HHS Public Access

Author manuscript

Wiley Interdiscip Rev Comput Mol Sci. Author manuscript; available in PMC 2020 July 01.

Published in final edited form as:

Wiley Interdiscip Rev Comput Mol Sci. 2019 ; 9(4): . doi:10.1002/wcms.1402.

Molecular Dynamics Simulations of Macromolecular Crystals

David S. Cerutti, **David A. Case**

Department of Chemistry and Chemical Biology, Rutgers University, 174 Frelinghuysen Road, Piscataway, NJ 08854-8066

Abstract

The structures of biological macromolecules would not be known to their present extent without X-ray crystallography. Most simulations of globular proteins in solution begin by surrounding the crystal structure of the monomer in a bath of water molecules, but the standard simulation employing periodic boundary conditions is already close to a crystal lattice environment. With simple protocols, the same software and molecular models can perform simulations of the crystal lattice, including all asymmetric units and solvent to fill the box. Throughout the history of molecular dynamics, studies of crystal lattices have served to investigate the quality of the underlying force fields, correlate the simulated ensembles to experimental structure factors, and extrapolate the behavior in lattices to behavior in solution. Powerful new computers are enabling molecular simulations with greater realism and statistical convergence. Meanwhile, the advent of exciting new methods in crystallography, including femtosecond free-electron lasers and image reconstruction for time-resolved crystallography on slurries of small crystals, is expanding the range of structures accessible to X-ray diffraction. We review past fusions of simulations and crystallography, then look ahead to the ways that simulations of crystal structures will enhance structural biology in the future.

Keywords

molecular simulation; dynamics; biomolecular crystals; X-ray diffraction; neutron diffraction

1 Introduction

The study of biological macromolecules in a crystalline environment has many attractive features. Crystals are often of high purity, and individual molecules see homogenous environments. Most macromolecular crystals consist of 30 to 70% solvent, and the usual solvent is water and ions. These structures mimic the crowded conditions in cells. In spite of the constrained motion imposed by the crystal lattice and notable effects on the conformations of polar surface side chains, there are remarkable similarities in the conformations and dynamics of crystals and solvated macromolecules.[1–3] Many enzymes even maintain activity in the crystalline state.[4–6] Results from X-ray diffraction and neutron scattering experiments on crystal structures, along with nuclear magnetic resonance (NMR) studies and other forms of spectroscopy, provide unparalleled structural and dynamical

*To whom correspondence should be addressed. Phone:(732) 445–0334. Fax: (732) 445–5958. david.cerutti@rutgers.edu.

information. The best known collection of this wealth of data on macromolecules is the Protein Data Bank, containing more than 140,000 structures.[7]

Historically, the majority of X-ray crystallography measurements were carried out at low temperatures to reduce radiation damage. Today, the development of serial and time-resolved crystallography, where many samples are illuminated in succession and femtosecond free-electron laser pulses can outrun radiation damage, has led to a rebirth of interest in room-temperature studies. With improvements in Laue crystallography[8–11] and later X-ray free-electron lasers,[12–14] crystallographic "snapshots" can now be collected with femtosecond to nanosecond time delays. Both of these advances bring experimental measurements within the time and temperature regimes where biomolecular simulations are mature.

The favorable aspects of crystals have long been evident to those carrying out molecular dynamics simulations of proteins and nucleic acids. Initial simulations of the small protein BPTI in vacuum[15] were soon followed by studies of the same protein in a crystal-like environment.[16] Crystal simulations of tens of picoseconds soon followed, making use of group-based cutoffs and attempts to infer water density where the X-ray data did not provide evidence for specific protein atoms.[17] After another decade, at the turn of the millennium, simulations of a few nanoseconds had become feasible, and simulations of protein crystals could measure the quality of the underlying molecular models.[18] Early molecular dynamics simulations of staphylococcal nuclease (SNase) in its crystalline state revealed some of the connections between protein dynamics and hydration levels. These simulations corroborated neutron diffraction experiments on the powdered protein.[19,20] Modern simulations, millions of times the length of the pioneering studies, rest on the emergence of powerful computers and efficient, accurate methods to deal with long-ranged electrostatic interactions in periodic systems.[21–24]

The early experiments and corresponding calculations offer a glimpse of the potential in crystal simulations: an ability to monitor molecular motions in atomic detail set against experiments that interrogate molecular structure in atomic detail. This confluence of theory and experiment promises to resolve details in the experimental observations and harness the experiments to refine molecular models. Advances in the experimental techniques themselves are dependent on much of the same hardware innovation, including data storage methods[25] and vector accelerated computing through graphics processing cards and field-programmable gate arrays.[26] The need for automation in refinement of experimental data will reinforce the demand for better molecular models and simulations.[27,28] The dynamics simulations see in explicit terms can explain aspects of the experimental X-ray diffraction patterns that could otherwise only be inferred, or heterogeneity in natural crystals that would have been overlooked.[29] The anticipated role of simulations in the coming years is to bring ensemble averages to the interpretive process of structure refinement, driving both fields forward.

2 Practical Considerations in Crystal Simulations

Molecular simulations in periodic boundary conditions are ideal for simulations of molecular and biomolecular crystal lattices. The periodic electrostatics models commonly used in molecular simulations impose an inherent periodicity that has frequently been analyzed as a liability in simulations of isolated molecules in solution.[30,31] In crystal simulations, these approaches are ideal and even necessary.[32] Nonetheless, copies of corresponding asymmetric units in neighboring unit cells impose a degree of artificial correlation by mirroring one another's motions. For this reason, many crystal simulations replicate the crystallographic unit cell within the simulation cell to provide an additional layer of independence among the various symmetry-related copies. The fluctuations of individual asymmetric units give rise to Debye-Waller factors, diffuse scattering, and alternate crystallographic states.

Individual asymmetric units of the crystal are equivalent to one another *on average*, and are arranged by symmetry operations in a particular space group to form the crystallographic unit cell. Multiple unit cells are related by three-dimensional translations. From the vantage point of any particular asymmetric unit, the crystal lattice looks the same: no matter its orientation in the unit cell as the crystallographic or simulation frame of reference defines it, all of the neighboring asymmetric units and the tiled pattern of unit cells look the same. In other words, one could align the crystallographic frame of reference to any asymmetric unit and see the same lattice of symmetry-related copies through the crystallographic rotation matrices and translations. Once a crystal simulation begins, the asymmetric units will diverge and the technical symmetry will become P1 like any other periodic simulation (including conventional "solution" MD simulations). However, the symmetry operations of the space group can still guide approximate superpositions of the asymmetric units, generating a larger ensemble of states. When applying symmetry operations to a divergent set of asymmetric units, the choice of reference can influence the result. The advisable convention is to choose the reference that minimizes overall deviations of the superimposed copies: treat the ensemble of asymmetric units as a cluster and take the central member as the reference.

The composition of the lattice carries important consequences for convergence in biomolecular simulations. Biomolecules tend to make up 30 to 70% of the material in their respective crystals, whereas the molecule of interest comprises only 5 to 20% of a typical protein in water simulation. The higher density of particles in a crystal simulation, coupled with increases in the number of complex terms (particularly dihedral potentials), adds a marginal cost to crystal simulations. In contrast, the ability to simulate multiple independent copies of the biomolecule at once, with small amounts of solvent per copy, improves the sampling rate for the crystalline environment by several fold over a simulation in the solution phase. The sampling problem is further mitigated by the moderate constraints that a crystal lattice places on molecular motion. Figure 1 shows how water is distributed over the surfaces of monomeric subunits to limit tight protein:protein contacts even in a crystal with low hydration (35% solvent by mass, in the case of a pectin lyase β-barrel). The convergence of phase space sampling in the lattice is a separate problem from sampling in the solution

phase, but most crystal lattice simulations would be expected to approach equilibrium sampling faster than their counterparts in solution.

The perception of convergence and sampling milestones in crystal simulations has tracked the growth of computational power: in 1995, Clarage and co-workers deduced that the Bragg intensity calculated from a trajectory of myoglobin in the solution phase had converged on the time scale of 150ps, but that the diffuse scattering intensity had not.[33] In 2005, Meinhold and Smith examined 10ns trajectories of a simulation of staphylococcal nuclease crystal, found convergence of the Bragg intensity for two thirds of the atoms on the time scale of 1 ns, and predicted that the diffuse scattering would converge on the timescale of 1 μs,[34] which is in rough agreement with more recent estimates.[35] Cerutti and colleagues simulated streptavidin crystals for 250ns,[36] but still, arguably, did not achieve a converged backbone positional root mean squared deviation from the original X-ray structure, suggesting that the Bragg pattern would still not be converged even when replicating the crystallization conditions. The underlying problem in these sampling questions is that the force field guiding the simulations may not recognize the native configuration, and continue to move the system towards new energy minima even after some perceived equilibration phase is complete. This process can make some arbitrary sampling interval look relatively converged, even though a much larger amount of sampling would reveal different configurations and therefore different values of the Bragg intensity peaks. Better force fields that identify the native minima of large biomolecules will improve convergence in the Bragg intensity and then provide realistic estimates of the simulation time to obtain the diffuse scattering intensity of these crystals. The fact that individual unit cells in a multi-unit cell simulation do not have equivalent average structures is another indication of lack of convergence.[37,38]

3 Simulation Setup

Standard periodic molecular dynamics engines will run crystal simulations, but unless the crystal structure displays all atoms of the asymmetric unit and solvent molecules, the utilities for preparing standard simulations of biomolecules in boxes of water are not suited to building crystal lattices. Utilities to replicate the asymmetric unit and then copy the unit cell itself are easy to write. What is more difficult to construct is a reliable program for packing solvent into the voids between lattice components that will have the unit cell maintain its volume under standard pressure, and then to determine the precise amounts of each solvent molecule to add.

