Correlated Motions in Structural Biology

*Da Xu ‡, Steve P. Meisburger‡, Nozomi Ando**

Department of Chemistry and Chemical Biology, Cornell University, 259 East Avenue, Ithaca, NY 14853, USA.

‡ These authors contributed equally.

* To whom correspondence should be addressed.

Phone: 607-255-9454, E-mail: nozomi.ando@cornell.edu

KEYWORDS: conformational dynamics, correlated motions, diffuse scattering, allostery, catalysis, evolution

ABSTRACT

Correlated motions in proteins arising from the collective movements of residues have long been proposed to have fundamental importance to key properties of proteins, from allostery and catalysis to evolvability. Recent breakthroughs in structural biology have made it possible to capture proteins undergoing complex conformational changes, yet intrinsic correlated motions within a conformation remain one of the least understood facets of protein structure. For many decades, the analysis of total X-ray scattering held the promise of animating crystal structures with correlated motions. With recent advances in both X-ray detectors and data interpretation methods, this long-held promise can now be met. In this perspective, we will introduce how correlated motions are captured in total scattering and provide guidelines on data collection, interpretation, and validation. As structural biology continues to push the boundaries, we see an opportunity for gaining atomistic insight into correlated motions using total scattering as bridge between theory and experiment.

Introduction

The past decade has seen remarkable developments in structural biology that were once simply unimaginable. A confluence of hardware and software advances in cryo-electron microscopy (cryo-EM) has now made it possible to determine structures that were previously considered unattainable[1]. In the X-ray field, the emergence of 4th generation light sources[2,3] has led to a renaissance of room-temperature studies at both X-ray free electron lasers (XFELs)[4] and synchrotrons[5]. The past year also witnessed a breakthrough in the accuracy of protein structure prediction by machine learning algorithms[6]. On all fronts, structural biology techniques have reached a new level of sophistication. We are arguably experiencing the second major watershed since the "Big Bang"[7] of protein X-ray crystallography.

The goal of structural biology has always been to relate structure to function. Yet since the beginning, it has been known that protein function can be better deduced by visualizing a change in structure. The seminal structures of hemoglobin in different states were critical in revealing how conformational change can be interpreted as inter-subunit allostery and explain cooperative oxygen binding[8]. Likewise, comparisons of lysozyme structures with and without inhibitor bound were the first example in which the active site of an enzyme and its interactions with a ligand were deduced with minimal prior knowledge[9,10]. These foundational studies laid out an extremely successful template for structural biology – if multiple structures of a protein can be solved, we can deduce much about its specific function by careful inspection of the changes. Today, the development of novel approaches to visualize proteins in action is at the forefront of structural biology. In crystallography, there has been a concerted effort to probe room-temperature dynamics either kinetically in response to a perturbation[11–13] or at equilibrium by modeling multiple conformers[14–16]. In cryo-EM, we have seen the development of techniques not only aimed

at classification of single particles in different conformations[17,18] but also in the estimation of a continuous manifold that describes a protein's structure[18–21]. We expect these recent advances, especially in cryo-EM, to revolutionize structural analyses of large-scale conformational changes, such as domain motions.

What remains to be done? Although much can be learned from observing changes in structure, there are collective structural fluctuations within a single conformation, known as correlated motions, that are difficult to visualize but are important to understand. In the context of allostery, subtle correlated motions are implicated in sensing local changes, such as ligand binding, and propagating a signal through the protein to alter its activity. For many proteins, like hemoglobin, the outcome of signal propagation is a large-scale conformational change that can be interpreted to explain changes in activity[22]. However, even when high-resolution structures of different allosteric states are known, the mechanism by which small-scale changes generate a global change is not readily apparent. Moreover, there is a phenomenon known as "dynamic allostery" or "fluctuation-induced allostery" which does not invoke any conformational change. First described by Cooper and Dryden[23], it was argued from statistical mechanics that allosteric signaling can occur solely through different movement patterns within a protein without any discernible changes to the average structure. To date, various instances of dynamic allostery have been identified experimentally, such as inter-subunit communication in catabolite activator protein (CAP)[24] and tuning of ligand-binding affinity by a peripheral alpha helix in PDZ domains[25].

