperature gas, and the crystals were manually plunged at ~1 m/s (estimated from video recordings) [49]. This gave cooling rates of order 10^4 K/s. Finally, two crystals were cooled at ~10^{-1} K/s by, first, meticulously removing all surface solvent and replacing it with thick layer of NVH oil (to prevent ice formation during the ~half hour required to cool to cryogenic temperature), and then inserting in an N_2 gas stream whose temperature was programmed to decrease at that rate [32]. Ice-free high resolution diffraction data were obtained at 100 K for each cooling rate.

The first two cooling rates of 10^2 K/s and 10^4 K/s are estimates based on thermocouple measurements, adjusting for difference in thermal mass between the crystals and thermocouple junctions. Given uncertainties in crystal size, shape, surrounding liquid volume, and in details of gas stream insertion, cold gas removal, and plunge cooling, these estimates are likely accurate to within a factor of 3 . Since observed changes in electron density / occupancy are more nearly logarithmic than linear functions of cooling rate, these uncertainties do not significantly affect conclusions.

2.2.4 Urea and guanidine soaked crystal preparation and data collection

KAU crystals do not show significant enzymatic activity [50]. This has been attributed to constraints on conformational changes involved in urea binding and product release in the crystal lattice [50]. The available data do not rule out limited urea diffusion as a possible cause, although urea's small size and the crystal's solvent content (~51% v/v) and solvent channel size (51 Å) would seem to make that unlikely. To investigate whether urea has any effects on structure including flap conformation or partial active site occupancy, urease crystals were soaked in 0.1 and 1 M urea before data collection at 300 K and 100 K. To compare with the effects of another common denaturant, crystals were soaked in 0.5, 1, 2 and 4 M guanidine before data collection at 300 K.

2.2.5 Data processing and analysis

Diffraction frames were indexed and scaled using HKL2000 [51]. Molecular replacement and refinement were performed in PHENIX [52]. To quantify conformational changes of the flap, we built a two-state model in which the flap – spanning residues 310 to 335 in chain C – could be in the open and closed conformations with different occupancies. Using Coot [53], multiple room temperature structures were used to define and manually build the closed flap state, and multiple $T = 100$ K structures were used to build the open flap state. At the two extreme temperatures, one of the flap states is well defined while the other is dimmed, enabling us to easily trace the backbone and the sidechains of the flap residues. We have chosen cryoprotectant-free data sets to build our two-state model as they have a greater contrast between the two flap states at the extreme temperatures compared to the glycerol-soaked samples. Although determined

independently, this open flap conformation is in a good agreement with the open flap conformations observed in SPU and HPU (PDB codes 4CEU and 1E9Y respectively).

The two-state model, initialized with an open/closed occupancy of 0.5/0.5, was used in molecular replacement and structural refinement for all data sets. The open/closed occupancy was refined as a whole and the sum of occupancies constrained to 1 in PHENIX, yielding a single metric to characterize the state preference at each temperature, cooling rate, etc. To prevent open and closed flap conformations from substantially changing in real space during the refinement, TLS and simulated annealing in PHENIX were turned off; otherwise, default refinement settings were used. The number of refinement macro-cycles was adjusted to ensure convergence of the refined flap occupancies, with 5–8 macrocycles generally being sufficient. At the end of refinement, refined flap conformations were inspected to verify that they were only slightly modified (e.g., at side chain rotamers) relative to the initial two-state template used in molecular replacement. Placement of ordered water molecules was performed in a separate round of refinement using default PHENIX parameters after the flap occupancies were determined. PyMOL was used for preparing structural visuals.

In the above refinement protocol, the open/closed occupancy of the whole flap and the B-factors of individual atoms were refined, so that the occupancy contained most information about the flap state. As a check on these results, a second refinement protocol was used in which open and closed state occupancies were assumed equal (0.5/0.5) and only the individual atomic B factors in the two states were refined. An average B-factor over the atomic positions in the open and closed flap configurations, respectively, was then calculated and used to characterize the flap state: the smaller a state's average B-factor, the more the flap tends to adopt that state. For example, in a native room temperature structure, the average B-factors for the closed and open flap states were ~17 Å^2

and ~62 $Å^2$, respectively, indicating that the closed configuration is favored at room temperature; in a 100 K native structure, the open flap state had a smaller B-factor.

An MD simulation study of KAU [54] revealed a previously unobserved "wide-open" flap state which allowed easier access to the bi-nickel site. The "wide-open" flap state was observed in a subsequent MD study of HPU urease [55]. These studies suggested that the "open" flap state previously identified in crystallographic studies was not open enough to allow urea to enter the active site, due to partial occlusion by the sidechain of His320 (KAU numbering). In the wide-open state, His320 is further away from the active site and its sidechain points away from the bi-nickel site, so that urea can access the bi-nickel site and products can be released after catalysis.

These studies noted that the wide-open state should have much lower occupancy than the other two flap states. But in an attempt to detect it, we built a three-state model for the flap using the coordinate file for the wide open state provided in supplementary information [54]. We then refined our data against the three-state model, constraining the sum of the occupancies to unity.

To explore structural changes other than those of the flap as a function of temperature, cooling rate or soak type/concentration, we used the program PanDDA, which can detect small changes and weak signals within a set of electron density maps, such as those caused by weakly bound ligands [56]. Our collection of more than 50 urease data sets was large enough for statistically meaningful analyses. However, no significant structural changes aside from those of the flap and the nickel active site or evidence of allosteric interactions were observed in PanDDA maps.

Finally, we used the program Ringer [20], which samples and maps electron density around dihedral angles to identify peaks associated with low-occupancy alternative

conformers of sidechain rotamers. However, our data set resolution was typically in the range 1.7 – 2.0 Å, and preliminary analysis indicated only ~5% of residues showed alternative conformers at room temperature with that map resolution. This was not deemed sufficient to draw statistically meaningful conclusions about the evolution of sidechain heterogeneity with temperature, cooling rate, or soak conditions.

2.3 Results

2.3.1 Temperature dependence of active site

High-resolution (1.7–2.0 Å), ice free data sets of KAU urease were determined at temperatures of 100, 150, 180, 200, 220, 240, 270, 300, 320, and 340 K, all using as-grown crystals with no penetrating cryoprotectants aside from the LiCl precipitant used in crystallization. These data sets were refined using a two state model for the active site flap as discussed in Section 2.2.5. Refined structures at all temperatures were similar, exhibiting only small variations in backbone (backbone rmsd between 100 K and 300 K of ~0.26 Å), almost no variation in the inter-nickel distance (~3.67 Å), a near-perfect tetrahedral water cluster ligating the bi-nickel center, and the same side chain conformations for important active site residues. The only major structural variation with temperature was the occupancy distribution of flap conformation. The 100 K flap conformation and density are consistent with those of previously deposited structures by Karplus (PDB codes: 1EJX and 1EJW).

The behavior of the flap as a function of temperature is shown in Figure 2.2 . The flap is dominantly closed at 300 K (closed occupancy ~0.72) and open at 100 K (open occupancy ~0.75). The two flap states are equally occupied at ~270 K. Sample-to-

sample variations in occupancy are likely due to variations in data quality / data set resolution.

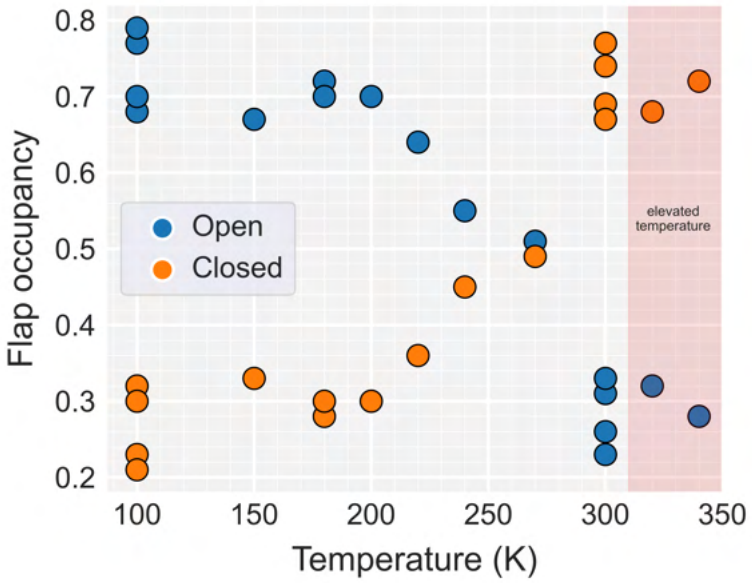

Figure 2.2: Flap occupancy as a function of temperature for cryoprotectant-free data series. Flap is dominantly closed at room temperature, open at 100 K and equally favored at 270 K. Two samples above the room temperature also follows the same trend. The fact that the flap occupancy evolves rapidly between room temperature and ~200 K but significantly slowly below ~200 K signifies a protein dynamical transition.

The occupancies vary rapidly with temperature above ~200 K, and are roughly independent of temperature below ~200 K. Similar temperature dependent behavior has been observed in overall protein crystal B-factors [57], in the rate of radiation damage with dose [8, 58] and other properties. This behavior is usually associated with the protein-solvent glass transition near 200 K, below which most protein motions and enzymatic activity cease [23, 25, 59].

The average B-factors, calculated over all atoms in the flap, of the open and closed flap conformations obtained from the two-state refinement are comparable, with closed state values being only slightly lower (Fig. A.1). These B-factors show no obvious trend with temperature, but do correlate with the B-factor obtained by averaging over

the whole protein, which is determined by data quality / resolution.

In another round of refinement, we refined only the B-factors for the open and closed flap conformations when the occupancies of these states were both fixed at 0.5 during refinement. Preliminary analysis showed that at room temperature the closed state has a much lower B-factor than the open state, indicating that the closed state is the dominant conformation present, while at 100 K the open state has the lowest B-factor. The average flap B-factors determined in this way do not show the smooth variation with temperature of the occupancies in Fig. 2.2. This is most likely because B-factors also depend on overall crystal order. The largest overall (i.e., over all atoms in the unit cell) B-factors and lowest overall data set resolution were observed near 200 K, where the open and closed state B-factors show maxima.

2.3.2 Glycerol dependence of active site and flap conformation

The active site flap is highly mobile and its conformation is modulated by temperature and by the pH of the crystallization buffer [60]. To investigate possible effects of penetrating cryoprotectants, specifically of glycerol which is one of the most commonly used cryoprotectants, crystallographic data was collected from crystals soaked in 25% w/v glycerol at 11 temperatures between 100 K and 300 K. These data were then refined using a two state model to yield occupancies as in Fig. 2.2. Glycerol-soaked samples diffracted to slightly higher resolution, typically 1.7 Å, than glycerol-free crystals. As with glycerol-free crystals, the refined structures showed little change in backbone or active site conformation with temperature, except for the flap region. However, Figure 2.3 shows that the flap occupancy versus temperature is substantially different from that of glycerol free crystals. At 300 K, the glycerol soak increases occupancy of the

open conformation, so that it is nearly equal that of the closed conformation. At 100 K, the glycerol soak increases occupancy of the closed conformation, from 0.24 to 0.32. Glycerol seems to destabilize the majority flap conformation in both cases. The reasons for this are unclear. Glycerol may disrupt some hydrogen bonds that stabilize the major flap conformation at each temperature, but these are difficult to reliably model. Furthermore, the open occupancy reaches its maximum near 200 K instead of at 100 K and shows evolution between 200 K and 160 K, in contrast with behavior observed in glycerol-free crystals.

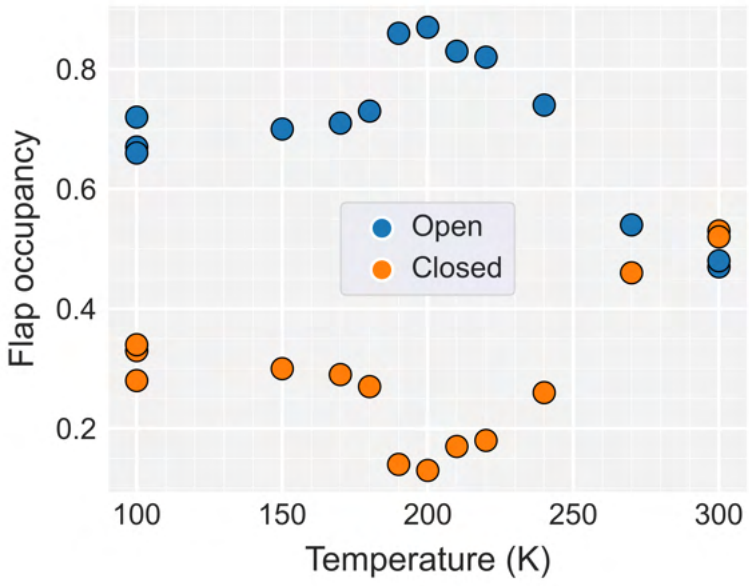

Figure 2.3: Flap occupancy as a function of temperature for 25% glycerol-soaked samples. Similar to the glycerol-free data series, closed flap is favored at room temperature while open flap is favored at low temperatures. There is a clear peak/dip at ~200 K which is not observed in the glycerol-free data series (Fig. 2.2) suggesting that soaking samples in glycerol shifts the flap occupancy equilibrium.

2.3.3 Cooling rate dependence of active site and flap conformation

The temperature dependence of protein structure provides information about equilibrium occupancies and the relative depths of minima in the free energy landscape. The cooling rate to, e.g., $T = 100$ K, determines the time available for the structure to evolve toward its (temperature-dependent) equilibrium and so can provide information about free energy barriers and protein dynamics. Complete structural data sets were collected at 100 K on otherwise identically prepared glycerol-free urease crystals cooled at nominal rates of 10^{-1} K/s, 10^{2} K/s, and 10^{4} K/s, spanning a factor of 10^{5} in rate. These data sets were then modeled using the two-state flap model to determine flap occupancies.

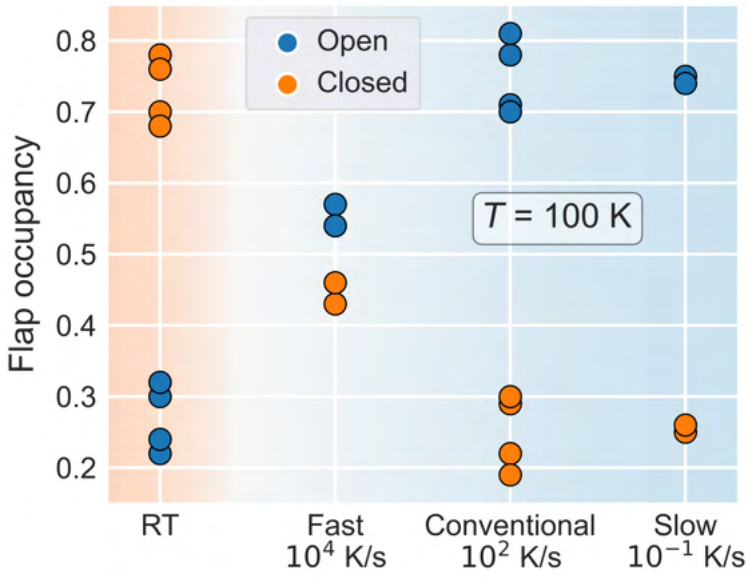

Figure 2.4: Flap occupancy as a function of cooling rate for cryoprotectant-free samples cooled from room temperature to 100 K. Room temperature data points are also shown on the left for comparison. Fast cooling produces significantly different flap occupancy from conventional or slow cooling suggesting that the flap opening/closure occurs at the timescale of 10 ms. Thus, fast cooled structure is a better representation of the room temperature–biologically relevant–structure than the conventional or slow cooled structures.

As shown in Figure 2.4, as the cooling rate increased, the occupancy of the closed conformation at 100 K increased toward its room temperature value, from ~0.26 at a

cooling rate of 10^{-1} K/s to ~0.44 at a cooling rate of 10^4 K/s. Slow and conventional cooling give similar occupancies, suggesting that the timescale of the flap motion is much shorter than the 1 s cooling time in conventional cooling, whereas a 10 ms cooling time is sufficient to partially trap the room temperature conformation. Conservatively assuming a logarithmic extrapolation based on the 10^2 and 10^4 K/s cooling rate occupancies gives an upper bound on the cooling rate needed to fully trap the room temperature conformation of $\sim 10^7$ K/s, corresponding to a cooling time of $\sim 10\,\mu s$. Cooling rates of 10^6 K/s are feasible when using microcrystals.

2.3.4 Effects of denaturants (urea and guanidine) on the active site and flap conformations

Destabilization of the mobile flap

Urease is an interesting target for experiments since it is structurally very flexible and extremely rigid at the same time. The mobile flap is so flexible that its occupancy distribution is modulated by changes in pH, temperature, and glycerol soaks. At the same time, its active site architecture other than the flap is rigid, being insensitive to all these perturbations. Apo-urease of KAU, completely lacking the catalytically critical bi-nickel site, shows only a very small alpha carbon rms deviation (0.2 Å) with respect to the holoenzyme (including the flap conformation) [38]. Thus, identifying ways to perturb the rigid active site geometry to gain insight into the enzymatic mechanism is of interest.

We collected complete datasets from samples soaked in 0.1 and 1 M concentrations of urea at 300 and 100 K; and 0.5, 1, 2 and 4 M concentrations of guanidine at 300 K.

Even though urea is the natural substrate of urease, urea-bound urease without any other inhibitor or cofactor has never been observed in a resting state of the enzyme due to the short lifetime of urease-urea complex. A recent study used fluoride in addition to urea in the crystallization recipe. The fluoride replaced a nickel-bridging hydroxide, preventing the enzyme from proceeding to the catalytic step and allowing structure determination of a minimally perturbed urea-bound urease complex [44].

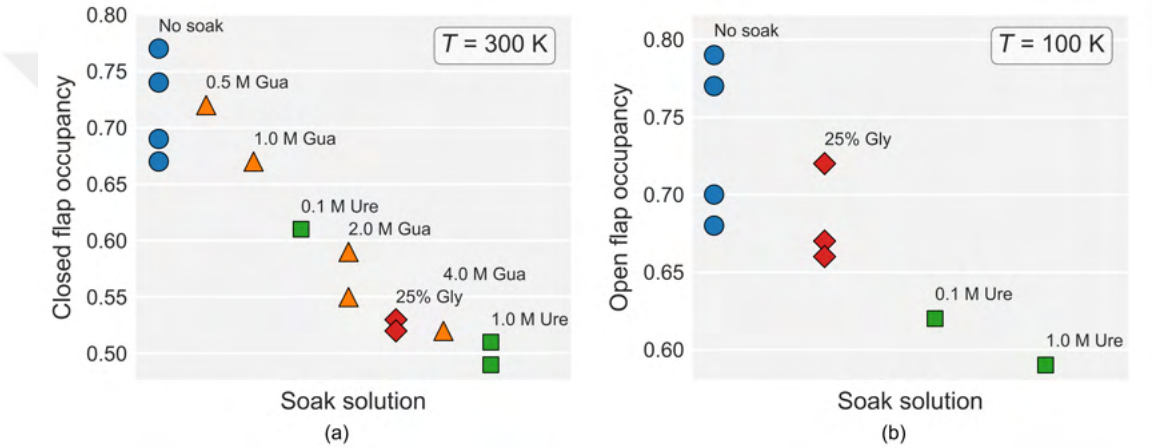

Figure 2.5: Flap occupancy vs. denaturant soak types and concentrations at 300 K (a) and 100 K (b). Soaking samples in denaturant or cryoprotectant solutions shifts the equilibrium occupancies with impact being systematically larger with increasing concentrations of soak solutions. Soaking samples in denaturant or cryoprotectant solutions does not favor or disfavor a particular flap conformation, but rather, they always disrupt the dominant flap at a given temperature i.e., closed flap at room temperature and open flap at 100 K.

Figure 2.5 shows the flap's closed state occupancy at (a) 300 K and (b) 100 K for different soak types and concentrations. The soak conditions are ordered horizontally by effect on flap occupancy. At 300 K both urea and guanidine soaks destabilize the dominant closed state of the flap, with urea being roughly four times as effective as guanidine on a molar basis. A 25% w/v glycerol (corresponding to ~2.7 M) soak has a similar impact on flap occupancies as ~4.0 M guanidine soak or ~1.0 M urea soak. At 100 K, soaking in urea and glycerol again destabilize the dominant – now open – flap state, but urea has a larger effect, and both have a smaller effect than at $T = 100$ K. Thus,

it seems that soaking in urea, glycerol or guanidine does not favor particularly open or closed flap; instead, they destabilize the dominant flap conformation at that temperature or condition.

Remodeling of tetrahedral water network

The tetrahedral water network around the bi-nickel site is invariantly found in KAU and SPU crystallographic structures as long as the resolution of the data allows ordered water placement. This network of four waters are so rigid and stable that they are only replaced or distorted by binding of inhibitors such as AHA, BME, PHO and fluoride to the bi-nickel site or by completely getting rid of the nickels in the case of apo-urease. Additionally, it has been proposed that the mobile flap might be coupled with the bi-nickel site through W_3, one of the four water molecules constituting the tetrahedral water network and the furthest from the nickels [41]. Thus, major changes in the tetrahedral water network may be relevant in terms of understanding their contribution to the enzyme structure and dynamics.

While the active site geometry was almost invariant across our samples soaked in urea, guanidine, and glycerol, the position of W_3 shifted significantly in the urea and 25% glycerol-soaked samples. As shown in Figure 2.6(a) for the 1 M urea-soaked sample, W_3 moves away from other three waters by ~1.0 Å. Figure 2.6(b) compares room temperature electron density maps of the soak-free native structure (on the left) and the 1 M urea-soaked structure (on the right). A similar shift is observed in KAU mutants His320Asn, His320Gln and Asp221Ala (PDB codes 1EJU, 1EJV and 1EJR, respectively). For the 0.1 M urea-soaked sample and the 25% v/v glycerol soaked sample, the W_3 displacement is ~0.4 Å, smaller than for 1 M urea.

Guanidine-soaked samples do not show any displacement of W_3, even at 4 M concentration. Both guanidine and urea have similar effects on the flap conformation, but they apparently interact differently with the nickel active site. One plausible explanation could be that urea is able to enter the active site so it can remodel the tetrahedral water network while guanidine is unable to do so. This is further supported by our finding that the inter-nickel distance systematically increases with increasing urea concentration from 3.68 Å (native, soak-free) to 3.76 Å (0.1 M urea) and 3.80 Å (1 M urea) at room temperature, while all guanidine-soaked samples had the same inter-nickel distance of 3.70 ± 0.01 Å, almost the same as of the native structure. As with urea, the 25% glycerol soak increases the inter-nickel distance to 3.74 Å.

Another notable perturbation in the tetrahedral network caused by soaking in urea but not in guanidine is the average inter-oxygen distance between W_B, W_1 and W_2

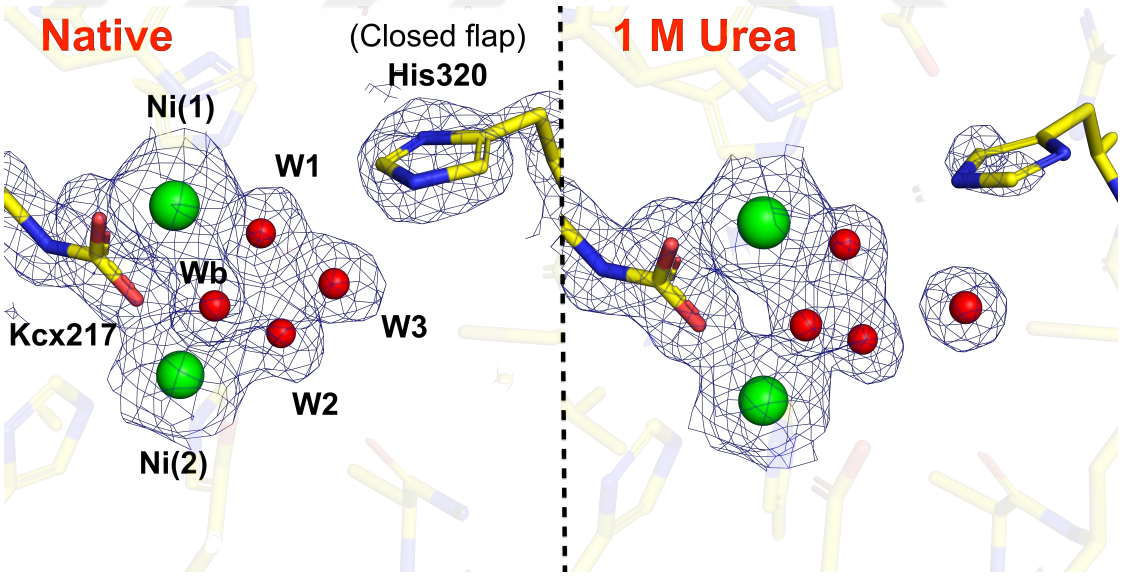

Figure 2.6: Displacement of the W_3 (the farthest of the four tetrahedral waters ligating the nickel site) away from the nickel site by 1 Å in 1 M urea-soaked sample with respect to the native sample. The native structure (left) and 1 M urea-soaked structure (right) are both at room temperature and $2mF_o - DF_c$ maps shown are rendered at 1.5 σ. His320 sidechain of the closed flap which is in H-bonding distance to W_3 has a weaker density map in 1 M urea-soaked sample as soaking the sample in urea destabilized the closed flap.

(nickel-bridging water or hydroxide, Ni(1)-bound water, and Ni(2)-bound water, respectively). The average inter-oxygen distance between these three waters in the native soak-free structure at room temperature is 2.17 Å, shorter than typical water-water distances of ~2.4 Å to ~3.2 Å with a sharp peak around ~2.8 Å found in crystallographic structures [61, 62]. With urea soaking it gets even shorter, 2.09 Å in 0.1 M urea and 2.04 Å in 1 M urea. Similarly, at 100 K, inter-oxygen distance decreases from 2.14 Å in native to 2.04 Å in 1 M urea soaked sample.

These findings may be evidence that flap conformation is not directly correlated with the nickel site architecture: displacement of W_3 by 1 Å, a decrease in the separation of the other three waters, and an increase in the inter-nickel distance in the case of urea soak produce a similar flap occupancy distribution as guanidine soak which does not alter the nickel site at all. Conversely, changes in temperature and cooling rate affect the flap conformation but not the active site water configuration, and a study of SPU showed that pH similarly changed flap occupancy without remodeling these waters [60]. Flap and active site conformations may still show correlated changes during substrate binding.

2.3.5 Search for the "wide open" flap conformation state

All of our data, at different temperatures, cooling rates, and soak conditions, was re-refined using the three-flap state model, including the "wide open" state identified in MD simulations. The average occupancy of the wide-open state across our data sets was 0.16 ± 0.03, and the maximum observed occupancy was 0.23 for a 200 K structure (without any soak) and for a 4 M guanidine-soaked sample at room temperature. These values are largely within the uncertainties of the refinement process, and showed no

clear trend with any experimental variable. Direct inspection of maps showed no unambiguous evidence of the wide-open conformation. These results are consistent with the MD simulations. Using the MD-estimated free energy ΔG of the closed, open and wide-open flap states at room temperature, the predicted occupancies are 0.83/0.16/0.01 for the closed, open, and wide-open states, respectively, the latter value being below our detection limit. The first two values are in agreement, within refinement uncertainties, with our 300 K occupancies of 0.76/0.24 obtained from the two state model.

2.4 Discussion and conclusions

The full evolution of the structure of urease with temperature between 340 K and 100 K suggests a delicate balance underlies opening and closing of the active site flap. The cooling rate dependence of flap conformation suggests conformational relaxation occurs on millisecond timescales. Together, these suggest that a dynamical equilibrium between the open and closed states exists at 300 K (and at biological temperature), with the closed state being preferred. However, even as the flap undergoes a large change in conformation, the active site geometry and key bond lengths, including those of four critical waters displaced by urea, are remarkably insensitive to temperature. These results provide information relevant to understanding the enzymatic mechanism.

The strong temperature dependence and high mobility of the active site flap, which contains the enzymatically essential His320 residue, suggest that there is an important entropic contribution to its free energy. This contribution likely plays a role in stabilizing the transition state of urea hydrolysis. The strong temperature dependence of the flap conformation also explains consistent differences between the flap states of KAU vs. SPU structures. As summarized in Table 1, the overwhelming majority of KAU

structures, including native structures, were collected at room temperature, and the flap exhibits a primarily closed conformation, except for KAU mutant variants with large mutations at the active site, whereas all SPU structures, including native structures, were collected at 100 K, and the flap exhibits a primarily open conformation, except for SPU complexes with transition state analogs (PDB codes: 3UBP, 5OL4, 6H8J), as summarized in Table 2.

The two state model we use is just a model for describing differences between data sets; it does not provide the best possible fit to the data, and does not accurately capture the full conformational ensemble of the enzyme at all temperatures. At high and low temperatures, where the refined occupancies of the closed and open state, respectively, are near 1, the model provides a reasonable fit to the density, as indicated by typical real space correlation coefficients (RSCC) in the flap region of ~0.94 (average RSCC of the closed flap at 300 K) and ~0.90 (average RSCC of the open flap at 100 K). But when occupancies of the closed and open states are near 0.5, the density typically shows little real space structure that can be identified as belonging to either state (RSCC of open or closed flap is ~0.79), so the flap may best be considered as disordered. On the other hand, inspection of previously deposited $T = 300$ K structures where the flap state was labeled as "disordered" [63] usually (when data set resolution is high enough) show density that can be clearly identified with open flap as the major state and closed flap as the minor state.

Our attempts to use a three-state model, including the "wide open" state identified in MD simulations [54, 55] as being essential for urease to gain access to the active site, to fit our data did not give statistically significant occupancy of the "wide open" state at any temperature. This is in line with the MD simulation results. The free energy difference between the closed state and the open and wide open states were estimated to be

less than ~1 kcal/mol (with an activation barrier separating the states of ~1.5 kcal/mol) and ~2.5 kcal/mol (with an activation barrier of less than ~4 kcal/mol), respectively. Assuming occupancies are proportional to $exp(-G/RT)$, these give predicted room temperature occupancies of the closed, open, and wide-open states of 0.83, 0.16 and 0.01. The first two values can be compared with a closed/open ratio of 0.73/0.27 obtained by refining our data against two-state model, which correspond to a free energy difference of ~0.6 kcal/mol. At $T = 100$ K, our closed/open occupancy ratio of 0.24/0.76 indicates that the closed state has become the higher energy state with an energy difference of ~0.23 kcal/mol.

The MD-predicted wide-open-state occupancy, while small, is non-zero, and has been claimed to be essential for enzymatic activity. Urease in crystal form has negligible enzymatic activity [50]. Substantial crystal solvent content (51%) and large solvent channels (up to 50 Å) should allow easy urea diffusion into the crystals, and this is borne out by substantial effects of urea, guanidine, and glycerol soaks on the in *crystallo* flap conformation seen here. This suggests that quenching of activity is associated with crystal contact imposed constraints on conformational fluctuations – e.g., into the wide-open state – required for activity. The timescale for quenching by cooling of the room-temperature closed conformation is ~10 ms, much longer than the ~200 μs required for enzymatic turnover in solution, which could be evidence for differences between solution and crystal environments.

Glycerol, urea, and guanidine soaks all have qualitatively similar effects on the active site flap: within our two-state model, all tend to destabilize the dominant conformation at a given temperature. At room temperature, the closed-state occupancy decreases, whereas at 100 K the open state occupancy decreases. Real space correlation coefficients between the density and both states decrease. Inspection of the flap density sug-

gests that as urea and guanidine concentrations are increased, the well-defined features corresponding to the closed and open conformations are gradually washed out. So, the primary effect of these soaks can be viewed as an increase in flap disorder, which tends to shift occupancies in the two state model toward 0.5/0.5, rather than a lowering of the free energy difference between two well defined structural states.

The lability of the active site flap contrasts with the rigidity of the rest of the active site. No significant change is observed in the native enzyme versus temperature between 340 K to 100 K or versus cooling rate in the distance between the catalytically active Ni atoms, between waters W_B, W_1 and W_2, and between these waters and the Ni atoms, even as the flap and its catalytically important His320 residue undergo a substantial shift from the closed toward the open conformation. Small variations in inter-water and water-Ni distances between data sets can be attributed to variations in dataset and map resolution. The distances between these four waters are much smaller than in bulk water. But refined $T = 300$ K occupancies are ~1 (> 0.95) for W_1, W_2 and W_3, and ~0.7–0.9 for W_B, suggesting that they are concurrently occupied. The enthalpy associated with the resulting large strain is released on urease binding and release of these waters into the bulk.

The only appreciable changes in active site are observed in response to soaks. The native Ni-Ni distance is ~3.68 ± 0.02 Å at all temperatures. At room temperature, soaking in 25% glycerol increases this to 3.74 ± 0.02 Å, soaking in 1 M urea increases this to 3.80 ± 0.01 Å, while soaking in 4 M guanidine has no significant effect. The corresponding shifts in the position of W_3 are 0.65 ± 0.03 Å and 0.94 ± 0.03 Å, respectively (insignificant displacement for 4 M guanidine), away from the other waters and the Ni sites. Only urea appreciably affects the average separation between W_B, W_1 and W_2, decreasing it from 2.15 ± 0.03 Å to 2.05 ± 0.01 Å. At 100 K, the Ni-Ni distance shows

no significant deviation from the native value for any of these soak conditions. The corresponding shifts in the position of W_3 for 25% glycerol and 1 M urea are 0.40 ± 0.06 Å and 0.71 Å, respectively; 1 M urea decreases the average separation between W_B, W_1 and W_2 from 2.24 ± 0.06 Å to 2.04 Å.

These findings may be evidence that flap conformation is not directly correlated with the nickel site architecture: displacement of W_3 by ~1 Å, a decrease in the separation of the other three waters, and an increase in the inter-nickel distance in the case of urea soak produce a similar flap occupancy distribution as guanidine soak which does not alter the nickels or waters at all. Conversely, changes in temperature and cooling rate affect the flap conformation but not the active site water configuration. A study of SPU showed that pH similarly changed flap occupancy without remodeling these waters [60]. Flap and active site conformations may still show correlated changes during substrate binding.

CHAPTER 3

EFFECTS OF PROTEIN-CRYSTAL HYDRATION AND TEMPERATURE ON SIDE-CHAIN CONFORMATIONAL HETEROGENEITY IN MONOCLINIC LYSOZYME CRYSTALS

This chapter is identical to the published manuscript (doi: 10.1107/S2059798318000207) with contributions of David W. Moreau under the supervision of Robert E. Thorne. Robert E. Thorne conceived of the presented idea. Hakan Atakisi and David W. Moreau carried out the experiments. Hakan Atakisi performed the analyses. Robert E. Thorne and Hakan Atakisi wrote the manuscript. All authors discussed the results and contributed to the final manuscript.

3.1 Introduction

Water, with its unique physicochemical properties, plays a central role in determining the structure and function of proteins and other biomolecules. Water is essential to protein folding/unfolding, to stabilization of higher-order structures, and in determining catalytic efficiencies [64]. The internal and large-scale conformational dynamics of proteins are coupled to the dynamics of surface-ordered (hydration) and bulk solvent [65, 66]. Protein function ceases when surrounding solvent is depleted to only one or two hydration layers [66].