A standard practice in solvating biomolecules or other chemical systems in simulation boxes of specified dimensions is to tile a box of pre-equilibrated solvent, such as water or methanol, throughout the simulation. Next, new solvent molecules which clash with the original system are culled. If the initial configuration of the system is to be maintained, restraints can be placed on the system to keep it in place until solvent molecules relax around it. Regardless of the relaxation protocol, the process of removing solvent molecules that clash forms vacuum bubbles if the system relaxes at a constant volume. These vacuum bubbles are removed by further dynamics to equilibrate the system at constant temperature and pressure. It is not uncommon for a simulation of solvated biomolecule to lose 15% or more of it original volume during such a procedure. Crystal simulations, however, must

maintain the volume prescribed by the original unit cell throughout equilibration and unrestrained production dynamics.

In order to keep crystal simulations at the correct volume without vacuum bubbles, it is necessary to add a precise amount of solvent to fill the interstitial voids between the observed asymmetric units of the lattice. The molecular models of both the lattice and the added solvent are important: as shown by Cerutti and colleagues for a scorpion venom toxin,[36] different force fields can require between 681 and 750 water molecules per unit cell, a range of 7% of the overall solvent content, to solvate a protein crystal holding 30% water content by mass. A uniform distribution of solvent should be added throughout the box in a single pass: if added and equilibrated in stages, the solvent can be drawn to one side of the lattice, sweeping ions or other mobile components of the observed asymmetric unit along with it and forcing the system far out of equilibrium when the right amount of solvent is finally reached. In order to add solvent in one pass, some trial and error is necessary. The amount of solvent that will fill out the unit cell after equilibration will have clashes either with the protein of interest or with other solvent particles when it is first introduced. A program for adding solvent should tolerate clashes among newly added solvent particles, anticipating relaxation via energy minimization, but avoid clashes between new solvent particles and the components of the lattice structure that are observed in the X-ray data. The crystal lattice may be kept in place with positional restraints on some or all of its heavy atoms during this process. Initially, the restraints maintain the structure of the asymmetric unit during energy minimization, but the restraints should also be tapered during the first few nanoseconds of dynamics to maintain the unit cell volume and aspect ratios. If the proper amount of solvent has been added, until it equilibrates under constant pressure the tendency will be to expand the simulation cell beyond the prescribed volume, but over time settling will reduce the simulation volume to some asymptotic value. Most systems with the correct amount of solvent can have restraints removed after ten nanoseconds of dynamics, then remain within 0.3% of the correct unit cell volume for the rest of the simulation.[39,40]

Another consideration is the content of the solvent itself, if not pure water. While many studies have simply filled the excess unit cell volume with water molecules,[34,41] the crystallization conditions often imply that there are additional non-water solvent molecules in the system. In some cases these molecules can be observed in the asymmetric unit, but the total content of solvent is uncertain if there is any empty space where the electron density is too flat to identify particular atoms. The crystallization conditions present in the structure file or the related publication may be helpful to fill this portion of the unit cell, but care must be taken to distinguish concentrations of reagents in the crystallization buffer, as originally prepared, from concentrations of those reagents in the final crystal. If hanging drop diffusion was used, the concentrations will be more like those of a higher-molality reservoir, and if evaporation was used to grow the crystal the concentrations of crystallization reagents would likewise be much higher than in the original buffer.[42] The propensity of solvent particles to reside near the surface of a protein or other macromolecule, their chemical potential in these regions, is not simple to determine. Grand canonical Monte-Carlo (GCMC) simulations offer a rigorous theoretical framework for addressing the problem of salt concentrations near biomolecules, but even with advances to the implementation[43,44] the force fields put bounds on the accuracy of these simulations.

When creating solvents with multiple components, mimicking crystallographic salt concentrations, or adding sufficient numbers of counterions to neutralize the unit cell, the compatibility of each molecular model becomes an important issue. Multiple studies have examined the best monovalent ion parameters[45,46] to use with each water model, and others have noted the consonance of cosolvents such as glycerol with water.[39] The common protein crystallization agent methyl-(2,4)-pentanediol (MPD) is miscible with water, but like glycerol, simulations of aqueous MPD mixtures can separate into phases on the timescale of nanoseconds. The balance of forces that keep these cosolvents hydrated, like the excess chemical potential of each species in the crystal lattice, is another important force field problem.

4 Improving Force Fields

X-ray diffraction images of molecular crystals provide a wealth of structural data for refining molecular models. It is possible to imagine situations where over-stabilization of some local interactions, such as protein side chain rotational profiles or salt bridges, could maintain a crystal structure against serious deficiencies elsewhere in the model. It is harder to imagine models with sparse repertoires of correct behavior stabilizing many unique crystal structures held together by weak non-bonded interactions. Biomolecular heteropolymers taking an abundance of unique folds and crystal packing arrangements could not be stabilized by a force field with many imbalanced features. The reasoning dates to some of the earliest attempts to fit intermolecular potential functions.[47,48] Even for DNA, a biopolymer with less diversity in its structure than RNA or polypeptides, the lattice packing exerts a modest reciprocal influence on the crystallized forms[49–52], pushing the canonical A- and B-forms into different sub-classifications. The ability to reproduce conformations in different crystal space groups therefore provides additional checks on the nuances of a molecular model.

The computation time needed to confirm that a given force field represents the structure and thermal fluctuations of numerous crystal structures is not trivial: periodic simulations of many systems with 5,000 or more atoms, each running tens to hundreds of ns. Compounding the problem, the parameter search space is huge: a description of the standard twenty amino acids comprises 250 partial charges, numerous van-der Waals parameters, dozens of bond lengths and equilibria, scores of angle parameters, and perhaps hundreds of dihedral parameters. A viable approach to refine force fields with crystal structure data is to start with an established model such as one of the Amber, CHARMM, or OPLS force fields[53–59] with parameters developed from quantum mechanical calculations and proceed to test small changes in selected parameters. This has been attempted in the line of YASARA force fields.[41]

Another strategy for validating force fields with X-ray structures is a combinatorial study of distinct lineages of biomolecular force fields and solvent models. This approach posits that each of the force fields exists in a different region of a parameter subspace that stabilizes natural protein configurations. The combinations of different three- and four-point water models with distinct protein models did vary in their ability to stabilize the X-ray structure of a scorpion venom protein wherein the solvent was only water with a low concentration of

ammonium acetate.[36] The most successful combination in the study was the Amber ff99SB force field (which has since been succeeded by the ff14SB force field) with any of the three-point water models tested. More recent protein force fields have been compared in simulations of the triclinic form of lysozyme.[38] The most pervasive problem was not the integrity of the monomer itself, but the inability to maintain weak interactions between contacting side chains of the lattice and the proper aspect ratios of the unit cell. Figure 1, coupled with the varying amount of water required by multiple force fields seen in simulations of the scorpion venom toxin, outlines the problem: even as protein force fields have improved their treatment of intra-molecular interactions to stabilize secondary structures and maintain the correct folds, there is another problem with the way they stabilize the film of water molecules on their surfaces that will be critical in maintaining crystal lattices.

Even with molecular dynamics on modern Graphics Processing Units (GPUs), crystal simulations are an expensive way to navigate parameter space. Whereas the 2010 study[36] took weeks on multiple clusters of eight to 64 processors, repeating this in 2018 would take only four days, using one Volta V100 GPU and the most recent version of the Amber molecular dynamics code. Even so, the number of protein force fields that could be examined would be quite limited. What is needed is a means for reducing the parameter search space. One aspect of biomolecular models that will have great impact on the structural properties of crystal simulations is the balance of interactions between the biopolymers themselves and the interactions with water or other solvent models. A subset of parameters that govern these interactions could be tractable to testing with long timescale simulations of multiple biomolecular crystals. Another strategy that may be helpful is adherence to quantum potential energy surfaces. Biomolecular parameter sets are seldom unique solutions for mimicking quantum potential energy surfaces. An extension of current development techniques[55,60,61] that determines multiple parameter sets could also map out a tractable search space for crystal simulations seeking parameters that best preserve structural details. However, it is beyond the scope of this review to discuss force field development strategies in depth.

Models of simpler molecules like water may get structural properties with greater accuracy than dynamical properties,[62] often as a result of the latter demanding more computation to quantify. Models developed for biopolymers based on crystal structure data could fall into the same trap, lacking information about the transition state barriers between stable conformations and the kinetics of folding. However, if the biomolecular force fields draw their own details from quantum mechanical energy surfaces, the transition states of side chain rotations and kinetics of secondary structure formation might be reasonable extrapolations.

There may not be simple, non-polarizable models for biopolymers that maintain all native crystal packing arrangements in the context of water models tuned to reproduce bulk properties of the liquid. More advanced force fields such as Amber's Amoeba[63,64] or the CHARMM polarizable Drude oscillator model[65,66] may be beneficial, as most of the water in a protein crystal lies within the first or second solvent layers of one or more proteins (see Figure 1). In lieu of a perfect model, molecular simulations can take place in the context of

restraints towards the X-ray or neutron scattering data. One study of ubiquitin[67] using the Amber ff99SB-ILDN force field[68] introduced ensemble-averaged restraints to hold the collection of asymmetric units in simulations towards the protein's X-ray structure. The results show that restraints applying energy penalties of 0.2 kcal/mol per residue will keep the ensemble near the crystal structure. While a modest amount of structural rearrangement still occurs (reflecting inaccuracies in the force field that are only countered by slight departures from the target structure), the convergence time for these rearrangements is tractable, on the order of a hundred nanoseconds. Analysis of the equilibrium ensemble could suggest changes to atomic parameters that would reinforce the native lattice packing and therefore improve the underlying force field.