Furthermore, the potential roles of subtle protein motions in catalysis has been one of the greatest debates in modern enzymology[26–32]. For the structural biologist, it is useful to consider how motions may contribute to different parts of the rate constant as described by transition-state

theory[26]. In this framework, the reactant and transition state are in a quasi-equilibrium, such that the rate constant is proportional to a Boltzmann distribution that has an exponential dependence on the activation free energy. The pre-exponential factor in the rate constant includes the transmission coefficient, which can account for phenomena such as quantum mechanical tunneling in reactions that involve the transfer of light particles. This factor is thought to be sensitive to fluctuations in the donor-acceptor distances[30]. For the best-studied system of dihydrofolate reductase (DHFR), a diverse set of experimental and theoretical tools has been applied that suggest that a network of residues undergo sub-angstrom to angstrom-scale correlated motions that are relevant to catalysis[33–37] and that this network is preserved throughout evolution[33,38,39]. Yet it has also been argued that the dominant contribution to enzymatic rate enhancements comes from the structure itself, i.e. that the polar environment of the active site is pre-organized in a way that water molecules are not[40].

To achieve a physical understanding of the unique properties of proteins, it is clear that correlated motions must be investigated in addition to structure (Fig. 1). Thus far, a number of computational approaches have been developed to predict networks of residues that may display correlated motion[22]. These include sequence-based methods that infer the coevolution between residues from multiple sequence alignments (MSA)[41], as well as the physical modeling of structural dynamics by elastic network models (ENM)[42–44] or molecular dynamics (MD) simulations[45] and graph-theory analysis based on proximity and bonding between residues in protein structures[46,47]. Direct measurements of correlated motions, on the other hand, are challenging, and thus far, largely limited to nuclear magnetic resonance (NMR) spectroscopy, where different methods can be applied to cover a wide range of time scales from ps to s[48].

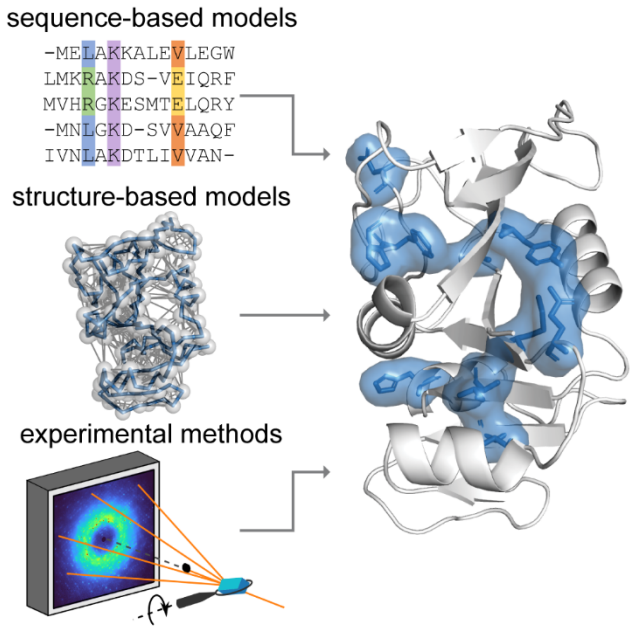

Figure 1. Combining theoretical and experimental approaches to determine how motions are correlated in proteins. Theoretical models include those based on sequence (represented by an MSA) and structure (represented by an ENM). Among experimental methods, total X-ray scattering from crystals stands out for its ability to measure high-resolution structure and correlated motions simultaneously. Shown on the right is a set of residues that was predicted to coevolve in DHFR (PDB: 1RX2) by statistical coupling analysis[41] (shown in blue).

In this perspective, we will focus on a new approach to X-ray crystallography that provides a direct measure of correlated motions: the analysis of total scattering from protein crystals. The total scattering signal contains both Bragg diffraction and diffuse scattering, from which both atomic detail and correlated motions can be obtained[49]. The measurement is unique in that data for both the average structure and its fluctuations can be collected in a single experiment on the same sample, and thus, the two can be compared directly in a self-consistent manner. Furthermore, the total scattering signal provides a powerful restraint for bridging the gap

between experiment and theory. Here, we will begin with a brief introduction on how atomic displacements are represented in protein crystallography (B-factors) in order to explain how information on correlated displacements is contained in diffuse scattering. We will then provide guidelines on data collection, processing, and interpretation. Finally, we will discuss grand challenges and opportunities, particularly in emerging areas of interest, such as relating correlated motions to sequence and evolution.