X-ray crystallography has long been used to study the effects of crystal and protein hydration [67–71]. Protein crystals contain large amounts of solvent, typically 30–70% by volume [72, 73] that fills intermolecular spaces within the unit cell as well as internal protein cavities. Solvent molecules preferentially occupy specific sites near the protein surface, especially within the first hydration shell and especially at low temperatures,

allowing them to be crystallographically detected, but most solvent tends to be positionally disordered and so contributes only to the average solvent electron density.

Crystal solvent content depends on relative humidity (r.h.) or equivalently, the activity of water, a thermodynamic variable. Relative humidity determines how proteins pack within the crystal, the structure of individual proteins, and also the solvent structure. Dehydration – both intentional and accidental – decreases the unit cell volume, and crystal order and diffraction resolution often degrade but sometimes improve. Commercial and homemade systems for adjusting hydration levels to optimize diffraction properties are now in regular use [74–76], and diffraction improvements have been obtained at relative humidities as low as 74% [77]. Dehydration below ~85% r.h. often leads to first-order structural phase transitions involving cooperative loss of a finite amount of water and the appearance of a new lattice [68, 70, 71, 78]. Diffraction outcomes are strongly affected by the rate of dehydration, which affects the ability of the protein and crystal to adjust to the evolving equilibrium state. Rapid dehydration often leads to spatial inhomogeneities and plastic lattice failure, while slow dehydration can "supercool" a given lattice to relative humidities well below that at which it becomes thermodynamically unstable [79]. Disorder created by dehydration can often be largely or fully reversed and the original crystal packing restored by rehydration [76, 78, 79].

Previous crystallographic studies of the effects of hydration on protein structure have mainly focused on changes to the main chain conformation. For example, orthorhombic and monoclinic crystals of bovine pancreatic ribonuclease-A dehydrated to r.h. showed rearrangement of surface loops and a small change in the hinge angle between subdomains relative to the native hydrated structure [5, 6]. Dehydration of cytomegalovirus immediate-early 1 protein triggered a structural transition involving remodeling of tertiary structure in addition to a change in space group [7]. The most detailed studies

have been performed using various crystal forms of hen egg white lysozyme, which can retain crystalline and molecular order at humidities as low as [78, 80–83]. The monoclinic form diffracts to highest resolution over the widest range of relative humidities. The most pronounced main-chain changes occur in regions involved in substrate binding [81]. However, the effects of dehydration on side chain conformations and solvent structure have not been systematically studied.

The importance of a protein's conformational ensemble in folding [84], molecular recognition [85], catalysis [86], and allosteric regulation [87] has long been recognized [88, 89]. As of 2010, only ~5% of PDB-deposited crystallographic structures were refined against multiple conformer models [20]. In the last decade, this situation has improved as methods for identifying and modeling main and side-chain conformational heterogeneity have been developed and become easier to use. Low electron density levels previously assumed to be noise in fact can contain abundant information about minority occupancy conformers. Analysis software including Ringer [20, 90], which samples and maps electron density around dihedral angles to identify peaks associated with low-occupancy alternative conformers of side chain rotamers, and qFit [19, 91], which auto builds up to four conformations per residue, have facilitated generation of multiconformer models. Bond order parameters S^2, originally developed to describe NMR data [92], have been generalized for crystallography to characterize both harmonic and nonharmonic (conformational) side chain fluctuations [21].

A primary application of these tools to date has been to characterize the effects of temperature. A comparison of PDB-deposited structures of 30 proteins at room temperature and $T = 100$ K indicated that cryocooling causes substantial remodeling of side chain conformations [1]. For at least 24 of these proteins cryoprotectants used in cryogenic data collection were absent from crystals used in room temperature mea-

surements, complicating interpretation of structural differences. Conformational heterogeneity of a Michaelis complex of DHFR was found to be significantly distorted by cryocooling [3]. Analysis of crystallographic data from the human proline isomerase cyclophilin A (cypA) acquired at 7 temperatures between 298 K and 100 K has revealed how side chain conformations and bond order parameters evolve with temperature across the entire molecule. The evolution was found to be highly heterogeneous: bond disorder parameters have a "hockey stick" shape and the "transition" temperatures above which order decreases more rapidly have a broad distribution between ~180 and ~270 K. Analysis of diffraction data collected to high doses from three different proteins (including cypA) at room temperature indicates that radiation damage has only small effects on conformational heterogeneity compared with those of temperature [93]. However, the full physical and biological significance of temperature-dependent conformational changes remains unclear.

Water activity (relative humidity), like temperature, is a thermodynamic variable. Some of the gross effects of dehydration, such as unit cell contraction, are similar to those of cooling. Moreover, unintentional dehydration routinely occurs during crystal handling, mounting, and storage in both room temperature and $T = 100$ K crystallographic practice. Unintentional dehydration is a primary cause of nonisomorphism of identically prepared crystals at room temperature and of unit cell variations at cryogenic temperatures [94]. Could dehydration account for some conformational changes that have been attributed to cooling? To what extent does dehydration mimic the effects of cooling? How do the effects of these two thermodynamic perturbations on protein conformational ensembles and solvent structure as they exist in crystals differ?

Here we report high resolution (1.1 to 1.9 Å) measurements and multiconformer analysis of hydration-dependent side chain conformations in monoclinic lysozyme crys-

tals from relative humidities between 99% to 11% r.h., corresponding to crystal solvent contents ranging from 38% to 18% (v/v), respectively, at both 298 K and 100 K. Detailed structural comparisons were facilitated by using carefully prepared and characterized crystals of otherwise identical composition. Substantial sidechain remodeling occurs on dehydration. Solvent exposed residues are significantly more affected than buried residues, in part due the large increase in the number of crystal contacts as the protein's packing becomes denser. The largest side chain conformation changes occur at humidities near ~75%, where significant depletion of water within the first hydration shell begins. Dehydration to 93% r.h has similar effects to cooling to 100 K: dehydration causes a 5.4% contraction of the unit cell volume and remodels 12.6% of side chains, compared with a 6.8% cell contraction and 13.6% remodeling on cooling, with only solvent-exposed residues affected in both cases. These results complicate interpretation of differences between room temperature and cryogenic temperature conformational ensembles when the precise crystal history is not known and/or is not carefully controlled; they suggest that modest dehydration can serve as a proxy for studying many of the effects of temperature; and they emphasize the importance of packing environment in selecting targets for variable temperature crystallographic studies.

3.2 Materials and methods

3.2.1 Crystallization

Monoclinic lysozyme crystals were grown at room temperature by the hanging-drop vapor diffusion method. Lyophilized lysozyme powder (Sigma Aldrich, St. Louis, Missouri USA) was dissolved in 50 mM sodium acetate buffer solution at

pH 4.5 at a concentration of 10–15 mg/ml, and reservoir solutions were comprised of 2.8–3.0 % (w/v) (~0.35 M) sodium nitrate in the same buffer solution. Crystallization drops of 6–10 μL volume, prepared by mixing equal volumes of protein and reservoir solution, were suspended over ~500 μL of reservoir solution, and rectangular plate-like crystals, mostly in clusters, were obtained within a week. The largest crystals, 300–600 μm in the long dimension, were used to facilitate collection of complete crystallographic data sets at room temperature with minimal radiation damage.

3.2.2 Crystal dehydration

Crystals were dehydrated to seven relative humidities between 99% and 11% by equilibrating them at $T = 293-298$ K with mother liquor reservoir solution (~99.2±0.5% r.h.) and saturated solutions of potassium nitrate (93.4 ± 0.8%), potassium chloride (85.3 ± 0.7%), sodium chloride (75.4 ± 0.3%), sodium bromide (57.7 ± 0.9%), magnesium chloride (33.1 ± 0.2%) and lithium chloride (11.3 ± 0.5%) [95–97]. Each crystal was harvested from its mother liquor using a microfabricated polymer loop, and excess mother liquor carefully removed from crystal and loop by blotting. The crystal+loop was then inserted into a polymer capillary (MicroRT system, MiTeGen, LLC, Ithaca, NY) containing ~20 μL of saturated salt solution and equilibrated for 24–48 h. The distance between the crystal and solution was 1 cm, and a few extra salt grains were added to the solution to ensure that it remained saturated.

Crystals that are rapidly dehydrated to 75% r.h. and below often undergo catastrophic plastic failure leading to loss of diffraction [79]. This most likely occurs due to inhomogeneous stresses that develop as the crystal lattice undergoes its structural phase transition near 85% r.h. To reduce the rate of equilibration during dehydration to

75% r.h. and below, crystals were transferred to and then harvested from oil (NVH oil (Cargille, Cedar Grove, NJ) or LV cryo-oil (MiTeGen, LLC, Ithaca, NY)) before dehydration, and this was found to reproducibly yield high quality diffraction. On the other hand, crystal dehydration to 85% and 93% r.h. can be excessively slowed by surrounding mother liquor, so excess mother liquor was carefully removed. Crystal equilibration times in the capillary tubes were 24–48 hours in all cases, which was experimentally determined to be sufficient for equilibration and to yield unit cell volumes consistent with those from previous studies [83, 98].

3.2.3 X-ray data collection

X-ray diffraction data was collected at the Cornell High Energy Synchrotron Source (CHESS) on MacCHESS station A1 using a 19.6 keV X-ray beam with a diameter of 100 μm. Diffraction images were recorded using either an ADSC Quantum-210 or Dectris Pilatus 6M detector. Complete room temperature data sets were collected from 2–5 different crystals at each r.h., and models refined using the highest resolution data sets at each r.h.

To compare the effects of dehydration with the effects of cryocooling, additional X-ray data sets were collected at $T = 100$ K using a nitrogen gas cryostream from native monoclinic lysozyme crystals (99% r.h.) and from crystals dehydrated at 298 K to 85%, 75%, 58%, and 11% r.h. No penetrating cryoprotectants were used, to ensure that the crystal composition was identical for 298 K and 100 K data collection. Each crystal was looped from its drop and excess mother liquor removed by blotting. The crystal was then transferred to LV CryoOil, harvested using a microfabricated polyimide loop, and inserted into a salt-solution-containing capillary tube. After equilibration for

	r.h. 99%	r.h. 93%	r.h. 85%	r.h. 75%	r.h. 58%	r.h. 33%	r.h. 11%
Wavelength (Å)	0.6299	0.6309	0.6309	0.6309	0.6309	0.6309	0.6299
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
a	28.13	27.62	27.40	26.73	26.36	25.32	25.33
b	63.33	62.84	62.61	58.44	57.21	56.84	56.39
c	60.72	58.96	58.79	31.10	31.00	30.75	30.46
β(°)	90.61	91.27	90.86	111.81	111.54	109.55	109.76
Solvent content (%)	34.9	31.2	30.2	22	19.1	15.6	14.1
Total reflections	706703	174070	253172	112196	53402	28031	43782
Redundancy	12.4(5.3)	3.3(2.4)	3.4(2.6)	3.3(2.7)	3.2(2.2)	2.8(2.1)	5.1(2.8)
Completeness (%)	98.9 (98.9)	99.3 (93.2)	98.0 (98.0)	99.6 (96.9)	97.9(82.1)	95.2 (84.1)	96.6 (78.8)
Mosaicity (°)	0.09 – 0.56	0.13 – 0.66	0.14 – 0.28	0.33 – 0.51	0.48 – 0.64	0.57 – 2.45	0.53 – 1.14
$CC_{1/2}$	1.000 (0.589)	0.996 (0.593)	1.000 (0.725)	0.996 (0.804)	0.980 (0.242)	0.957 (0.163)	0.996 (0.788)
Mean I/σ	45.3(1.0)	19.5(0.9)	20.7(1.1)	24.8(2.0)	19.5(0.5)	9.3(0.5)	35.2(20.6)
Resolution (Å) (at I/σ = 2)	1.35	1.31	1.19	1.11	1.54	1.94	1.70
R-merge / R-meas (%)	5.6/5.8	10.4/12.3	4.3/5.1	3.5/4.2	5.7/6.6	16.4/19.4	7.0/7.8
Wilson B-factor ($Å^2$)	20.4	14.5	11.9	11	22.6	26.3	28.9
Refinement resolution (Å)	28.13-1.26	27.61-1.24	24.69-1.14	16.09-1.11	16.00-1.38	23.86-1.61	23.84-1.69
R-work / R-free (%)	12.9/15.8	14.2/17.2	12.2/15.2	10.5/13.2	17.3/22.6	23.8/29.1	22.0/26.6
Test set for R-free (%)	5	5	5	5	5	5	5
RMS bonds (Å) / angles (°)	0.015/1.46	0.013/1.36	0.026/1.46	0.011/1.37	0.005/0.73	0.003/0.76	0.003/0.73
Ramachandran favored / outliers (%)	97.3/0.3	98.3/0.0	96.9/0.0	96.1/0.0	95.4/0.5	96.3/0.7	94.6/0.7
Rotamer outliers (%)	2.6	0.2	0.9	0.4	0.4	2.61	6.3
Clashscore	3.8	0.9	4.2	3.4	1.1	3.8	6.2
B-factor ($Å^2$)	28.4	19.7	14.8	13.6	28.6	32.7	38.7
Protein B-factor ($Å^2$)	27.9	18.9	13.9	13	28.2	32.5	38.6
Solvent B-factor ($Å^2$)	35.1	32.8	30.4	25.9	39	38	40.8
Waters (first shell / total)	128/137	177/180	212/215	90/97	53/54	40/42	15/18

Table 3.1: Data collection and refinement statistics for room temperature data collected using monoclinic lysozyme crystals.

24–48 hours, a few frames of X-ray data were first collected at 298 K, with the capillary tube in place and with the nitrogen cryostream blocked using an “air blade” shutter, to ensure that a unit cell consistent with the target dehydrated state had been obtained and that diffraction was of high resolution. The capillary tube was then removed and the cryostream unshuttered to cool the sample, and a complete and redundant data set collected. The maximum time the sample was exposed to ambient humidity (20–45% r.h.) air or dry N2 while its temperature was warm enough (>230 K) for dehydration to

	r.h. 99% cryo	r.h. 85 % cryo	r.h. 75 % cryo	r.h. 58 % cryo	r.h. 11 % cryo
Wavelength (Å)	0.6307	0.6307	0.9768	0.9768	0.6307
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
a	27.39	26.74	26.22	26.23	25.03
b	62.09	62.09	58.01	57.16	56.41
c	59.28	58.21	30.67	30.69	30.30
β (°)	90.36	90.32	111.64	111.35	109.14
Solvent content (%)	30.2	27.2	20.9	16.7	13
Total reflections	52483	74056	28918	23476	7418
Redundancy	6.7(4.0)	7.8(4.5)	8.0(4.2)	6.0(3.4)	2.5(2.1)
Completeness (%)	97.0(97.0)	97.0(97.0)	96.9(95.1)	98.4(96.8)	96.0(96.0)
Mosaicity (°)	0.43 – 0.61	0.41 – 1.34	0.46 – 0.65	0.32 – 0.85	0.74 – 1.56
$CC_{1/2}$	0.996(0.762)	0.996(0.858)	1.000(0.957)	0.996(0.909)	0.992(0.885)
Mean I/σ	34.0(1.7)	36.5(1.8)	40.3(4.4)	34.7(3.6)	10.2(0.6)
Resolution (Å) (at I/σ = 2)	1.26	1.10	1.17	1.25	1.80
R-merge / R-meas (%)	6.8/7.4	6.5/6.9	5.5/6.1	7.3/8.1	9.2/9.8
Wilson B-factor (Å^2)	11.9	7.5	11.7	14.5	20.9
Refinement resolution (Å)	27.50 – 1.26	29.10 – 1.10	23.24 – 1.16	28.58 – 1.24	28.62 – 1.77
R-work / R-free (%)	13.4/16.7	12.9/15.2	11.2/14.3	14.3/17.9	18.2/23.1
Test set for R-free (%)	5	5	5	5	5
RMS bonds (Å) / angles (°)	0.032/1.77	0.017/1.68	0.010/1.14	0.008/1.00	0.0087/1.16
Ramachandran favored / outliers (%)	98.2/0.0	98.6/0.0	96.0/0.4	97.4/0.0	95.3/2.3
Rotamer outliers (%)	0.5	0.9	0.83	0.59	3.7
Clashscore	3.6	5.5	5.2	4.7	5.6
B-factor (Å^2)	16.8	11	13.9	19.5	22.7
Protein B-factor (Å^2)	15.6	9.2	13.1	18.6	22.4
Solvent B-factor (Å^2)	27.8	20.6	24	29.1	29.7
Waters (first shell / total)	320/326	363/384	133/143	102/106	35/41

Table 3.2: Data collection and refinement statistics for data collected from monoclinic lysozyme crystals at T = 100 K.

occur was less than 3 seconds. Given the relatively large crystal sizes, this was too short for appreciable changes in crystal hydration. At the lowest r.h. of 11%, the 100 K unit cell for each crystal was smaller than the room temperature unit cell recorded before the capillary tube was removed, indicating that no significant rehydration occurred. High quality, ice-free T = 100 K data with consistent unit cells was obtained for all r.h. values without use of penetrating cryoprotectants. Data-collection and refinement statistics are

given in Tables 3.1 and 3.2.

3.2.4 Data processing

Data frames were indexed and scaled using SCALEPACK and DENZO [51], and molecular replacement and model refinement were performed using PHENIX [52]. PDB entries 2D4K and 2D4J were used as initial models for crystals with r.h. above and below the ~85% r.h. structural phase transition, respectively. Water molecules were modeled in all structures, and for data with resolution better than 1.4 Å anisotropic B-factors were refined and H atoms modeled.

Standard refinement programs, by default, fit a single conformer model to the experimental electron density map. Initial multiconformer models were generated using qFit [19], which builds up to four alternative conformations per residue where they are locally supported by the real space electron density map. Ringer was used to detect additional rotameric conformations of each side chain, and then the qFit models were manually inspected, modified, and validated using COOT [53]. Final model validation was performed using MolProbity in the PHENIX software package [99]. The resulting multiconformer models led to substantial reductions in R-free values (up to ~2%), especially for high resolution structures.

Protein volumes within the unit cell were calculated as the volume enclosed by the solvent excluded surface (SES) using the program *3VEE* with a probe size of 1.4 Å and grid size of 0.07 Å. The SES is calculated by "rolling" a fixed-diameter probe over the surface of a voxelized 3D model of the protein [100]. Hydrogens were added (if absent) and alternate conformations of side chains were removed (as the software assumes fixed occupancy of 1 for all sidechains) from the model before using *3VEE*.

The fractional solvent accessible surface area for each residue was calculated using the web server at http://cib.cf.ocha.ac.jp/bitool/ASA/. The solvent volume within the unit cell was calculated by subtracting the SES volume of protein from the unit cell volume. These results were compared with those obtained using a Matthews coefficient analysis.

3.3 Experimental results and analysis

3.3.1 Unit-cell, protein and solvent volumes

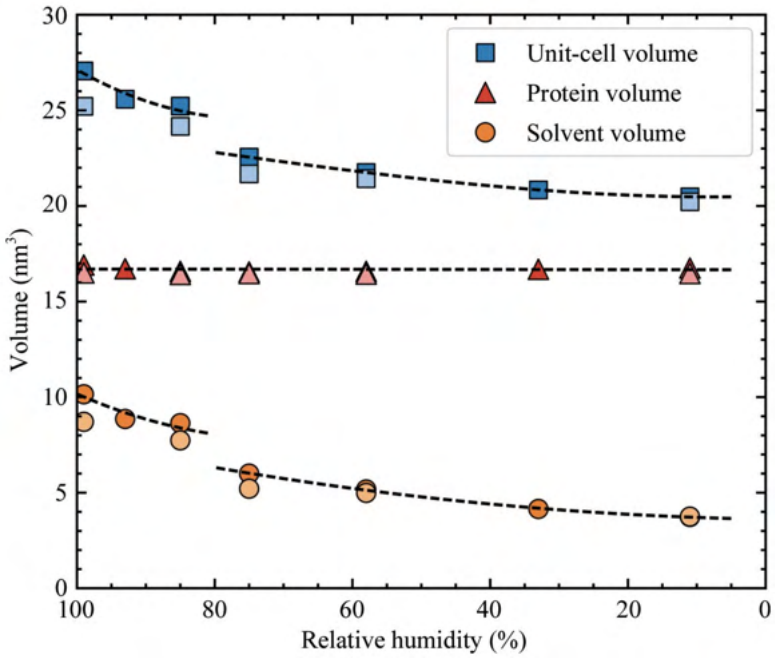

Figure 3.1: Crystallographic cell volume (square), protein volume (triangle) (calculated as the volume enclosed by solvent excluded surface (SES) area of the protein molecule), and solvent volume (circle) (given by the difference between cell and protein volume) per protein molecule for monoclinic lysozyme crystals versus relative humidity. Dark and light symbols are for room temperature and $T = 100$ K values, respectively. The discontinuity in cell and solvent volume between 85% and 75% r.h. is due to a structural transition in which the unit cell volume is roughly halved. The dashed lines are guides to the eye through the room temperature data points.

Monoclinic crystals of hen egg-white lysozyme undergo a structural transition near 85% r.h., in which the two crystallographically independent molecules in the native asymmetric unit become equivalent, resulting in a halving of the unit cell [78]. The crystals can maintain excellent order down to 5% r.h., corresponding to a Matthews coefficient derived solvent content of 9% by volume [82]. Figure 3.1 shows how the unit cell volume, protein volume, and solvent volume per protein molecule vary with relative humidity at 298 K and 100 K. At 298 K the unit cell volume decreases by 24% between 99% and 11% r.h. On cooling to 100 K, the cell volume contracts by 6.8% at 99% r.h., decreasing to only 1.3% at 11% r.h. The SES protein volume (Appendix Table C.1) decreases by only 1–2% over this r.h. range, and by 0.5–2.4% at fixed r.h. (with no systematic dependence on r.h.) on cooling to 100 K. At 99% r.h., nearly all the 6.8% cell contraction on cooling to $T = 100$ K is due to the solvent, whereas at 11% r.h. the cell, protein, and solvent have comparably small contractions. Note that cooling the native (99%) structure to 100 K has a similar effect on cell and solvent volume as does dehydration to ~85% r.h. Note that the Matthews coefficient analysis overestimates the solvent content at all r.h. and temperatures, with the largest fractional errors occurring at low r.h. and $T = 100$ K. These errors result because our calculated $T = 298$ K lysozyme density (1.41 g/ml for the native structure) is larger than the assumed density (1.35 g/ml) in, e.g., PHENIX [101,102]; because the protein density is larger (1.44 g/ml) at 100 K; and because errors in protein density have a larger effect on solvent fraction at low solvent fractions.

3.3.2 Solvent content and solvent ordering

Figure 3.2 shows the crystal solvent content, as determined from the unit cell volume and *3VEE* analysis, as a function of relative humidity at both 298 K and 100 K, as

well as the resolution of the refined crystallographic data sets. The T = 298 K solvent content decreases from 0.37 to ~0.34 v/v between 99% and 85% r.h., abruptly drops to 0.26 v/v as a result of the structural transition, and then decreases to ~0.18 v/v at 11% r.h. Dehydration to 75% r.h. leads to a small improvement in resolution. Further dehydration degrades resolution by a variable amount that tends to decrease as the rate of dehydration decreases, and that also depends on how each crystal's shape and initial perfection affects its response to stresses during the structural transition near 85% r.h. [79].

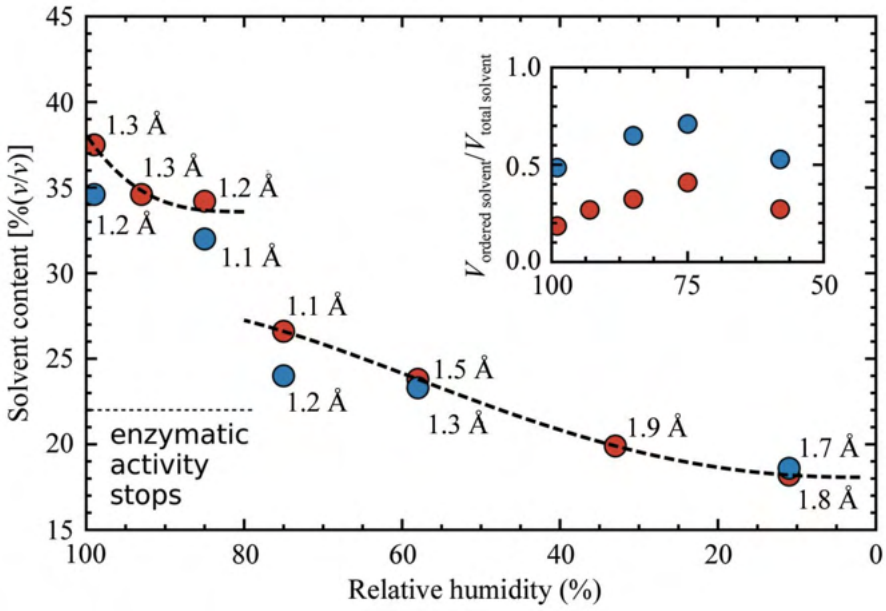

Figure 3.2: Crystal solvent content (derived from analysis of the solvent excluded protein volume as in Figure 3.1) and data set resolution as a function of relative humidity. Red and blue symbols represent room temperature and 100 K values, respectively, and the dashed line is a guide to the eye. Lysozyme's enzymatic activity ceases below 0.2 g of water per gram of protein molecule, corresponding to 21% v/v solvent content. Inset: Ratio of the crystallographically detected hydration shell volume to the total solvent volume as a function of r.h. for the five highest resolution data sets, assuming unit occupancy for all modelled waters. Near 75% r.h., the ratio of the hydration solvent volume to total solvent volume at cryogenic temperature has a maximum. The decrease in hydration shell volume between 75% and 58% r.h. is due in part to an increase in overall and hydration shell B-factors so that fewer waters were modelled. Truncation of all data sets to 1.5 Å resolution had little effect on hydration shell volumes. The dashed lines are guides to the eye through the room temperature data points.

Internal crystal solvent can be crudely separated into hydration water – water contained within the first hydration shell that interacts strongly with the protein and stabilizes its secondary and tertiary structure – and bulk-like water. Dehydration first depletes bulk-like water before affecting hydration water. Lysozyme's enzymatic activity ceases at below 0.2 g of water per gram of lysozyme molecule [64, 103]. This corresponds to a ~22% v/v solvent content that, for monoclinic lysozyme, occurs below the structural transition at ~75% r.h.

In crystallography, hydration waters are more likely to be found at particular sites near the protein surface, especially at $T = 100$ K. Many hydration shell water sites in lysozyme are invariant with respect to crystal space group [104]. The inset in Figure 3.2 shows the ratio of the crystallographically detected hydration shell volume to the total solvent volume as a function of r.h. for the five highest resolution data sets (99%–58% r.h.) at both 298 K and 100 K. Waters were modeled only if the resulting B-factor was less than 60 Å^2. The hydration shell volume was determined using the *3VEE* server by subtracting the solvent-excluded volume of the protein + modeled waters (at 100% occupancy) and of the protein only. The detected hydration shell volume roughly doubles on cooling to 100 K at a fixed r.h. and the average B-factor of the modeled waters decreases from ~35 Å^2 to ~28 Å^2 for the native crystal. This reflects increased solvent ordering and localization when thermal fluctuations are reduced [105, 106]. At ~75% r.h., corresponding to the threshold for enzymatic activity, almost all disordered "bulk" solvent has been depleted, and further dehydration depletes hydration water.

Table 3.3 indicates the extent to which modeled water sites are conserved on cooling at fixed r.h. A water site is deemed to be "conserved" between two structures if it is displaced by no more than a cutoff distance of 1.2 Å after least-squares superposition of the structures. The cutoff distance to identify conserved water sites in the literature

varies from 1 Å to 2.4 Å, with values between 1 Å and 1.5 Å being most common [107–109]. For the present data, varying the threshold between 1 Å and 1.5 Å produced only small fractional increases/decreases in the number of conserved sites. At fixed r.h., the fraction of water sites detected at 298 K that are conserved at 100 K drops from ~85% at 99% r.h. to ~74% at 58% r.h., and then drops to only ~50% at 11% r.h. Waters at the active site are mostly conserved down to 75% r.h., and then begin to disappear on further dehydration.

Relative humidity (%)	Fraction of waters conserved (298 → 100 K)
99	0.85
85	0.81
75	0.84
58	0.74
11	0.50

Table 3.3: Fraction of water sites modeled at 298 K that are conserved on cooling to 100 K, versus r.h.

3.3.3 Crystal contacts

Increased protein packing due to removal of crystal solvent increases the number of and area of intermolecular contacts. Various criteria can be used to determine if two atoms in different molecules within a crystal are in contact, e.g., if their separation is below a threshold of 4–6 Å [110–112]. Figure 3.3 shows the evolution of the number and fraction of residues involved in crystal contacts versus relative humidity at T = 298 and 100 K, assuming that atoms are in contact if their center-to-center distance is less than the sum of their van der Waals radii plus 0.25 Å [14]. As shown in Appendix Figure B.2 qualitatively similarly behavior is observed if a fixed 4 Å separation threshold is used.

At 298 K, the fraction of residues involved in crystal contacts increases from ~32%

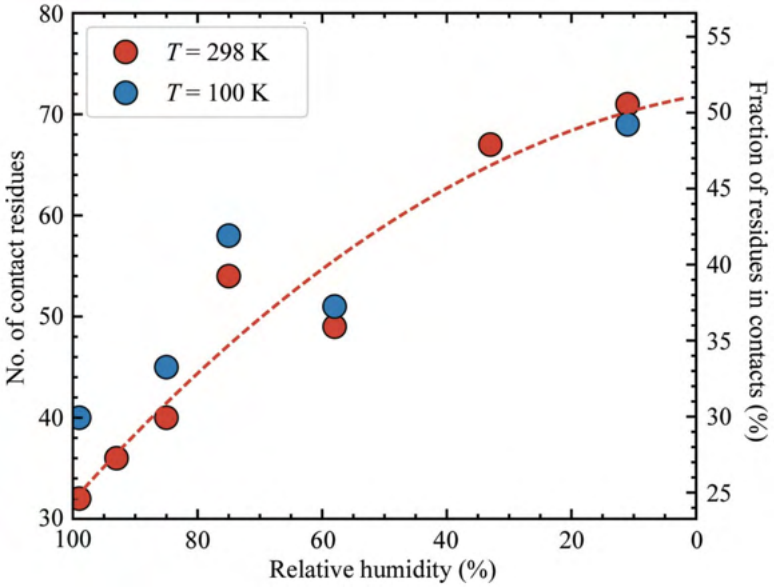

Figure 3.3: Number of residues involved in crystal contacts versus relative humidity at room temperature (red) and T = 100 K (blue). The dashed red line is a guide to the eye for the room temperature data. As r.h. decreases, the protein molecules become more densely packed, increasing the number of contact residues. The number of contacts detected depends on the assumed cutoff separation between residues. Here, two atoms are considered to be in contact if their center to center distance is less than the sum of their van der Waals radii plus some small distance e.g. 0.25 Å. Appendix Figure B.2 shows the number of contact residues determined using a fixed 4 Å cutoff. The dashed line is a guide to the eye for the room temperature data points.

at 99% r.h. to ~50% at 11% r.h. Cooling to 100 K increases the 99% r.h. contact fraction to 40%, but has no significant effect at 11% r.h. Comparing the effects of cooling and dehydration on the native (99% r.h.) structure, cooling to 100 K produces a comparable increase in crystal contacts as dehydration to 85% r.h., consistent with the similar changes in cell volume these perturbations produce.

3.3.4 Side-chain remodeling: *Ringer* analysis

Intermolecular crystal contacts affect sidechain heterogeneity as contacts formed by electrostatic interactions impose some local restrictions on the sidechain conformations

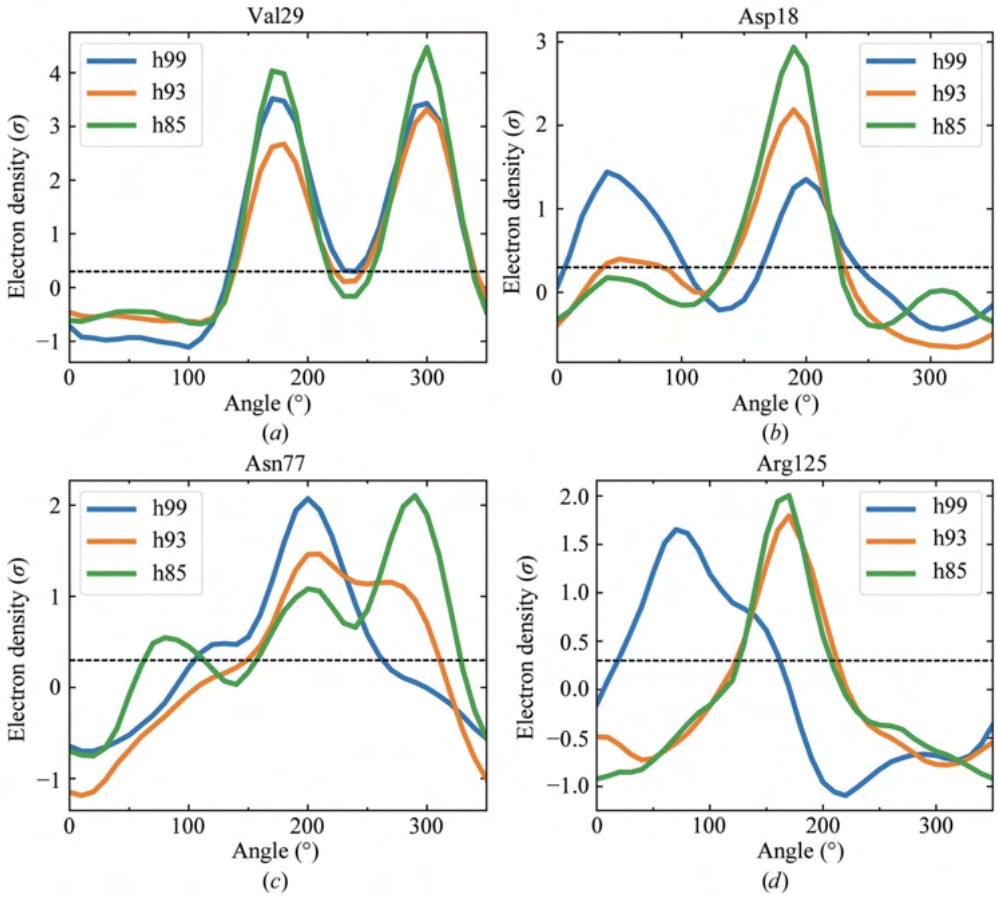

Figure 3.4: Example plots of $2mF_o - DF_c$ electron density (σ) for χ_1 rotamers as a function of dihedral angle determined using Ringer, for three humidities above the unit cell transition at room temperature. The assumed 0.3 σ noise level for each map is indicated by the dashed solid line. (a) VAL29 maintains its conformation during dehydration. (b) An alternative conformation of ASP18 in the native structure (99% r.h.) is systematically suppressed with decreasing humidity. (c) An alternate conformation of ASN77 is enhanced with decreasing humidity. (d) ARG125 develops a new conformation at lower humidity.

of residues involved. Previous analysis using Ringer has shown that cryocooling induces changes in electron density sampled in rings around side chain dihedral angles χ and these have been interpreted as evidence for conformational remodeling of side chain rotamers. As shown in Figure 3.4 for r.h. between 99% and 85%, dehydration induces similar changes, including suppression of minor (low-occupancy) conformers (Figure 3.4(b)), enhancement of minor conformers (Figure 3.4(c)), and appearance of

	Suppressed	Enhanced	New conformer
99 → 93% r.h.	11	2	4
99 → 85% r.h.	14	6	3
99 → 75% r.h.	14	10	5
99% r.h. → T = 100 K	19	3	3
85% r.h. → T = 100 K	12	6	2
75% r.h. → T = 100 K	8	7	3
58% r.h. → T = 100 K	11	10	3
11% r.h. → T = 100 K	11	9	12

Table 3.4: Modifications of the 103 lysozyme residues having χ_1 rotamers due to dehydration and cryocooling, relative to the native (99% r.h.) room temperature structure. Modifications were determined by visual inspection of Ringer plots as in Figure 3.4. Minor conformers above the 0.3 σ electron density threshold were deemed suppressed if their density fell below threshold or decreased substantially relative to the major conformer. A minor conformer was deemed enhanced if it became detectable above threshold or if its above-threshold density increased while that of the major conformer decreased. A new conformation was identified if there was a major change in position of the major conformer or of a high density minor conformer. This qualitative assessment more readily identifies changes in minor conformations that does the Pearson correlation coefficient of Figure 3.6, as the latter weights large electron density features more heavily.

new conformations (Figure 3.4(d)). Lysozyme has 129 residues, of which 103 have side chain rotamers with a dihedral χ_1 angle and rotamer. As shown in Table 3.4, on dehydration from 99% to 85% r.h., 12 residues showed suppression of a minor conformer, 7 showed systematic enhancement of a minor conformer, 3 residues adopted new conformers, and the remaining 81 residues showed no change. Cryocooling of the native (99% r.h.) crystal causes suppression of more minor conformers, enhancement of fewer minor conformers, and generation of a comparable number of new conformations. As relative humidity is decreased, cryocooling still causes substantial side chain remodeling, even though the change in cell volume on cooling becomes smaller and smaller.