5 How much physical realism should we expect from crystal simulations?

The most straightforward comparisons between simulations and atomic models refined against experimental X-ray diffraction data involve comparisons of the average structure, and of the mean-square atomic deviations about that average structure. It is not yet possible to give general rules about what to expect, but here we look at one recent example of a well-studied globular protein, triclinic lysozyme, comparing a recent simulation[38] with a room-temperature model refined against 0.95Å diffraction data.[69]

- Fig. 2 compares the backbone for triclinic hen egg-white lysozyme as determined from a 0.95 Åambient-temperature crystal structure (PDB id 4LZT) with the average structure determined from a 2 μs simulation of 12 unit cells of the crystal.[38] The differences are quite small, with a Cα average difference of 0.35Å. Instead of computing an average structure from MD snapshots, one might better construct an average electron density, since this is what the diffraction experiment is probing. In this case, refining a structural model into this average electron density yields a nearly identical structure, (with a Cα difference of 0.39Å), but bigger differences are seen in fluctuations. This level of agreement is better than seen in most macromolecular crystal simulations, although not as good as can be achieved for peptides.[70]
- Fig. 3 compares the isotropic Debye-Waller (or "B") factors derived from experiment with two different ways of analyzing the MD simulation. The red curve plots Cα B-factors for each residue from the structure refined against the computed electron density. These values are in excellent agreement with those derived in the same way from experiment, except for the N-terminal 23 residues, where the simulation is rather more flexible than in experiment. Debye-Waller factors are nominally related to mean-square fluctuations,

$$B = \left(8\pi^2/3\right)\left\langle u^2\right\rangle$$

and for the most part this is borne out by the close agreement of purple and red curves. However, for the regions of highest flexibility (near residues 47, 85 and 101), the refined B-factors significantly underestimate the actual fluctuations. This behavior has been noted before,[29,71] and highlights the fact that B-factors

are fitting parameters in a model. The B-factors are computed in way to optimize the agreement of observed and calculated Bragg intensities, and may not always provide a good measure of local fluctuations.

- As mentioned above, a major challenge for simulations is to preserve the protein-protein interactions that stabilize the crystal lattice. Fig. 4 plots the positions of the center of mass of the protein chain, for each of the 12 unit cells in the simulation. Colors represent individual unit cells, points show 100 snapshots for each, and the plot is arranged such that the origin is at the position in an undistorted lattice. Two main features are evident from this figure. First, the distribution of lattice positions varies from one unit cell to another, indicative of a lack of convergence in the simulation (which started from the fully symmetric experimental atomic model.) But the deviations are not large, on the order of 0.Å. Second, within each unit cell, there are thermal motions, so that the instantaneous position of the protein center of mass fluctuates about an average position. This sort of thermal motion, perhaps arising from lattice vibrations, would persist even in a "perfect" simulation. Again, the mean amplitude of these fluctuations are small (about 0.15Å), but as we discuss below, may represent an important contribution to disorder that is reflected in diffuse X-ray scattering intensities.

6 Making connections to experiment

6.1 Crystallographic refinement

Since the early days of molecular dynamics simulations, the prospects of combining simulations with pseudo-energy restraints based on a comparison of observed and calculated Bragg intensities (or the corresponding structure factor amplitudes) have been explored in the interest of improving atomic potentials.[72] In a more recent use of simulation methodology, the Boltzmann weight corresponding to force field energies (derived from a crystal simulation) can serve as a Bayesian "prior" distribution (encoding collective information about macromolecules in general), to be combined with experimental data from a particular crystal, in order to maximize a "posterior" likelihood of a refined model given the observed data.[73]

One promising prospect from using simulations to drive crystallographic refinement is the possibility of constructing "ensemble" models with more than a single atomic model. Such ensembles can describe motion and disorder beyond what is available from Debye-Waller atomic displacement parameters[74–76] Molecular dynamics simulations can also support time averaging and well as ensemble averages; these have seen wide adoption in the analysis of NMR data,[77–80] but fewer applications (so far) in crystallography.[81–83]

6.2 Time-resolved crystallography

Motivated by a desire to see evidence of reaction intermediates and to mitigate the effects of sample destruction by ionizing X-ray radiation, researchers have turned to detectors with shorter and faster X-ray pulses. In an example of a progenitor technique getting new attention with advances in the apparatus and computing capabilities, Laue diffraction with a

polychromatic X-ray source and static crystals afforded the first examples of time-resolved crystallography.[9,84,85] Indeed, in the typical "pump-probe" experiment, the destruction of the sample owes more to the laser activation pump than the 100ps polychromatic X-ray "probe." The time resolution of these studies is as low as 2μs, set by the speed of opening and closing the X-ray shutter. X-ray emitting free-electron lasers (XFELs), affording pulses with femtosecond duration, can resolve events as short as a few hundreds of ps. These studies aim to catch tiny crystals passing through a micrometer-sized detector, whether on a stage or diffusing free in liquid. The time limit in this latter case is set by the timing of the pump laser and X-ray probe pulses.[86–88]

Simulation and experimental technologies have met in the middle, and then passed one another: each can now perform studies on the timescale to which the other was limited fifteen years ago. However, another limitation has prevented molecular dynamics from modeling processes observed by time-resolved X-ray crystallography: the processes involve chemical reactions. For these problems, a quantum mechanical regime must be embedded within the molecular mechanics simulation, which sets back the simulation timescales by two to three orders of magnitude and removes the benefit of GPU accelerators as most of the codes do not handle hybrid QM and MM simulations at this time. Further complications arise since some of the fastest events take place on excited state potential energy surfaces. Even with all these current limitations, QM/MM simulations have been used to interpret time-resolved changes in bacteriorhodopsin[89] and myoglobin,[28] and more such studies are sure to appear.

Simulations attempting to mimic time-resolved crystal studies can be more difficult than they first appear. In such studies, it is necessary to average simulations over the distribution of starting configurations present in the crystal prior to the experiment. Also, individual unit cells in the crystal will inevitably get out of "sync" with their neighbors. This complicates both the interpretation of observed electron densities and potential comparisons between simulation and experiment. In spite of this, the potential for using MD simulations in the interpretation of time-resolved events is very promising.

6.3 Diffuse scattering of X-rays

In a perfect, static crystal, where the electron density is identical in every unit cell, X-ray scattering would be confined to intense and infinitely sharp "Bragg" peaks. Even for real crystals, where the Bragg peaks have finite width and there is additional scattering at other locations, data reduction is almost always limited to estimating the integrated intensities of the Bragg reflections and thereby obtaining atomic models of the structure. Advances in the field have brought scrutiny to the "diffuse" intensity in between the Bragg peaks. This diffuse intensity holds details about crystalline imperfections and correlated motions that cause divergence among unit cells.[90–92] New pixel array detectors enable accurate measurements of diffuse scattering,[93] sparking a renewed interest in mining this source of information.

The basic calculation of diffuse intensity from an MD simulation is straightforward. One collects snapshots from the simulation and computes the variance of calculated structure factors from the mean:

$$I_{diffuse}(\mathbf{q}) = \left\langle |F(\mathbf{q}) - \overline{F(\mathbf{q})}|^2 \right\rangle$$

Practical simulations of macromolecular crystals assume periodicity. The contributions of defects (such as mosaic blocks) would therefore not be captured by simulations, unless the calculations modeled a very large block of multiple repeating unit cells containing one such defect. However, simulations always capture thermal motion, and it is of interest to see how well these contributions can account for observed scattering. A "supercell" approach, in which more than a single unit cell forms the repeating unit in the simulation,[35] is again preferable as correlated fluctuations ("lattice modes") between unit cells are likely to be important.[92,94–98]

There have been numerous attempts to model diffuse scattering for organic crystals that resemble biological macromolecules.[99–103] The analysis of diffuse scattering through molecular dynamics simulations of proteins includes early studies on lysozyme[104] as well as ongoing simulations of staph nuclease.[35,105,106] For the latter, an old but comprehensive data set was collected using CCD detectors.[107] The simulations have provided helpful information in designing and assessing models for correlations in protein motion,[92] but the extent of direct agreement with experimental data has been limited by the paucity of detailed data and the limited size and time-scale of the simulations, but new technologies are expected to overcome both of these limitations. Pixel-array detectors for both synchrotron[92,93] and X-ray free electron laser[108] light sources should yield fine-grained reciprocal-space maps of scattering intensities. At the same time, developments in GPU technology make microsecond timescale simulations of macromolecular crystals feasible.

The physical origins of diffuse scattering are understood in part, but are expected to vary from one system to the next. Scattering adjacent to the Bragg peaks (in reciprocal space) is often modeled by a distribution of unit-cell translational dislocations,[92,97,108,109] similar to those seen in Fig. 4. The physical origins of these dislocations are not yet clear: they might arise from the thermal excitation of lattice vibrations, or from static sources of disorder, such as slight changes in effective protein size or conformation. MD simulations suggest that contributions from "disordered" solvent molecules appear play an important role in more slowly varying parts of diffuse scattering intensities.[105,106] A fuller understanding must await further studies on the both the experimental and the computational side.