Understanding conformational disorder in terms of correlated motions

In crystallography, structure determination relies on the integrated intensities of the Bragg reflections – the bright spots that are captured in diffraction images (Fig. 2A). This dataset, which we call "Bragg diffraction," reports on the mean electron density of the unit cell of the crystal, where the average is over time and space. Although the lattice imposes more orderliness on proteins than if they were completely free in solution, the inside of a protein crystal is a crowded but watery environment with non-covalent contacts between neighboring molecules. As a result, individual molecules in a protein crystal sample a conformational ensemble that share similarities with the ensemble in solution[50]. Often, evidence for distinct conformations can be seen in the mean electron density and modeled by accounting for their partial occupancies[51]. In addition, apparent fluctuations in atomic positions result in a local blurring of the electron density. This blurring is quantified during structure refinement by fitting atomic B-factors that specify the width of a Gaussian probability distribution for each atom's displacement from the average position[52].

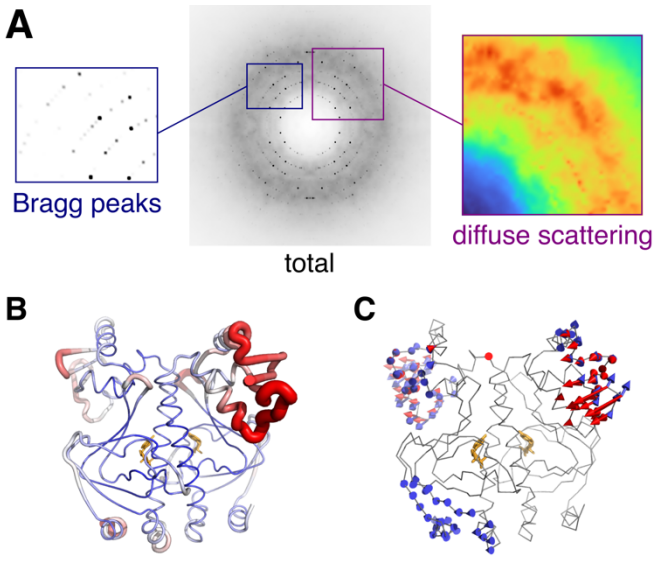

Figure 2. Components of total scattering illustrated using the experimental X-ray structure of CAP (PDB: 1g6n[53]) with dynamics added using ENM simulations[54] of one unit cell. (**A**) The simulated diffraction image contains two signals: Bragg peaks (left) that depend on the average structure, and diffuse scattering (right) that arises from correlated atomic displacements (in this case, vibrations of the ENM). (**B**) B-factors refined to experimental Bragg data vary along the polypeptide chain (blue to red). (**C**) Normal modes of the ENM seem to explain regions of high experimental B-factor. The reality of such collective motions can be verified by diffuse scattering analysis.

Crystallographic B-factors often display a distinct pattern along a polypeptide chain that, for a given protein crystal, is largely reproducible from experiment to experiment (Fig. 2B). Thus, although Bragg data contain no information about displacement correlations, B-factors can help identify potential regions of flexibility and mobility in a protein[55,56]. In addition, motions may be inferred from more sophisticated disorder models applied during crystallographic structure refinement, including normal mode[57], TLS[58], and multi-conformer[14] approaches. In computational

studies of allostery and correlated motions, B-factors are often used for experimental validation[59]. B-factors are particularly relevant to ENMs, whose normal modes naturally predict Gaussian-distributed fluctuations that can be compared directly with experiment[60–62] (Fig. 2C). B-factors are also commonly compared with mean-squared fluctuations in MD simulations, both for validation and for benchmarking force fields[63].

However, there are two major limitations to B-factors. First, although B-factors are used to restrain dynamical models, they are not a particularly powerful restraint. In fact, very different models can account for B-factors equally well, making interpretation ambiguous. For example, it has been questioned whether the correlation between B-factors and ENM displacements is physically meaningful, or if it merely reflects the tendency for greater disorder on the surface of a protein compared with the core[64]. Second, B-factors contain contributions from multiple sources of disorder, and thus, without accounting for these other sources, they overestimate the atomic motions arising from protein motions. As we will describe below, the contribution of non-protein dynamics cannot be treated as a constant, and thus, the use of so-called normalized B-factor[55] is insufficient.