No correlation is observed between nitrate ion position and occupancy and side chain remodeling. The number of modeled nitrates per lysozyme molecule varied with r.h., at T = 300 K increasing from 2.5 (5 in the ASU) at 99% r.h. to 4 at 85% r.h., dropping

to 3 at 75% and 0 at 11% r.h. This variation largely follows the variation in diffraction resolution, and so does not necessarily reflect variations in actual nitrate numbers or their ordering.

Contact versus non-contact residues

To understand how a side chain's local environment affects its remodeling during dehydration or cooling, residues were first categorized as contact or non-contact, based upon their final (dehydrated and/or cooled) side chain state. To assess whether the χ_1 side chain rotamer was remodeled, the Pearson correlation coefficient between Ringer plots for the initial and final states was calculated for each residue [22]. Pearson correlation coefficients can range from -1 to 1, and are independent of the absolute value of the electron density. They are only weakly dependent on B-factors and resolution, and blurring or sharpening of the electron density has only a small effect. Correlation coefficients below 0.85 but well above 0 generally indicate remodeling of minority conformations, whereas negative values usually correspond to changes in majority conformation. Sidechains with correlation coefficients below 0.85, a conservative cut-off, were thus considered remodeled.

Figure 3.5(a) gives the number of altered contact and non-contact residues, relative to the native (99% r.h., 298 K) structure, versus relative humidity at 298 K and 100 K. At room temperature, the majority of altered residues are involved in crystal contacts, although the fraction of altered noncontact residues grows substantially below 58% r.h. Cooling the native crystal to 100 K alters roughly the same number of residues as does dehydrating to 93% r.h., although cooling affects more non-contact residues.

Solvent-exposed and buried residues

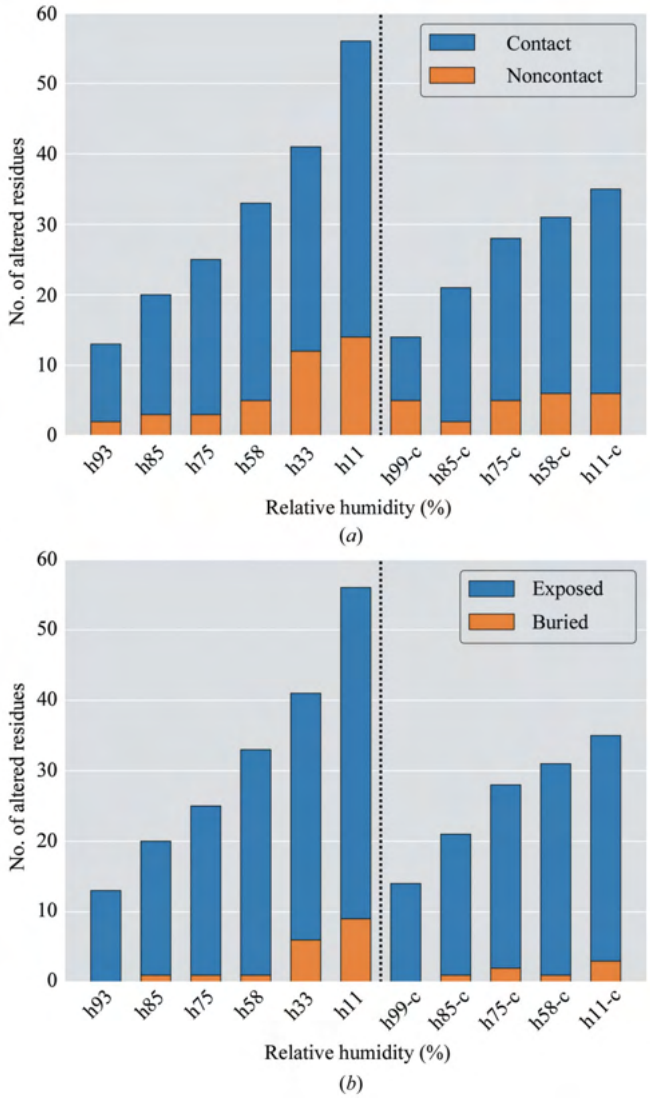

Figure 3.5: Number of altered residues, of the 103 lysozyme residues having χ_1 rotamers, due to dehydration and cryocooling, relative to the native (99% r.h.) structure, categorized according to (a) whether or the not the residues were involved in crystal contacts in the final, non-native structure; and (b) whether the residues were solvent exposed or buried. Here, side chains were deemed to be altered if the Pearson correlation coefficient between Ringer curves for the native and non-native structures was less than 0.85. A large majority of the altered side chains are involved in crystal contacts and/or are solvent exposed in non-native room temperature and 100 K structures, although non-contact and buried residues are substantially perturbed at r.h. of 33% and 11%. Dehydrating the native crystal to 93% r.h. alters a comparable number of residues as does cryocooling to 100 K.

Solvent accessibility might also be expected to influence which residues are altered. Solvent accessibility was assessed by calculating the solvent accessible surface area (SASA) using the PyMOL function "get_area" with a default probe size of 1.4 Å, and then normalizing by the maximum surface area of each residue type to obtain the fractional SASA. The maximum surface area of each residue type was estimated using an isolated tripeptide of Gly-X-Gly in which X is the residue of interest. In previous work, fractional SASA values below 5% and below 20% were most commonly used to define the cut-off for buried residues [113–115]. A survey of 37 different protein structures found that 15% of all residues in small proteins and 32% in large proteins were buried assuming a 5% cut-off [114]. Lysozyme is a small protein and Appendix Figure B.3 plots the number of buried residues versus fractional SASA cut-off. The number of buried residues versus cut-off exhibits a change in slope near 3%. With a 3% cutoff, buried residues comprise 22–24% of all residues for all humidities, consistent with the survey. Residues with fractional SASA values smaller than 3% were thus defined as buried.

Figure 3.5(b) gives the number of altered residues that are solvent accessible and buried in the final state. Dehydration to r.h. values between 93 and 58%, as well as cryocooling of those dehydrated samples, has almost no effect on buried residues.

Overall side-chain correlations

Figure 3.6(a) shows overall pairwise Pearson correlation coefficients for sidechain rotamers, which provide an overall measure of similarity of side chain conformations. These were obtained by averaging pairwise side chain correlations calculated from Ringer plots over the 103 residues having sidechain rotamers, between 298 and 100 K structures at different relative humidities. As one would expect, correlations decrease

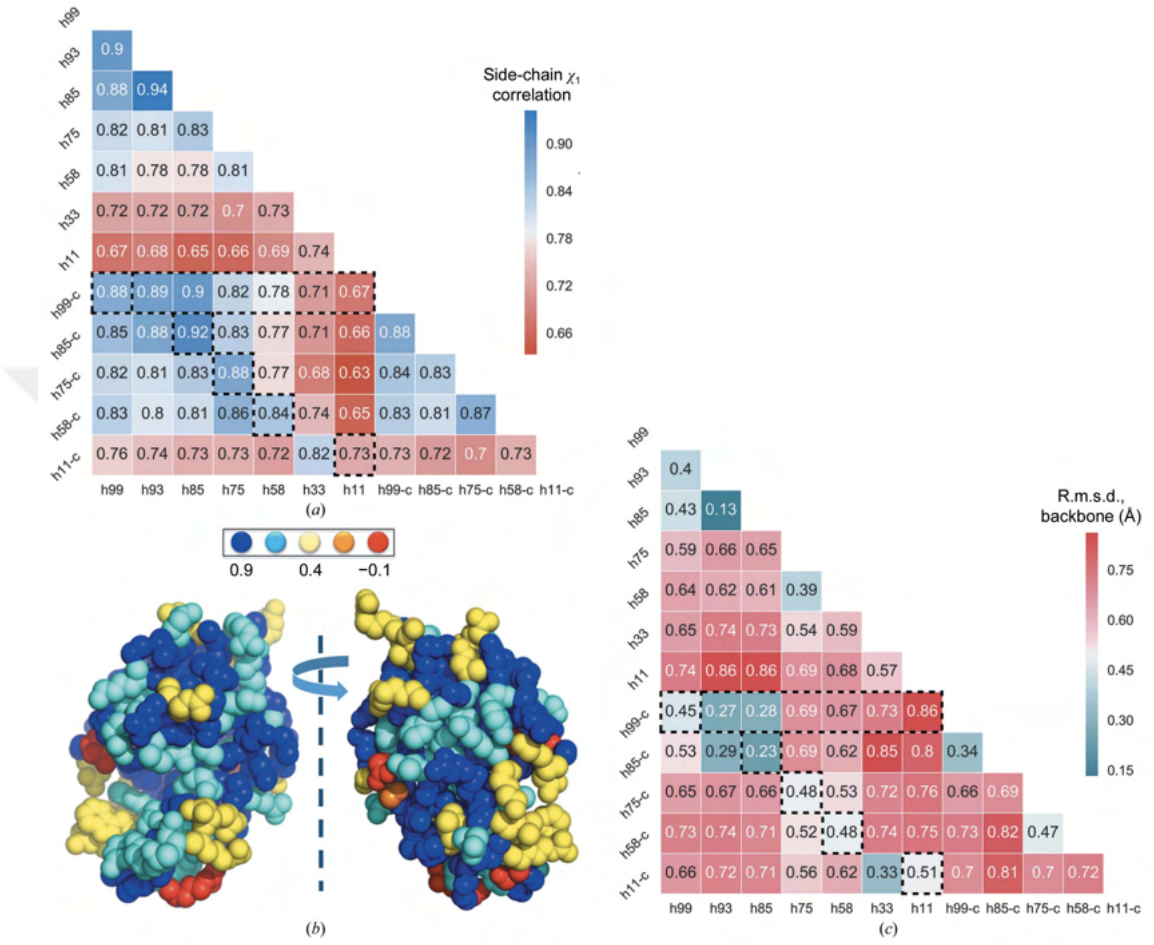

Figure 3.6: (a) Pairwise Pearson correlation coefficients for Ringer plots at χ_1 as in Figure 3.4, averaged over all 103 residues, for all possible pairs of dehydrated and cryo-cooled structures. Data sets collected at 100 K are labelled with "-c". The leftmost column indicates correlations with the native structure at room temperature; the diagonal boxed entries indicate correlations between room temperature and 100 K structures for crystals dehydrated to the same r.h., and the horizontal rectangle highlights correlation between the 100 K native structure and dehydrated structures at room temperature. (b) Lysozyme structure color coded according to the Pearson correlation coefficient between the native and 75% r.h. room temperature structures, calculated using each residue's χ_1 Ringer plot. All residues with lower correlation coefficients are located on the surface while buried residues generally have high correlation coefficients. The active site and its vicinity exhibit intermediate correlation. (c) Backbone RMSD between all possible pairs of dehydrated and cryo-cooled structures. Only backbone atoms were used for alignment of the two structures and calculation of RMSD.

as the difference in relative humidity increases. As seen in the leftmost column of Figure 3.6(a), cooling the native structure ("h99") to 100 K ("h99c") gives the same correlation coefficient as dehydrating the native structure at 298 K to 85% r.h. ("h85"). As seen by inspecting the diagonal boxed entries in Figure 3.6(a), for crystals dehydrated to the same r.h., the largest correlation between room T and 100 K structures occurs for 85% r.h, which is also the humidity that yields the highest resolution 100 K data set. Similar patterns are evident in Appendix Figure B.4 for χ_2 which is calculated by averaging over 72 residues having χ_2 rotamers. Similarly, as indicated by the entries within the horizontal rectangles in Figure 3.6(a) and Appendix Figure B.4, the χ_1 rotamers of the native (99% r.h.) structure at 100 K are most strongly correlated with those in the 85% r.h. room temperature structure, and for χ_2 rotamers with the 93% r.h. room temperature structure. This demonstrates the structural similarity between the effect of cryocooling and dehydration to 85−93% r.h. of the native crystal in terms of the side chain conformations and heterogeneity. Figure 3.6(b) shows the room temperature molecular structure in the native state color coded according to each residue's χ_1 correlation with the 75% r.h. structure.

Backbone r.m.s.d.

Lysozyme does not exhibit major structural rearrangement of its backbone on dehydration or cooling, but smaller changes are evident, especially at turn regions. Figure 3.6(c) shows the pairwise backbone root-mean-squared deviation (RMSD) between all dehydrated and cooled structures. RMSDs increase with increasing humidity difference. As with sidechain correlations, cryocooling the native structure produces overall changes in backbone that are comparable to those obtained by dehydration to 85% r.h. at room temperature. Surprisingly, 298 K backbones at 58, 33, and 11% r.h. have smaller rmsd

values with the native structure than with 93%–75% r.h. structures. A similar although less strong trend is seen in Figure 3.6(a) for side chain correlations. As is the case with sidechain correlations discussed above, the smallest backbone perturbation caused by cryocooling at a fixed r.h. is observed for 85% r.h.

3.3.5 Multiconformer modeling and side-chain order parameters

Overall side chain conformational heterogeneity can be characterized using bond order parameters S^2, which include contributions from both harmonic and non-harmonic motions as well as from static disorder. The order parameter S^2 corresponding to a dihedral angle χ_i can be decomposed into two components: S^2_{ortho} , associated with displacements around a conformer at a given χ_i value, and S^2_{angular} , associated with displacements between conformers at different χ_i values. At each r.h., qFit, Ringer, and manual building were used to construct a multiconformer model. Order parameters S^2 were then calculated for χ_1 rotamers about bond vectors $C_\beta - X_\gamma$ where X is C, O, or S for most residues, about $C_\alpha - C_\beta$ for Ala, and about $C_\alpha - H_{\alpha 2}$ for Gly. Appendix Figure B.5 shows sample plots of order parameters S^2 and its component S^2_{ortho} versus relative humidity r.h. for several residues near lysozyme's active site, for r.h. between 99% and 75% for which data set resolution was high and nearly constant. Truncation of all data sets to 1.4 Å resolution had little effect on these results. Of the 103 residues with side chain rotamers, 76 showed an increase in S^2 with decreasing r.h., and nearly all showed an increase in S^2_{ortho}.

3.3.6 Modulation of the active site

The active site of lysozyme, which in the native state is largely hydrated by ordered water molecules, undergoes significant remodeling on dehydration. The most pronounced change occurs at TRP62, which undergoes motion in substrate binding [116, 117]. TRP62 is somewhat disordered in the native crystal, and on dehydration to 85% r.h. it loses one of its minor conformers while maintaining its major conformation. At r.h. values below that of the crystal lattice transition, TRP62 undergoes a large conformational change that has been suggested to be similar to its behavior during substrate binding [80, 118].

The present data show that this conformation change results from a steric clash with

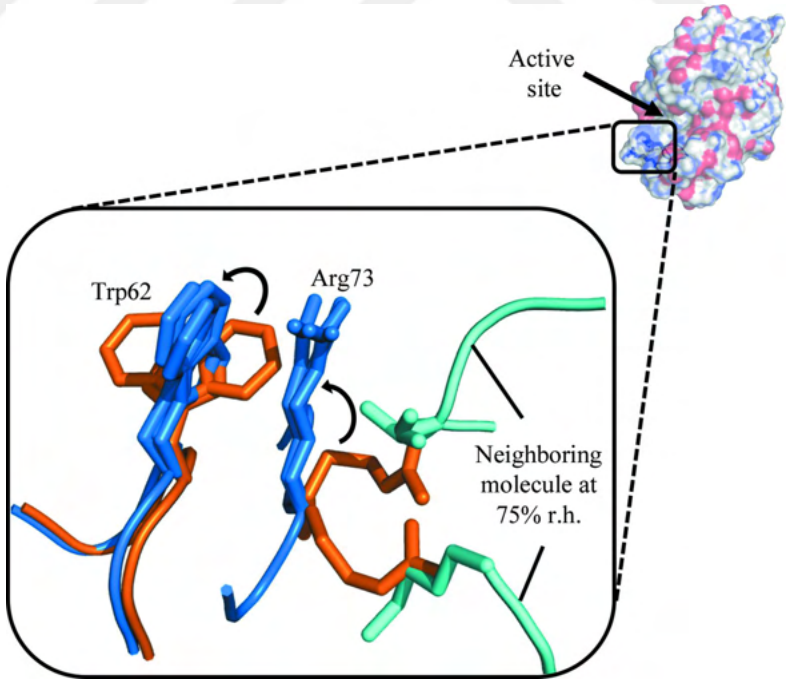

Figure 3.7: The important active site residue TRP62 undergoes a significant conformational change between the native (orange) and 75% r.h. (blue) structures. In the native crystal, ARG73 freely extends away from the active site. Dehydration brings the protein molecules closer together. In the 75% r.h. structure, steric clash with the sidechains of an adjacent molecule (cyan) changes the conformation of ARG73, which causes a 90° rotation of TRP62.

a symmetry mate. Figure 3.7 shows the two side chains TRP62 and ARG73 at the active site of native (orange) and of 75% r.h. (blue) structures, as well as nearby sidechains of a symmetry mate at 75% r.h. (cyan). In the native crystal, ARG73 has one alternate conformation and both conformations freely extend away from the center of the reference protein molecule. On dehydration from 85% r.h. to 75% r.h., significant loss of solvent associated with the lattice transition leads to tighter packing of protein molecules that forces ARG73 to change its conformation. This conformation chain causes a ~90° rotation in TRP62's sidechain to prevent a steric clash. The low humidity conformation, which stacks ARG73 onto the TRP62 ring, is also stabilized by a favorable cation-π interaction.

3.3.7 Dehydration- and cryocooling-induced changes in tetragonal thaumatin crystals

A more limited investigation of the effects of dehydration and cryocooling was conducted using crystals of tetragonal thaumatin. These crystals were grown in hanging drops comprised of equal volumes of 40 mg/ml protein in 0.1 M sodium acetate buffer at pH 6.5 and a reservoir solution of 0.5 M potassium sodium tartrate in the same buffer. The buffer gave an r.h. for the native crystal of 99%, and dehydration using saturated solutions of potassium nitrate, zinc sulfate, and sodium chloride gave r.h. values of 93%, 89% and 75%, respectively. Complete and redundant X-ray data sets were collected from these crystals and from cryoprotectant-free native crystals cooled to 100 K, and these data were modeled and refined as for lysozyme. Table C.2 gives the resulting refinement statistics.

Thaumatin has 206 residues compared with lysozyme's 129, and its tetragonal crys-

tals have a Matthews coefficient-derived solvent content of 58.5%, compared with monoclinic lysozyme's 37.5%. As shown in Appendix Figure B.6, the unit cell contracts by about 18% between 99% and 75% r.h., and by about 4% on cooling to 100 K, compared with contractions of ~20% and 7% for monoclinic lysozyme. For both proteins, nearly all contraction is due to a reduction in solvent volume, and this contraction is associated with an increase in the fraction of residues involved in crystal contacts (Appendix Figure B.7).

As shown in Appendix Figure B.8, the absolute number and fraction of residues whose rotamers are altered by dehydration and by cooling are much smaller for tetragonal thaumatin than for monoclinic lysozyme; at 75% r.h., the altered fractions are ~0.07 and ~0.19, respectively. The fraction of altered residues that are involved in crystal contacts – 0.6 vs 0.88 at 75% r.h. – is much smaller for tetragonal thaumatin. As shown in Appendix Figure B.9, the pairwise Pearson correlation coefficients for Ringer plots at , averaged over all residues with side chain rotamers, are larger for all r.h. pairs and all room-cryo temperature pairs for thaumatin than for lysozyme. Moreover, an inspection of Ringer plots and correlation coefficients for residues flagged as altered shows that side chain conformational changes are on average smaller for thaumatin than for lysozyme. These results are consistent with the much larger solvent content (0.49 v/v vs. 0.22 v/v at 75% r.h.) and less intimate contact between molecules in thaumatin.

Appendix Figures B.8 and B.9 show that cryocooling the native state of thaumatin to 100 K produces a comparable increase in the number of contact residues and a comparable fraction of altered residues as does dehydration to 89% r.h. Furthermore, Appendix Figure B.8(a) shows that the ratio of non-contact to contact residues altered by cooling to 100 K and by dehydration to 89% r.h. are similar. This suggests a connection between conformational perturbations caused by cooling and dehydration.

3.4 Discussion

3.4.1 Unit-cell, protein and solvent volumes

The use of carefully prepared, cryoprotectant-free crystals with identical solvent compositions for room temperature and 100 K measurements allows quantitative comparison of unit cell, protein, and solvent volumes. Both dehydration from 99% r.h. to as low as 11% r.h. and cooling from 298 to 100 K lead to only small (1% and 2%, respectively) changes in protein volume, as determined from the calculated volume of the SES. On the other hand, the calculated solvent volume at 99% r.h. decreases on cooling to 100 K by ~14% (from the SES volume) or ~19% (from the Matthews coefficient analysis). At 11% r.h., the calculated solvent volume contractions are much smaller (~0% from the SES or ~9% from the Matthews coefficient) and have much larger uncertainties (~5–7% of solvent volume) because solvent volumes are much smaller. SES protein volumes are calculated assuming a single conformer with no thermal motions, and so do not include effects of alternate conformers or finite atomic B-factors. To a first approximation (assuming fixed, temperature-dependent bond lengths and no overlap of non-bonded atomic VDW spheres), protein motions should not affect the instantaneous and thus time-averaged protein volume. As an upper bound on the possible effects of conformational heterogeneity, SES volumes were calculated assuming all side chain conformers were populated with unit occupancy. This increased 298 K SES volumes by ~5% at 99% r.h. and ~1% at 11% r.h.

3.4.2 Comparison of dehydration and cooling

Previous work has explored how cryocooling modifies unit cell and protein volumes [14], overall protein structure [25, 59], and a protein's ensemble of side chain conformations [1, 3, 20, 22]. The present work shows that the effects of modest dehydration on these properties are similar, both superficially and in many details. Cooling native (99% r.h.) monoclinic lysozyme crystals to 100 K has similar effects on unit cell and protein volume, on the number of crystal contacts, on backbone structure (as measured by RMSD), and on side chain structure (as measured by χ_1 and χ_2 correlation coefficients) as does dehydration to 85% r.h. Cooling native crystals to 100 K modifies a comparable fraction of side chain conformations at χ_1 as dehydration to 93%; both overwhelmingly affect solvent-exposed residues, while cooling may have slightly more effect on non-contact residues. Cooling and dehydration both cause a similar variety of side chain modifications, including suppression of minor conformers, enhancement of minor conformers, and appearance of new conformations. The same general trends are observed in tetragonal thaumatin crystals, although side chain modifications at a given r.h. are less pronounced due to the much large solvent content and reduced protein packing.

This similarity complicates interpretation of differences between structures and especially side chain conformational ensembles determined at room and cryogenic temperatures [1], especially when the detailed crystallization, harvesting, mounting, and cryocooling histories of each crystal are either not known or are not identical. Protein crystals regularly undergo modest dehydration during harvesting, mounting, storage, and data collection at room temperature. Anecdotal evidence, such as substantial variations in deposited unit cell volumes between similarly prepared crystals, suggests that dehydration was particularly frequent when room-temperature data collection involved drawing crystals into glass capillaries, wicking away excess liquid, and carefully seal-

ing both ends of the capillary with wax. Dehydration has likely been a primary cause of room-temperature crystal non-isomorphism; crystals that are reproducibly hydrated (by insertion in humidity controlled gas streams or by equilibration with saturated salt solutions) show highly reproducible (to within 0.1%) cell parameters [94]. Moreover, dehydration to 93% r.h. and somewhat lower typically has no visible effects on crystals (e.g., they don't crack as long as dehydration occurs gradually) and usually does not cause major degradation of diffraction, and so can be hard to detect. Consequently, careful control of crystal hydration for both room and cryogenic temperature data collection is essential if the effects of these two perturbations are to be deconvolved.

3.4.3 Why do cooling and dehydration generate similar structural perturbations?

Temperature affects chemical properties such as pH, the pKas of ionizable groups, and their charges, and "mechanical" properties such as the amplitude of thermal vibrations about each conformation, the occupancy of alternative conformations, and the rate of interconversion between them. As temperature decreases, average bond lengths shrink, higher energy conformations are depopulated, and the protein becomes more compact. More flexible regions at the protein surface that tend to be involved in crystal contacts contract more than the protein as a whole, decreasing the volume of the unit cell more than the protein volume, and increasing the number and area of contacts between molecules. Both protein compaction and increased protein-protein contacts stiffen the protein as whole, which further reduces the amplitude of thermal motions. These also introduce steric hindrances that can block minor or major conformations or lead to new conformations.

Dehydration affects chemical properties including the concentrations of solutes within the solvent channels and pH. By removing solvent, dehydration increases the number and area of contacts between molecules. These stiffen the protein as whole, reducing the amplitude of thermal motions at fixed temperature. They introduce steric hindrances that can block minor or major conformations or lead to new conformations. As solvent is depleted from the first hydration layer, protein function ceases, in part due to a loss of conformational flexibility facilitated by mobile surface waters.

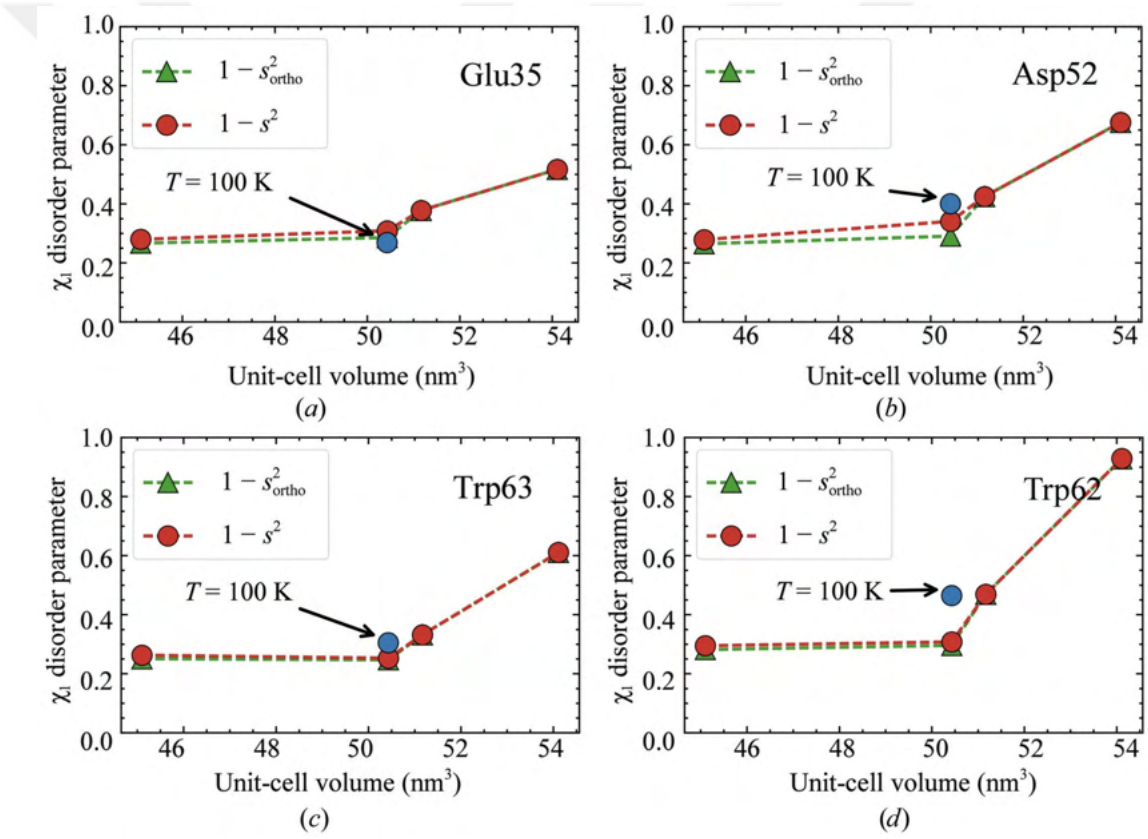

Figure 3.8: Disorder parameters $1 - S^2$ and $1 - S^2_{\text{ortho}}$ versus unit cell volume, for a selection of residues at χ_1, as deduced using multiconformer refinement from room temperature data sets collected at r.h. of 99%, 93%, 85%, and 75%. The same data, plotted as S^2 versus r.h., is shown in Appendix Figure B.5.

The strong similarity in the protein's conformational responses to cooling and to modest dehydration suggests that, at least for monoclinic lysozyme crystals, the global character of these responses is dominated by the effects of increased molecular packing.

If this is the case, then one might expect that plots of harmonic and overall conformational heterogeneity versus unit cell size would be similar, regardless of whether the cell changes were caused by temperature or dehydration. Figure 3.8 shows example plots of $1-S^2$ and $1-S^2_{\mathrm{ortho}}$ for χ_1 rotamers at $T = 298$ K versus unit cell volume, for r.h. values between 99% and 75%, using the same data as in Appendix Figure B.5. The right-most data point corresponds to the native 99% r.h. crystal at 298 K. For comparison, $1 - S^2$ and the unit cell volume of the native crystal at 100 K is indicated. Plots for the majority of side chains have a "hockey stick" shape, with the rate of increase of disorder with cell volume being larger at larger cell volumes. This is similar to the behavior observed when side chain order parameters of cyclophilin A were plotted versus temperature between 300 and 100 K [22]. In both proteins, the majority of χ_1 rotamers have a single dominant conformation that does not change, so S^2 and S^2_{ortho} are similar. Some do show conformational rearrangements that affect S^2_{angular} and S^2, but S^2_{ortho} still generally increases with decreasing unit cell. These results are consistent with the notion that cell volume and protein packing dominates global aspects of the effects of both cooling and dehydration on protein conformation.

As previously established (e.g., [6]), the fact that unit cell contraction may drive many conformational changes observed on cooling or dehydration in crystallography does not imply that these changes are biologically irrelevant. Crystal contacts can mimic some effects of ligand binding at active sites or of regulator binding at remote sites, and chains of steric hindrance triggered by increased contact may illuminate allosteric networks. Since crystal dehydration can generate much larger changes in cell volume and in area of intermolecular contact than can cooling, it can be a highly useful probe of conformational interactions within proteins when adequate diffraction resolution is preserved.

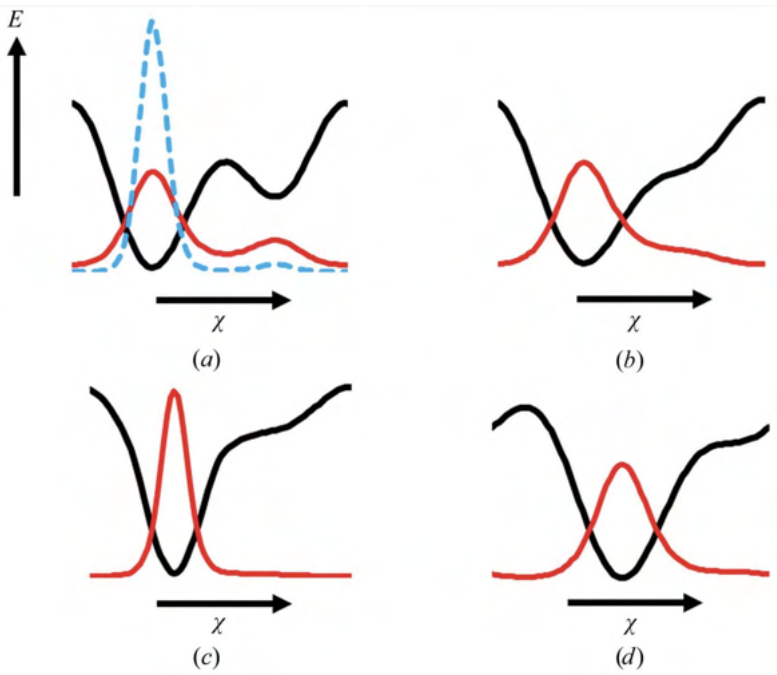

Figure 3.9: The effects of dehydration and cooling on side chain energy landscapes, shown as plots of free energy versus side chain rotamer angle χ (black lines). Red lines indicate occupancy (electron density) at room temperature. The dashed blue line in (a) indicates occupancy at 100 K. (a) A potential with two wells. (b) A steric clash eliminates one well. (c) Stiffening of the potential of (b) due to, e.g., increased packing density and steric encroachment. (d) A potential with a well at a new position relative to (a)-(c). Dehydration from state (a) at fixed temperature may lead to (b) elimination of a minor conformer, increasing S^2_{angular}, (c) reduced amplitude of thermal motion about the major conformation, increasing S^2_{ortho}, and (d) appearance of a new major conformation. Cooling at fixed hydration can lead to similar changes in energy landscape associated with increased packing, and in addition to depopulation of the minor conformer in (a) and reduced amplitudes of thermal motion in (a)-(d), without affecting the underlying energy landscape.