6.4 NMR on microcrystalline samples

NMR has, for many years, provided a balance of information about the structure and dynamics of macro-molecules. It is unique among spectroscopic probes for its global atomic detail (often with assigned peaks corresponding to nearly all hydrogen, carbon and nitrogen nuclei) and for its ability to study systems near ambient temperature in a variety of environments. Both solution and solid-state samples can be investigated. Among solids, microcrystalline preparations often yield superior spectra,[110–113] offering the benefits of high purity and homogeneity that have aided crystallography. Although solid-state NMR can yield three-dimensional structures, this remains an emerging field. At the time of this writing, only 112 solid-state NMR structures have been deposited in the PDB, compared to

12,000 from solution NMR and 129,000 from X-ray crystallography. High spinning speeds and proton-detection offers the promise of increased automation.[114] However, much of the appeal of NMR on crystalline samples lies in the fact that it shines a different light on dynamics and disorder than does X-ray crystallography.

NMR relaxation in solids follows much of the basic formalism used for liquids,[115] with the recognition that the overall tumbling (or rotational diffusion) that is characteristic of macromolecules in solution is greatly restricted in a crystal lattice, and that account needs to be taken of the effects of magic angle spinning on the way molecular motions are averaged.[116–118] As in solution, longitudinal relaxation can provide detailed insight into the amplitudes and time scales of the rotations of particular spin-spin (i.e. atom-atom) vectors, although information about motions on longer time scales is present in solids because anisotropies are not averaged to zero by the overall tumbling that occurs in solution. For this reason, chemical exchange effects on microsecond to millisecond time scales contribute to order parameters, which presents both an opportunity to explore motions on a wide range of time scales, but also a challenge to simulation methods to account for all these possibilites.[2,118–120]

The analysis of transverse relaxation in macromolecules in the solid-state presents additional significant challenges: in the absence of the overall tumbling that dominates such relaxation in solution, it can be hard to identify all the remaining relevant interactions. One contribution that is of particular relevance to the study of crystalline proteins (and to comparisons to motional models extracted from X-ray crystallography) is the “rocking” or restricted librational motions of protein chains within the lattice of neighboring molecules. Detailed comparisons of experiment and crystal simulations have been carried out for microcrystalline forms of the small globular proteins GB1[119,121,122] and ubiquitin;[123–127] the case of ubiquitin is of particular interest, since it has been possible to study this behavior in different crystal forms.[127] NMR relaxation offers a useful complement to X-ray diffuse scattering here, since the former is sensitive to rotational motions, whereas the latter detects translational dislocations seen in Fig. 4. Both types of analysis place stringent requirements on convergence of MD simulations, and continued efforts will be required to probe the limits of such comparisons.

6.5 TeraHertz spectroscopy on protein crystals

Another promising method to study correlated motions in biological macromolecules is through vibrational spectroscopy in the teraHertz region, roughly below 300 cm^{-1}. As with NMR, the restrictions that lattice contacts have on both protein motions and water mobility are expected to be important determinants of what is observed. However, infrared absorption in this frequency range is strongly dominated by motions of water and mobile ions (salt) that may be present,[128–131] so that most macromolecular crystals give fairly featureless spectra that can be difficult to interpret.[132–135]

As a result, much of what has been learned about biological molecules in crystals has been restricted to smaller systems, like carbohydrates and peptides, that contain fragments of macromolecules but which can also be stripped of water.[136–140] In these cases, molecular dynamics simulations of crystals can give a good account of the observed spectra. Because

of the periodicity, the computation of the time-correlation function of the system dipole moment that is conventionally used to compute vibrational spectra[141] cannot be used: water molecules or other species that diffuse across a unit cell boundary appear to enter on the other side, so that a conventional calculation of the dipole moment is not appropriate. This difficulty can be circumvented by moving to a formulation based on velocity time-correlation functions.[136]

One intriguing idea to obtain information in protein crystals, which almost inevitably have a lot of hydration, is to measure absorption as a function of direction within the crystal lattice.[134,135] The hope, already realized to a certain extent for lysozyme[135] is that water contributions will be largely isotropic, and that the dependence on orientation can be interpreted in terms of protein contributions.

7 Conclusions

As a bridge between theory and experiment, simulations of protein crystals demonstrate computationalists making accommodations to replicate in detail the things that structural biologists observe. It is much easier for the computationalists to design simulations that mimic experimental conditions than for biochemists to perform measurements more amenable to simulation, although some uncertainty remains in the exact solvent composition of most macromolecular crystals. For biological systems that do not undergo chemical changes, equilibrium molecular dynamics simulations on the required timescales for convergence of the Bragg peaks and diffuse scattering are within reach. The problem, instead, is now to design force fields that will sample the correct native structures rather than take the crystal lattice on a slow march towards some non-native state that the model finds lower in energy. This may yet require more detailed molecular models whose implementations on graphics card accelerators, the preferred computing platforms, are not yet mature. However, accurate simulations of static X-ray structures will enable better crystallographic refinement as well as molecular models that simulations can extrapolate X-ray structures into the behavior of proteins in bulk water.

Computational studies of enzymatic reactions, such as those studied by time-resolved crystallography, are further off. The molecular models are not ready, at least to perform on the timescales that free-electron laser setups can capture. However, the hardware needed by these experiments will advance in step with accelerated computing, and the necessary spatial reconstruction will be aided by molecular mechanics force fields, reducing the amount of data needed to obtain structures with high confidence. The proven synergy of simulations and X-ray diffraction studies will continue to grow in the coming decade, to give clearer pictures of the relationships between structure and function in nature’s most intricate molecules.

Acknowledgments

This work was supported by NIH grant GM122086. We thank James Holton, George Phillps and Janowski for many useful discussions.

References

[1]. Jacobson MP, Friesner RA, Xiang Z, Honig B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. J Mol Biol 2002, 320:597–608. [PubMed: 12096912]

[2]. Torchia DA. NMR studies of dynamic biomolecular conformational ensembles. Prog NMR Spectr 2015, 84–85:14–32.

[3]. Ahlstrom LS, Vorontsov II, Shi J, Miyashita O. Effect of the Crystal Environment on Side-Chain Conformational Dynamics in Cyanovirin-N Investigated through Crystal and Solution Molecular Dynamics Simulations. PLOS One 2017, 12:1–17.

[4]. Moncrief MBC, Hom LG, Jabri E, Karplus PA, Hausinger RP. Urease activity in the crystalline state. Protein Sci 1995, 4:2234–2236. [PubMed: 8535259]

[5]. Yeremenko S, van Stokkum IHM, Moffat K, Hellingwerf KJ. Influence of the Crystalline State on Photoinduced Dynamics of Photoactive Yellow Protein Studied by Ultraviolet-Visible Transient Absorption Spectroscopy. Biophys J 2006, 90:4224–4235. [PubMed: 16513787]

[6]. Schmidt M, Saldin DK. Enzyme transient state kinetics in crystal and solution from the perspective of a time-resolved crystallographer. Struct Dyn 2014, 1:024701. [PubMed: 26798774]

[7]. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucl Acids Res 2000, 28:235–242. [PubMed: 10592235]

[8]. Hajdu J, Machin PA, Campbell JW, Greenhough TJ, Clifton IJ, Zurek S, et al. Millisecond X-ray diffraction and the first electron density map from Laue photographs of a protein crystal. Nature 1987, 329:178–181. [PubMed: 3114644]

[9]. Srajer V, Ren Z, Teng TY, Schmidt M, Ursby T, Bourgeois D, et al. Protein Conformational Relaxation and Ligand Migration in Myoglobin: A Nanosecond to Millisecond Molecular Movie from Time-Resolved Laue X-ray Diffraction. Biochemistry 2001, 40:13802–13815. [PubMed: 11705369]

[10]. Schotte F, Lim M, Jackson TA, Smirnov AV, Soman J, Olson JS, et al. Watching a protein as it functions with 150-ps time-resolved X-ray crystallography. Science 2003, 300:1944–1947. [PubMed: 12817148]

[11]. Schotte F, Cho HS, Kaila VRI, Kamikubo H, Dashdork N, Henry ER, et al. Watching a signaling protein function in real time via 100-ps time-resolved Laue crystallography. Proc Natl Acad Sci USA 2012, 109:19256–19261. [PubMed: 23132943]

[12]. Aquila A, Hunter MS, Doak RB, Kirian RA, Fromme P, White TA, et al. Time-resolved protein nanocrystallography using an X-ray free-electron laser. Opt Express 2012, 20:2706–2716. [PubMed: 22330507]

[13]. Kupitz C, Basu S, Grotjohann I, Fromme R, Zatsepin NA, Rendek KN, et al. Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. Nature 2014, 513:261–EP. [PubMed: 25043005]

[14]. Hunter MS, Segelke B, Messerschmidt M, Williams GJ, Zatsepin NA, Barty A, et al. Fixed-target protein serial microcrystallography with an x-ray free electron laser. Sci Rep 2014, 4:6026. [PubMed: 25113598]

[15]. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature 1977, 267:585–590. [PubMed: 301613]

[16]. Van Gunsteren WF, Karplus M. Protein dynamics in solution and in a crystalline environment. Biochemistry 1982, 21:2259–2274. [PubMed: 6178423]

[17]. Avbelj F, Moult J, Kitson DH, James MNG, Hagler AT. Molecular dynamics study of the structure and dynamics of a protein molecule in a crystalline ionic environment, Streptomyces griseus protease A. Biochemistry 1990, 29:8658–8676. [PubMed: 2125469]

[18]. Stocker U, Spiegel K, van Gunsteren WF. On the similarity of properties in solution or in the crystalline state: A molecular dynamics study of hen lysozyme. J Biomol NMR 2000, 18:1–12. [PubMed: 11061223]

[19]. Joti Y, Nakagawa H, Kataoka M, Kitao A. Hydration Effect on Low-Frequency Protein Dynamics Observed in Simulated Neutron Scattering Spectra. Biophys J 2008, 94:4435–4443. [PubMed: 18310244]

[20]. Joti Y, Nakagawa H, Kataoka M, Kitao A. Hydration-Dependent Protein Dynamics Revealed by Molecular Dynamics Simulation of Crystalline Staphylococcal Nuclease. J Phys Chem B 2008, 112:3522–3528. [PubMed: 18293961]

[21]. Hockney RW, Eastwood JW. Computer Simulation using Particles. New York: McGraw-Hill; 1981.