Fortunately, diffraction images contain a second signal called “diffuse scattering” (Fig. 2A). This signal is a direct consequence of disorder: photons lost from the high-resolution Bragg reflections are scattered instead, leaving a faint pattern everywhere in the diffraction image, between and underneath Bragg peaks. As a result, diffuse scattering is directly related to B-factors, but the information content differs. While B-factors come from the average electron density, diffuse scattering depends intimately on how atomic displacements are correlated with each other[49]. The two are thus complementary. By analyzing the diffuse scattering and Bragg

diffraction simultaneously (total scattering analysis), a more complete view of the structural fluctuations in a crystal can be gained[54].

Historically, protein diffuse scattering was difficult to measure accurately, especially in the vicinity of intense Bragg peaks[65]. The use of a direct detector recently enabled the first clear measurement of the three-dimensional diffuse scattering from a protein crystal, allowing for a detailed analysis of the various contributions to the signal (further described below and in Fig. 3)[54]. Importantly, intense near-Bragg features were resolved for the first time, which revealed that the crystals contained phonon-like displacement waves extending over hundreds of angstroms. A key take-away is that lattice disorder resulting from displacement waves accounted for the bulk of the B-factors in the most-ordered atoms in the protein. Thus, it was necessary to subtract this lattice contribution from the B-factors to reveal the true contribution from protein motions.

Now that historical limitations have largely been overcome, we expect biological applications of total scattering to become increasingly common. Thus, it will be important to establish guidelines and best practices to ensure data quality and robust interpretations, as occurred with crystallography, small-angle X-ray scattering, and cryo-EM. In the following, we outline the issues based on our experience so far, both to guide design of future experiments and to help evaluate published results.

How to measure total scattering

With modern detectors, total scattering experiments can be measured in much the same way as conventional, single-crystal diffraction. It may even be possible to extract diffuse scattering patterns from conventional diffraction images deposited in public databases[66,67]. However, without optimizing the experiment, the kind of quantitative, total scattering analysis we describe here

will be difficult. One reason is that proper treatment of background is extremely important for diffuse scattering as it competes with the signal of interest. Unlike in conventional diffraction experiments, total scattering measurements require a separate measurement of the background or its effective elimination. Furthermore, most diffraction data are currently collected at temperatures of 100 K to mitigate radiation damage. Although cryo-cooled crystals also exhibit diffuse scattering[66,68], room temperature is preferred for dynamic studies as the conformational ensembles best resemble the physiological state[15]. Additionally, the cryo-cooling process may add strain to the crystal lattice[69–71], which increases mosaicity and constitutes an additional perturbation to be accounted for. As we detail here, high-quality data can be obtained without cryocooling if extra care is taken in sample selection and experimental setup.

The first and most obvious consideration is signal-to-noise ratio. In terms of the number of photons recorded, diffuse scattering is actually comparable to Bragg diffraction[65]. However, the diffuse signal is much weaker because those photons are spread over the entire detector, while Bragg diffraction is concentrated in sharp peaks. Moreover, much of the diffuse signal is isotropic and relatively featureless, with the more informative signal present as small rapidly varying features. Crystal scattering also competes with background from other sources that can be a challenge to eliminate, such as air in the beam path, liquid on the crystal surface, and other mounting materials like loops and capillaries. Finally, if the data are collected at room temperature, the total exposure budget is severely limited by susceptibility to radiation damage[72,73]. For all of these reasons, it helps considerably if the crystals are large by current standards. While excellent Bragg data may be collected from small ($< 50\ \mu$m) crystals using cryo- or serial crystallography, room-temperature diffuse scattering practically requires larger crystals (smallest

dimension > 100 μm). The possibility of using XFELs for total scattering from microcrystals is revisited in the final section.