Figure 3.9 shows how the effects of dehydration and cooling can be understood and distinguished in terms of energy landscapes and occupancies. Both cause increased packing density, molecular stiffening, and steric clashes that translate into changes in the local free-energy landscape and into changes in occupancies within that landscape. However, there are no obvious hydration-related analogs to the Boltzmann factor, or to thermal depopulation of minor conformers and reduced amplitudes of thermal motions within fixed energy landscapes. In general, distinguishing purely thermal effects from

those associated with energy landscape changes, and distinguishing landscape changes intrinsic to the protein from those associated with the crystalline environment, require careful analysis of each residue's environment.

3.5 Conclusions

Improved computational and modelling tools are facilitating study of protein conformational ensembles using crystallography, and the analysis of these ensembles to identify functionally important motions and interactions. The electron density associated with minor conformations is generally weak and so can be difficult to positively identify, especially at resolutions near and above 2 Å typical of many targets of biological interest. Dehydration and cooling (as well as hydrostatic pressure [119]) provide easily-controlled continuous crystal and molecular perturbations, and the resulting systematic changes in electron density should in many cases allow "real" weak density features to be distinguished from artefacts.

The present results, obtained using crystals in carefully prepared reference states and subject to well-defined perturbations, allow detailed and quantitative comparison of the local and global effects of dehydration and cryocooling on protein structure and conformational heterogeneity. In low-solvent-content, relatively tightly packed monoclinic lysozyme crystals, the effects of dehydration and cooling are strongly similar, suggesting that changes in inter- and intramolecular packing interactions associated with unit cell contraction dominate the observed behavior. Only modest dehydration, as might be expected to occur in past crystallographic practice (e.g., when using glass capillaries) or current practice (e.g., when using small crystals), is required to mimic many of the effects of cooling from room temperature to 100 K. Dehydration combined with differ-

ences / uncertainties in cryoprotectant concentrations thus complicates interpretation of differences between room temperature and 100 K structures.

Weak density features associated with minor conformers are more likely to be perturbed by functionally interesting interactions, but they are also more likely to be disrupted by changes in crystal packing. Here, temperature- rather than hydration-dependent crystallography has a clear advantage, in that dehydration effects are dominated by crystal packing interactions while temperature also directly affects occupancies and motion amplitudes. High solvent content, weakly packed crystals are ideal targets for temperature-dependent structural studies, assuming crystals with adequate order can be obtained and cooled without ice formation. When these are not available, comparison of cryocooled and dehydrated structures of cryoprotectant-free crystals having comparable unit cell contractions relative to the native, room-temperature state may highlight aspects of the protein's conformational landscape that are not specific to the crystalline environment.

CHAPTER 4

RESOLUTION AND DOSE DEPENDENCE OF RADIATION DAMAGE IN BIOMOLECULAR SYSTEMS

This chapter is identical to the published manuscript (doi: 10.1107/S2052252519008777) with contributions of Lauren Conger, David W. Moreau under the supervision of Robert E. Thorne. Robert E. Thorne conceived of the presented idea. Hakan Atakisi and Lauren Conger developed the theoretical formalism, performed the analytic calculations and performed the numerical simulations. Robert E. Thorne and Hakan Atakisi wrote the manuscript. All authors discussed the results and contributed to the final manuscript. Chris Myers gave useful advice on the numerical simulations.

4.1 Introduction

Radiation damage is a key issue in all diffraction and imaging methods that illuminate biological samples with energetic particles such as x-ray photons, electrons, neutrons, and positrons. Absorption and inelastic scattering processes transfer energy to the sample (quantified as dose D, the energy deposited per unit sample mass, in gray=J/kg), with initial energy deposition from each scattering event confined to a small volume that depends on the energy transfer [10, 120, 121]. Energetic electrons and reactive atomic and molecular species are generated and diffuse and react, causing additional chemical, bond-scale damage [122, 123]. Accumulation of bond-scale damage causes degradation of sample order on larger and larger length scales [124]. In the absence of macroscopically disruptive damage processes such as eruption of hydrogen bubbles [125, 126] (e.g., at very large dose rates [127]) or fracturing, this accumulation of damage manifests in imaging as a loss of image contrast or 'blurring', evident first in short length scale fea-

tures and then moving to larger and larger scales. In diffraction, damage manifests as a decrease in diffracted intensity at large angles 2θ or large diffraction wavevectors q that progresses to smaller and smaller q. In addition to these 'global' effects of radiation damage, metal sites, disulphide bonds and other structures within the sample may be particularly sensitive to damage, giving rise to 'site-specific' damage [128, 129] [130, 131].

In biomolecular crystallography, radiation damage limits the amount of diffraction data that can be collected per unit sample volume and introduces errors in experimental structure factors. The effects of radiation damage on individual Bragg peak intensities can be complex [127]. Aside from global q-dependent intensity decay and differential decays associated with site specific damage, measured Bragg peak intensities depend on radiation damage-induced broadening of crystal mosaicity and strain distributions, which can cause partially illuminated reflections to initially brighten with dose [127]. They also depend on the spatial pattern of sample irradiation during data collection [132], which can lead to heterogeneous sample dose states within the illuminated volume and to measured diffraction intensities that reflect a complex convolution of spatially (and, at high dose rates, temporally) nonuniform damage effects [127]. Current crystallographic processing software [133, 134] [135, 136] incompletely accounts for these effects. Refined structural models based on data with typical maximum resolutions of ~2 Å are at least somewhat insensitive to these experimental errors, as models obtained from nominally zero-dose XFEL data sets show good general agreement with those from finite-dose synchrotron data sets (e.g., [22, 137]).

The most commonly used metric for characterizing the radiation sensitivity of biomolecular crystals is the half-dose $D_{1/2}$, equal to the dose at which the integrated intensity within all observed diffraction peaks drops to one-half its zero-dose value. Typical reported half-dose values for protein crystals are ~10–30 MGy at 100 K and

100–400 kGy at room temperature [11]. Half-doses depend on the initial diffraction resolution of the crystal [138], which depends on crystal quality and size, x-ray beam size, and amount of background scatter from air and surrounding liquid. Experimental half-doses also depend on the spatial distribution of dose within the crystal, which depends on the beam intensity profile, the crystal's size and shape, and on whether the crystal is fixed or rotated during data collection [127, 132] and on how dose is defined when irradiation is spatially nonuniform [127]. Furthermore, the way in which the half-dose weights the contribution of diffraction intensities at different q values is not obviously related to their information content or to their utility in defining and constraining the final structural model. Consequently, half-doses are at best a crude metric of radiation sensitivity and provide at best a rough rule-of-thumb limit in crystallographic data collection. Other metrics of radiation sensitivity such as the change in scaling B-factor [11, 139], scale [11] and decay R-factor [136] are used, but have related limitations.

Radiation damage has been more fully characterized using the q dependent integrated intensity $I(q)$, obtained by integrating sample diffraction within a wavevector shell of width Δq, with values of diffraction resolution $d = \lambda/2\sin(\theta) \propto 1/q$ rather than q typically quoted [140–142]. Plots of experimental $I(q)$ for a given q (or resolution) shell versus time, time-integrated flux density (in ph/cm^2) or nominal dose typically show an initial linear or exponential decay and then a more gradual decay at larger times/doses, giving an overall 'hockey-stick' shape when plotted on semilog axes (e.g., as in [127, 143]). The rate of the initial intensity decay with dose increases with increasing q (increasing resolution, corresponding to decreasing numerical value of d). Kinetics-inspired models [143–145] have been used for nearly 50 years to obtain good fits to measured $I(q)$ vs exposure time / nominal dose, reproducing both the overall "hockey stick" shape at fixed q and its q resolution dependence. However, the physical significance of the models and the fit parameters obtained have been unclear.

Recent experiments and analyses have emphasized the profound effects of nonuniform crystal irradiation during data collection on measured integrated intensity-dose curves [127]. For thaumatin and lysozyme crystals held in fixed orientation and illuminated using an x-ray beam with a Gaussian intensity profile, the measured integrated (over all q) intensity versus time, time integrated flux density, or nominal dose has the "hockey stick" shape. Simulations show that the Gaussian illumination profile generates this shape even if the underlying relation between diffracted intensity and dose is strictly exponential. Since previous radiation damage experiments have seldom if ever provided perfectly uniform crystal irradiation (even when the X-ray beams had nominally 'top-hat' profiles), interpretation of their intensity-dose curves using kinetics-inspired models is now suspect.

Our goal here is to determine the underlying 'local' Fourier-space relationship between damage and dose, $\tilde{I}(q, D)$, as it would be measured from a uniformly illuminated crystal. This relationship plays an analogous role in radiation damage studies to the local relation between conductivity and electric field, $\sigma(E)$, of conducting materials: it is the key to detailed understanding of damage and its mechanisms. Beginning with previously published experimental data at $T = 100$ K, we account for spatially nonuniform illumination during data collection and use this to estimate $\tilde{I}(q, D)$. We then consider a simple model for radiation damage involving random, local disordering interactions. This model predicts a purely exponential $\tilde{I}(q, D)$ relation, and a decay constant that scales with q in a manner roughly consistent with experiment. Fits of previous models [143–145] to experimental intensity vs resolution and dose data are invalidated as the primary trends have a physical origin different than was assumed. Experimental deviations from the present model's predictions should illuminate how damage mechanisms evolve between cryogenic and room/biological temperatures.

4.2 Materials and methods

4.2.1 Modeling and fitting experimental data for intensity versus dose and resolution

In an ideal experiment to measure the local $\tilde{I}(q, D)$ relation in a bulk crystal [where photoelectron escape from the crystal [10, 120, 146] can be neglected], the entire crystal volume is illuminated with a fixed and uniform flux-density X-ray beam, and the crystal thickness along the incident beam direction is small compared with the X-ray absorption length. Under these illumination conditions, every crystal region receives the same dose D regardless of whether the crystal is rotated or held in a fixed orientation.

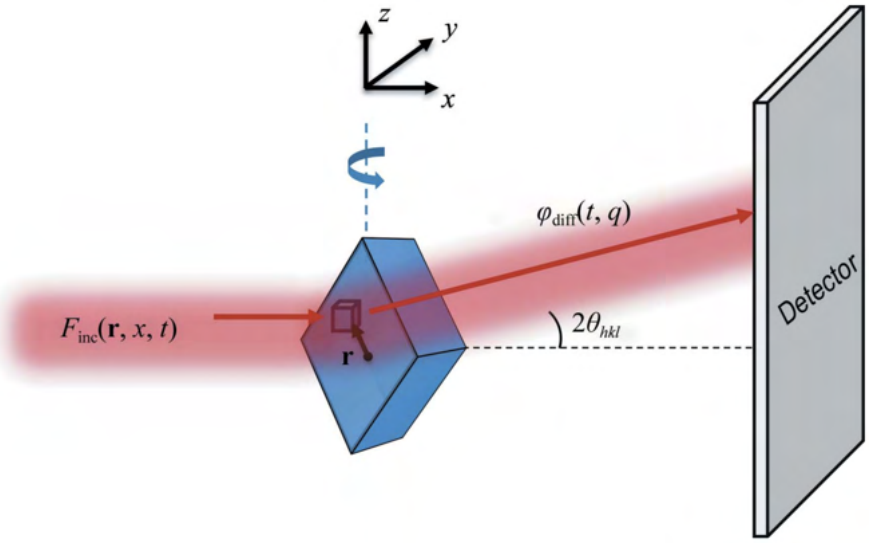

Figure 4.1: Experiment and simulation setup. A crystal (simulated as a rectangular prism) is located at $x, y, z = 0$ and a volume element (voxel) of the crystal is located at position $\mathbf{r}$. The crystal is illuminated by an X-ray beam with a top-hat or Gaussian profile, and the incident photon flux at the voxel position is $F_{\text{inc}}(\mathbf{r}, x, t)$. Crystal planes with spacing d_{hkl} generate Bragg scattering at angle $2\theta_{hkl}$, corresponding to a scattering wavevector $q_{hkl} = 2\pi/d_{hkl}$. The scattered flux (in photons s^{-1}) from each voxel is $\varphi_{\text{diff}}(t, q)$.

In actual experiments (Fig. 4.1), the X-ray beam may be smaller than the crystal, the

incident flux density within the beam may be nonuniform (even when nominally flat or 'top-hat' beams are used) and the crystal may be rotated during data collection. These result in different regions of the crystal experiencing different (and time-dependent) incident photon flux densities and dose rates, and accumulating different total doses $D(x, y, z, t)$ [132, 136]. This leads to spatially nonuniform damage. The measured diffracted intensity at any time is determined by both the incident X-ray flux-density distribution and by the distribution of damage states within the X-ray-illuminated sample volume [127, 136] [which can be characterized using an incident flux-density-weighted dose [147] or, more meaningfully, by a diffraction-weighted dose [127]]. Because the local relation between diffracted intensity and dose is in general nonlinear, knowing only the measured $I(q)$ versus exposure time or nominal fluence (in photons mm^{-2}) or nominal dose and the spatiotemporal pattern of crystal irradiation during data collection is not sufficient to uniquely determine the local $\tilde{I}(q, D)$ relation.

Thus, to analyze previous experimental $I(q)$ data, we (i) calculate the spatial distribution of sample irradiation and dose from given experimental details, (ii) define a (somewhat) general expression for the local $\tilde{I}(q, D)$ relation, with adjustable parameters, (iii) calculate the diffracted intensity $I(q)$ versus nominal fluence or dose using this relation and the calculated dose distribution within the sample, and (iv) refine the parameters to optimize the quality of the fit to the $I(q)$ data.

4.2.2 Calculating experimental dose distributions

To calculate the diffracted flux from a crystal after a given exposure time, we first need to know the X-ray beam size and flux-density profile, the crystal size, shape, initial orientation and location in the X-ray beam, and how the crystal is rotated during the

exposure. Using this information, we can calculate the total dose (in J kg^{-1}) delivered to each volume element (voxel) of the crystal after time t : $D(\mathbf{r}, t)$. This dose determines the damage state of the voxel and the diffracted flux (in photons s^{-1}) that it will produce per unit incident flux density. The diffracted flux from the crystal at time t is then obtained by summing the product of the incident flux density at each voxel and the diffracted flux per unit flux density of that voxel.

As shown in Fig. 4.1, a crystal of arbitrary shape is illuminated by an X-ray beam propagating along the x direction. During irradiation/data collection, the crystal may be rotated about an axis perpendicular to the beam direction, and we define a (stationary with respect to beam and detector) coordinate system oriented as shown with $x = 0$ located on the rotation axis and with $y = z = 0$ at the beam center. $F_{\text{inc}}[\mathbf{r}(t), x, t]$ is the incident X-ray photon flux density (photons m^{-2} s^{-1}), where $\mathbf{r}$ is a vector pointing from the coordinate system origin to the position of a voxel within the crystal at time t. The incident X-ray flux density decreases owing to scattering and absorption as it propagates through the crystal, and so in general is a function of the position x along the propagation direction.

Each voxel within the crystal is labeled with its initial ($t = 0$) Cartesian coordinates $\mathbf{r}_0$. As the crystal is rotated about the axis during data collection, a voxel at location $\mathbf{r}_0$ will rotate to location $\mathbf{r}(t) = \mathbf{M}[\varphi(t)]\mathbf{r}_0$ at time t, where $\mathbf{M}$ is a linear rotation matrix and $\varphi(t)$ is the rotation angle about the axis. The dose that has been delivered at time t to a sample voxel initially located at $\mathbf{r}_0$ is given by

$$D(\mathbf{r}_0, t) = k \int_0^t F_{\text{inc}}\{\mathbf{M}[\varphi(t)]\mathbf{r}_0, x, t\}\, \mathrm{d}t, \tag{4.1}$$

where the constant k depends on the X-ray energy and on the atomic composition

and density of the crystal, and can be calculated using *RADDOS E–3D* [132] or standard tables [148]. We assume that the X-ray beam flux density is time-independent and that the crystal is thin compared with the X-ray attenuation length [a reasonable approximation when using 10–15 keV X-rays and 50–300 μm crystals; at 10 keV the beam is attenuated by an average of 17% on passing through 300 μm crystals of the proteins (lysozyme, thaumatin, apoferritin, HLA, λ3 and US2) of relevance here]. The incident flux density $F_{\text{inc}}(\mathbf{r}, x, t)$ can then be written as $F_{\text{inc}}(\rho)$, where $\rho = \left(y^2 + z^2\right)^{1/2}$ is the radial distance from the beam center.

4.2.3 Calculating diffracted intensities

The crystal diffraction pattern measured on the detector consists of a large number of bright 'spots' corresponding to Bragg diffraction at angles $(2\theta)_{hkl}$ for which the scattering wavevector satisfies $q_{hkl} = 2\pi/d_{hkl}$, where d_{hkl} is the spacing of a diffracting crystal lattice plane. Integrating the detector photon counts about the incident beam direction in radial bins of fixed width dq (variable width in detector coordinates) and dividing by the detector exposure time then gives a photon flux $\varphi_{\text{diff}}(q, t)$. At $t = 0$ the decay of $\varphi_{\text{diff}}(q, 0)$ with increasing q owing to thermal atomic motions and to static crystal disorder can be approximated by a Debye-Waller factor DWF(B, q).

At time $t > 0$, the diffraction at scattering wavevector q from each crystal voxel will be proportional to the incident flux density $F_{\text{inc}}(\mathbf{r}, x, t)$ at the position of the voxel and to the diffraction in Bragg peaks due the voxel (the number of diffracted photons per incident photon), which will depend on the dose received by the voxel from $t = 0$ to time t. Let $\Gamma\left[D(\mathbf{r}_0, t), q\right]$ be the factor by which the diffraction in Bragg peaks due to a voxel at initial position $\mathbf{r}_0$ is reduced by radiation damage. The total diffracted flux from

all crystal voxels at wavevector q per unit volume of reciprocal space, $\varphi_{\text{diff}}(q,t)$, is then

$$\varphi_{\text{diff}}(q,t) \propto \int_{\text{crystal volume}} F_{\text{inc}}\ \{\mathbf{M}[\varphi(t)]\mathbf{r}_0\} \times \text{DWF}(B,q) \times \Gamma\left[D\left(\mathbf{r}_0,t\right),q\right] \mathrm{d}V_0. \quad (4.2)$$

We assume that the diffraction in Bragg peaks at a given q decays exponentially with dose,

$$\Gamma\left[D\left(\mathbf{r}_0,t\right),q\right] \propto \exp\left[-D\left(\mathbf{r}_0,t\right)/D_{\text{e}}(q)\right], \quad (4.3)$$

where $D_{\text{e}}(q)$ determines the decay rate. Experimentally, the scattered intensity decays more rapidly as q and 2θ increase, corresponding to a more rapid decay of short-wavelength Fourier components of the unit cell's electron density. Based on a fit to experimental data from crystallography and X-ray imaging, Howells et al. (2009) [138] suggested that $D_{\text{e}}(q) \propto 1/q$ holds and that the resolution-dependent half-dose is given by $D_{1/2}(d) \simeq d \times 10\ \text{MGy/Å}$. We assume a more general relation

$$D_{\text{e}}(q) = K/q^{\alpha}, \quad (4.4)$$

and then fit experimental data from crystallography to deter- mine the exponent α.

The total diffracted photon flux of the crystal Φ_{diff} is obtained by integrating (4.3) over all q as

$$\Phi_{\text{diff}}(t) \propto \int_{q_{\text{min}}}^{q_{\text{max}}} \varphi_{\text{diff}}(t,q)\ q\ \mathrm{d}q. \quad (4.5)$$

4.2.4 Fitting experimental data

Fitting of reported data for diffracted intensity versus resolution d and nominal dose/fluence from prior experiments [140–142] was performed as follows, using the parameters given in Appendix Tables C.1 and C.2.

We used either the actual reported beam profile (measured by scanning a slit across the beam) or else a profile matching the stated shape and width parameters (Appendix

Table C.1). A 'top-hat' profile was represented as $F_{\mathrm{inc}}(\mathbf{r}) = F_0$ for $\rho < \rho_{\max}$ (or $|y| < y_{\max}, |z| < z_{\max}$) and 0 otherwise, and a Gaussian profile by

$$F_{\mathrm{inc}}(\mathbf{r}) = F_0 \exp\left[-\left(\frac{y^2}{2\sigma_y^2} + \frac{z^2}{2\sigma_z^2}\right)\right]. \tag{4.6}$$

Each crystal was divided into cubic voxels, with roughly 100 voxels in each dimension. Crystals were assumed to be rectangular prisms with reported dimensions (Appendix Table C.1). Prisms were oriented as reported, and if no orientation was specified the orientation was adjusted to obtain the best fit to the intensity data (Appendix Section C.2). As a check, calculations were also performed assuming cylindrical crystals with their axes corresponding to the rotation axis.

The reported experiments either repeatedly oscillated the crystal through a small angle (*e.g.* 2°), collecting one or more frames for each oscillation and returning to the starting orientation before collecting the next set of frames, or else continuously rotated the crystal during data collection (*e.g.* by 60° with 1° rotation per diffraction frame). The former method gives a more uniform dose distribution in the irradiated and diffracting crystal regions. For simulations with repeated oscillations through an angular wedge, the simulation time step was set to 1/1000 of the total exposure time for the entire set of oscillations, the crystal was rotated 1/10 of its total oscillation in the wedge (*e.g.* by 0.5° for a 5° oscillation) in each step, and after every ten steps the rotation angle was reset to the starting angle of the wedge. For simulations with continuous rotations, the crystal was rotated by 1/100 of its maximum rotation in each simulation step. The diffracted flux versus q, $\varphi_{\mathrm{diff}}(t_n, q)$, the total diffracted flux $\varphi_{\mathrm{diff}}(t_n)$ and a nominal dose (an average over the voxels that have received nonzero dose) corresponding to the reported dose were calculated at each step. Appendix Fig. C.2 shows example dose distributions for a cylindrical crystal held in a fixed orientation and when rotated, when illuminated with flat-top and isotropic Gaussian beams.

Reported Bragg intensity data were integrated within resolution shells bounded by upper and lower d values. Reported incident beam profiles and diffracted intensity versus resolution shell and fluence/dose plots were digitized using the software WebPlotDigitizer (A. Rohatgi; https://automeris.io/WebPlotDigitizer/). Diffracted intensities in each shell were normalized by their zero-dose extrapolation, eliminating both the Debye-Waller factor DWF(B, q) in (4.2) as well as the Lorentz-polarization correction to the measured intensities, which was not consistently applied in all studies. The resulting normalized plots only show variations with resolution shell (q) and fluence or dose owing to radiation damage.

4.2.5 A simple physics-based model for radiation damage

The primary model used to analyze global radiation damage to protein crystals for the last 50 years is due to [143] and [144], with an additional extension proposed later [145]. As shown in Appendix Fig. C.1, an undamaged crystal becomes disordered at 'rate' k_1 (proportional to the volume fraction of crystal disordered per unit dose). This damaged crystal continues to exhibit Bragg diffraction, but its intensities decrease with increasing q as $I(q) = I_0 \exp\left(-B_{\text{disorder}}\, q^2\right)$, where B_{disorder} is a (fixed) average B-factor increase in the disordered regions. Disordered crystal becomes completely amorphous and ceases to generate Bragg diffraction [$I(q) = 0$] at 'rate' k_2. Undamaged crystal can also proceed directly to the amorphous state at rate k_3. The resulting diffracted intensity versus q and dose D is given by

$$\frac{I(q,D)}{I(q,0)} = e^{-(k_1+k_3)D} + \frac{k_1}{k_1+k_3-k_2}\, e^{-k_2 D}\, \left(1 - e^{-(k_1+k_3-k_2)D}\right)\, e^{-B_{\text{disorder}}\, q^2}. \tag{4.7}$$

With four adjustable parameters, this model has provided good fits to reported $I(q)$ versus nominal dose data, including to the 'hockey-stick' dose dependence evident for data

at larger q or in higher (smaller numeric) resolution shells [140, 143, 144] [58, 141] [127, 142, 149]. However, the connection of this model to underlying damage processes is opaque and the physical significance of the obtained fit parameters is unclear.

We thus considered a very simple model that captures some essential physical aspects of radiation damage. Incident X-ray photons interact with the sample at random times and locations. Each interaction ejects a photoelectron, which then generates lower energy secondary electrons within a volume (determined by the photoelectron mean free path) of a few micrometres. These secondary electrons then diffuse, break bonds and create free radicals that can diffuse and cause additional damage. Experiments using X-ray microbeams [120, 121] show that the resulting perturbations to electron densities from those of the original, undamaged crystal, as reflected in the degradation of the diffraction properties, are confined to within a length comparable in magnitude to the photoelectron mean free path, even at room temperature [127] where diffusion occurs freely.

To model damage, we thus assume that X-ray photon interactions occur at random locations in the crystal. We model the effect of each interaction as a modest 'blurring' of the real-space electron density within a finite region around each interaction point. The number of such interactions per unit crystal volume determines the dose D. We calculate the Fourier transform of the electron density of the crystal and evaluate the integrated Bragg intensity within q (or resolution) shells and plot this versus dose D. Simulations were performed in two dimensions, using $m \times m$ arrays of unit cells containing a grayscale image of a flea (Fig. 4.5), and in three dimensions, using unit cells obtained by discretizing the protein in PDB entry 3E4H [150] tetragonal crystals of the 29-residue plant protein cyclotide varv F at 1.8 Å resolution (Appendix Fig. C.6).

In two dimensions, for each X-ray hit a Gaussian spatial filter was applied to a small

$n \times n$ pixel interaction region centered at a randomly chosen crystal location $\left(x_i, y_j\right)$ as shown in Appendix Fig. C.7. FFTs of the crystal were periodically calculated as hits accumulated. The simulations were continued until the random hits caused the diffraction peaks in the highest resolution shell to fall below the background level, which corresponded to roughly 5–10 hits per pixel. A similar procedure was followed for the computationally more intensive three-dimensional simulations. Details of the simulations are given in Section C.2 of the Appendix.

4.3 Results

4.3.1 Fitting experimental intensity versus dose data

Experimental data for intensity in resolution shells versus dose at $T \simeq 100$ K from three previous studies were analyzed, modeled and fitted using the approach in Sections 4.2.1-4.2.5. Appendix Tables C.1 and C.2 give the experimental details for each reference and our model parameters. We focused on cryogenic temperature data because crystals of different proteins are comparably radiation sensitive, damage is independent of dose rate and because free-radical diffusion and relaxation of protein and lattice structure following bond-scale damage are strongly constrained by the frozen solvent matrix, so that the overall behavior should be simpler and more consistent between protein crystals than at 300 K.

Liebschner et al. (2015)

Liebschner et al. [142] reported the most optimally executed experiments of those examined here. Data were collected from thaumatin crystals at 100 K by repeatedly oscillating the crystals through the same 2° range. Their measured beam profile (their Fig. 1, reproduced here as Appendix Fig. C.3) was nominally flat-topped but had significant tails, and the full widths at half maximum (FWHMs) were much smaller than the crystal dimensions. Fig. 4.2 shows their data for normalized integrated intensity in resolution shells versus nominal dose. As the resolution of the shell increases, the initial decay rate with dose becomes more rapid, and deviations above exponential behavior become evident at smaller doses.

The solid lines in Fig. 4.2(a) show the calculated intensities assuming a top-hat incident beam profile, a 2° oscillation and an exponent $\alpha = 1$ for the q (resolution) dependence of the diffracted intensity decay with dose in (4.4). The calculated dose variation is nearly perfectly exponential and thus does not capture the large deviations from exponential behavior at higher resolutions and doses. By using the measured beam profile [Fig. 4.2(b)], the non-exponential behavior at higher resolutions is qualitatively reproduced. However, in both Figs. 4.2(a) and 4.2(b) the choice of $\alpha = 1$, motivated by the results of Howells et al. (2009), seriously underestimates the observed increase in decay rate with increasing resolution. As shown in Fig. 4.2(c), relaxing this constraint yields a best-fit value of $\alpha \simeq 1.7$ and agreement with the data that is generally excellent in all resolution shells.

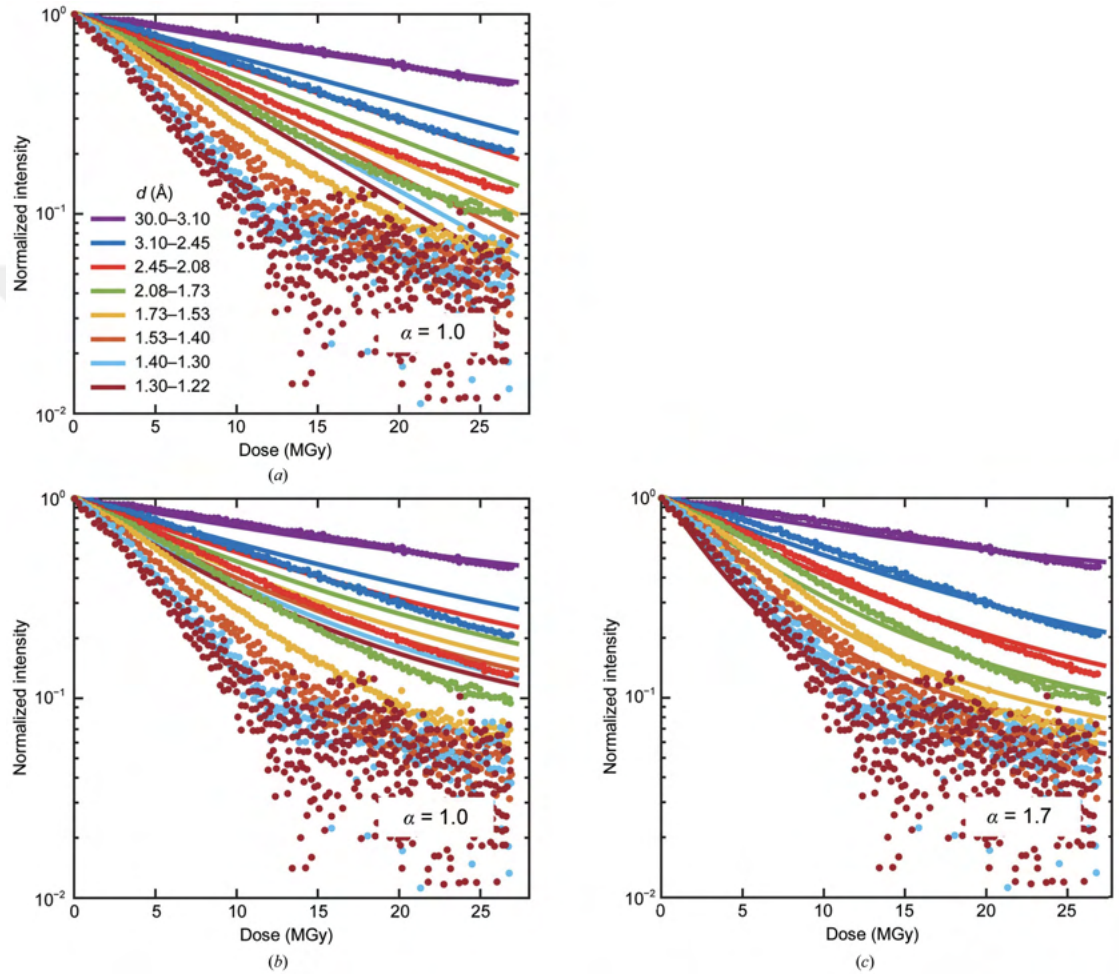

Figure 4.2: Experimental data (solid circles) for integrated intensity in resolution shells versus dose for thaumatin crystals at 100 K, measured by Liebschner et al. [142]. Absolute intensities in each resolution shell (Fig. 4 in the original manuscript) have been normalized by the first ($\sim$ zero dose) intensity point; non-normalized data are shown in Fig. C.4. The solid lines indicate results from simulations assuming (a) a perfect top-hat incident x-ray beam profile and an exponent $\alpha = 1$ in Eq. 4; (b) the measured beam profile (Fig. 1 in the original manuscript, reproduced in Fig. C.3) and $\alpha = 1$; and (c) the measured beam profile and a best-fit exponent $\alpha = 1.7$.

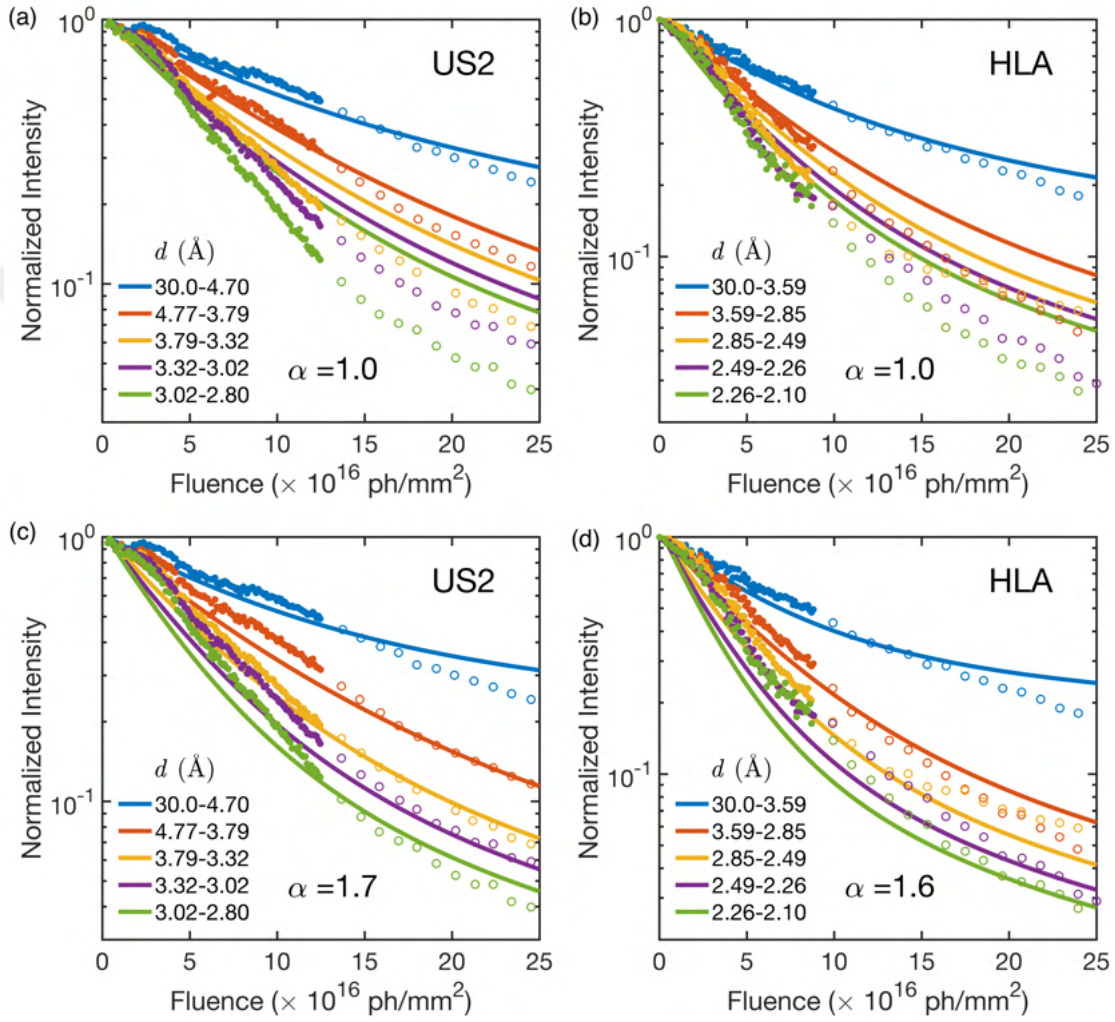

Figure 4.3: Experimental data in the low dose (solid circles) and high dose (open circles) regions for integrated intensity in resolution shells versus incident fluence (in ph/mm^2, proportional to dose) reported by Sliz et al. [140] (Fig. 1 in the original manuscript.) for crystals of US2 and HLA at 100 K. The solid lines indicate results from simulations assuming a top-hat incident beam profile in the horizontal direction and a Gaussian profile in the vertical direction (based on descriptions of the experimental setup), with (a), (b) $\alpha = 1$ and (c), (d) "best-fit" values chosen based on visual comparison.