[22]. Darden T, York D, Pedersen L. Particle mesh Ewald–an Nlog(N) method for Ewald sums in large systems. J Chem Phys 1993, 98:10089–10092.

[23]. Sagui C, Darden TA. Molecular dynamics simulations of biomolecules: Long-range electrostatic effects. Annu Rev Biophys Biomol Struct 1999, 28:155–179. [PubMed: 10410799]

[24]. Hünenberger P Lattice-sum methods for computing electrostatic interactions in molecular simulations In: Pratt LR, Hummer G, editors. Simulation and Theory of Electrostatic Interactions in Solution Melville, NY: American Institute of Physics; 1999p. 17–83.

[25]. Barty A, Kirian RA, Maia FRNC, Hantke M, Yoon CH, White TA, et al. Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. J Appl Crystallogr 2014, 47:1118–1131. [PubMed: 24904246]

[26]. Choe H, Gorfman S, Heidbrink S, Pietsch U, Vogt M, Winter J, et al. Multichannel FPGA-Based Data-Acquisition-System for Time-Resolved Synchrotron Radiation Experiments. IEEE Trans Nucl Sci 2017, 64:1320–1326.

[27]. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. To-wards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr 2012, 68:352367.

[28]. Barends TRM, Foucar L, Botha S, Doak RB, Shoeman RL, Nass K, et al. De novo protein crystal structure determination from X-ray free-electron laser data. Nature 2014, 505:244–247. [PubMed: 24270807]

[29]. Kuzmanic A, Pannu NS, Zagrovic B. X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals. Nature Commun 2014, 5:3220. [PubMed: 24504120]

[30]. Hummer G, Pratt LR, Garca AE, Berne BJ, Rick SW. Electrostatic potentials and free energies of solvation of polar and charged molecules. J Phys Chem B 1997, 101:3017–3020.

[31]. Hub JS, de Groot BL, Grubmüller H, Groenhof G. Quantifying Artifacts in Ewald Simulations of Inhomogeneous Systems with a Net Charge. J Chem Theory Comput 2014, 10:381–390. [PubMed: 26579917]

[32]. Walser R, Hünenberger PH, van Gunsteren WF. Comparison of Different Schemes to Treat Long-Range Electrostatic Interactions in Molecular Dynamics Simulations of a Protein Crystal. Proteins 2001, 43:509–519. [PubMed: 11340666]

[33]. Clarage JB, Romo T, Andrews BK, Pettitt BM, Phillips GN Jr. A sampling problem in molecular dynamics simulations of macromolecules. Proc Natl Acad Sci USA 1995, 92:3288–3292. [PubMed: 7724554]

[34]. Meinhold L, Smith JC. Fluctuations and Correlations in Crystalline Protein Dynamics: A Simulation Analysis of Staphylococcal Nuclease. Biophys J 2005, 88:2554–2563. [PubMed: 15681654]

[35]. Wall ME. Internal protein motions in molecular-dynamics simulations of Bragg and diffuse X-ray scattering. IUCrJ 2018, 5:172–181.

[36]. Cerutti DS, Freddolino PL, Duke RE Jr, Case DA. Simulations of a protein crystal with a high resolution X-ray structure: Evaluation of force fields and water models. J Phys Chem B 2010, 114:12811–12824. [PubMed: 20860388]

[37]. Liu C, Janowski PA, Case DA. All-atom crystal simulations of DNA and RNA duplexes. Biochim Biophys Acta 2015, 1850:1059–1071. [PubMed: 25255706]

[38]. Janowski PA, Liu C, Deckman J, Case DA. Molecular dynamics simulation of triclinic lysozyme in a crystal lattice. Prot Sci 2016, 25:87–102.

[39]. Cerutti DS, LeTrong I, Stenkamp RE, Lybrand TP. Simulations of a Protein Crystal: Explicit Treatment of Crystallization Conditions Links Theory and Experiment in the Streptavidin-Biotin Complex. Biochemistry 2008, 47:12065–12077. [PubMed: 18950193]

[40]. Cerutti DS, Duke RE, Darden TA, Lybrand TP. Staggered Mesh Ewald: An Extension of the Smooth Particle-Mesh Ewald Method Adding Great Versatility. J Chem Theory Computat 2009, 5:2322–2338.

[41]. Krieger E, Darden T, Nabuurs SB, Finkelstein A, Vriend G. Making Optimal Use of Empirical Energy Functions: Force-Field Parameterization in Crystal Space. Proteins 2004, 57:678–683. [PubMed: 15390263]

[42]. Talreja S, Kim DY, Mirarefi AY, F ZC, Kenis PJA. Screening and optimization of protein crystallization conditions through gradual evaporation using a novel crystallization platform. J Appl Crystallogr 2005, 38:988–995.

[43]. Sun D, Lakkaraju SK, Jo S, MacKerell AD. Determination of Ionic Hydration Free Energies with Grand Canonical Monte Carlo/Molecular Dynamics Simulations in Explicit Water. J Chem Theory Comput 2018,.

[44]. Ross GA, Rustenburg AS, Grinaway PB, Fass J, Chodera JD. Biomolecular Simulations under Realistic Macroscopic Salt Conditions. J Chem Phys B 2018, 122:5466–5486.

[45]. Joung IS, Cheatham TE III. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. J Phys Chem B 2008, 112:9020–9041. [PubMed: 18593145]

[46]. Chen AA, Pappu RV. Parameters of monovalent ions in the AMBER-99 forcefield: Assessment of inaccuracies and proposed improvements. J Phys Chem B 2007, 111:11884–11887. [PubMed: 17887792]

[47]. Hagler A, Lifson S. Energy functions for peptides and proteins. II. Amide hydrogen bond and calculation of amide crystal properties. J Am Chem Soc 1974, 96:5327–5335. [PubMed: 4851861]

[48]. Bernstein J, Hagler AT. Conformational Polymorphism. The Influence of Crystal Structure on Molecular Conformation. J Am Chem Soc 1978, 100:673–681.

[49]. Jain S, Sundaralingam M. Effect of Crystal Packing Environment on Conformation of the DNA Duplex. J Biol Chem 1989, 264:12780–12784. [PubMed: 2753886]

[50]. Shakked Z, Guerstein-Guzikevich G, Eisenstein M, Frolow F, Rabinovich D. The conformation of the DNA double helix in the crystal is dependent on its environment. Nature 1989, 342:456–460. [PubMed: 2586615]

[51]. Lipanov A, Kopka ML, Kaczor-Grzeskowiak M, Quintana J, Dickerson RE. Structure of the B-DNA decamer C-C-A-A-C-I-T-T-G-G in two different space groups: Conformational flexibility of B-DNA. Biochemistry 1993, 32:1373–1389. [PubMed: 8448146]

[52]. Dickerson RE, Goodsell DS, Neidle S. "…the tyranny of the lattice…". Proc Natl Acad Sci USA 1994, 91:3579–3583. [PubMed: 8170950]

[53]. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 2006, 65:712–725. [PubMed: 16981200]

[54]. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theor Comput 2015, 11:3696–3713.

[55]. Debiec KT, Cerutti DS, Baker LR, Gronenborn AM, Case DA, Chong LT. Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. J Chem Theory Comput 2016, 12:3926–3947. [PubMed: 27399642]

[56]. Foloppe N, MacKerell AD Jr. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. J Comput Chem 2000, 21:86.

[57]. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom biological force fields. J Comput Chem 2010, 31:671–690. [PubMed: 19575467]

[58]. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone and Side-Chain Dihedral Angles. J Chem Theory Comput 2012, 8:3257–3273. [PubMed: 23341755]

[59]. Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput 2016, 12:281–296. [PubMed: 26584231]

[60]. Wang LP, Martinex TJ, Pande VS. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. J Phys Chem Lett 2014, 11:18851891.