The second consideration is diffraction quality. It is sometimes assumed that diffuse scattering requires poorly diffracting crystals. While the diffuse signal is most obvious when Bragg peaks are weak[74], the evidence so far is that all protein crystals produce strong diffuse scattering[49], even those that diffract to exceptionally high resolution[54]. Poor diffraction quality often results from high mosaicity, a kind of macroscopic disorder often described in terms of distortions and misalignment of so-called mosaic blocks[70], as opposed to the microscopic lattice disorder measured by diffuse scattering. Mosaicity broadens both Bragg peaks and diffuse scattering features, setting a fundamental limit on how finely the diffuse map can be sampled. Thus, although a mosaicity of 1° is sometimes tolerable for Bragg datasets, it would be problematic for diffuse scattering. The highly detailed diffuse maps from triclinic lysozyme[54] benefited from the low apparent mosaicity of ~0.02°. In a similar vein, X-ray sources optimized for high flux often have high energy bandwidth and beam divergence and therefore produce a similar broadening effect[75].

The final consideration is data collection strategy. The diffuse map results from merging individual diffraction images, which represent slices through the 3D reciprocal space in different orientations. The merging process is an opportunity to estimate and correct for uncontrolled variables in the measurement, such as the volume of the crystal in the X-ray beam and variations in response across the detector. As in anomalous diffraction experiments, the best collection strategy is to aim for high redundancy[76], so that regions of reciprocal space are observed multiple times independently. Collecting from several different crystals or tilting the spindle axis also helps fill in blind spots in reciprocal space caused by Ewald sphere curvature and physical gaps

between detector panels. The background scattering from the diffraction instrument should also be measured and subtracted from each image. It may be necessary to collect backgrounds as a function of spindle angle if the mounting materials cast a shadow on the detector[54]. Some background scattering can also be removed computationally during merging, as long as it is isotropic and present in only some rotation angles[54].

The internal consistency of the data (i.e. 3D diffuse map) can be assessed using metrics commonly in use for Bragg data such as $CC_{1/2}$, the correlation coefficient between random half-datasets binned by resolution shell[77]. The map quality sets an upper limit on model-data agreement, and this can be calculated from $CC_{1/2}$ using the CC* estimate[54,77]. Note that statistics such as $CC_{1/2}$ depend on how finely the map is sampled, which is not an issue for Bragg data, so care must be taken to always specify the sampling and to compare datasets sampled on the same grid.

How to interpret total scattering

As described above, one set of diffraction images produces two datasets: the integrated Bragg intensities and the 3D diffuse map. For total scattering analysis, the datasets first need to be placed on the same intensity scale relative to each other. Both datasets can then be placed on an absolute intensity scale of electron units[16,54]. The inelastic scattering component, which contains no structural information, can then be subtracted from the diffuse map, leaving the elastic scattering for the remainder of the analyses (Fig. 3A).