Sliz et al. (2003)

Sliz et al. [140] collected data at 100 K from crystals of three different proteins: the ternary US2-HLA-A2-Tax peptide complex (referred to as 'US2'), HLA-A2 with a bound melanoma decamer peptide (referred to as 'HLA') and viral polymerase $\lambda 3$ from reovirus (referred to '$\lambda 3$') The focused and collimated X-ray beam was assumed to have a top-hat form in the collimated horizontal direction and a Gaussian form in the focused vertical direction. The crystals were oscillated by only 1°, and all were larger than the beam. Intensities were plotted versus incident fluence (photons mm^{-2}, proportional to dose) and data at low and high fluences were separately reported.

As shown in Fig. 4.3 and Appendix Fig. C.5, these data again show a faster increase in decay with resolution than can be accounted for with $\alpha = 1$. Fit values were 1.7 for US2, 1.6 for HLA and 1.2 for $\lambda 3$. Poorer fits at all resolutions and much larger uncertainties in 'best-fit' α values than for the data of Liebschner and coworkers result because of obvious problems with the original data, and because the beam profiles and initial crystal orientations were not reported. Intensities for all three proteins show an initial plateau (US2) or reduced slope (HLA and $\lambda 3$) versus fluence. Similar behavior observed for thaumatin and lysozyme crystals has been attributed to the effects of dose-dependent mosaicity broadening and cell expansion [127], which is not accounted for by scaling algorithms or by our modeling.

Bourenkov & Popov (2010)

Bourenkov & Popov [141] collected data at 100 K from crystals of insulin, P19siRNA, FAE and FtsH. Crystals of the first three were rotated during exposure by a total angle of between 35 and 300°, with 0.5–1° rotation per frame. Large rotations are not ideal for

our modeling because the dose distribution within and diffraction from the illuminated volume will have a larger dependence on the detailed crystal shape and initial orientation than when crystals are oscillated through a small angle. Crystals of FtsH were both rotated and translated (after each 30°, by an unknown amount) perpendicular to the beam direction, and their data were not modeled. Crystals were illuminated by a nominally

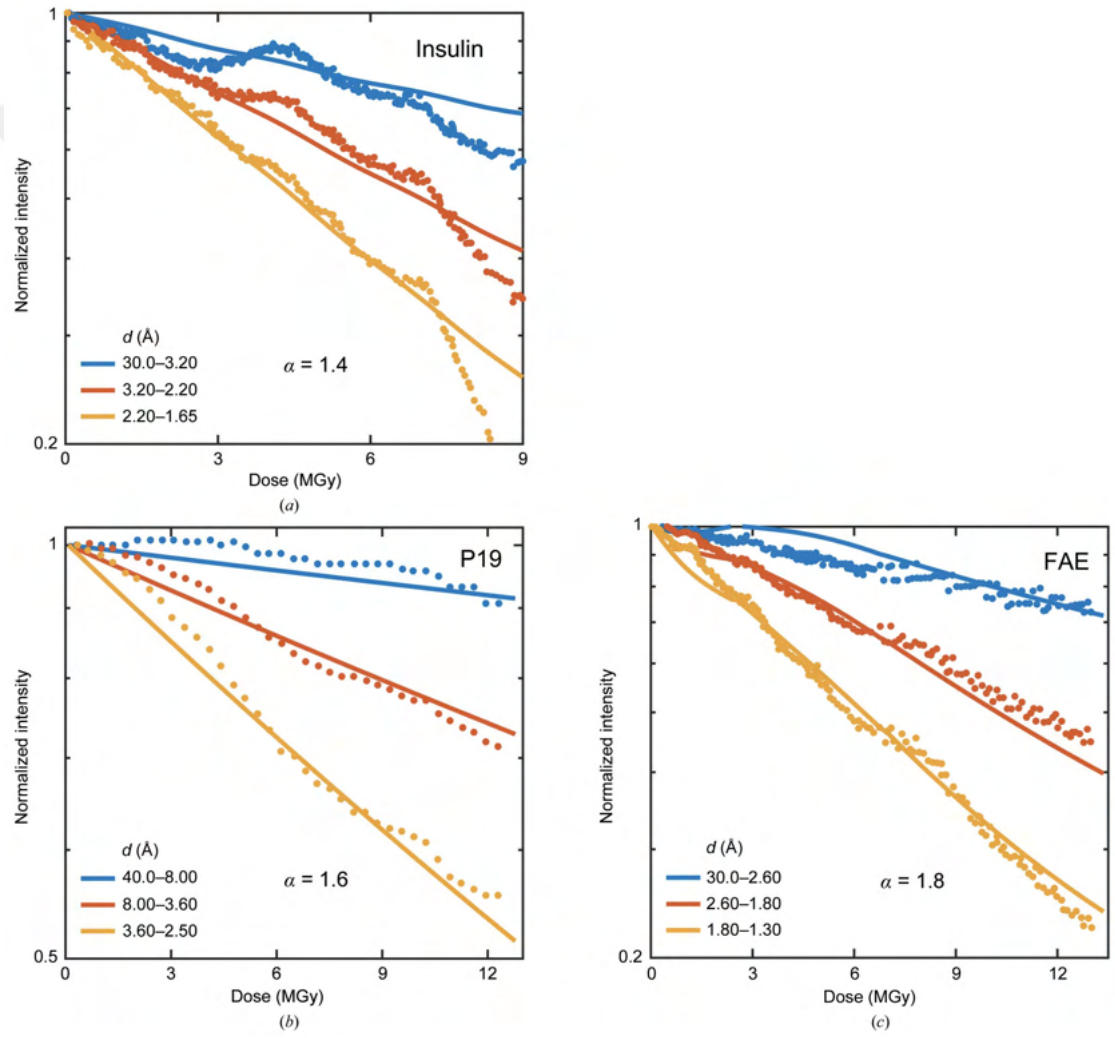

Figure 4.4: Experimental data (solid circles) for integrated intensity in resolution shells versus dose for crystals of insulin, P19 and FAE at 100 K, measured by Bourenkov & Popov [141] (Figs. 3, 4 and 6 in the original paper). Crystals were rotated continuously during irradiation as for crystallographic data collection. Oscillations in the data and may be due to irregular crystal shapes and large rotations that produce complex dose distributions within the x-ray illuminated volume. The solid lines indicate the results from simulations with best fit exponents α of (a) 1.4, (b) 1.6, and (c) 1.8.

Gaussian beam with dimensions that were equal to or smaller than the largest crystal dimension. Beam profiles and initial crystal orientations were not reported. Despite many uncertainties the model calculations yield good fits to the data for all three crystals (Fig. 4.4), with best-fit exponents of $\alpha = 1.4, 1.6$ and 1.8 for insulin, P19 and FAE, respectively.

4.3.2 Simulations of radiation damage

Figs. 4.5 and 4.6 show the results of simulations of our model for radiation damage as a sequence of random, local Gaussian blurs and corresponding to a condition of spatially uniform irradiation. Fig. 4.5(a) shows four unit cells of an initial, undamaged 16×16 cell two-dimensional crystal. Its diffraction (proportional to the square of the FFT amplitudes) has strong peaks extending out to the maximum q or resolution of the initial image. After some large number of hits, the real-space density is blurred throughout the crystal and its diffraction decays much more rapidly with q. Movies of the evolution of the electron density of the crystal and its diffraction are provided as Appendix Movies C.9 and C.10.

Figs. 4.6(a) and 4.6(c) give results in two and three dimensions, respectively, for the predicted diffracted intensity in a resolution shell versus dose. In both two and three dimensions, the diffracted intensity within a resolution shell has a strictly exponential decay with dose, consistent with our assumption for the behavior of the local $\tilde{I}(q, D)$ in fitting the data in Figs. 4.2, 4.3 and 4.4. Figs. 4.6(b) and 4.6(d) give results in two and three dimensions, respectively, for the half-dose in a given resolution shell versus resolution, determined from plots as in Figs. 4.6(a) and 4.6(c) as the dose at which the intensity in a given resolution shell drops to half of its initial value. Except at the high-

est (lowest numerical value) resolutions, in both two and three dimensions the half-dose varies with resolution d approximately as $D_{1/2}(d) \propto d^{\alpha}$ with an apparent low-resolution asymptote of $\alpha \simeq 2$; best-fit values to the near-linear regions in the two- and three-dimensional results are $\alpha \simeq 1.96$ and 1.86, respectively. These results do not change when the Gaussian blur is replaced by a uniform blur (Appendix Fig. C.8). Deviations from simple power-law behavior at the highest (lowest numerical value) resolutions depend on the Gaussian width, with larger widths causing deviations at lower resolutions.

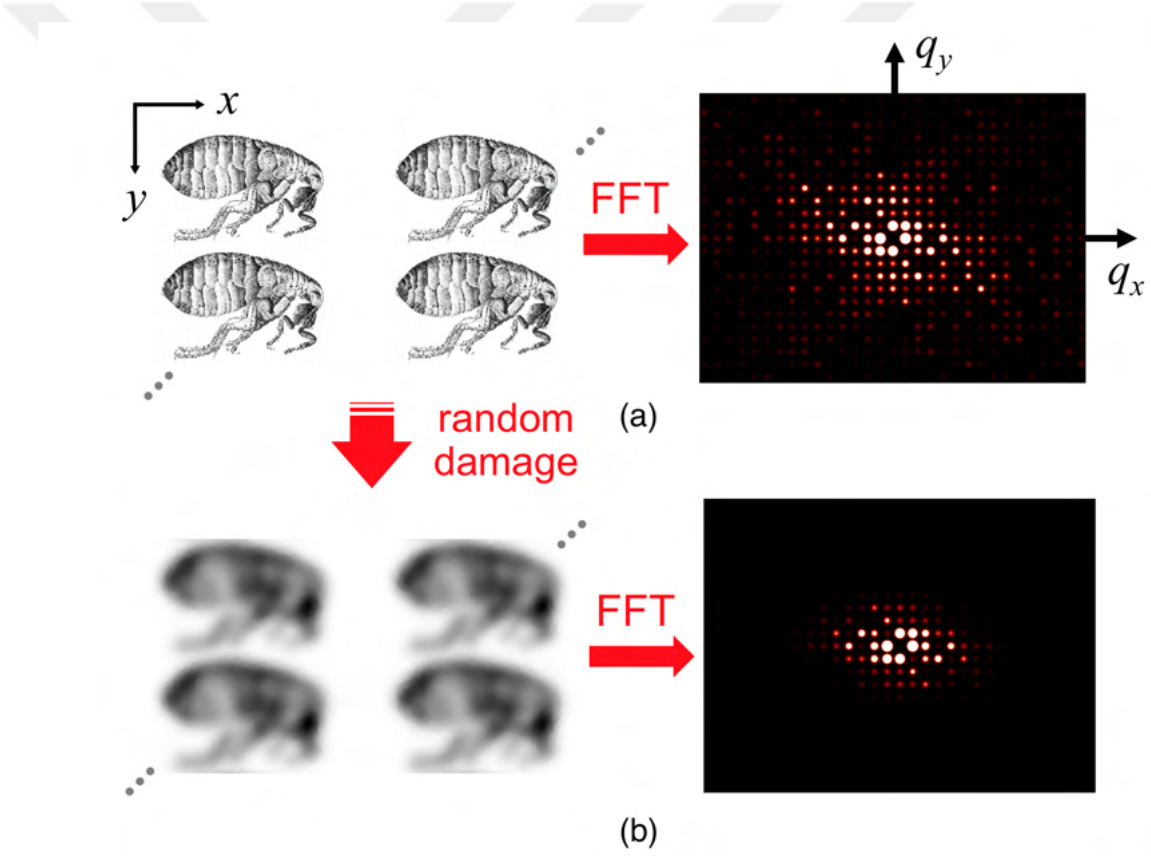

Figure 4.5: Simulation of radiation damage to a 2D crystal formed of images of a flea (from Robert Hooke's *Micrographia*). Each "hit" corresponds to application of a local Gaussian blur at a randomly selected location (Fig. C.7). (a) Four unit cells (each 1024× 512 pixels) of the undamaged 2D crystal (left) and the square of the FFT amplitude, proportional to the diffracted intensity, of a 16×16 cell crystal (right). (b) After receiving a large number of hits (>> 1/ pixel) the electron density has been blurred and the high resolution (large q) diffraction peaks have faded out. A full video of the evolution of the crystal and its diffraction pattern with dose is given in the Appendix (Movies C.9 and C.10).

Similar results were obtained using other twodimensional images and using other PDB entries as the basis for the three-dimensional unit cell.

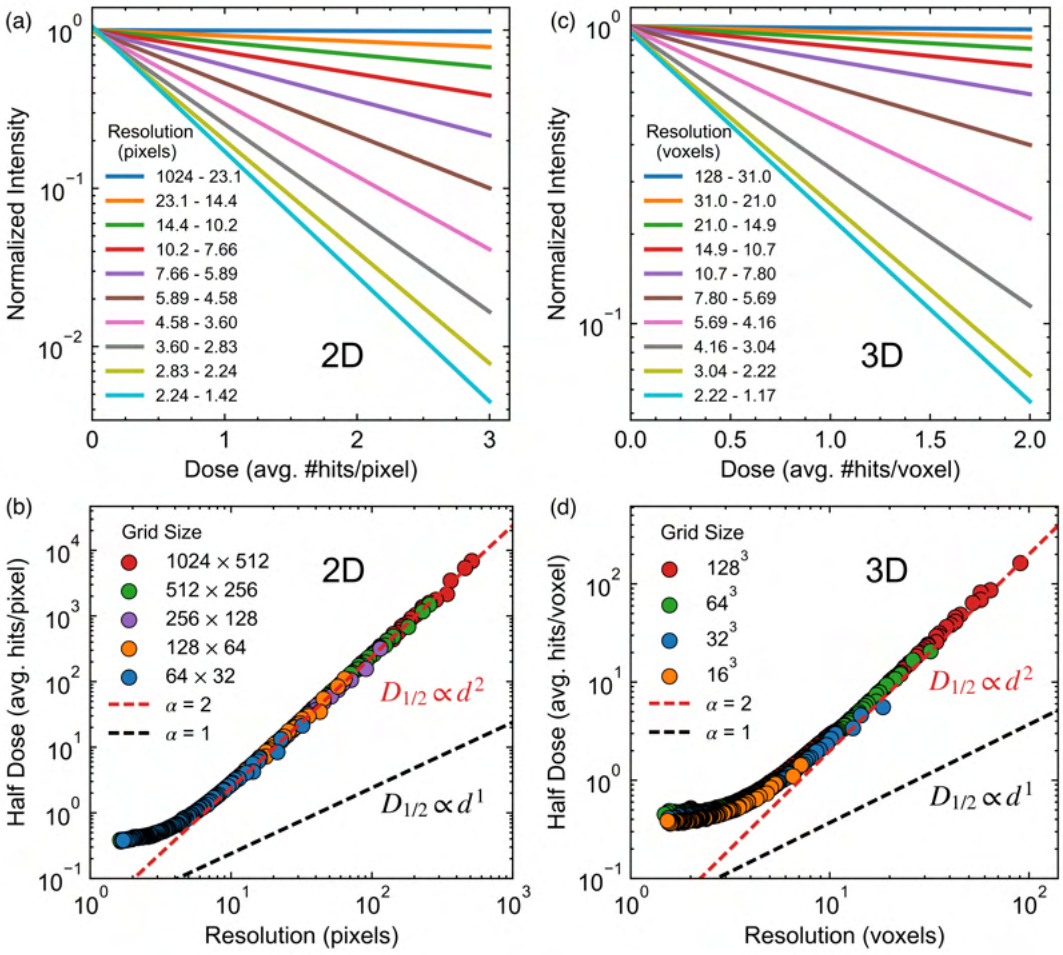

Figure 4.6: Simulation results for the random Gaussian blur model of radiation damage, applied to (a, b) a 2D crystal formed of pixelated images of a flea, and (c, d) a 3D crystal formed of a pixelated electron density map for PDB entry 3E4H. In both 2D and 3D, the intensity within a given resolution shell decays exponentially with dose, and the half-dose varies with resolution d approximately as $D_{1/2}(d) \propto d^{\alpha}$ with α approaching 2 at low resolutions (large numeric values of resolution). The best fit values in the near-linear region are $\alpha = 1.95$ and $\alpha = 1.86$ in (c) and (d), respectively. Simulation details are given in the Appendix C.2.

4.4 Discussion

4.4.1 The local $\tilde{I}(q, D)$ relation: dose dependence

By assuming a purely exponential local dependence of diffracted intensity on dose of the form $\tilde{I}(q, D) = I_0(q) \exp\left[-D/D_e(q)\right]$ with $D_e(q) = K/q^{\alpha}$, and accounting for the nonuniform pattern of crystal irradiation during data collection, we obtain good fits to experimental $I(q)$ versus nominal dose/fluence relations measured for several protein crystals at $T = 100$ K under diverse data-collection conditions. Deviations of the calculated dose/fluence dependence from the data may arise because the actual crystal shapes and initial crystal orientations (which were not given or adequately described) deviate from those assumed, and because of issues in data collection and processing that cause measured intensities to deviate from the actual dose-dependent structure factors [127]. Consequently, based on the available data at 100 K, there is no reason to believe that the local $\tilde{I}(q, D)$ relation at $T \simeq 100$ K is anything but purely exponential in all resolution shells over the resolution and dose range relevant in biomolecular crystallography.

The present analysis also shows that even relatively small deviations of the profile of an X-ray beam from an 'ideal' top-hat form can have a substantial effect on the dose dependence of the intensity at larger doses and higher resolutions. This is particularly evident in the fits to the data of Liebschner and coworkers in Fig. 4.2 using the actual profile of the nominally top-hat beam dramatically improves the fit quality.

4.4.2 The local $\tilde{I}(q, D)$ relation: q dependence

Howells et al. [138] presented a summary of available data for resolution-dependent maximum tolerable doses, obtained from published half-dose values in biomolecular crystallography and from X-ray and electron imaging studies. These results, spanning resolutions from ~2 to 700 Å, show large scatter but are roughly consistent with a linear resolution dependence corresponding to $D_e(q) = K/q^{\alpha}$ with $\alpha = 1$.

The present analysis shows that data for protein crystallography with resolutions between ~1 and 10 Å are unambiguously inconsistent with $\alpha = 1$, and yield best-fit values of between ~1.4 and 1.8, with the most 'ideal' data of Liebschner and coworkers yielding a value of ~1.7.

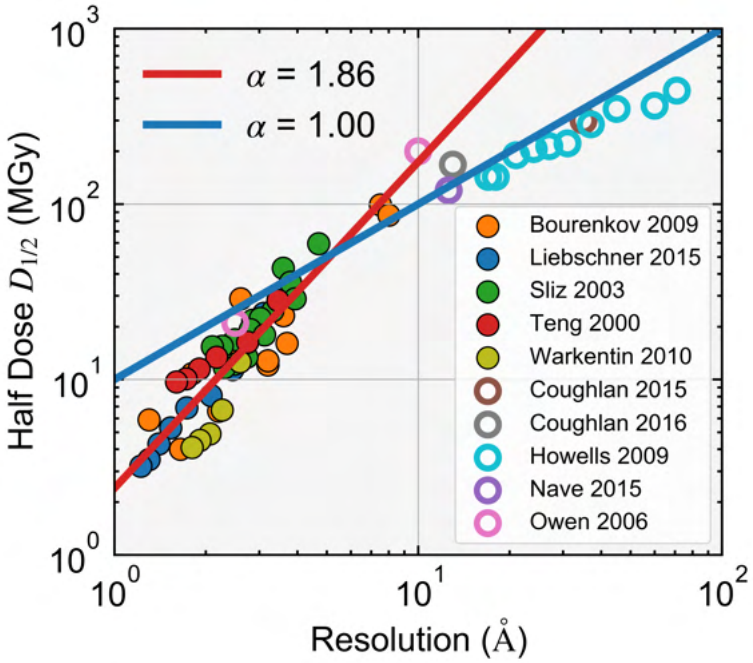

Figure 4.7: Experimental half-doses versus resolution obtained from several previous crystallographic studies [140–142] [9, 58, 138, 151] [152, 153], analyzed accounting for their dose distribution (solid circles) or used as reported (open circles). Aside from the data of Howells et al. (2009) at resolution beyond 10 Å, the overall trend is in good agreement with $\alpha = 2$, with a best-fit exponent of 1.86.

Fig. 4.7 summarizes the results for half-dose versus resolution at 100 K, deduced from previous $I(q)$ measurements using the methods of Sections 4.2.1-4.2.5 (solid symbols) or as originally reported (open symbols). These half-dose values are for diffraction within resolution shells, rather than overall halfdose values obtained by integrating the entire diffraction pattern over all resolutions. The data between 1 and 10 Å are well described by $\alpha \simeq 2$, with a best-fit value of 1.86; the only data that appreciably deviate from this fit are those first reported by Howells and coworkers. Extrapolating the $\alpha = 1.86$ fit from 1 to 600 Å yields a half-dose of $\sim 3 \times 10^{11}$ Gy and using $\alpha = 2$ gives $\sim 7 \times 10^{11}$ Gy. This compares with a reported overall half-dose value (obtained by integrating over all resolutions) in X-ray imaging of cells to this resolution of 5×10^{11} Gy [154] and with a value of only $\sim 1 \times 10^{10}$ Gy based on the best fit in Howells and coworkers (their Fig. 3) with $\alpha = 1$.

4.4.3 The local $\tilde{I}(q, D)$ relation: connection to "kinetic" models

The 'kinetic' models of Blake & Phillips [143], Hendrickson [144] and Sygusch & Allaire [145] all implicitly assume uniform sample irradiation, and so should not have been used to fit experimental data that were collected under conditions of substantially nonuniform illumination. The present analysis shows that the local $\tilde{I}(q, D)$ is consistent with a purely exponential dose dependence for all q at 100 K and that the dose scale for intensity at a given q varies as a power of q. Equation (4.7) cannot replicate these features with any sensible parameter choices. The deviations from exponential behavior that these models have proved so successful at fitting are owing to nonuniform sample irradiation (and possibly also to data-processing errors), which these models do not include. These models are thus inconsistent with experiment at $T = 100$ K and should no longer be used.

4.4.4 The "dose limit" in biomolecular crystallography

Based upon experience in cryoelectron microscopy, Henderson [155] suggested that the maximum tolerable dose in X-ray cryocrystallography, beyond which diffraction would be seriously degraded, would be roughly 20 MGy [155]. Teng & Moffat [9, 156], using perhaps the most nearly ideal irradiation conditions to date – a beam with a 2σ width much larger than their crystal size (250 μm versus 100 μm), giving nearly uniform illumination of the entire crystal volume – obtained a T = 100 K half-dose of ~17 MGy for lysozyme crystals diffracting to 1.6 Å resolution. Based on diffraction statistics they suggested a dose limit of ~10 MGy be used in macromolecular crystallography. Burmeister [130] obtained a T = 100 K half-dose of ~21 MGy for myrosinase crystals diffracting to 2.0 Å resolution. Owen et al. [153] obtained half doses of 40 and 48 MGy for holoferritin and apoferritin crystals diffracting to ~2.3 Å, and based on examination of diffraction statistics and electron-density maps suggested that a maximum dose of ~30 MGy be used. Liebschner et al. [142] using a nearly flat-top beam, obtained a half-dose of 18.5 MGy for thaumatin crystals diffracting to 2.1 Å. Warkentin et al. [127], using a Gaussian microbeam with no crystal oscillation and correcting for the effects of the Gaussian beam profile, obtained half-doses of 10 MGy for lysozyme crystals diffracting to 1.4 Å and 13 MGy for thaumatin crystals diffracting to 1.6 Å. Many other half-dose measurements have been reported but have generally involved substantially nonuniform crystal irradiation, which can make measured half-doses substantially larger than the true, local half-dose that would be measured under conditions of purely uniform irradiation [127].

Our conclusion based these previous studies is that diffraction half-doses at T = 100 K for crystals diffracting to ~1.5–2 Å are ~15–20 MGy, and that the ~10 MGy dose limit suggested by Teng and Moffat is appropriate. However, the 30 MGy 'Garman

limit' reported by Owen et al., rather than the 10 MGy limit of Teng and Moffat, has been by far the most widely cited, and has become the accepted standard dose limit. It far exceeds the dose that would be required to severely degrade diffraction in the highest resolution shells of, for example, the overwhelming majority of PDB entries, which have a median refined resolution of ~2.0 Å.

As shown in Fig. 4.7, using the local half-dose in a given resolution shell as a more meaningful and robust metric, we find that the dose limit at $T = 100$ K increases from ~2–3 MGy at 1 Å to 8 MGy at 2 Å, 16 MGy at 3 Å and 30 MGy at 4 Å. Half-doses obtained by integrating over all resolutions up to the maximum available resolution are somewhat larger than but track these values; for lysozyme crystals diffracting to a maximum resolution of 1.4 Å the half-dose is ~10 MGy [127]. These resolution-dependent dose limits should be used as rules of thumb in place of the previous 20 or 30 MGy limit when determining exposure strategies in crystallographic measurements.

Why were the half-doses reported by Owen et al. so much larger than were obtained in the other measurements? As noted previously [8], the data sets analysed had a resolution limit of 2.3 Å, somewhat lower than those used in other studies; this can account for roughly half the difference with half-dose values measured by Teng and Moffat [9] and Warkentin et al. [127]. Owen et al. used a 100×100 μm X-ray beam, and stated that the beam profile on the beamline used for their studies 'has been determined to be a top-hat shape', citing Arzt et al. [157] for the profile. However, the beamline profiles reported in Fig. 4 of Arzt et al., are not top hat (normally produced through collimation or slitting of a defocused beam). They are standard focused profiles. With tight focusing to $31(V) \times 47(H)$ μm [Arzt et al., [157]; Fig. 4(a)], the beam profile was roughly Gaussian in the vertical and somewhat flattened in the horizontal; with a 100×100 μm FWHM spot size as used by Owen et al., the profile may have been more nearly Gaussian in both hor-

izontal and vertical [Arzt et al., [157]; Fig. 4(b)]. Since the beam size was much smaller than the ∼200 μm of the holo- and apoferritin crystals examined, the crystals may thus have been nonuniformly irradiated, and this may have increased the apparent half-dose relative to the true, local half-dose by a factor close to two. Which experimental dose should be compared with these limits, for a crystal diffracting to a given maximum resolution? In nearly all crystallographic data collection the crystal is nonuniformly irradiated owing to nonuniform flux density in the beam and owing to crystal rotation. Using *RADDOSE–3D* or simple code (written in, for example, *MATLAB*) based on (4.1), and knowing the incident X-ray flux-density profile, the approximate crystal dimensions and the crystal rotation or oscillation pattern, the dose distribution within a crystal during a given data collection can easily be calculated. *RADDOSE–3D* currently calculates the maximum dose (at any position) received within the X-ray-illuminated crystal volume, the average dose within the illuminated volume, the average dose within the crystal and an incident flux-densityweighted dose [147]. The most conservative choice is to use the maximum dose. The average dose within the illuminated volume and within the crystal can both yield problematically small dose estimates when the crystal is larger than the beam and when the beam has a non-top-hat (*e.g.* Gaussian) profile.

A more robust measure of average dose is the diffraction-weighted dose [127], which weights the dose received at each location after an exposure time t by its contribution to the measured diffraction at time t, and thus appropriately downweights contributions from regions that, either owing to weak incident illumination or owing to radiation damage, contribute little to the measured diffraction. The diffraction-weighted dose is given by

$$\mathrm{DWD}(t) = \frac{\int_0^t \int_{\mathrm{crystal}} D\left(\mathbf{r}, t'\right) S\left(\mathbf{r}, t'\right) F_{\mathrm{inc}}\left(\mathbf{r}, t'\right) \mathrm{d}^3 r \cdot \mathrm{d}t'}{\int_0^t \int_{\mathrm{crystal}} S\left(\mathbf{r}, t'\right) F_{\mathrm{inc}}\left(\mathbf{r}, t'\right) \mathrm{d}^3 r \cdot \mathrm{d}t'}. \tag{4.8}$$

Here, $S(\mathbf{r}, t) = S[D(\mathbf{r}, t)]$ is the diffracted flux (in all reflections) per unit illuminated crystal volume per unit incident flux density at position $\mathbf{r}$ and time t, and the denominator

gives the total number of scattered photons up to time t (not the total diffracted intensity as stated by Warkentin and coworkers). As assumed by Warkentin and coworkers and supported by the present analysis, $S(\mathbf{r}, t)$ decays exponentially with dose D. However, since the highest resolution diffraction peaks fade out at the smallest doses, they are down-weighted in the definition of (4.8). If the effects of dose within a given resolution shell centered at some q are of primary interest, then the weighting diffraction can be restricted to that q, *e.g.*

$$\mathrm{DWD}(q,t) = \frac{\int_0^t \int_{\mathrm{crystal}} D(\mathbf{r},t')\, s(\mathbf{r},q,t')\, F_{\mathrm{inc}}(\mathbf{r},t')\, \mathrm{d}^3 r \cdot \mathrm{d}t'}{\int_0^t \int_{\mathrm{crystal}} s(\mathbf{r},q,t')\, F_{\mathrm{inc}}(\mathbf{r},t')\, \mathrm{d}^3 r \cdot \mathrm{d}t'} \tag{4.9}$$

where

$$S(\mathbf{r},t) = \int_q s(\mathbf{r},q,t) \times q \, \mathrm{d}q$$

and, averaging over reflections at a given q as in our model in Section 4.2.3, $s(\mathbf{r},q,t) \equiv \Gamma[D(\mathbf{r},t),q] \times \mathrm{DWF}(B,q)$.

4.4.5 Mechanisms underlying the local $\tilde{I}(q, D)$ relation

Global radiation damage to biomolecular crystals at 100 K is robust: it does not have an appreciable dependence on any properties of the biomolecule (*e.g.* primary sequence and fold) or of the crystal (e.g. packing density, solvent content and composition). Each X-ray absorption or inelastic scattering event generates secondary electrons and damage in a volume of many cubic micrometres containing a large number of biomolecules and unit cells. The frozen solvent network prevents relaxation of the structure following each damage event on any but the shortest length scales. As a result, one might expect that a fairly simple physical damage model would be required to reproduce the essential features of the decay of diffraction with dose at 100 K.

This expectation is borne out by the present results. A model of sequential random damage interactions that cause local blurring of the electron density reproduces all salient trends of the available data for global radiation damage, including both its dose and q dependence (to within uncertainties arising from how available data were collected).

The apparent asymptote of the simulation results to $\alpha = 2$ at large q in Fig. 4.6 can be readily understood. In the limit of each pixel or voxel having received a large number of hits N, the fractional fluctuations in the number of hits (or blurs) per voxel $N^{1/2}/N$ will become small, and the blurring of the electron density will become nearly uniform throughout the sample. The uniformity of the blurring will be greater for long-wavelength (large numeric resolution, small q) Fourier components of the electron density, since these average over fluctuations in a larger volume. Instead of our elementary damage event corresponding to a local blurring in a small volume of the crystal, in this large-dose, low-resolution limit we can assume a simpler model in which our elementary damage event delivers a uniform dose D_0 that produces a uniform blurring throughout the crystal. After n of these events, the total dose received by the sample is $D_n = nD_0$, and the electron density satisfies $\rho(\mathbf{r}, D_n) = \rho(\mathbf{r}, D_{n-1}) \otimes G(\mathbf{r})$ where the blur kernel $G(\mathbf{r})$ is convolved with the real-space density. The Fourier transform of the electron density and thus the diffracted intensity will have the form $I(q, D_n) = I(q, 0) \times [G'(q)]^{2n}$. Taking $G(\mathbf{r})$ and thus $G'(\mathbf{q})$ as isotropic Gaussians, the diffracted intensity can be written as

$$\begin{aligned} I(q, D_n) &\propto I(q, 0) \times \exp\left(-4\pi^2\sigma^2 q^2 n\right) \\ &\propto I(q, 0) \times \exp\left[-D_n/D_e(q)\right], \end{aligned} \tag{4.10}$$

where

$$D_e(q) \propto 1/q^2. \tag{4.11}$$

Thus, this simplified model predicts an exponential dependence of intensity on dose and an exponent $\alpha = 2$ in (4.4).

Above the protein-solvent glass transition near 200 K, damage processes involving relaxations on large length scales and longer, temperature-dependent timescales may qualitatively change the evolution of disorder with dose, especially at large doses. Deviations of the local $\tilde{I}(q, D)$, determined by deconvolving the effects of nonuniform irradiation, from the predictions of the simple model used here should provide a useful starting point for the study of these damage processes.

4.4.6 Implications for crystallographic data analysis

Diffraction scaling programs used in crystallography attempt to correct for changes in Bragg intensities owing to radiation damage. An early approach assumed a linear variation of peak intensity with frame number or dose and used measurements of equivalent peaks in different frames to extrapolate back to the zero-dose intensity. A second approach assumes that the decay of intensities with dose can be described by a linearly increasing B factor, $I(\theta, n) = I_0 \left[B(n) \sin^2(\theta)/\lambda^2 \right]$, where $B(n) = (1 - n)B_0 + nB_1$ and n, the frame number, is proportional to dose [133, 158]. This can be rewritten as

$$\begin{aligned} I(q, n) &= I_0 \exp\left(-\frac{B_0 q^2}{16\pi^2}\right) \exp\left[-\frac{(B_1 - B_0)\, n q^2}{16\pi^2}\right] \\ &= I(q, 0) \exp\left[-n/n_e(q)\right] \end{aligned} \tag{4.12}$$

with $n_e(q) \propto 1/q^2$. This matches the result of our model and is consistent (within experimental uncertainties) with the local $\tilde{I}(q, D)$ relation we deduced from previous experiments.

This correction can be calculated separately for each set of symmetry-related peaks and their Friedel mates, assuming that all peaks in this set have the same decay rate [136]. This requires that the multiplicity of the data set be sufficiently high and that observations of equivalent reflections are well spaced over the data set.