[61]. Vanommeslaeghe K, Yang M, MacKerell AD. Robustness in the fitting of molecular mechanics parameters. J Comput Chem 2015, 36:1083–1101. [PubMed: 25826578]

[62]. Wu Y, Tepper HL, Voth GA. Flexible simple point-charge water model with improved liquid-state properties. J Chem Phys 2006, 124:024503. [PubMed: 16422607]

[63]. Schnieders MJ, Fenn TD, Pande VS, Brunger AT. Polarizable atomic multipole X-ray refinement: application to peptide crystals. Acta Cryst D 2009, 65:952–965. [PubMed: 19690373]

[64]. Fenn EE, Wong DB, Fayer MD. Water dynamics in small reverse micelles in two solvents: Two-dimensional infrared vibrational echoes with two-dimensional background subtraction. J Chem Phys 2011, 134:054512–11. [PubMed: 21303143]

[65]. Baker CM, Lopes PEM, Zhu X, Roux B, MacKerell AD Jr. Accurate Calculation of Hydration Free Energies using Pair-Specific Lennard-Jones Parameters in the CHARMM Drude Polarizable Force Field. J Chem Theor Comput 2010, 6:1181–1198.

[66]. Lopes PEM, Huang J, Shim J, Luo Y, Li H, Roux B, et al. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. J Chem Theory Comput 2013, 9:5430–5449. [PubMed: 24459460]

[67]. Xue Y, Skrynnikov NR. Ensemble MD simulations restrained vi crystallographic data: Accurate structure leads to accurate dynamics. Prot Sci 2014, 23:488–507.

[68]. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis J, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 2010, 78:1950–1958. [PubMed: 20408171]

[69]. Walsh MA, Schneider TR, Sieker LC, Dauter Z, Lamin VS, Wilson KS. Refinement of triclinice hen egg-white lysozyme at atomic resolution. Acta Cryst 1998, D54:522–546.

[70]. Janowski PA, Cerutti DS, Holton J, Case DA. Peptide crystal simulations reveal hidden dynamics. J Am Chem Soc 2013, 135:7938–7948. [PubMed: 23631449]

[71]. Kuriyan J, Wilz S, Karplus M, Petsko GA. X-ray structure and refinement of carbon-monoxy Fe(II) myoglobin at 1.5 resolution. J Mol Biol 1986, 192:133–154. [PubMed: 3820301]

[72]. Brünger AT. Refinement of three-dimensional structures of proteins and nucleic acids In: Goodfellow JM, editor. Topics in Molecular Biology London: Macmillan Press, Ltd.; 1991p. 137–178.

[73]. Rupp B Biomolecular Crystallography: Princples, Practice, and Application of Structural Biology. New York: Garland Science; 2010.

[74]. Kuriyan J, Ösapay K, Burley SK, Brünger AT, Hendrickson WA, Karplus M. Exploration of disorder in protein structures by X-ray restrained molecular dynamics. Proteins: Str Func Genet 1991, 10:340–358.

[75]. Kohn JE, Afonine PV, Ruscio JZ, Adams PD, Head-Gordon T. Evidence of Functional Protein Dynamics from X-Ray Crystallographic Ensembles. PLoS Computational Biology 2010, 6:e1000911. [PubMed: 20865158]

[76]. Burnley BT, Afonine PV, Adams PD, Gros P. Modelling dynamics in protein crystal structures by ensemble refinement. eLife 2012, 1:e00311. [PubMed: 23251785]

[77]. Torda AE, Scheek RM, VanGunsteren WF. Time-dependent distance restraints in molecular dynamics simulations. Chem Phys Lett 1989, 157:289–294.

[78]. Pearlman DA, Kollman PA. Are time-averaged restraints necessary for nuclear magnetic resonance refinement? A model study for DNA. J Mol Biol 1991, 220:457–479. [PubMed: 1856868]

[79]. Scott WRP, Mark AE, van Gunsteren WF. On using time-averaged restraints in molecular dynamics simulations. J Biomol NMR 1998, 12:501–508. [PubMed: 20012761]

[80]. Hansen N, van Gunsteren WF. Practical Aspects of Free-Energy Calculations: A Review. J Chem Theory Comput 2014, 10:2632–2647. [PubMed: 26586503]

[81]. Gros P, van Gunsteren WF, Hol WGJ. Inclusion of thermal motion in crystallographic structures by restrained molecular dynmaics. Science 1990, 249:1149–1152. [PubMed: 2396108]

[82]. Clarage JB, Phillips GN Jr. Cross-validation tests of time-averaged molecular dynamics refinements for determination of protein structures by X-ray crystallography. Acta Cryst 1994, D50:24–36.

[83]. Romo TD, Clarage JB, Sorensen DC, Phillips GN Jr. Automatic identification of discrete substates in proteins: Singular value decomposition analysis of time averaged crystallographic refinements. Proteins 1995, 22:311–321. [PubMed: 7479706]

[84]. Moffat K Time-Resolved Biochemical Crystallography: A Mechanistic Perspective. Chem Rev 2001, 101:1569–1582. [PubMed: 11709992]

[85]. Bourgeois D, Schotte F, Brunori M, Vallone B. "Time-resolved methods in biophysics. 6. Time-resolved Laue crystallography as a tool to investigate photo-activated protein dynamics. Photochem Photobiol Sci 2007, 6:1047–1056. [PubMed: 17914477]

[86]. Moffat K Time-resolved crystallography and protein design: signalling photoreceptors and optogenetics. Phil Trans R Soc B 2014, 369:20130568. [PubMed: 24914168]

[87]. Srajer V, Schmidt M. Watching proteins function with time-resolved x-ray crystallography. J Phys D: Appl Phys 2017, 50:373001. [PubMed: 29353938]

[88]. Kupitz C, Olmos JL, Holl M, Tremblay L, Pande K, Pandey S, et al. Structural enzymology using X-ray free electron lasers. Struct Dynam 2017, 4:044003.

[89]. Nogly P, Weinert T, James D, Carbajo S, Ozerov D, Furrer A, et al. Retinal isomerization in bacteriorhopsin captured by a femtosecond x-ray laser. Science 2018, 361:145.

[90]. Benoit JP, Doucet J. Diffuse scattering in protein crystallography. Quart Rev Biophys 1995, 28:131–169.

[91]. Welberry TR. Diffuse X-ray Scattering and Models of Disorder. Oxford: Oxford University Press; 2004.

[92]. Meisburger SP, Thomas WC, Watkins MB, Ando N. X-ray Scattering Studies of Protein Structural Dynamics. Chem Rev 2017, 117:7615–7672. [PubMed: 28558231]

[93]. Gruner SM, Lattman EE. Biostructural Science Inspired by Next-Generation X-Ray Sources. Annu Rev Biophys 2015, 44:33–51. [PubMed: 25747590]

[94]. Meinhold L, Merzel F, Smith JC. Lattice dynamics of a protein crystal. Phys Rev Lett 2007, 99:138101. [PubMed: 17930640]

[95]. Riccardi D, Cui Q, Phillips GN. Application of Elastic Network Models to Proteins in the Crystalline State. Biophys J 2009, 96:464–475. [PubMed: 19167297]

[96]. Riccardi D, Cui Q, Phillips GN Jr. Evaluating Elastic Network Models of Crystalline Biological Molecules with Temperature Factors, Correlated Motions, and Diffuse X-Ray Scattering. Biophys J 2010, 99:2616–2625. [PubMed: 20959103]

[97]. Polikanov YS, Moore PB. Acoustic vibrations contribute to the diffuse scatter produced by ribosome crystals. Acta Cryst D 2015, 71:2021–2031. [PubMed: 26457426]

[98]. Van Benschoten AH, Liu L, Gonzalez A, Brewster AS, Sauter NK, Fraser JS, et al. Measuring and modeling diffuse scattering in protein X-ray crystallography. Proc Natl Acad Sci USA 2016, 113:4069–4074. [PubMed: 27035972]

[99]. Goossens DJ, Heerdegen AP, Welberry TR, Beasley AG. The molecular conformation of Ibuprofen,C13H18O2, through X-ray diffuse scattering. Int J Pharmaceut 2007, 343:59–68.

[100]. Chan EJ, Welberry TR, Heerdegen AP, Goossens DJ. Diffuse scattering study of aspirin forms (I) and (II). Acta Cryst B 2010, 66:696–707. [PubMed: 21099031]

[101]. Chan EJ, Welberry TR, Goossens DJ, Heerdegen AP, Appl Cryst J. A refinement strategy for Monte Carlo modelling of diffuse scattering from molecular crystal systems, vol. 43; 2010.

[102]. Goossens DJ, Heerdegen AP, Chan EJ, Welberry TR. Monte Carlo Modeling of Diffuse Scattering from Single Crystals: The Program ZMC. Metall Mat Trans A 2011, 42A:23–31.

[103]. Welberry TR, Chan EJ, Goossens DJ, Heerdegen AP. Diffuse Scattering as an Aid to the Understanding of Polymorphism in Pharmaceuticals. Metall Mat Trans A 2012, 43A:1434–1444.

[104]. Hery S, Genest D, Smith J. X-ray Diffuse Scattering and Rigid-body Motion in Crystalline Lysozyme Probed by Molecular Dynamics Simulation. J Mol Biol 1998, 279:303. [PubMed: 9636718]

[105]. Meinhold L, Smith JC. Protein dynamics from X-ray crystallography: Anisotropic, global motion in diffuse scattering patterns. Proteins 2007, 66:941–953. [PubMed: 17154425]

[106]. Wall ME, Van Benschoten AH, Sauter NK, Adams PD, Fraser JS, Terwilliger TC. Conformational dynamics of a crystalline protein from microsecond-scale molecular dynamics simulations and diffuse X-ray scattering. Proc Natl Acad Sci USA 2014, 111:17887–17892. [PubMed: 25453071]

[107]. Wall ME, Ealick SE, Gruner SM. Three-dimensional diffuse X-ray scattering from crystals of Staphy-lococcal nuclease. Proc Natl Acad Sci USA 1997, 94:6180–6184. [PubMed: 9177191]

[108]. Ayyer K, Yefanov OM, Oberthür D, Roy-Chowdhury S, Galli L, Mariani V, et al. Macromolecular diffractive imaging using imperfect crystals. Nature 2016, 530:202–206. [PubMed: 26863980]

[109]. Barabash RI, Chung JS, Thorpe MF. Lattice and continuum theories of Huang scattering. J Phys Condens Mat 1999, 11:3075–3090.