The difficulty in these approaches in accounting for radiation damage is that the measured peak intensities (even when fully recorded) may not have an exponential dependence on frame number [127, 149], the dose state of the diffracting crystal region does not in general vary linearly with frame number [127], and both of these effects are by far the largest for the highest resolution data. Because the crystal is nonuniformly irradiated and nonuniformly damaged, diffraction in any frame reflects a nonlinear weighting of structure factors from crystal regions in various states of decay. Since the local $D_e(q) \propto 1/q^2$, the effects of nonuniform irradiation are much more pronounced, and become evident at much smaller exposures, for the highest resolution structure factors. The size, shape and position in reciprocal space of a structure-factor peak evolve with dose, because mosaicity, the spread in lattice constants within the illuminated volume, and the average lattice constant generally increase with dose. Depending on the incident beam divergence and energy spread, the initial crystal mosaicity and its rate of increase with dose, the incident beam fluence or dose per frame and the sample rotation per frame, the resulting intensity variations with exposure time can introduce large errors in nominally fully recorded peak intensities. Recording high-multiplicity data can average over these effects, but there is no reason to expect that the 'average' for a set of equivalent reflections, or its extrapolation to frame $n = 0$ based on damage models implemented in current scaling programs, will correspond to the $n = 0$ structure factor, especially in the highest resolution shells. This may contribute to the rapid degradation of R-factors within each shell as the resolution limit of a data set is approached [159].

The good news, evident from the present work, is that our understanding of radiation damage is improving. It should soon be possible to implement much more sophisticated models, and perhaps also improved crystallographic datacollection protocols that include the measurement of key damage-related parameters, to allow more accurate correction of measured intensities. This could help close the *R*-factor gap between protein

and small-molecule structures. Whether the corrections are large enough to significantly impact structural models and mechanistic understanding remains to be determined.

4.5 Conclusions

We have shown that the experimentally observed diffracted intensity decays and their resolution or q dependence, arising from radiation damage to biomolecular crystals, can be explained by assuming a locally exponential relation $\tilde{I}(q, D)$ between diffracted intensity and dose with a half-dose $D_{1/2}(q) \propto 1/q^{\alpha}$ where $\alpha \simeq 1.7$, and by accounting for the effects of non-uniform irradiation, damage and diffraction during data collection. The very strong dependence of $D_{1/2}(q)$ on q increases the effects of both radiation damage and of nonuniform irradiation on measured intensities (and the structure factors derived from them) in the highest resolution shells. Consequently, the 20–30 MGy Henderson or Garman dose limit, which has long been used as a rule of thumb in crystallography, should be replaced with a metric that depends on the initial maximum resolution of a data set, and the application of this metric should account for nonuniform irradiation and diffraction during data collection. Radiation-damage models that have been long used to fit data for $I(q)$ versus dose implicitly assume uniform sample irradiation and do not apply. Both an exponential dose dependence for $\tilde{I}(q, D)$ and $D_{1/2}(q) \propto 1/q^{\alpha}$ with $\alpha \simeq 2$ follow from perhaps the simplest physically plausible model, in which damage events cause random, local blurring of the electron density. Experimental deviations from these model predictions, especially at temperatures above the protein-solvent glass transition where radiation sensitivity rapidly increases, should guide the development of a more complete model.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Very recently, the difference in flap state between KAU and SPU structures was interpreted as arising from differences in pH during crystallization was attempted to explain by the difference in the pH values of the crystallization solutions – KAU entries had pH of ~6.5 whereas SPU crystals had pH of 7.0–7.5 (except only a few KAU mutant variations) [60]. In order to test the hypothesis, they cocrystallized SPU in the presence of N-(n-butyl)phosphoric triamide (NBPTO) at two different pH values 6.5 and 7.5 and the structures were determined at 100 K (PDB code: 6RP1 and 6RKG respectively). NBPTO is in situ hydrolyzed to diamidophosphate (DAP), a tetrahedral moiety replacing all four active site water molecules, which has been previously proposed to be transition state analog of urease [41]. They observed that the flap adopts closed conformation at high pH and near-equal mixture of open and closed conformations at low pH, thus seemingly explaining the difference between the flap conformations of KAU vs. SPU structures. However, this is far from depicting the whole story about the different conformations adopted by mobile flap in different urease structures. First, in that study, the structure of DAP-bound, not native urease is solved, so whether the conclusion drawn there can be generalized to the native structure or other cases remains to be elucidated. Moreover, in an earlier study, Cys319Ala variants of KAU were soaked in buffers of four different pH values ranging from 6.5 to 9.4 for about a day and all structures determined at room temperature turned out to exhibit closed flap without any noticeable difference between their maps [160]. The contradictory results may or may not be due to the soaking of crystals in the buffer solutions in the earlier study vs. co-crystallization in the recent study. While pH dependence of the conformational state of such a flexible flap is not surprising, presumably even more important experimental parameter is the temperature at which the structure is determined.

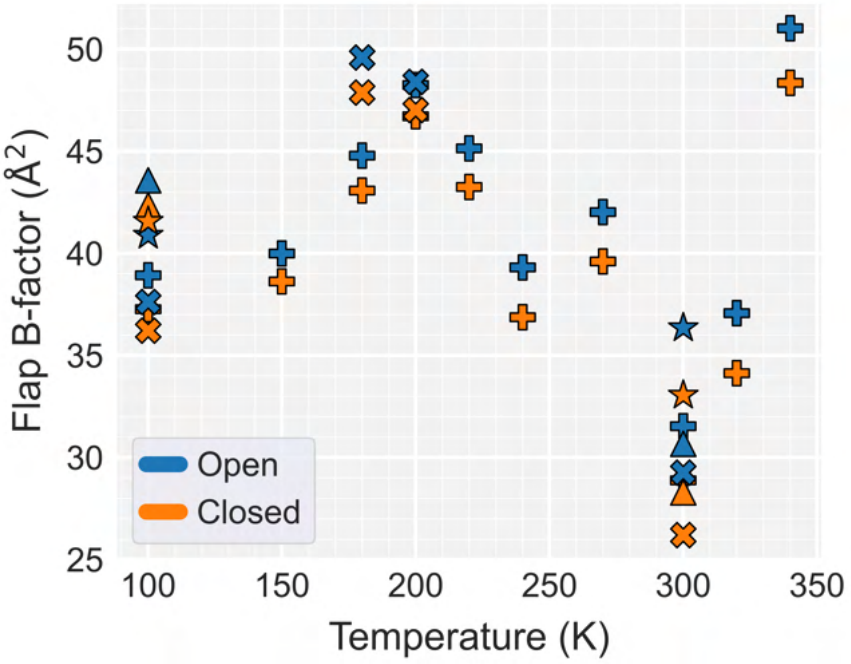

Figure A.1: Average B-factors, calculated over all atoms in the flap, of the open and closed flap conformations as a function of temperature for cryoprotectant-free data series. Average flap B-factors obtained from the two-state refinement are comparable, with closed state values being only slightly lower. These B-factors show no obvious trend with temperature, but do correlate with the B-factor obtained by averaging over the whole protein, which is determined by data quality / resolution.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

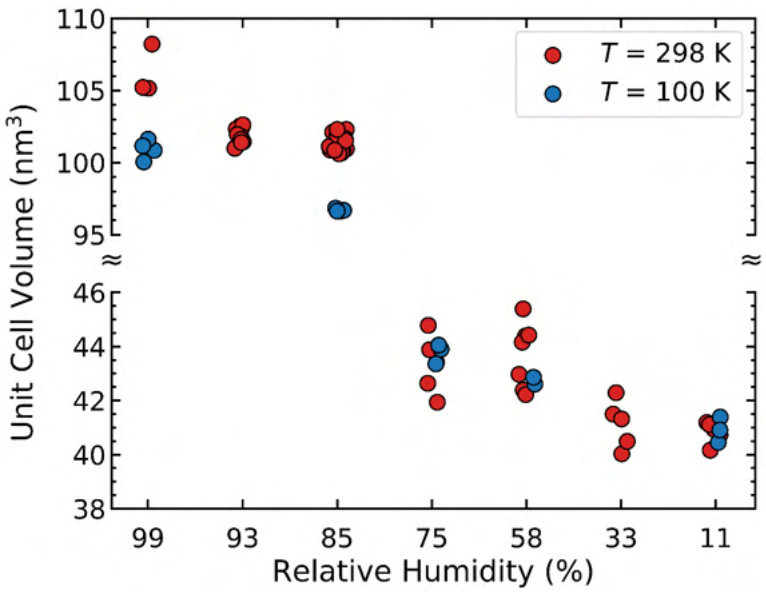

Figure B.1: Measured unit cell volumes for all crystals examined in this study at room temperature (red symbols) and at $T = 100$ K (blue symbols). Cell volumes obtained at relative humidities below that of the structural transition that occurs between 86% and 75% r.h. show more scatter, perhaps reflecting differences in unit cell evolution through the transition and in the lattice disorder created by the transition.

		Matthews coefficient method		Rolling probe method (*3VEE*)			
	Cell volume (Å^3)	Solvent content (%)	Solvent volume (Å^3)	Protein volume (Å^3)	Solvent volume(Å^3)	Solvent content(%)	Ordered solvent volume (Å^3)
h99	27055	34.9	9442	16908	10148	37.5	1880
h93	25583	31.2	7982	16722	8861	34.6	2374
h85	25215	30.2	7615	16581	8634	34.2	2789
h75	22550	22.0	4961	16552	5998	26.6	2451
h58	21750	19.1	4154	16575	5175	23.8	1404
h33	20840	15.6	3251	16689	4151	19.9	994
h11	20475	14.1	2887	16747	3728	18.2	413
h99-c	25213	30.2	7614	16496	8717	34.6	4230
h85-c	24160	27.2	6572	16421	7740	32.0	5025
h75-c	21681	20.9	4531	16472	5209	24.0	3697
h58-c	21428	16.7	3578	16441	4987	23.3	2634
h11-c	20215	13.0	2628	16465	3750	18.6	912

Table B.1: Protein volume, solvent volume, and ordered solvent volume per protein molecule for monoclinic lysozyme crystals as determined by the rolling probe method using the program 3VEE with a probe size of 1.4 Å (radius of water) and a grid size of 0.07 Å. The calculated volumes for lysozyme become nearly constant for a grid size below ~0.1. For comparison, the solvent content and volume derived from the Matthews coefficient are also given, where the solvent content (volume solvent/volume unit cell) is calculated as $1 - 1.23/V_m$. This latter value assumes a protein density of 1.35 g/cm^3. The *3VEE*-determined protein volume at 298 K is 16908 Å^3, corresponding to a protein density of ~1.41 g/cm^3, for the native structure, and these values do not vary appreciably with dehydration. and without a systematic dependence on hydration level. The solvent excluded volume of lysozyme calculated by *3VEE* is consistent with results we calculated with other widely used programs based on rolling probe method such as MSMS [161] (16637 Å^3 or protein density of ~1.43 g/cm^3, using a probe radius of 1.4 Å and the highest density of pixel points of 10) and *EDTSurf* [162] (16441 Å^3 or protein density of 1.45 g/cm^3, using a probe radius of 1.4 Å, a scale factor of 20, and MC triangulation.) Previous analysis using the Connolly algorithm and properly accounting for the surface atoms obtained an average protein density near 1.43 g/cm^3 [101], and found that protein density increases with decreasing molecular weight, especially for proteins with MW below 20 kDa [102]. Based on the reported exponential fit, the predicted density for lysozyme with molecular weight of 14.3 kDa is ~1.46 g/cm^3 (volume of 16270 Å^3), consistent with our results.

	r.h. 99%	r.h. 93%	r.h. 89%	r.h. 75%	r.h. 99% cryo
Wavelength (Å)	0.6299	0.6309	0.6309	0.6309	0.9770
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$
a	58.80	58.32	57.79	53.15	57.91
b	58.80	58.32	57.79	53.15	57.91
c	152.37	152.22	151.65	152.48	150.19
Solvent content (%)	58.5	57.8	56.8	49.3	56.6
Total reflections	57979	63573	39139	39416	46060
Redundancy	3.3(1.9)	4.9(2.7)	4.4(1.8)	8.5(4.0)	6.8(4.5)
Completeness (%)	99.4(96.1)	97.7(77.4)	92.6(58.2)	99.2(90.6)	99.6(93.4)
Mosaicity (°)	0.04 – 0.08	0.16 – 0.47	0.04 – 0.58	0.07 – 0.50	0.28 – 0.40
$CC_{1/2}$	0.996(0.503)	0.988(0.154)	0.996(0.197)	0.992(0.323)	1.000(0.756)
Mean I/σ	9.0(0.8)	8.6(0.5)	6.9(0.5)	7.7(0.6)	10.0(1.2)
Resolution (Å) (at $I/\sigma = 2$)	1.52	1.55	1.77	1.67	1.51
R-merge / R-meas (%)	0.068/0.074	0.110/0.119	0.135/0.137	0.201/0.211	0.065/0.075
Wilson B-factor ($Å^2$)	16.8	20.2	20.1	17.6	15.6
Refinement resolution (Å)	27.06 – 1.36	27.23 – 1.30	28.38 – 1.49	23.76 – 1.45	45.86 – 1.45
R-work / R-free (%)	12.0/14.2	12.9/16.1	13.6/18.1	14.0/17.8	13.2/15.5
Test set for R-free (%)	5	5	5	5	5
RMS bonds (Å) / angles (°)	0.007/0.938	0.008/0.934	0.008/0.951	0.009/0.982	0.007/0.941
Ramachandran favored / outliers (%)	96.5/0.0	98.0/0.0	97.8/0.0	98.6/0.0	98.2/0.0
Rotamer outliers (%)	1.1	1.5	2.8	2.1	0.0
Clashscore	2.4	3.3	4.3	6.1	2.8
B-factor ($Å^2$)	22.1	27.5	25.6	23.8	18.6
Protein B-factor ($Å^2$)	21.2	26.5	24.9	23.2	16.6
Solvent B-factor ($Å^2$)	35.0	41.9	38.0	34.7	31.2
Observed waters	157	158	133	120	325

Table B.2: Data collection and refinement statistics for room temperature and $T = 100$ K data collected using tetragonal thaumatin crystals.

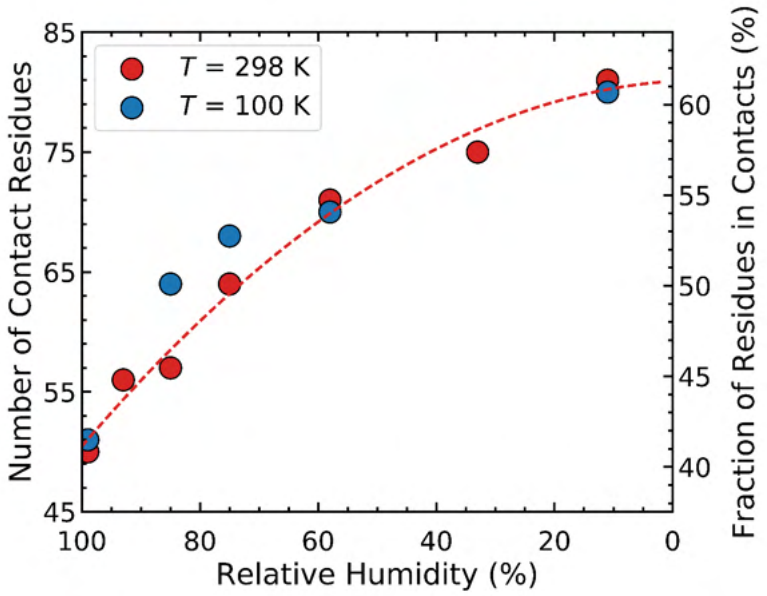

Figure B.2: Evolution of number of residues involved in crystal contacts as a function of relative humidity. Two atoms are considered to be in contact if their center to center distance is less than 4 Å.

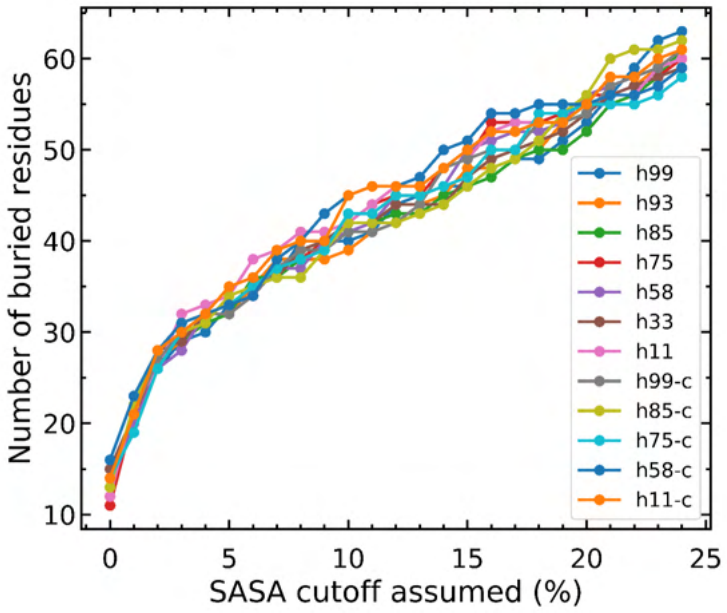

Figure B.3: Number of buried residues as a function of the cut-off percent assumed for the fraction of a given residue's "free" solvent accessible surface area (SASA) that is solvent accessible in the crystal.

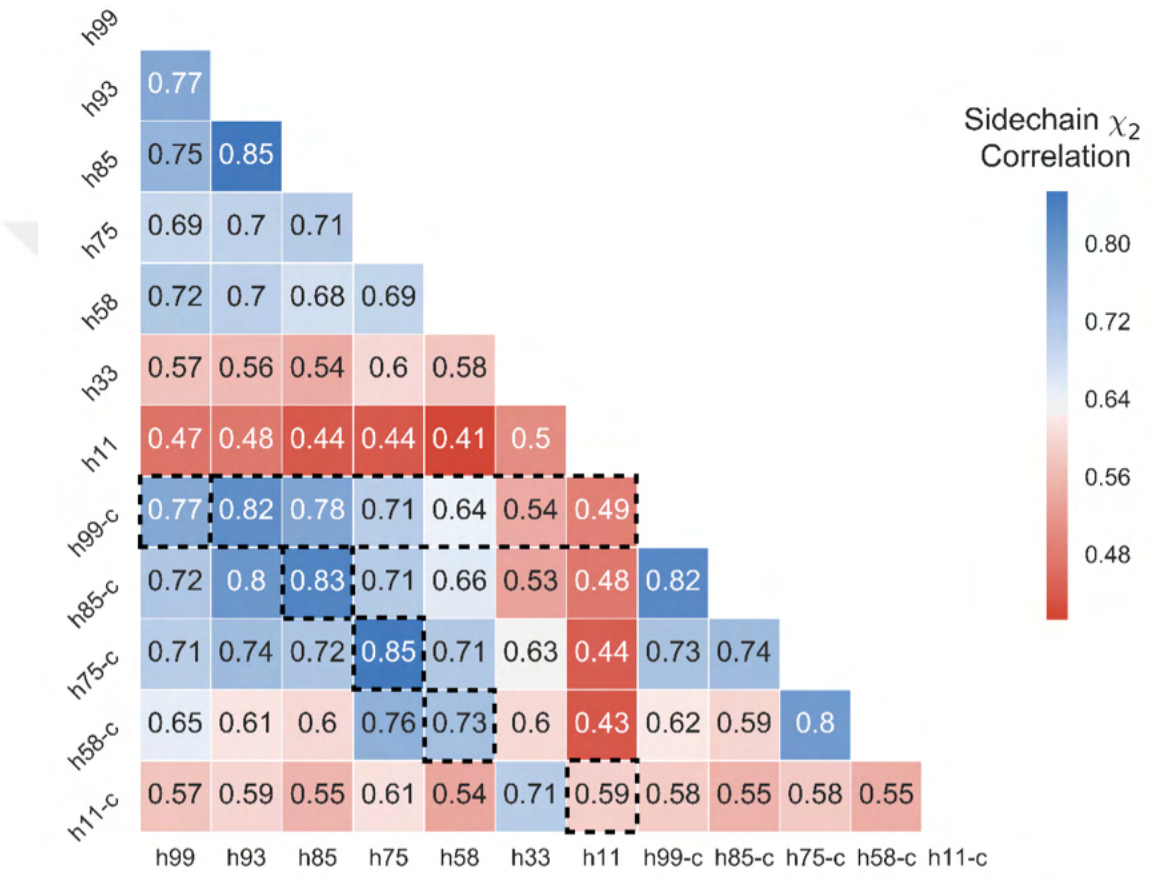

Figure B.4: Pairwise Pearson correlation coefficients for Ringer plots at χ_2, averaged over all 72 residues with χ_2 rotamers in lysozyme, for all possible pairs of dehydrated and cryocooled structures. Data sets collected at 100 K are labelled with "-c". The leftmost column indicates correlations with the native structure at room temperature; the diagonal boxed entries indicate correlations between room temperature and 100 K structures for crystals dehydrated to the same r.h., and the horizontal rectangle highlights correlation between the 100 K native structure and dehydrated structures at room temperature.

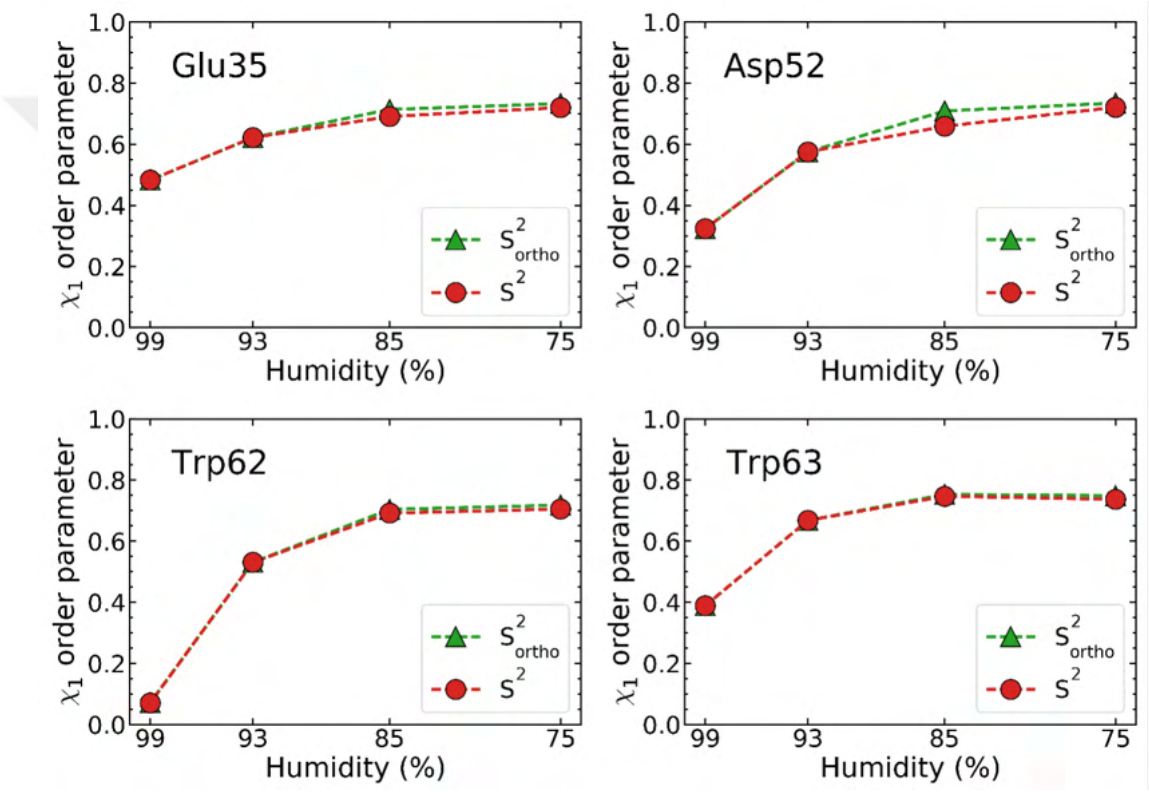

Figure B.5: Order parameter S^2 and its component S^2_{ortho} versus relative humidity, for a selection of lysozyme residues at χ_1, as deduced using multiconformer refinement from room temperature data sets. The same data plotted as $1 - S^2$ versus unit cell volume is shown in Figure 3.8.

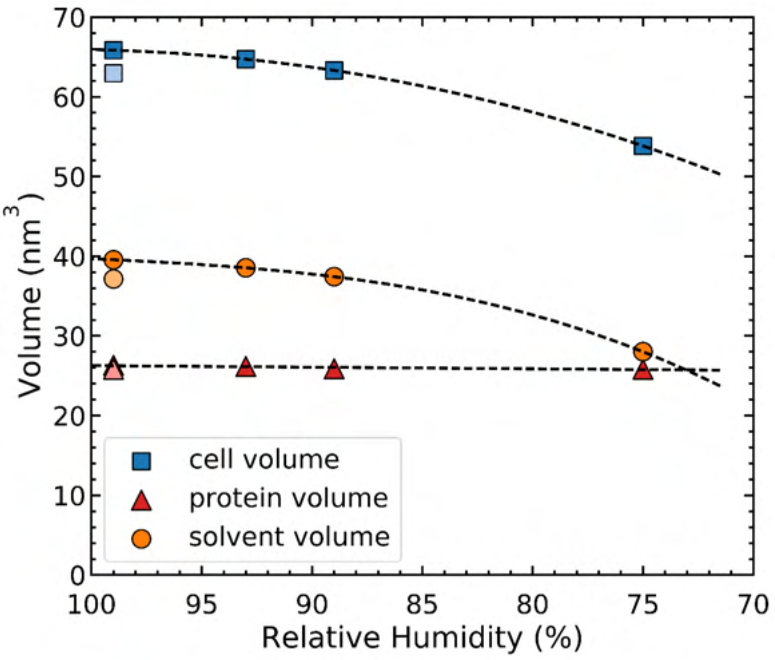

Figure B.6: Crystallographic cell volume (square), protein volume (triangle) (calculated as the volume enclosed by solvent excluded surface (SES) area of the protein molecule), and solvent volume (circle) (given by the difference between cell and protein volume) per protein molecule for tetragonal lysozyme crystals versus relative humidity. Dark and light symbols are for room temperature and $T = 100$ K values, respectively. The dashed lines are guides to the eye through the room temperature data points.

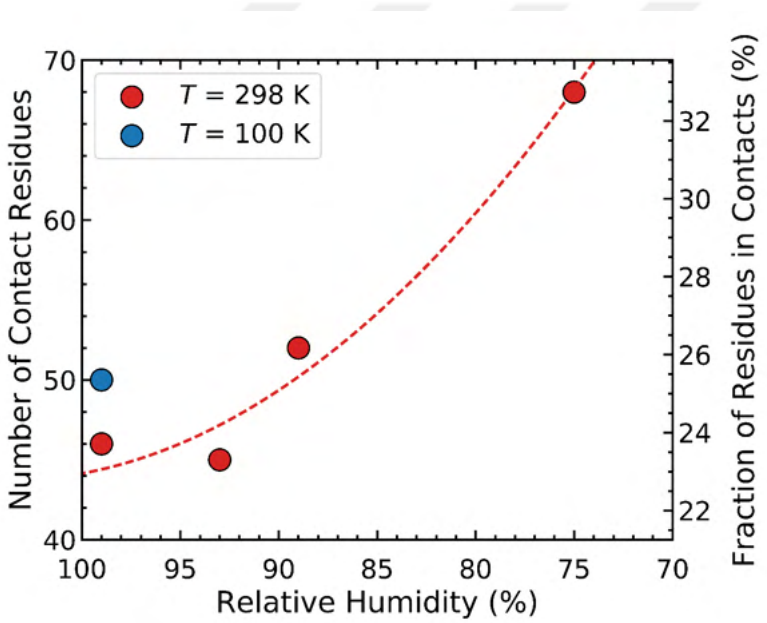

Figure B.7: Number of residues involved in crystal contacts versus relative humidity at room temperature (red) and $T = 100$ K (blue) for tetragonal thaumatin crystals. The dashed red line is a guide to the eye for the room temperature data.

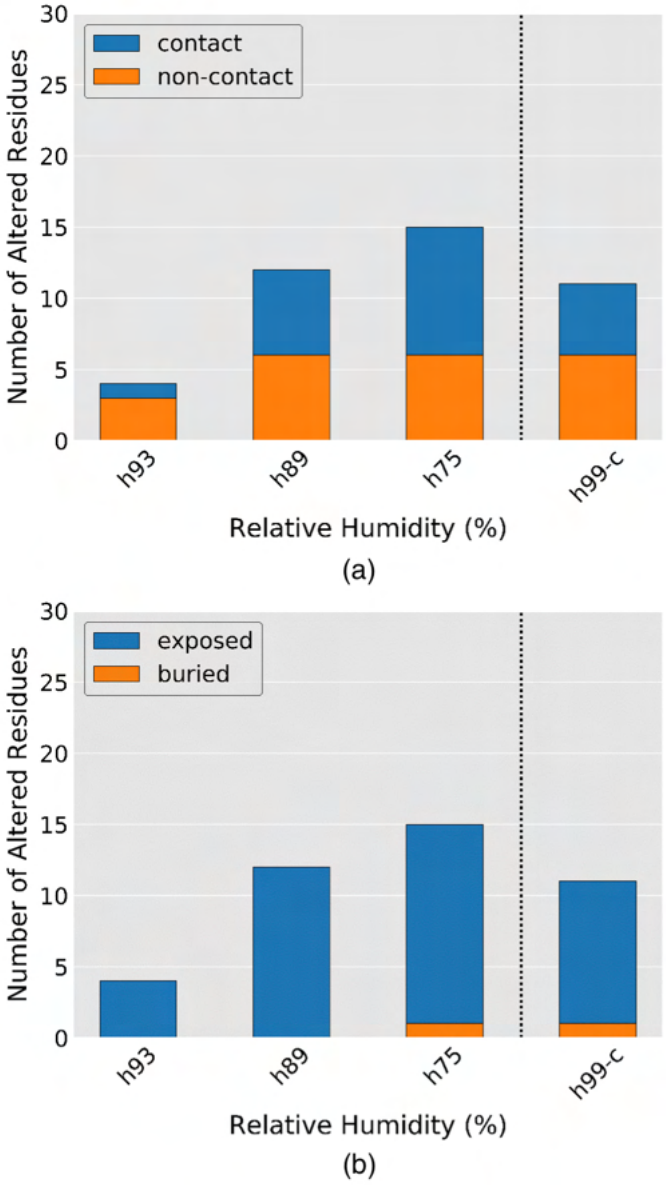

Figure B.8: Number of altered residues, of the 156 thaumatin residues having χ_1 rotamers, due to dehydration and cryocooling, relative to the native (99% r.h.) structure, categorized according to (a) whether or the not the residues were involved in crystal contacts in the final, non-native structure; and (b) whether the residues were solvent exposed or buried. Here, side chains were deemed to be altered if the Pearson correlation coefficient between Ringer curves for the native and non-native structures was less than 0.85.

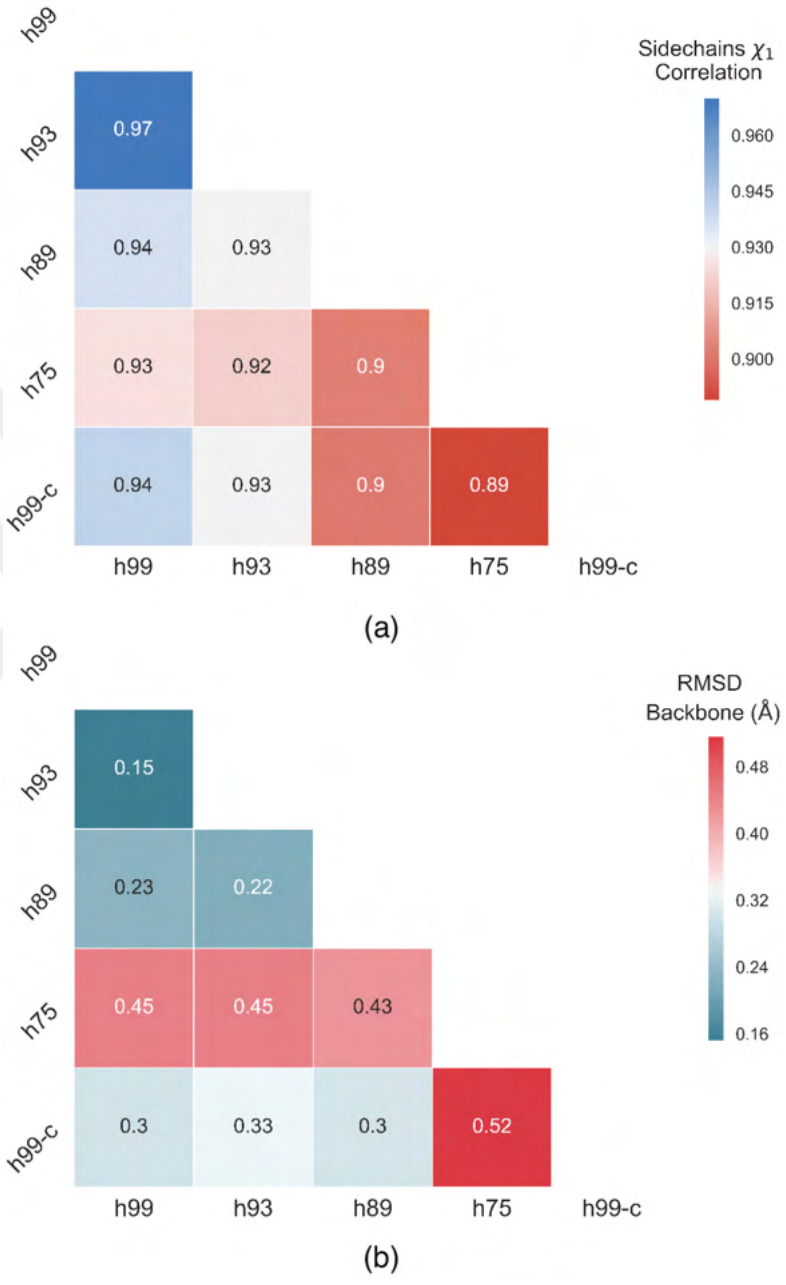

Figure B.9: a) Pairwise Pearson correlation coefficients for Ringer plots at as in Figure 3.4, averaged over all residues in thaumatin with χ_1 rotamers, for all possible pairs of dehydrated and cryo-cooled structures. Data sets collected at 100 K are labeled with "-c". (b) Backbone RMSD between all possible pairs of dehydrated and cryo-cooled structures. Only backbone atoms were used for alignment of the two structures and calculation of RMSD.

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

C.1 Fitting of experimental intensity vs resolution and dose data when the crystal orientation was unknown

When the crystal orientation was not given, we assumed that one of the three crystal axes was aligned with the beam direction, and performed three separate fits, assuming alignment of each axis. If the crystal dimensions along the axes were substantially (e.g., an order of magnitude) different, we used the orientation that gave the best fit; the other orientations gave substantially worse fits, regardless of the choice of α. When the crystal dimensions along the three axes were comparable (e.g., within a factor of 2), then the best-fit α would change with assumed orientation by at most ± 0.2, varying with the ratio of the beam FWHM to the crystal size.

C.2 Simulations of random blur model for radiation damage

C.2.1 2D simulations

For 2D crystal simulations, we used as a unit cell a grey-scale image with a resolution of 1024×512 pixels and with pixel values from 0 to 512 (and a maximum value in the image of 300). The pixel values correspond to the number of electrons per unit pixel area. Effects of initial static crystal disorder were modeled by randomly rotating the image in each unit cell by $\pm 0.5 - \pm 1^{\circ}$.

For each X-ray hit, a small $n \times n$ (typically 5×5) pixel interaction region centered at $\left(x_i, y_j\right)$ on the crystal is randomly selected and copied into a matrix. To account for the edges of this region, the matrix is extended to a $3n \times 3n$ matrix by reflecting the original matrix about each edge; e.g., in 1D, an original array "a b c d" becomes "d c b a | a b c d | d c b a". This is the default edge handling method of the scipy library of Python. A Gaussian spatial filter is applied to each of the pixels within the $n \times n$ interaction region of the extended matrix. The 2D Gaussian kernel applied to a pixel at $\left(x_i, y_j\right)$ is given by

$$G\left(x_{i'} - x_i, y_{j'} - y_j\right) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x_i - x_i)^2 + \left(y_{j'} - y_j\right)^2}{2\sigma^2}\right), \tag{C.1}$$

with σ between 0.5 and 1, the small σ values causing the Gaussian to decay to near zero in only a few pixels. The intensity of a pixel within the interaction region before and after the kernel is applied to that pixel are $\rho\left(x_i, y_j\right)$ and

$$\rho'\left(x_i, y_j\right) = \sum_{i',j'} G\left(x_{i'} - x_i, y_{i'} - y_i\right) \rho\left(x_{i'}, y_{j'}\right), \tag{C.2}$$

where the sum is over all pixels in the extended matrix. The pixel value in the original image matrix is updated after the calculation, while the extended matrix values are not changed. Fig. C.7 shows the effect of the Gaussian blur.

Simulations were performed using $m \times m$ arrays of unit cells with m ranging between 4 and 128 , and using unit cells with grid sizes ranging from 64×32 to 1024×512 pixels. Results with smaller unit cells became largely independent of m for $m \geq 8$. Consequently, $m = 32$ was used in most simulations, with $m = 16$ used for the largest unit cell. Simulations were run until the random hits caused the diffraction peaks in the highest resolution shell to fall below the background level, which corresponded to roughly 5–10 hits/pixel. Each simulation was divided into 30–50 segments, and after each segment the FFT of the crystal was calculated. The intensity of each diffraction peak is calculated by subtracting the squared local background pixel count from the squared peak count. The

diffraction peaks were binned into resolution shells defined by upper and lower resolutions. The integrated intensity of the diffraction peaks, normalized by the undamaged intensity, in each resolution shell is plotted against the average number of hits per pixel.

Fig. 4.5 illustrates the model using a two-dimensional crystal. Each unit cell contains a grey-scale image of a flea, and the grey-scale value of each pixel corresponds to the electron density there. The Fourier Transform of the initial, undamaged crystal has strong diffraction peaks extending out to the maximum q/resolution of the initial image. After some large number of hits, the real space electron density is blurred throughout the sample, and the Fourier transform decays much more rapidly with q. Movies C.9 and C.10 show the evolution of the crystal's electron density and the squared amplitude of its FFT, proportional to the diffracted intensity, with dose.

C.2.2 3D simulations

Computationally much more intensive simulations of radiation damage to 3D crystals were performed to assess whether the whether the shapes of the dose curves $I(q, D)$ and the exponent α in Eq. 4 depended on dimension. PDB entry 3E4H, for tetragonal crystals of the 29 residue plant protein Cyclotide varv F at 1.8 Å resolution was chosen for the simulation. The 3D electron density map of the asymmetric unit was calculated from the PDB file (with hetero atoms removed) using *phenix.fmodel*, which created an output in *.mtz* format. This map was discretized using double-precision floating-point format and between 16 and 128 points in each dimension (limited by the available memory for the simulation) by reading from the *.mtz* file using the Phenix tool *phenix.map_value_at_point*, and is shown in Fig. C.4. The simulations were performed using the *m*2.2*xlarge* cluster instance of Cornell Advanced Computing's Red Cloud,

which has 28 cores and 192 GB of RAM. This allowed simulation of crystals with $16 \times 16 \times 8$ unit cells when using 128^3 voxels per cell; $16 \times 16 \times 16$ unit cells were used for all other simulations. The largest simulations took about 7 hours.

C.2.3 Other Supplementary Figures & Tables

Reference	Protein	Crystal size (μm) & initial orientation	Beam shape	Beam h× v FWHM (μm)	λ(Å)	Data collection	Comments	Original dose calculation
Liebschner 2015	Thaumatin	99×145×120, c-axis aligned with beam	top-hat profile given	51 × 29	0.979	oscillation 2° frames	Only fully recorded reflections used, no LP correction	RADDOSE v2, no rotation
Sliz 2003	US2	100×100×60, orientation not given	gaussian	65 × 20		oscillation 1° frames		Not calculated, fluence given
	HLA	150×150×40, orientation not given	gaussian	65 × 20	1.033	oscillation 1° frames	Low and high dose data don't align	Not calculated, fluence given
	λ3	300×300×20, orientation not given	gaussian	65 × 20		oscillation 1° frames	Low and high dose data don't align	Not calculated, fluence given
Bourenkov 2009	FAE	size >> beam orientation not given	gaussian	40 × 30	0.99	rotation 70°, 0.25 and 0.5° frames		RADDOSE v1, no rotation
	Insulin	35 × 35 × 35 orientation not given	gaussian	40 × 30	0.97	rotation 300°, 1° frames		RADDOSE v1, no rotation
	P19-siRNA-1B	25 × 25 × 200 orientation not given	gaussian	40 × 30	0.99	rotation 28.8°, 0.8° frames		RADDOSE v1, no rotation
Teng 2000	Lysozyme	100×100×40, orientation not given	gaussian	295 × 270	1	oscillation 1° frames	Crystal rotated during long exposure, then frame acquired	Manual
Owen 2006	Holoferritin ("holo1")	200 × 200 × 200 orientation not given	top-hat	100 × 100	0.939	rotation 30 per set up to 12 sets	Overall $D_{1/2}$ and $D_{1/2}$ for 2 res. shells given for "holo1" crystal	RADDOSE v1, no rotation
Warkentin 2010	Thaumatin	size > beam, orientation not given	top-hat	100 × 100	0.984	oscillation 5° frames		Manual, no rotation

Table C.1: Experimental parameters for the radiation damage data analysed here.

Reference	Protein	Crystal shape & dimensions (μm)	Beam shape & dimensions FWHM h×v (μm)	Data simulated	Comments
Liebschner 2015	Thaumatin "xtal 01"	rectangular prism $99 \times 145 \times 120$ beam along c	rectangular top-hat 51×29	oscillation 2° frames	
			actual beam shape (See Fig. C.5)	oscillation 2° frames	
Sliz 2003	US2	rectangular prism 100×100×60 beam along short dim.	h: top-hat v: Gaussian 65×20	oscillation 1° frames	
	HLA	rectangular prism 150×150×40 beam along short dim.	h: top-hat v: gaussian 65×20	oscillation 1° frames	
	λ3	rectangular prism 300×300×20 beam along short dim.	h: top-hat v: gaussian 65×20	oscillation 1° frames	
Bourenkov 2009	FAE	rectangular prism 100×100×75 beam along short dim.	gaussian 40×30	rotation 70°, 0.25 and 0.5° frames	Assumed 30Å for lowest res. No scaling to account for change in illuminated volume applied.
	Insulin	rectangular prism 35×35×35	gaussian 40×30	rotation 300°, 1° frames	Assumed 40Å for lowest res. No scaling to account for change in illuminated volume applied.
	P19	rectangular prism $25 \times 25 \times 250$ beam $\perp$ long dim	gaussian 40×30	rotation 28.8°, 0.8° frames	Assumed 30Å for lowest res. No scaling to account for change in illuminated volume applied
Teng 2000	Lysozyme	rectangular prism 100×100×40	Gaussian 295×270	oscillation 1° frames	Continuously rotated during long irradiation, before recording diffraction pattern from same 1° wedge
Owen 2006	Holoferritin	rectangular prism 200×200×200	rectangular top-hat 100×100	oscillation 30° frames	Fit parameters adjusted so half doses reported for two resoluton shells are correctly predicted and so that the reported overall intensity versus dose curve is well fit.
Warkentin 2010	Thaumatin	rectangular prism 300×300×300	rectangular top-hat 100×100	oscillation 5° frames	

Table C.2: Simulation parameters for the radiation damage data analysed here.

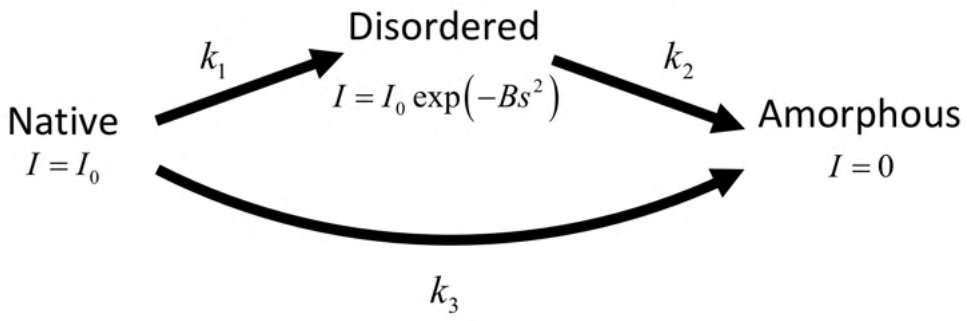

Figure C.1: The radiation-damage model of Blake & Phillips as modified by Hendrickson [143, 144]. An initially undamaged crystal region may become either disordered, corresponding to an increased average B-factor, or amorphous, corresponding to a complete loss of Bragg scattering. Undamaged crystal becomes disordered with a "rate constant" (fraction per unit dose) k_1 and this disordered crystal becomes fully amorphous with a rate constant k_2. Undamaged crystal can also directly proceed to the amorphous state at rate k_3. Previous fits to experimental data typically found $k_3 \approx 0$ [58, 140, 144].

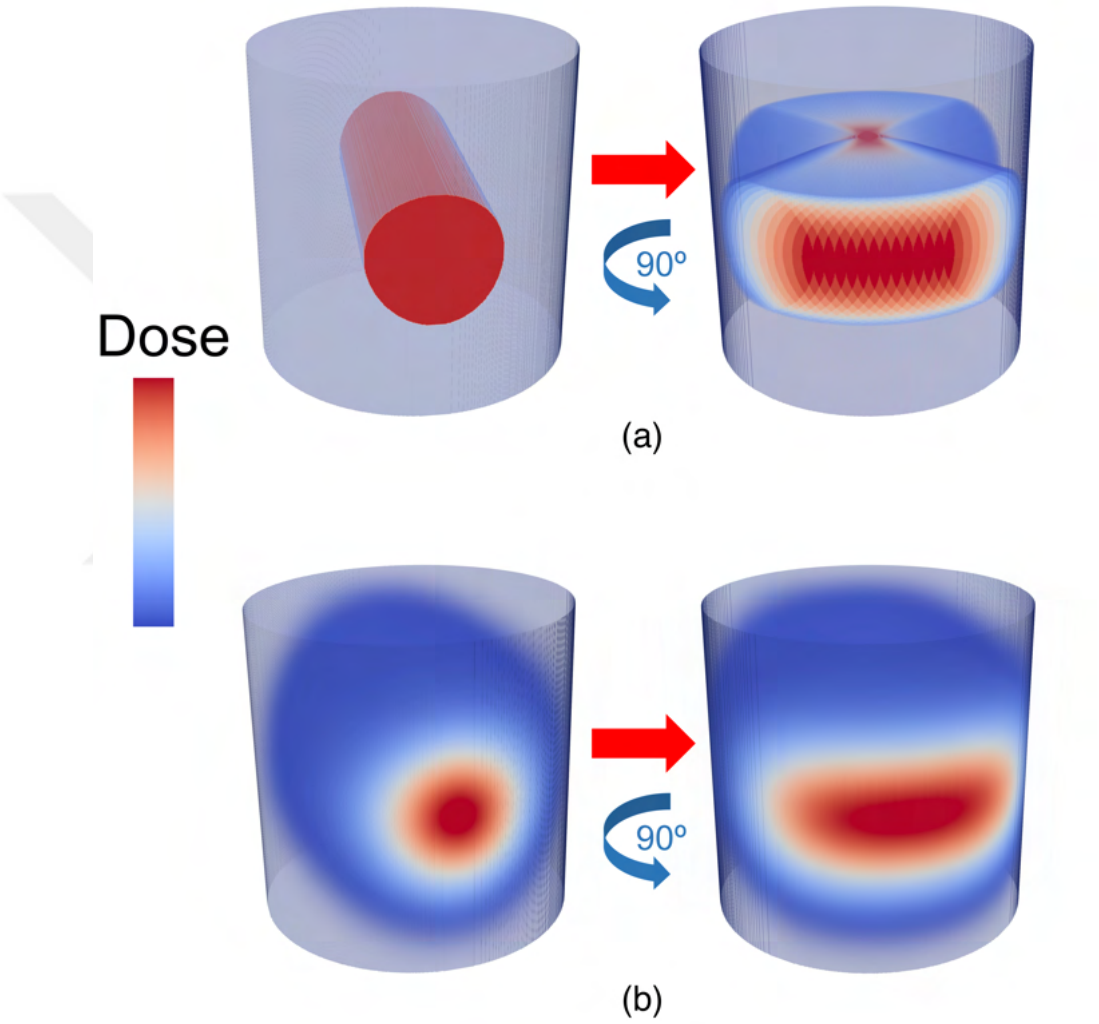

Figure C.2: Dose distribution received by a cylindrical crystal illuminated by an x-ray beam with (a) a flat-top flux profile and (b) an uncollimated Gaussian flux profile, when held in a fixed orientation (left) and when rotated by 90°. The crystal diameter and height are 100 μm and the FWHM of the Gaussian and the diameter of the flat-top beam are both 35 μm. Figures are prepared using the software ParaView.

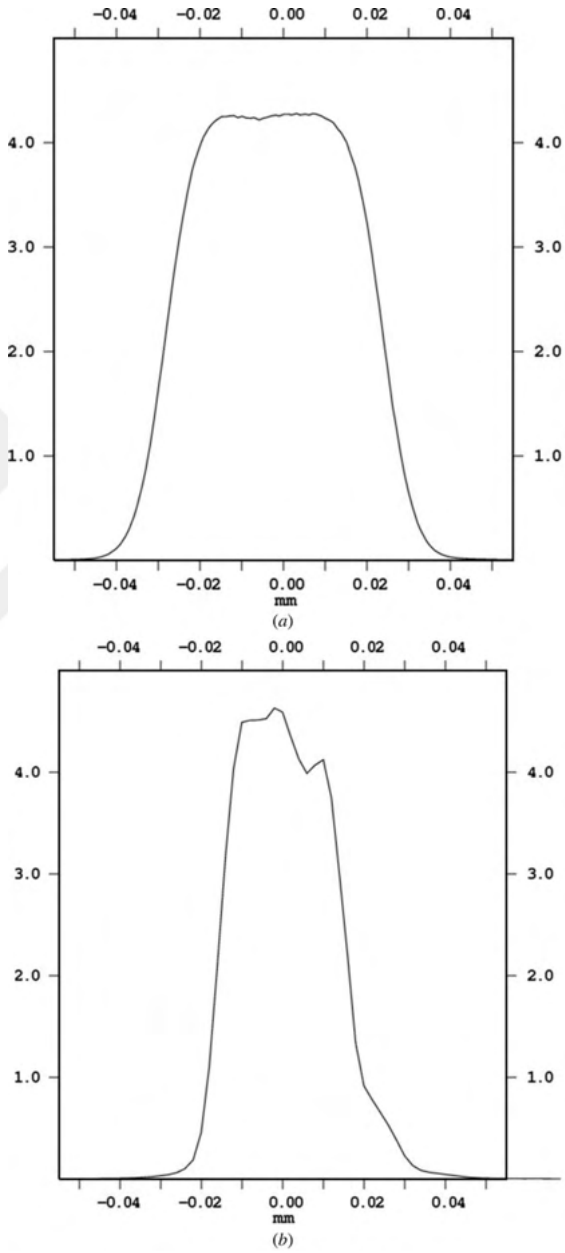

Figure C.3: Measured x-ray beam profile in horizontal (a) and vertical (b) reported by Liebschner et al. [142] in their Fig. 1, reproduced here for convenience.

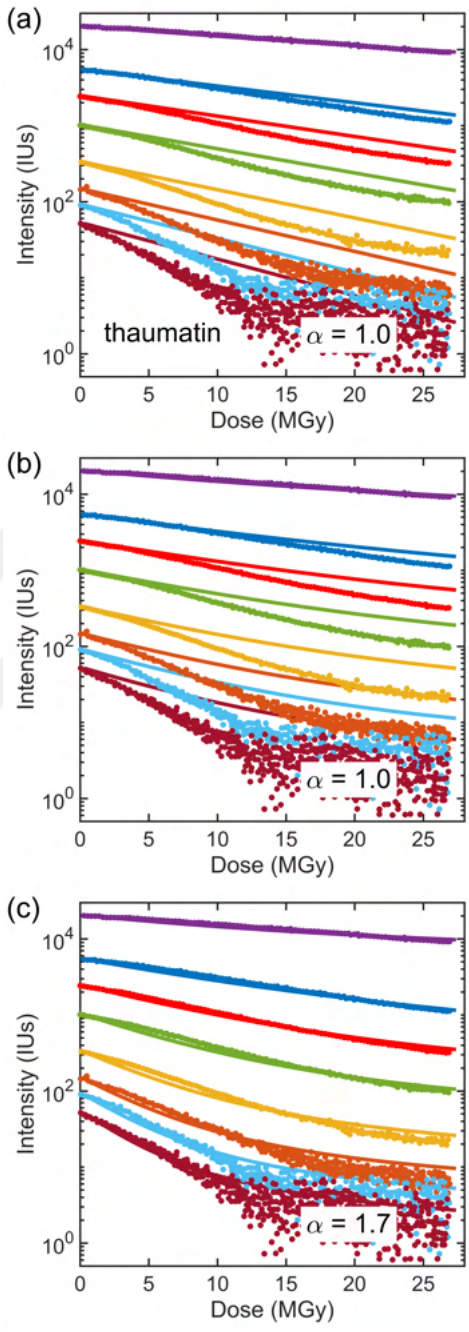

Figure C.4: Experimental data (solid circles) for integrated intensity in resolution shells versus dose for thaumatin crystals at 100 K, measured by Liebschner et al. (Fig. 4 in the original manuscript) [142]. Figure 4.2 shows the same data normalized in each resolution shell by the zero-dose intensity. The solid lines indicate results from simulations assuming (a) a perfect top-hat incident x-ray beam profile and an exponent $\alpha = 1$ in Eq. 4; (b) the measured beam profile (Fig. 1 in the original manuscript, reproduced in Fig. C.3) and $\alpha = 1$; and (c) the measured beam profile and a best-fit exponent $\alpha = 1.7$.

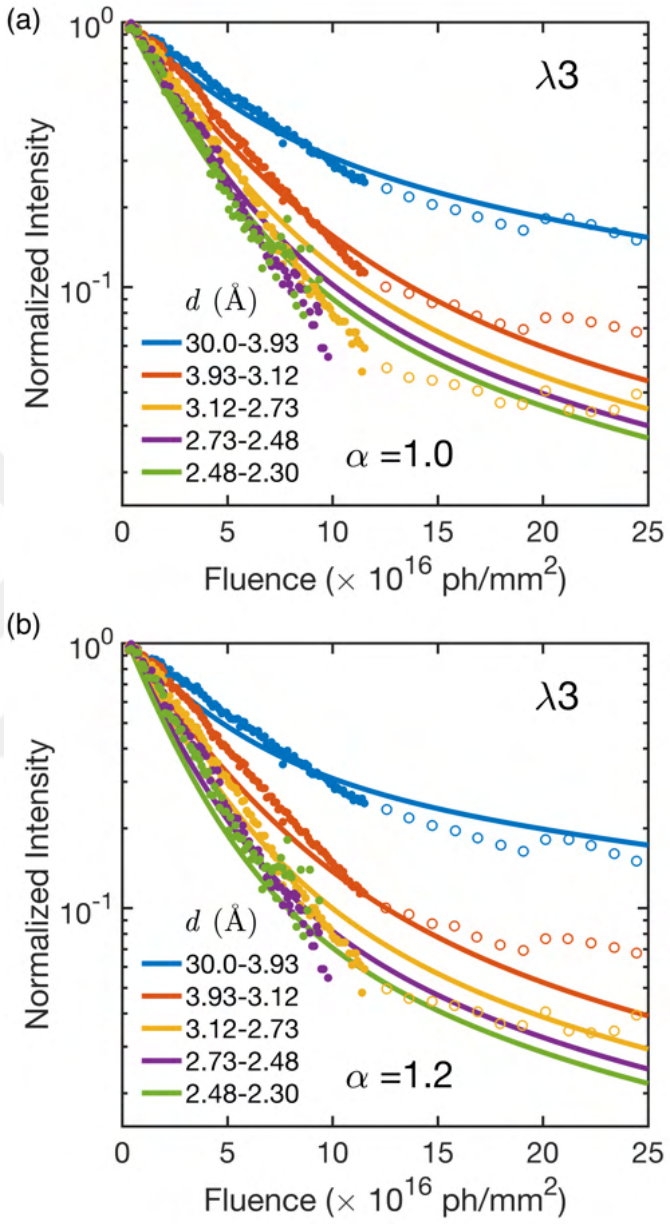

Figure C.5: Experimental data in the low dose (solid circles) and high dose (open circles) regions for integrated intensity in resolution shells versus incident fluence (in ph/mm^2, proportional to dose) reported by Sliz et al. [140] (Fig. 1 in the original manuscript) for a crystals of $\lambda 3$ at 100 K. The solid lines indicate results from simulations assuming a top-hat incident beam profile in the horizontal direction and a Gaussian profile in the vertical direction (based on descriptions of the experimental setup), with (a) $\alpha = 1$ and (b) a "best-fit" value chosen based on visual comparison. The discontinuity in slope between the low and high dose data suggests issues in data collection or analysis.

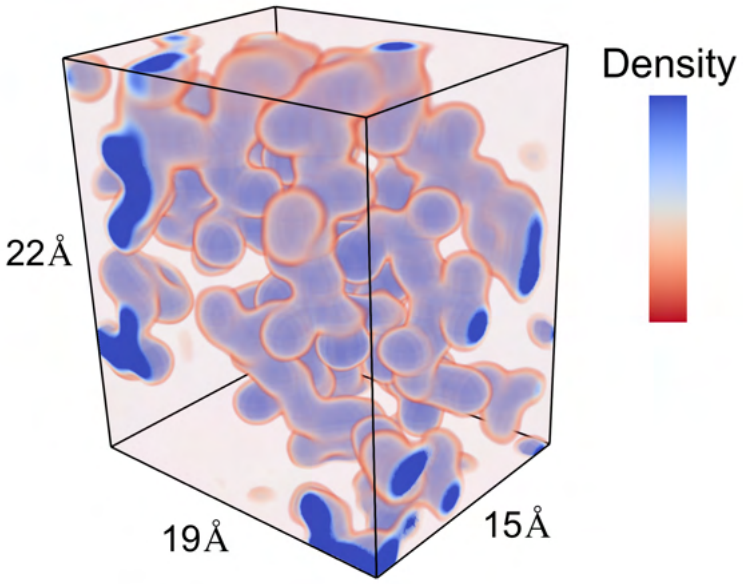

Figure C.6: Discretized electron density map based on PDB entry 3E4H using 128 grid points in each dimension. Shown is the smallest rectangular prism enclosing the protein chain in the asymmetric unit. This was used as the unit cell for simulations. The actual crystal of 3E4H has a cubic unit cell of size 82 Å and contains 48 copies of the ASU, and was too large for our simulations.

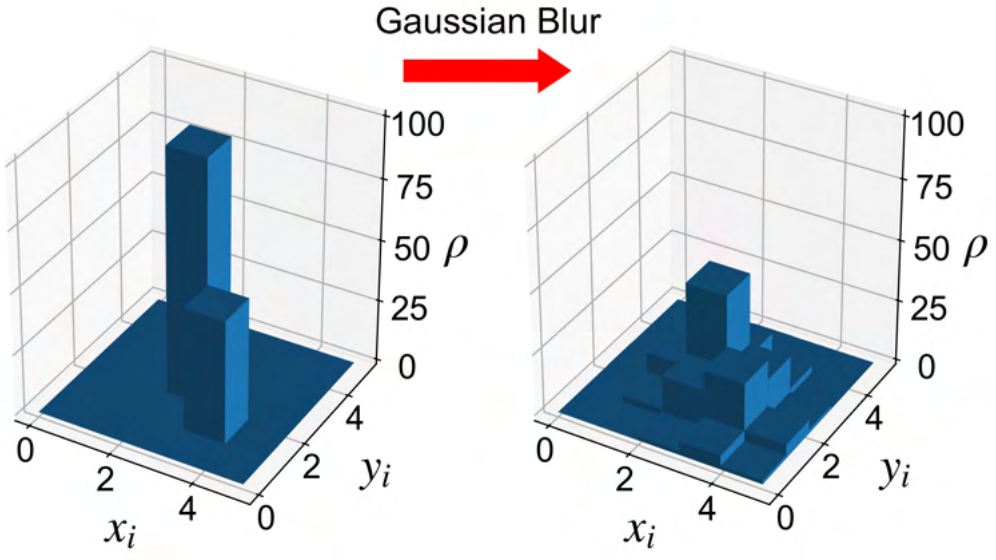

Figure C.7: Gaussian blur with $\sigma = 0.6$ applied to a 5×5 pixel region, with edge handling performed as described in Section C.2. Pixel counts correspond to electrons/pixel or electron density. The sum of pixel counts - corresponding to the total number of electrons - is conserved in the blurring. The Gaussian function vanishes beyond $\pm 3\sigma$ from the central pixel.

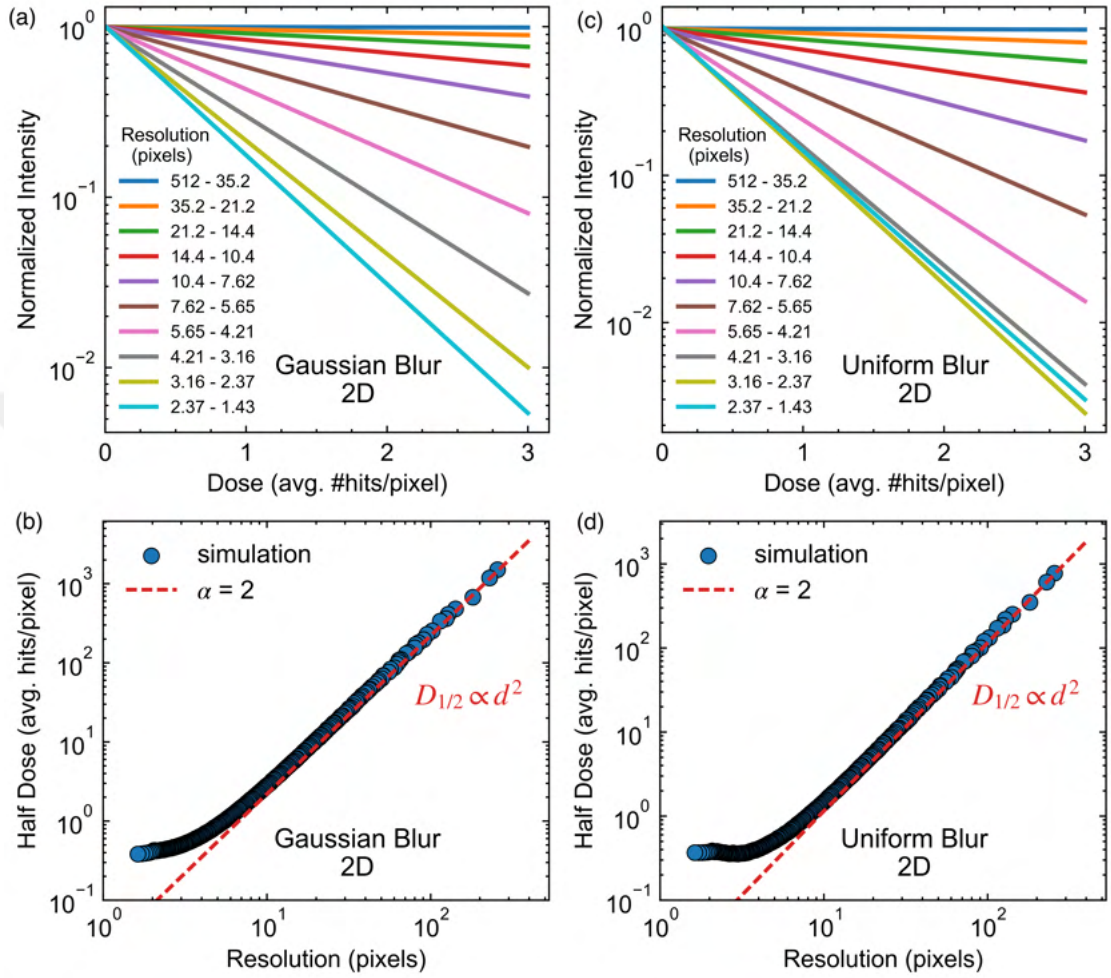

Figure C.8: Results of 2D simulations of radiation damage as described in Section C.2 using (a), (b) Gaussian and (c), (d) uniform blur kernels, for a uniformly irradiated sample. (a) and (c) show normalized integrated intensity in resolution shells versus dose. (b) and (d) show half dose $D_{1/2}(d)$ in a resolution shell vs resolution d. Both kernels yield purely exponential intensity decays with dose in each resolution shell, and the same exponent α (within errors) for $D_{1/2}(d) \propto d^{\alpha}$.

Figure C.9: Video (https://doi.org//10.1107/S2052252519008777/jt5036sup2.mp4) showing the evolution of the electron density of the 2D crystal in Fig. 4.5 as the number of "hits" increases, generated using our model of radiation damage as a series of Gaussian blurs applied at random locations. Audio as in Video C.10.

Figure C.10: Video (https://doi.org//10.1107/S2052252519008777/jt5036sup3.mp4) showing the evolution of the square of the amplitude of the FFT of the electron density shown in Video C.9, corresponding to the evolution of the crystal's diffraction pattern. The audio was generated by mapping the magnitude of the q vector of each FFT peak onto a frequency between 50 Hz (for $q = 0$) and 20 kHz, and generating tones for each with an amplitude proportional to the square of the FFT peak amplitude. The large q peaks / high frequency tones disappear rapidly, while the small q peaks / low frequency tones fade out very slowly.

BIBLIOGRAPHY

[1] J. S. Fraser, H. van den Bedem, A. J. Samelson, P. T. Lang, J. M. Holton, N. Echols, and T. Alber, "Accessing protein conformational ensembles using room-temperature x-ray crystallography," *Proceedings of the National Academy of Sciences*, vol. 108, no. 39, pp. 16 247–16 252, 2011.

[2] B. Halle, "Biomolecular cryocrystallography: structural changes during flash-cooling," *Proceedings of the National Academy of Sciences*, vol. 101, no. 14, pp. 4793–4798, 2004.

[3] D. A. Keedy, H. Van Den Bedem, D. A. Sivak, G. A. Petsko, D. Ringe, M. A. Wilson, and J. S. Fraser, "Crystal cryocooling distorts conformational heterogeneity in a model michaelis complex of dhfr," *Structure*, vol. 22, no. 6, pp. 899–910, 2014.

[4] E. Pozharski, C. X. Weichenberger, and B. Rupp, "Techniques, tools and best practices for ligand electron-density analysis and results from their application to deposited crystal structures," *Acta Crystallographica Section D: Biological Crystallography*, vol. 69, no. 2, pp. 150–167, 2013.

[5] J. A. Bell, "X-ray crystal structures of a severely desiccated protein," *Protein Science*, vol. 8, no. 10, pp. 2033–2040, 1999.

[6] R. Kishan, N. R. Chandra, C. Sudarsanakumar, K. Suguna, and M. Vijayan, "Water-dependent domain motion and flexibility in ribonuclease a and the invariant features in its hydration shell. an x-ray study of two low-humidity crystal forms of the enzyme," *Acta Crystallographica Section D: Biological Crystallography*, vol. 51, no. 5, pp. 703–710, 1995.

[7] S. Klingl, M. Scherer, T. Stamminger, and Y. A. Muller, "Controlled crystal dehydration triggers a space-group switch and shapes the tertiary structure of cytomegalovirus immediate-early 1 (ie1) protein," *Acta Crystallographica Section D: Biological Crystallography*, vol. 71, no. 7, pp. 1493–1504, 2015.

[8] M. Warkentin, J. B. Hopkins, J. B. Haber, G. Blaha, and R. E. Thorne, "Temperature-dependent radiation sensitivity and order of 70s ribosome crystals," *Acta Crystallographica Section D: Biological Crystallography*, vol. 70, no. 11, pp. 2890–2896, 2014.

[9] T. y. Teng and K. Moffat, "Primary radiation damage of protein crystals by an intense synchrotron x-ray beam," *Journal of synchrotron radiation*, vol. 7, no. 5, pp. 313–317, 2000.

[10] C. Nave and M. A. Hill, "Will reduced radiation damage occur with very small crystals?" *Journal of synchrotron radiation*, vol. 12, no. 3, pp. 299–303, 2005.

[11] R. M. F. Leal, G. Bourenkov, S. Russi, and A. N. Popov, "A survey of global radiation damage to 15 different protein crystal types at room temperature: A new decay model," *Journal of synchrotron radiation*, vol. 20, no. 1, pp. 14–22, 2013.

[12] B. Rupp, *Biomolecular crystallography: principles, practice, and application to structural biology*. Garland Science, 2009.

[13] J. Pflugrath, "Practical macromolecular cryocrystallography," *Acta Crystallographica Section F: Structural Biology Communications*, vol. 71, no. 6, pp. 622–642, 2015.

[14] D. H. Juers and B. W. Matthews, "Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions," *Journal of molecular biology*, vol. 311, no. 4, pp. 851–862, 2001.

[15] J. Pflugrath, "Macromolecular cryocrystallography—methods for cooling and mounting protein crystals at cryogenic temperatures," *Methods*, vol. 34, no. 3, pp. 415–423, 2004.

[16] S. Kriminski, C. Caylor, M. Nonato, K. Finkelstein, and R. Thorne, "Flash-cooling and annealing of protein crystals," *Acta Crystallographica Section D: Biological Crystallography*, vol. 58, no. 3, pp. 459–471, 2002.

[17] D. H. Juers and B. W. Matthews, "Cryo-cooling in macromolecular crystallography: advantages, disadvantages and optimization," *Quarterly reviews of biophysics*, vol. 37, no. 2, pp. 105–119, 2004.

[18] C. Farley and D. H. Juers, "Efficient cryoprotection of macromolecular crystals using vapor diffusion of volatile alcohols," *Journal of structural biology*, vol. 188, no. 2, pp. 102–106, 2014.