[110]. Cole HBR, Torchia DA. An NMR study of the backbone dynamics of staphylococcal nuclease in the crystalline state. Chem Phys 1991, 158:271–281.

[111]. Zech SG, Wand AJ, McDermott AE. Protein structure determination by high-resolution solid-state NMR spectroscopy: Application to microcrystalline ubiquitin. J Am Chem Soc 2005, 127:8618–8626. [PubMed: 15954766]

[112]. Lorieau JL, McDermott AE. Conformational Flexibility of a Microcrystalline Globular Protein: Order Parameters by Solid-State NMR Spectroscopy. J Am Chem Soc 2006, 128:11505–11512. [PubMed: 16939274]

[113]. Wylie BJ, Sperling LJ, Frericks HL, Shah GJ, Franks WT, Rienstra CM. Chemical-Shift Anisotropy Measurements of Amide and Carbonyl Resonances in a Microcrystalline Protein with Slow Magic-Angle Spinning NMR Spectroscopy. J Am Chem Soc 2007, 129:5318–5319. [PubMed: 17425317]

[114]. Andreas LB, Jaudzems K, Stanek J, Lalli D, Bertarello A, Le Marchand T, et al. Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. Proc Natl Acad Sci USA 2016, 113:9187–9192. [PubMed: 27489348]

[115]. Kowalewski J, Mäler L. Nuclear Spin Relaxation in Liquids: Theory, Experiments, and Applications. New York: Taylor & Francis; 2006.

[116]. Giraud N, Blackledge M, Goldman M, Bockmann A, Lesage A, Penin F, et al. Quantitative Analysis of Backbone Dynamics in a Crystalline Protein from Nitrogen-15 Spin-Lattice Relaxation. J Am Chem Soc 2005, 127:18190–18201. [PubMed: 16366572]

[117]. Kurbanov R, Zinkevich T, Krushelnitsky A. The nuclear magnetic resonance relaxation data analysis in solids: General R1/R1rho equations and the model-free approach. J Chem Phys 2011, 135:184104. [PubMed: 22088049]

[118]. Schanda P, Ernst M. Studying dynamics by magic-angle spinning solid-state NMR spectroscopy: Principles and applications to biomolecules. Prog NMR Spectr 2016, 96:1–46.

[119]. Lewandowski JR, Halse ME, Blackledge M, Emsley L. Direct observation of hierarchical protein dynamics. Science 2015, 348:578–581. [PubMed: 25931561]

[120]. Le Marchand T, de Rosa M, Salvi N, Sala BM, Andreas LB, Barbet-Massin E, et al. Conformational dynamics in crystals reveal the molecular bases for D76N beta-2 microglobulin aggregation propensity. Nature Comm 2018, 9:1658.

[121]. Lewandowski JR, Sein J, Blackledge M, Emsley L. Anisotropic Collective Motion Contributes to Nuclear Spin Relaxation in Crystalline Proteins. J Am Chem Soc 2010, 132:1246–1248. [PubMed: 19916496]

[122]. Mollica L, Baias M, Lewandowski JR, Wylie BJ, Sperling LJ, Rienstra CM, et al. Atomic-Resolution Structural Dynamics in Crystalline Proteins from NMR and Molecular Simulation. J Phys Chem Lett 2012, 3:3657–3662. [PubMed: 26291002]

[123]. Haller JD, Schanda P. Amplitudes and time scales of picosecond-to-microsecond motion in proteins studied by solid-state NMR: a critical evaluation of experimental approaches and application to crystalline ubiquitin. J Biomol NMR 2013, 57:263–280. [PubMed: 24105432]

[124]. Fasshuber HK, Lakomek N, Habenstein B, Loquet A, Shi C, Giller K, et al. Structural heterogeneity in microcrystalline ubiquitin studied by solid-state NMR. Prot Sci 2015, 24:592–598.

[125]. Ma P, Xue Y, Coquelle N, Haller JD, Yuwen T, Ayala I, et al. Observing the overall rocking motion of a protein in a crystal. Nature Commun 2015, 6:8361. [PubMed: 26436197]

[126]. Kurauskas V, Weber E, Hessel A, Ayala I, Marion D, Schanda P. Cross-Correlated Relaxation of Dipolar Coupling and Chemical-Shift Anisotropy in Magic-Angle Spinning R1'rho' NMR Measurements: Application to Protein Backbone Dynamics Measurements. J Phys Chem B 2016, 120:8905–8913. [PubMed: 27500976]

[127]. Kurauskas V, Izmailov SA, Rogacheva ON, Hessel A, Ayala I, Woodhouse J, et al. Slow conformational exchange and overall rocking motion in ubiquitin protein crystals. Nature Commun 2017, 8:145. [PubMed: 28747759]

[128]. Conti Nibali V, Havenith M. New Insights into the Role of Water in Biological Function: Studying Solvated Biomolecules Using Terahertz Absorption Spectroscopy in Conjunction with Molecular Dynamics Simulations. J Am Chem Soc 2014, 136:12800–12807. [PubMed: 25127002]

[129]. Xu Y, Rodger PM. Improved Estimation of Density of States for Monte Carlo Sampling via MBAR. J Chem Theory Comput 2015, 11:4565–4572. [PubMed: 26574248]

[130]. Conti Nibali V, Morra G, Havenith M, Colombo G. Role of Terahertz (THz) Fluctuations in the Allosteric Properties of the PDZ Domains. J Phys Chem B 2017, 121:10200–10208. [PubMed: 28991478]

[131]. Esser A, Forbert H, Sebastiani F, Schwaab G, Havenith M, Marx D. Hydrophilic Solvation Dominates the Terahertz Fingerprint of Amino Acids in Water. J Phys Chem B 2018, 122:1453–1459. [PubMed: 29281284]

[132]. Burnett AD, Kenkdrick J, Cunningham JE, Hargreaves MD, Munshi T, Edwards HGM, et al. Calculation and measurement of terahertz active normal modes in crystalline PETN. ChemPhysChem 2010, 11:368–378. [PubMed: 20049763]

[133]. Tych KM, Burnett AD, Wood CD, Cunningham JE, Pearson AR, Davies AG, et al. Applying broad-band terrahertz time-domain spectroscopy to the analysis of crystalline proteins: a dehydration study. J Appl Cryst 2011, 44:129–133.

[134]. Singh S, Chiu C, de Pablo JJ. Efficient Free Energy Calculation of Biomolecules from Diffusion-Biased Molecular Dynamics. J Chem Theory Comput 2012, 8:4657–4662. [PubMed: 26605621]

[135]. Niessen KA, Xu M, Paciaroni A, Orecchini A, Snell EH, Markelz AG. Moving in the right direction: Protein vibrational steering function. Biophys J 2017, 112:933–942. [PubMed: 28297652]

[136]. Agarwal M, Alam MP, Chakravarty C. Thermodynamic, Diffusional, and Structural Anomalies in Rigid-Body Water Models. J Phys Chem B 2011, 115:6935–6945. [PubMed: 21553909]

[137]. Ahmed Z, Chou SG, Siegrist K, Flusquellic DF. State-resolved THz spectroscopy and dynamics of crystalline peptide-water systems. Faraday Disc 2011, 150:175–192.

[138]. Chen C, Huang C, Waluyo I, Nordlund D, Weng T, Sokaras D, et al. Solvation structures of protons and hydroxide ions in water. J Chem Phys 2013, 138:154506–7. [PubMed: 23614429]

[139]. Takahashi M, Ishikawa Y. Terahertz vibrations of crystalline-D-glucose and the spectral change in mutual transitions between the anhydride and monohydrate. Chem Phys Lett 2015, 642:29–34.

[140]. Ruggiero MT, Zeitler JA. Resolving the Origins of Crystalline Anharmonicity Using Terahertz Time-Domain Spectroscopy and ab Initio Simulations. J Phys Chem B 2016, 120:11733–11739. [PubMed: 27766874]

[141]. McQuarrie DA. Statistical Mechanics. New York: Harper and Row; 1976.