[19] H. Van Den Bedem, A. Dhanik, J.-C. Latombe, and A. M. Deacon, "Modeling discrete heterogeneity in x-ray diffraction data by fitting multi-conformers," *Acta Crystallographica Section D: Biological Crystallography*, vol. 65, no. 10, pp. 1107–1117, 2009.

[20] P. T. Lang, H.-L. Ng, J. S. Fraser, J. E. Corn, N. Echols, M. Sales, J. M. Holton, and T. Alber, "Automated electron-density sampling reveals widespread confor-

mational polymorphism in proteins," *Protein Science*, vol. 19, no. 7, pp. 1420–1431, 2010.

[21] R. B. Fenwick, H. van den Bedem, J. S. Fraser, and P. E. Wright, "Integrated description of protein dynamics from room-temperature x-ray crystallography and nmr," *Proceedings of the National Academy of Sciences*, vol. 111, no. 4, pp. E445–E454, 2014.

[22] D. A. Keedy, L. R. Kenner, M. Warkentin, R. A. Woldeyes, J. B. Hopkins, M. C. Thompson, A. S. Brewster, A. H. Van Benschoten, E. L. Baxter, M. Uervirojnangkoorn *et al.*, "Mapping the conformational landscape of a dynamic enzyme by multitemperature and xfel crystallography," *Elife*, vol. 4, p. e07574, 2015.

[23] H. Frauenfelder, G. A. Petsko, and D. Tsernoglou, "Temperature-dependent x-ray diffraction as a probe of protein structural dynamics," *Nature*, vol. 280, no. 5723, pp. 558–563, 1979.

[24] B. F. Rasmussen, A. M. Stock, D. Ringe, and G. A. Petsko, "Crystalline ribonuclease a loses function below the dynamical transition at 220 k," *nature*, vol. 357, no. 6377, pp. 423–424, 1992.

[25] R. F. Tilton Jr, J. C. Dewan, and G. A. Petsko, "Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-a at nine different temperatures from 98 to 320k," *Biochemistry*, vol. 31, no. 9, pp. 2469–2481, 1992.

[26] D. A. Keedy, Z. B. Hill, J. T. Biel, E. Kang, T. J. Rettenmaier, J. Brandao-Neto, N. M. Pearce, F. von Delft, J. A. Wells, and J. S. Fraser, "An expanded allosteric network in ptp1b by multitemperature crystallography, fragment screening, and covalent tethering," *Elife*, vol. 7, p. e36307, 2018.

[27] K. Henzler-Wildman and D. Kern, "Dynamic personalities of proteins," *Nature*, vol. 450, no. 7172, pp. 964–972, 2007.

[28] S. Khodadadi and A. Sokolov, "Protein dynamics: from rattling in a cage to structural relaxation," *Soft Matter*, vol. 11, no. 25, pp. 4984–4998, 2015.

[29] G. P. Lisi and J. P. Loria, "Using nmr spectroscopy to elucidate the role of molecular motions in enzyme function," *Progress in nuclear magnetic resonance spectroscopy*, vol. 92, pp. 1–17, 2016.

[30] M. Grimaldo, F. Roosen-Runge, F. Zhang, F. Schreiber, and T. Seydel, "Dynamics of proteins in solution," *Quarterly Reviews of Biophysics*, vol. 52, 2019.

[31] G. Bhabha, J. T. Biel, and J. S. Fraser, "Keep on moving: discovering and perturbing the conformational dynamics of enzymes," *Accounts of chemical research*, vol. 48, no. 2, pp. 423–430, 2015.

[32] M. Warkentin and R. E. Thorne, "Slow cooling of protein crystals," *Journal of applied crystallography*, vol. 42, no. 5, pp. 944–952, 2009.

[33] D. W. Moreau, H. Atakisi, and R. E. Thorne, "Solvent flows, conformation changes and lattice reordering in a cold protein crystal," *Acta Crystallographica Section D: Structural Biology*, vol. 75, no. 11, pp. 980–994, 2019.

[34] B. P. Callahan, Y. Yuan, and R. Wolfenden, "The burden borne by urease," *Journal of the American Chemical Society*, vol. 127, no. 31, pp. 10 828–10 829, 2005.

[35] J. B. Sumner, "The isolation and crystallization of the enzyme urease. preliminary paper," in *A Source Book in Chemistry, 1900-1950*. Harvard University Press, 2013, pp. 322–327.

[36] N. E. Dixon, C. Gazzola, R. L. Blakeley, and B. Zerner, "Jack bean urease (ec 3.5. 1.5). metalloenzyme. simple biological role for nickel," *Journal of the American Chemical Society*, vol. 97, no. 14, pp. 4131–4133, 1975.

[37] F. Paulot and D. J. Jacob, "Hidden cost of us agricultural exports: particulate matter from ammonia emissions," *Environmental science & technology*, vol. 48, no. 2, pp. 903–908, 2014.

[38] E. Jabri and P. A. Karplus, "Structures of the klebsiella aerogenes urease apoenzyme and two active-site mutants," *Biochemistry*, vol. 35, no. 33, pp. 10 616–10 626, 1996.

[39] M. A. Pearson, I.-S. Park, R. A. Schaller, L. O. Michel, P. A. Karplus, and R. P. Hausinger, "Kinetic and structural characterization of urease active site variants," *Biochemistry*, vol. 39, no. 29, pp. 8575–8584, 2000.

[40] P. A. Karplus, M. A. Pearson, and R. P. Hausinger, "70 years of crystalline urease: what have we learned?" *Accounts of Chemical Research*, vol. 30, no. 8, pp. 330–337, 1997.

[41] S. Benini, W. R. Rypniewski, K. S. Wilson, S. Miletti, S. Ciurli, and S. Mangani,

"A new proposal for urease mechanism based on the crystal structures of the native and inhibited enzyme from bacillus pasteurii: why urea hydrolysis costs two nickels," *Structure*, vol. 7, no. 2, pp. 205–216, 1999.

[42] B. Krajewska, R. van Eldik, and M. Brindell, "Temperature-and pressure-dependent stopped-flow kinetic studies of jack bean urease. implications for the catalytic mechanism," *JBIC Journal of Biological Inorganic Chemistry*, vol. 17, no. 7, pp. 1123–1134, 2012.

[43] B. Krajewska, "A combined temperature-ph study of urease kinetics. assigning pka values to ionizable groups of the active site involved in the catalytic reaction," *Journal of Molecular Catalysis B: Enzymatic*, vol. 124, pp. 70–76, 2016.

[44] L. Mazzei, M. Cianci, S. Benini, and S. Ciurli, "The structure of the elusive urease–urea complex unveils the mechanism of a paradigmatic nickel-dependent enzyme," *Angewandte Chemie International Edition*, vol. 58, no. 22, pp. 7415–7419, 2019.

[45] L. Mazzei, F. Musiani, and S. Ciurli, "The structure-based reaction mechanism of urease, a nickel dependent enzyme: tale of a long debate," *JBIC Journal of Biological Inorganic Chemistry*, pp. 1–17, 2020.

[46] E. Jabri, M. B. Carr, R. P. Hausinger, and P. A. Karplus, "The crystal structure of urease from klebsiella aerogenes," *Science*, vol. 268, no. 5213, pp. 998–1004, 1995.

[47] D. H. Juers and J. Ruffin, "Map_channels: a computation tool to aid in the visualization and characterization of solvent channels in macromolecular crystals," *Journal of applied crystallography*, vol. 47, no. 6, pp. 2105–2108, 2014.

[48] D. W. Moreau, H. Atakisi, and R. E. Thorne, "Ice formation and solvent nanoconfinement in protein crystals," *IUCrJ*, vol. 6, no. 3, pp. 346–356, 2019.

[49] M. Warkentin, V. Berejnov, N. S. Husseini, and R. E. Thorne, "Hyperquenching for protein cryocrystallography," *Journal of applied crystallography*, vol. 39, no. 6, pp. 805–811, 2006.

[50] M. B. C. Moncrief, R. P. Hausinger, L. G. Hom, E. Jabri, and P. Andrew Karplus, "Urease activity in the crystalline state," *Protein Science*, vol. 4, no. 10, pp. 2234–2236, 1995.

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

[52] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve *et al.*, "Phenix: a comprehensive python-based system for macromolecular structure solution," *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 2, pp. 213–221, 2010.

[53] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, "Features and development of coot," *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 4, pp. 486–501, 2010.

[54] B. P. Roberts, B. R. Miller III, A. E. Roitberg, and K. M. Merz Jr, "Wide-open flaps are key to urease activity," *Journal of the American Chemical Society*, vol. 134, no. 24, pp. 9934–9937, 2012.

[55] M. S. Minkara, M. N. Ucisik, M. N. Weaver, and K. M. Merz Jr, "Molecular dynamics study of helicobacter pylori urease," *Journal of chemical theory and computation*, vol. 10, no. 5, pp. 1852–1862, 2014.

[56] N. M. Pearce, T. Krojer, A. R. Bradley, P. Collins, R. P. Nowak, R. Talon, B. D. Marsden, S. Kelm, J. Shi, C. M. Deane *et al.*, "A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density," *Nature communications*, vol. 8, no. 1, pp. 1–8, 2017.

[57] D. Ringe and G. A. Petsko, "The 'glass transition'in protein dynamics: what it is, why it occurs, and how to exploit it," *Biophysical chemistry*, vol. 105, no. 2-3, pp. 667–680, 2003.

[58] M. Warkentin and R. E. Thorne, "Glass transition in thaumatin crystals revealed through temperature-dependent radiation-sensitivity measurements," *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 10, pp. 1092–1100, 2010.

[59] M. Teeter, A. Yamano, B. Stec, and U. Mohanty, "On the nature of a glassy state of matter in a hydrated protein: relation to protein function," *Proceedings of the National Academy of Sciences*, vol. 98, no. 20, pp. 11 242–11 247, 2001.

[60] L. Mazzei, M. Cianci, S. Benini, and S. Ciurli, "The impact of ph on catalytically critical protein conformational changes: the case of the urease, a nickel enzyme," *Chemistry–A European Journal*, vol. 25, no. 52, pp. 12 145–12 158, 2019.

[61] L. L. Palese, "Oxygen-oxygen distances in protein-bound crystallographic water suggest the presence of protonated clusters," *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1864, no. 3, p. 129480, 2020.

[62] J. Lee and S.-H. Kim, "Water polygons in high-resolution protein crystal structures," *Protein Science*, vol. 18, no. 7, pp. 1370–1376, 2009.

[63] S. Ciurli, "Urease. recent insights in the role of nickel," *Nickel and its surprising impact in nature*, vol. 10, 2007.

[64] J. A. Rupley and G. Careri, "Protein hydration and function," *Advances in protein chemistry*, vol. 41, pp. 37–172, 1991.

[65] P. W. Fenimore, H. Frauenfelder, B. McMahon, and R. Young, "Bulk-solvent and hydration-shell fluctuations, similar to α-and β-fluctuations in glasses, control protein motions and functions," *Proceedings of the National Academy of Sciences*, vol. 101, no. 40, pp. 14 408–14 413, 2004.

[66] H. Frauenfelder, G. Chen, J. Berendzen, P. W. Fenimore, H. Jansson, B. H. McMahon, I. R. Stroe, J. Swenson, and R. D. Young, "A unified model of protein dynamics," *Proceedings of the National Academy of Sciences*, vol. 106, no. 13, pp. 5129–5134, 2009.

[67] D. Crowfoot and D. Riley, "Crystal structures of the proteins an x-ray study of palmar's lactoglobulin," *Nature*, vol. 141, no. 3568, pp. 521–522, 1938.

[68] M. Perutz, "The composition and swelling properties of haemoglobin crystals," *Transactions of the Faraday Society*, vol. 42, pp. B187–B195, 1946.

[69] M. F. Perutz, "The structure of haemoglobin iii. direct determination of the molecular transform," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 225, no. 1161, pp. 264–286, 1954.

[70] J. Einstein and B. Low, "Insulin. some shrinkage stages of sulfate and citrate crystals," *Acta Crystallographica*, vol. 15, no. 1, pp. 32–34, 1962.

[71] H. Huxley and J. Kendrew, "Discontinuous lattice changes in haemoglobin crystals," *Acta Crystallographica*, vol. 6, no. 1, pp. 76–80, 1953.

[72] K. A. Kantardjieff and B. Rupp, "Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, dna, and protein–nucleic acid complex crystals," *Protein Science*, vol. 12, no. 9, pp. 1865–1871, 2003.

[73] B. W. Matthews, "Solvent content of protein crystals," *Journal of molecular biology*, vol. 33, no. 2, pp. 491–497, 1968.

[74] M. W. Bowler, U. Mueller, M. S. Weiss, J. Sanchez-Weatherby, T. L. Sorensen, M. M. Thunnissen, T. Ursby, A. Gobbo, S. Russi, M. G. Bowler *et al.*, "Automation and experience of controlled crystal dehydration: results from the european synchrotron hc1 collaboration," *Crystal Growth & Design*, vol. 15, no. 3, pp. 1043–1054, 2015.

[75] J. Sanchez-Weatherby, M. W. Bowler, J. Huet, A. Gobbo, F. Felisaz, B. Lavault, R. Moya, J. Kadlec, R. B. Ravelli, and F. Cipriani, "Improving diffraction by humidity control: a novel device compatible with x-ray beamlines," *Acta Crystallographica Section D: Biological Crystallography*, vol. 65, no. 12, pp. 1237–1246, 2009.

[76] R. Kiefersauer, M. E. Than, H. Dobbek, L. Gremer, M. Melero, S. Strobl, J. a. M. Dias, T. Soulimane, and R. Huber, "A novel free-mounting system for protein crystals: transformation and improvement of diffraction power by accurately controlled humidity changes," *Journal of applied crystallography*, vol. 33, no. 5, pp. 1223–1230, 2000.

[77] I. Russo Krauss, F. Sica, C. A. Mattia, and A. Merlino, "Increasing the x-ray diffraction power of protein crystals by dehydration: the case of bovine serum albumin and a survey of literature data," *International journal of molecular sciences*, vol. 13, no. 3, pp. 3782–3800, 2012.

[78] D. Salunke, B. Veerapandian, R. Kodandapani, and M. Vijayan, "Water-mediated transformations in protein crystals," *Acta Crystallographica Section B: Structural Science*, vol. 41, no. 6, pp. 431–436, 1985.

[79] I. Dobrianov, S. Kriminski, C. Caylor, S. Lemay, C. Kimmer, A. Kisselev, K. Finkelstein, and R. Thorne, "Dynamic response of tetragonal lysozyme crystals to changes in relative humidity: implications for post-growth crystal treatments," *Acta Crystallographica Section D: Biological Crystallography*, vol. 57, no. 1, pp. 61–68, 2001.

[80] R. Kodandapani, C. G. Suresh, and M. Vijayan, "Crystal structure of low humidity tetragonal lysozyme at 2.1-a resolution. variability in hydration shell and its structural consequences." *Journal of Biological Chemistry*, vol. 265, no. 27, pp. 16 126–16 131, 1990.

[81] R. Kodandapani, M. Vijayan *et al.*, "Protein hydration and water structure: X-ray analysis of a closely packed protein crystal with very low solvent content,"

Acta Crystallographica Section D: Biological Crystallography, vol. 49, no. 2, pp. 234–245, 1993.

[82] H. Nagendra, N. Sukumar, and M. Vijayan, "Role of water in plasticity, stability, and action of proteins: the crystal structures of lysozyme at very low levels of hydration," *Proteins: Structure, Function, and Bioinformatics*, vol. 32, no. 2, pp. 229–240, 1998.

[83] K. Harata and T. Akiba, "Structural phase transition of monoclinic crystals of hen egg-white lysozyme," *Acta Crystallographica Section D: Biological Crystallography*, vol. 62, no. 4, pp. 375–382, 2006.

[84] M. Vendruscolo, "Determination of conformationally heterogeneous states of proteins," *Current opinion in structural biology*, vol. 17, no. 1, pp. 15–20, 2007.

[85] O. F. Lange, N.-A. Lakomek, C. Farès, G. F. Schröder, K. F. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, and B. L. De Groot, "Recognition dynamics up to microseconds revealed from an rdc-derived ubiquitin ensemble in solution," *science*, vol. 320, no. 5882, pp. 1471–1475, 2008.

[86] D. D. Boehr, D. McElheny, H. J. Dyson, and P. E. Wright, "The dynamic energy landscape of dihydrofolate reductase catalysis," *science*, vol. 313, no. 5793, pp. 1638–1642, 2006.

[87] K. Gunasekaran, B. Ma, and R. Nussinov, "Is allostery an intrinsic property of all dynamic proteins?" *Proteins: Structure, Function, and Bioinformatics*, vol. 57, no. 3, pp. 433–443, 2004.

[88] M. A. DePristo, P. I. de Bakker, and T. L. Blundell, "Heterogeneity and inaccuracy in protein structures solved by x-ray crystallography," *Structure*, vol. 12, no. 5, pp. 831–838, 2004.

[89] N. Furnham, T. L. Blundell, M. A. DePristo, and T. C. Terwilliger, "Is one solution good enough?" *Nature structural & molecular biology*, vol. 13, no. 3, pp. 184–185, 2006.

[90] P. T. Lang, J. M. Holton, J. S. Fraser, and T. Alber, "Protein structural ensembles are revealed by redefining x-ray electron density noise," *Proceedings of the National Academy of Sciences*, vol. 111, no. 1, pp. 237–242, 2014.

[91] H. Van Den Bedem and J. S. Fraser, "Integrative, dynamic structural biology at

atomic resolution—it's about time," *Nature methods*, vol. 12, no. 4, pp. 307–318, 2015.

[92] R. Brueschweiler and P. E. Wright, "Nmr order parameters of biomolecules: a new analytical representation and application to the gaussian axial fluctuation model," *Journal of the American Chemical Society*, vol. 116, no. 18, pp. 8426–8427, 1994.

[93] S. Russi, A. González, L. R. Kenner, D. A. Keedy, J. S. Fraser, and H. v. d. Bedem, "Conformational variation of proteins at room temperature is not dominated by radiation damage," *Journal of synchrotron radiation*, vol. 24, no. 1, pp. 73–82, 2017.

[94] C. Farley, G. Burks, T. Siegert, and D. H. Juers, "Improved reproducibility of unit-cell parameters in macromolecular cryocrystallography by limiting dehydration during crystal mounting," *Acta Crystallographica Section D: Biological Crystallography*, vol. 70, no. 8, pp. 2111–2124, 2014.

[95] L. B. Rockland, "Saturated salt solutions for static control of relative humidity between 5° and 40° c." *Analytical Chemistry*, vol. 32, no. 10, pp. 1375–1376, 1960.

[96] J. F. Young, "Humidity control in the laboratory using salt solutions—a review," *Journal of Applied Chemistry*, vol. 17, no. 9, pp. 241–245, 1967.

[97] L. Greenspan *et al.*, "Humidity fixed points of binary saturated aqueous solutions," *Journal of research of the national bureau of standards*, vol. 81, no. 1, pp. 89–96, 1977.

[98] H. Nagendra, C. Sudarsanakumar, and M. Vijayan, "Characterization of lysozyme crystals with unusually low solvent content," *Acta Crystallographica Section D: Biological Crystallography*, vol. 51, no. 3, pp. 390–392, 1995.

[99] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson, "Molprobity: all-atom structure validation for macromolecular crystallography," *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 1, pp. 12–21, 2010.

[100] N. R. Voss and M. Gerstein, "3v: cavity, channel and cleft volume calculator and extractor," *Nucleic acids research*, vol. 38, no. suppl_2, pp. W555–W562, 2010.

[101] M. L. Quillin and B. W. Matthews, "Accurate calculation of the density of pro-

teins," *Acta Crystallographica Section D: Biological Crystallography*, vol. 56, no. 7, pp. 791–794, 2000.

[102] H. Fischer, I. Polikarpov, and A. F. Craievich, "Average protein density is a molecular-weight-dependent function," *Protein Science*, vol. 13, no. 10, pp. 2825–2828, 2004.

[103] J. A. Rupley, E. Gratton, and G. Careri, "Water and globular proteins," *Trends in Biochemical Sciences*, vol. 8, no. 1, pp. 18–22, 1983.

[104] Madhusudan and M. Vijayan, "Rigid and flexible regions in lysozyme and the invariant features in its hydration shell," *Current science*, pp. 165–170, 1991.

[105] M. Nakasako, "Water–protein interactions from high–resolution protein crystallography," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 359, no. 1448, pp. 1191–1206, 2004.

[106] M. Nakasako, M. Odaka, M. Yohda, N. Dohmae, K. Takio, N. Kamiya, and I. Endo, "Tertiary and quaternary structures of photoreactive fe-type nitrile hydratase from rhodococcus sp. n-771: roles of hydration water molecules in stabilizing the structures and the structural origin of the substrate specificity of the enzyme," *Biochemistry*, vol. 38, no. 31, pp. 9887–9898, 1999.

[107] C. Poornima and P. Dean, "Hydration in drug design. 3. conserved water molecules at the ligand-binding sites of homologous proteins," *Journal of computer-aided molecular design*, vol. 9, no. 6, pp. 521–531, 1995.

[108] S. Shaltiel, S. Cox, and S. S. Taylor, "Conserved water molecules contribute to the extensive network of interactions at the active site of protein kinase a," *Proceedings of the National Academy of Sciences*, vol. 95, no. 2, pp. 484–491, 1998.

[109] X.-J. Zhang and B. Matthews, "Conservation of solvent-binding sites in 10 crystal forms of t4 lysozyme," *Protein Science*, vol. 3, no. 7, pp. 1031–1039, 1994.

[110] S. Dasgupta, G. H. Iyer, S. H. Bryant, C. E. Lawrence, and J. A. Bell, "Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers," *Proteins: Structure, Function, and Bioinformatics*, vol. 28, no. 4, pp. 494–514, 1997.

[111] D. Lorimer, A. Raymond, J. Walchli, M. Mixon, A. Barrow, E. Wallace, R. Grice, A. Burgin, and L. Stewart, "Gene composer: database software for protein

construct design, codon engineering, and gene synthesis," *BMC biotechnology*, vol. 9, no. 1, pp. 1–22, 2009.

[112] O. Carugo and P. Argos, "Protein—protein crystal-packing contacts," *Protein science*, vol. 6, no. 10, pp. 2261–2263, 1997.

[113] H. Zhang, T. Zhang, K. Chen, S. Shen, J. Ruan, and L. Kurgan, "On the relation between residue flexibility and local solvent accessibility in proteins," *Proteins: Structure, Function, and Bioinformatics*, vol. 76, no. 3, pp. 617–636, 2009.

[114] S. Miller, J. Janin, A. M. Lesk, and C. Chothia, "Interior and surface of monomeric proteins," *Journal of molecular biology*, vol. 196, no. 3, pp. 641–656, 1987.

[115] H. Chen and H.-X. Zhou, "Prediction of solvent accessibility and sites of deleterious mutations from protein sequence," *Nucleic acids research*, vol. 33, no. 10, pp. 3193–3199, 2005.

[116] J. A. Kelly, A. R. Sielecki, B. D. Sykes, M. N. James, and D. C. Phillips, "X-ray crystallography of the binding of the bacterial cell wall trisaccharide nam-nag-nam to lysozyme," *Nature*, vol. 282, no. 5741, pp. 875–878, 1979.

[117] L. N. Johnson and D. Phillips, "Structure of some crystalline lysozyme-inhibitor complexes determined by x-ray analysis at 6 å resolution," *Nature*, vol. 206, no. 4986, pp. 761–763, 1965.

[118] H. Nagendra, C. Sudarsanakumar, and M. Vijayan, "An x-ray analysis of native monoclinic lysozyme. a case study on the reliability of refined protein structures and a comparison with the low-humidity form in relation to mobility and enzyme action," *Acta Crystallographica Section D: Biological Crystallography*, vol. 52, no. 6, pp. 1067–1074, 1996.

[119] M. D. Collins, C. U. Kim, and S. M. Gruner, "High-pressure protein crystallography and nmr to explore protein conformations," *Annual review of biophysics*, vol. 40, pp. 81–98, 2011.

[120] R. Sanishvili, D. W. Yoder, S. B. Pothineni, G. Rosenbaum, S. Xu, S. Vogt, S. Stepanov, O. A. Makarov, S. Corcoran, R. Benn *et al.*, "Radiation damage in protein crystals is reduced with a micron-sized x-ray beam," *Proceedings of the National Academy of Sciences*, vol. 108, no. 15, pp. 6127–6132, 2011.

[121] Y. Z. Finfrock, E. A. Stern, R. Alkire, J. J. Kas, K. Evans-Lutterodt, A. Stein,

N. Duke, K. Lazarski, and A. Joachimiak, "Mitigation of x-ray damage in macromolecular crystallography by submicrometre line focusing," *Acta Crystallographica Section D: Biological Crystallography*, vol. 69, no. 8, pp. 1463–1469, 2013.

[122] J. M. Holton, "A beginner's guide to radiation damage," *Journal of synchrotron radiation*, vol. 16, no. 2, pp. 133–142, 2009.

[123] E. F. Garman, "Radiation damage in macromolecular crystallography: what is it and why should we care?" *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 4, pp. 339–351, 2010.

[124] M. Warkentin, J. B. Hopkins, R. Badeau, A. M. Mulichak, L. J. Keefe, and R. E. Thorne, "Global radiation damage: temperature dependence, time dependence and how to outrun it," *Journal of synchrotron radiation*, vol. 20, no. 1, pp. 7–13, 2013.

[125] W. H. Massover, "Radiation damage to protein specimens from electron beam imaging and diffraction: a mini-review of anti-damage approaches, with special reference to synchrotron x-ray crystallography," *Journal of synchrotron radiation*, vol. 14, no. 1, pp. 116–127, 2007.

[126] A. Meents, B. Dittrich, and S. Gutmann, "A new aspect of specific radiation damage: hydrogen abstraction from organic molecules," *Journal of synchrotron radiation*, vol. 16, no. 2, pp. 183–190, 2009.

[127] M. A. Warkentin, H. Atakisi, J. B. Hopkins, D. Walko, and R. E. Thorne, "Lifetimes and spatio-temporal response of protein crystals in intense x-ray microbeams," *IUCrJ*, vol. 4, no. 6, pp. 785–794, 2017.

[128] R. B. Ravelli and S. M. McSweeney, "The 'fingerprint'that x-rays can leave on structures," *Structure*, vol. 8, no. 3, pp. 315–328, 2000.

[129] M. Weik, R. B. Ravelli, G. Kryger, S. McSweeney, M. L. Raves, M. Harel, P. Gros, I. Silman, J. Kroon, and J. L. Sussman, "Specific chemical and structural damage to proteins produced by synchrotron radiation," *Proceedings of the National Academy of Sciences*, vol. 97, no. 2, pp. 623–628, 2000.

[130] W. P. Burmeister, "Structural changes in a cryo-cooled protein crystal owing to radiation damage," *Acta Crystallographica Section D: Biological Crystallography*, vol. 56, no. 3, pp. 328–341, 2000.

[131] S. Banumathi, P. H. Zwart, U. A. Ramagopal, M. Dauter, and Z. Dauter, "Structural effects of radiation damage and its potential for phasing," *Acta Crystallographica Section D: Biological Crystallography*, vol. 60, no. 6, pp. 1085–1093, 2004.

[132] C. S. Bury, J. C. Brooks-Bartlett, S. P. Walsh, and E. F. Garman, "Estimate your dose: Raddose-3d," *Protein Science*, vol. 27, no. 1, pp. 217–228, 2018.

[133] P. Evans, "Scaling and assessment of data quality," *Acta Crystallographica Section D: Biological Crystallography*, vol. 62, no. 1, pp. 72–82, 2006.

[134] Z. Otwinowski, W. Minor, D. Borek, and M. Cymborowski, "International tables for crystallography, vol," 2006.

[135] K. Diederichs, S. McSweeney, and R. B. Ravelli, "Zero-dose extrapolation as part of macromolecular synchrotron data reduction," *Acta Crystallographica Section D: Biological Crystallography*, vol. 59, no. 5, pp. 903–909, 2003.

[136] K. Diederichs, "Some aspects of quantitative analysis and correction of radiation damage," *Acta Crystallographica Section D: Biological Crystallography*, vol. 62, no. 1, pp. 96–101, 2006.

[137] K. Hirata, K. Shinzawa-Itoh, N. Yano, S. Takemura, K. Kato, M. Hatanaka, K. Muramoto, T. Kawahara, T. Tsukihara, E. Yamashita *et al.*, "Determination of damage-free crystal structure of an x-ray–sensitive protein using an xfel," *Nature methods*, vol. 11, no. 7, pp. 734–736, 2014.

[138] M. R. Howells, T. Beetz, H. N. Chapman, C. Cui, J. Holton, C. Jacobsen, J. Kirz, E. Lima, S. Marchesini, H. Miao *et al.*, "An assessment of the resolution limitation due to radiation-damage in x-ray diffraction microscopy," *Journal of electron spectroscopy and related phenomena*, vol. 170, no. 1-3, pp. 4–12, 2009.

[139] J. Kmetko, N. S. Husseini, M. Naides, Y. Kalinin, and R. E. Thorne, "Quantifying x-ray radiation damage in protein crystals at cryogenic temperatures," *Acta Crystallographica Section D: Biological Crystallography*, vol. 62, no. 9, pp. 1030–1038, 2006.

[140] P. Sliz, S. C. Harrison, and G. Rosenbaum, "How does radiation damage in protein crystals depend on x-ray dose?" *Structure*, vol. 11, no. 1, pp. 13–19, 2003.

[141] G. P. Bourenkov and A. N. Popov, "Optimization of data collection taking radi-

ation damage into account," *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 4, pp. 409–419, 2010.

[142] D. Liebschner, G. Rosenbaum, M. Dauter, and Z. Dauter, "Radiation decay of thaumatin crystals at three x-ray energies," *Acta Crystallographica Section D: Biological Crystallography*, vol. 71, no. 4, pp. 772–778, 2015.

[143] C. Blake and D. Phillips, "International atomic energy agency (vienna), biological effects of ionizing radiation at the molecular level, international atomic energy agency symposium," *Effects of X-irradiation on single crystals of myoglobin*, pp. 183–191, 1962.

[144] W. A. Hendrickson, "Radiation damage in protein crystallography," *Journal of molecular biology*, vol. 106, no. 3, pp. 889–893, 1976.

[145] J. Sygusch and M. Allaire, "Sequential radiation damage in protein crystallography," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 44, no. 4, pp. 443–448, 1988.

[146] Y. Z. Finfrock, E. A. Stern, Y. Yacoby, R. Alkire, K. Evans-Lutterodt, A. Stein, A. F. Isakovic, J. J. Kas, and A. Joachimiak, "Spatial dependence and mitigation of radiation damage by a line-focus mini-beam," *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 12, pp. 1287–1294, 2010.

[147] O. B. Zeldin, S. Brockhauser, J. Bremridge, J. M. Holton, and E. F. Garman, "Predicting the x-ray lifetime of protein crystals," *Proceedings of the National Academy of Sciences*, vol. 110, no. 51, pp. 20 551–20 556, 2013.

[148] J. Hubbell and S. Seltzer, "Tables of x-ray mass attenuation coefficients and mass energy-absorption coefficients (version 1.4). national institute of standards and technology, gaithersburg, md," *Gaithersburg, MD*, 2004.

[149] R. L. Owen, N. Paterson, D. Axford, J. Aishima, C. Schulze-Briese, J. Ren, E. E. Fry, D. I. Stuart, and G. Evans, "Exploiting fast detectors to enter a new dimension in room-temperature crystallography," *Acta Crystallographica Section D: Biological Crystallography*, vol. 70, no. 5, pp. 1248–1256, 2014.

[150] C. K. Wang, S.-H. Hu, J. L. Martin, T. Sjogren, J. Hajdu, L. Bohlin, P. Claeson, U. Goransson, K. J. Rosengren, J. Tang *et al.*, "Combined x-ray and nmr analysis of the stability of the cyclotide cystine knot fold that underpins its insecticidal activity and potential use as a drug scaffold," *Journal of Biological Chemistry*, vol. 284, no. 16, pp. 10 672–10 683, 2009.

[151] H. Coughlan, C. Darmanin, H. Kirkwood, N. Phillips, D. Hoxley, J. Clark, D. Vine, F. Hofmann, R. Harder, E. Maxey *et al.*, "Bragg coherent diffraction imaging and metrics for radiation damage in protein micro-crystallography," *Journal of synchrotron radiation*, vol. 24, no. 1, pp. 83–94, 2017.

[152] C. Nave, G. Sutton, G. Evans, R. Owen, C. Rau, I. Robinson, and D. I. Stuart, "Imperfection and radiation damage in protein crystals studied with coherent radiation," *Journal of synchrotron radiation*, vol. 23, no. 1, pp. 228–237, 2016.

[153] R. L. Owen, E. Rudiño-Piñera, and E. F. Garman, "Experimental determination of the radiation dose limit for cryocooled protein crystals," *Proceedings of the National Academy of Sciences*, vol. 103, no. 13, pp. 4912–4917, 2006.

[154] D. Tennant, "Soft x-ray microscopy with a cryo scanning transmission x-ray microscope: I. instrumentation, imaging and spectroscopy," *Journal of Microscopy*, vol. 197, no. Pt 1, pp. 68–79, 2000.

[155] R. Henderson, "Cryo-protection of protein crystals against radiation damage in electron and x-ray diffraction," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 241, no. 1300, pp. 6–8, 1990.

[156] T.-Y. Teng and K. Moffat, "Radiation damage of protein crystals at cryogenic temperatures between 40 k and 150 k," *Journal of synchrotron radiation*, vol. 9, no. 4, pp. 198–201, 2002.

[157] S. Arzt, A. Beteva, F. Cipriani, S. Delageniere, F. Felisaz, G. Förstner, E. Gordon, L. Launer, B. Lavault, G. Leonard *et al.*, "Automation of macromolecular crystallography beamlines," *Progress in biophysics and molecular biology*, vol. 89, no. 2, pp. 124–152, 2005.

[158] Z. Otwinowski Otwinowski, "Processing of x-ray diffraction data collected," *Methods in Enzymology*, vol. 276, pp. 307–326, 1997.

[159] J. M. Holton, S. Classen, K. A. Frankel, and J. A. Tainer, "The r-factor gap in macromolecular crystallography: an untapped potential for insights on accurate structures," *The FEBS journal*, vol. 281, no. 18, pp. 4046–4060, 2014.

[160] M. A. Pearson, L. O. Michel, R. P. Hausinger, and P. A. Karplus, "Structures of cys319 variants and acetohydroxamate-inhibited klebsiella aerogenes urease," *Biochemistry*, vol. 36, no. 26, pp. 8164–8172, 1997.

[161] M. F. Sanner, A. J. Olson, and J.-C. Spehner, "Reduced surface: an efficient way to compute molecular surfaces," *Biopolymers*, vol. 38, no. 3, pp. 305–320, 1996.

[162] D. Xu and Y. Zhang, "Generating triangulated macromolecular surfaces by euclidean distance transform," *PloS one*, vol. 4, no. 12, p. e8140, 2009.