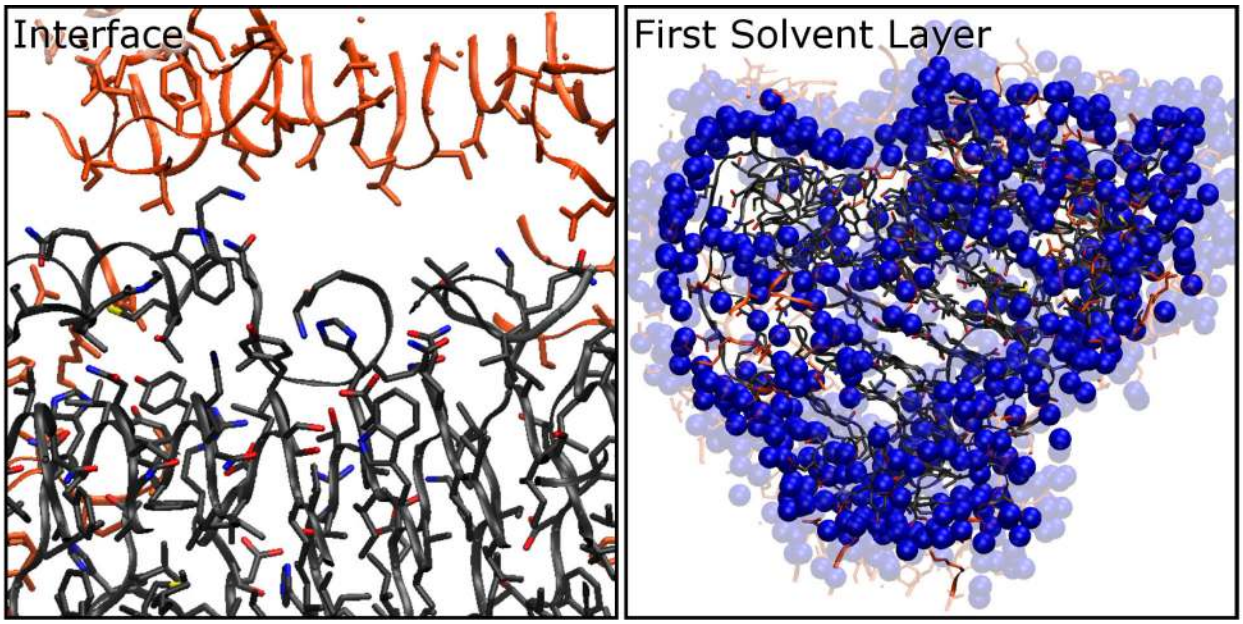

Figure 1:

A selected interface of the pectin lyase protein crystal from PDB structure 1QCX. In the left panel, the β-barrel structure is evident in the repeating loops of the full color asymmetric unit (black = backbone or carbon, red = oxygen, blue = nitrogen, yellow = sulfur) as well as the orange-colored neighboring unit. Near and far clipping was used to show this interface without occlusion by other nearby asymmetric units. This interface, typical of others in the lattice, does not involve tight protein:protein contacts. In the right panel, water molecules (both crystallographic and placed in an amount needed to fill the unit cell volume) within 4Å of both the central asymmetric unit and at least one of its neighbors are shown as solid blue spheres. Water molecules within 8Å of the central asymmetric unit and at least one of its neighbors are shown as faded blue spheres. As with the water, components of neighboring asymmetric units within 4Å or 8Å of the central unit are shown as solid or faded orange wiring, respectively. The "first solvent layer" enshrouds the central asymmetric unit, leaving little room for direct contacts with neighboring lattice proteins. This crystal, which is only 35% solvent (water and ions) by mass, shows how even a low amount of water is distributed to cover individual monomers in a protein crystal.

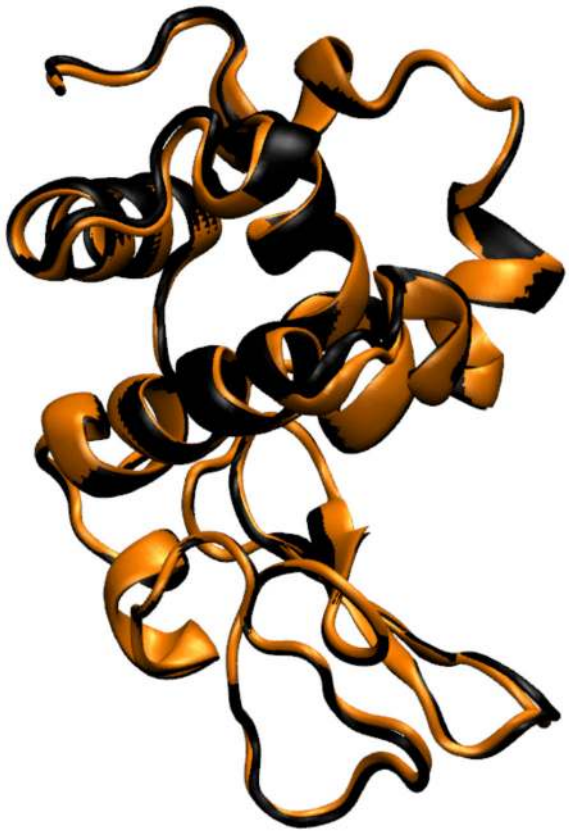

Figure 2:
Backbone cartoons for hen egg white lysozyme. Orange represents the average simulated structure for twelve independent copies of the protein simulated as part of the same crystal, while black represents the X-ray structure.

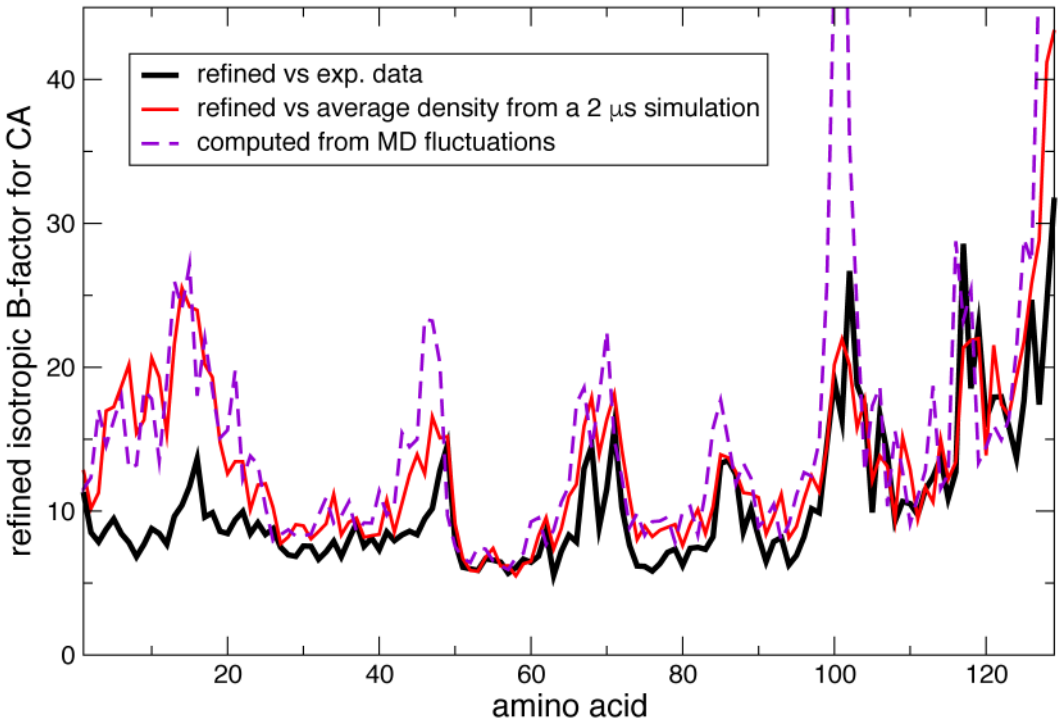

Figure 3:

Isotropic B-factors for Cα atoms in triclinic lysozyme. Black: refined against experimental data (PDB id 4LZT); red: refined against the average electron density from an MD simulation;[38] purple: computed from the mean-square structure fluctuations from the same simulation.

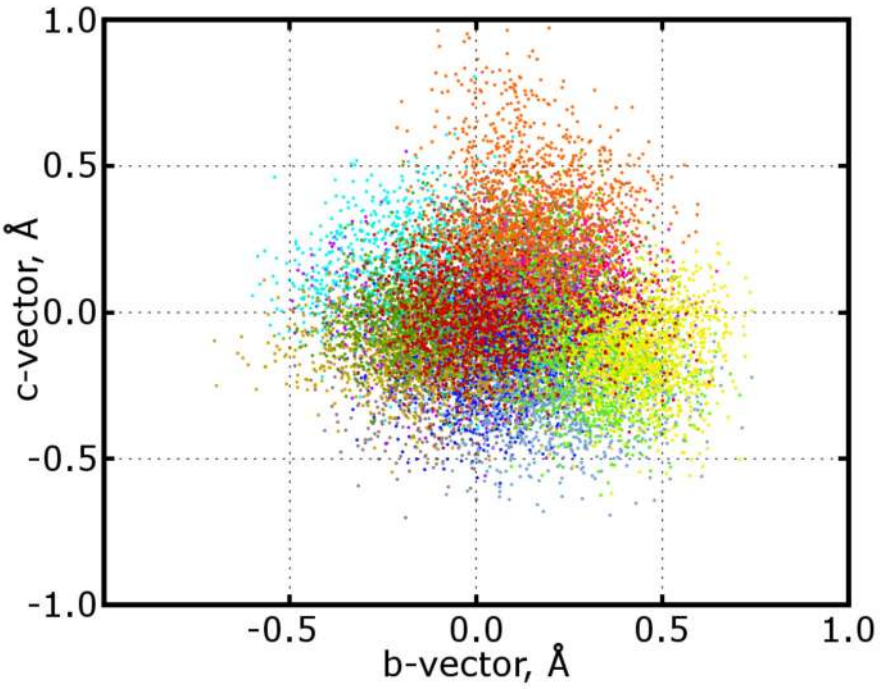

Figure 4:

Centers of mass of the protein chain, projected onto the *bc* crystalographic plane; the origin represents the location in the starting structure (from PDB id 4LZT). Different colors represent each of the chains in a 12 unit cell simulation,[38] and there are 100 equally-spaced snapshots.