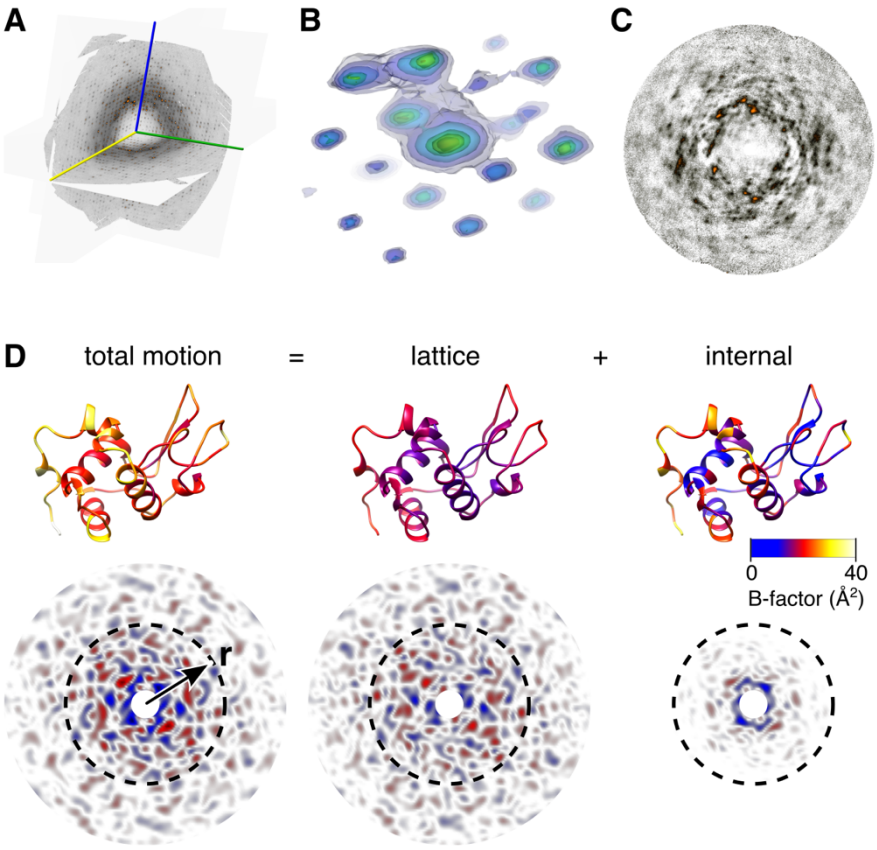

Figure 3. Total scattering analysis separates motion into lattice and internal components[54]. (A) Three-dimensional map of diffuse scattering from triclinic lysozyme shown as intersecting central slices. A movie showing this volume from multiple perspectives is available online. The scattering includes an intense isotropic ring that may be subtracted to better visualize the halo and cloudy features. (B) Around Bragg peaks are three-dimensional halos (shown here as transparent contours, blue to yellow) attributed to thermally excited lattice vibrations. (C) Cloudy features due to short-ranged correlated motion are most visible in sections mid-way between Bragg planes. (D) Total motion and correlations are quantified using B-factor (top, PDB 6o2h[54]) and diffuse Patterson maps (bottom), which report electron density fluctuations vs. inter-atomic vector, **r**. A lattice dynamics model fit to diffuse halos accounts for most of the B-factor for well-ordered atoms (total vs. lattice, top) and the correlated motions at large distances (total vs. lattice,

bottom), but underestimates those at short distance (r < 10 Å, dashed circles). An ENM describing protein dynamics was fit to the residual B-factors (top right). The simulated diffuse Patterson of the protein dynamics model (bottom right) explains the remaining short-ranged correlations.

The resulting diffuse map can be further divided into different components. The most noticeable feature is a broad ring-shaped background at ~3 Å resolution (Fig. 3A), which arises from both the disordered solvent and from the protein[78]. Although it is potentially informative and useful for evaluating dynamical models, it can be subtracted for visualization purposes. Once the background is "turned off," the fine features become visible. The most intense are typically found close to the Bragg peaks and are associated with lattice disorder. With sufficient sampling, halos with distinctive three-dimensional shapes can be observed (Fig. 3B). Different physical processes produce distinct intensity profiles[49]. For example, if the intensity decays away from the Bragg peak with a power-law exponent of -2, then acoustic phonon-like vibrations are a likely source[54]. Far from the Bragg peaks, it is common to see a cloudy pattern (Fig. 3C), which includes the scattering from collective motions of the protein[79]. After these initial inspections, the diffuse map can be interpreted by rigorous comparison with the simulated scattering from atomistic simulations.

Building an accurate model of the lattice disorder is the key to total scattering interpretation. The reasons are twofold. First, it allows for the lattice component of the B-factors to be estimated and subtracted (Fig. 3D, top). Because lattice dynamics involve rotations (not just translations) of molecules[80], its contribution to the B-factors cannot be assumed to be a constant value. Second, it allows the short-ranged correlations associated with lattice motion to be accounted for when analyzing the cloudy pattern. Thankfully, lattice vibration models have relatively few degrees of

freedom that can be fit using even just a subset of intense three-dimensional halos[54]. In contrast to previous methods, which relied on cloudy scattering exclusively[49,68], refining the lattice model to the halos allows for unambiguous separation of lattice motion from internal motion.

Once the lattice disorder is understood in detail, we can focus on the continuous cloudy pattern, which extends throughout reciprocal space and is associated with the correlated motions of atoms within the unit cell. Separating the contributions from lattice and internal motion to the cloudy pattern in reciprocal space is difficult. However, by calculating the Fourier transform of the diffuse scattering intensities, i.e. the "diffuse Patterson map"[54], the data are transformed into a readily interpretable form (Fig. 3D, bottom left). The diffuse Patterson represents electron density fluctuations as a function of pairwise distance between atoms. The part of this map near the origin (e.g. smaller than the size of the protein) is affected by correlations that are short-ranged, while further from the origin, the correlations between unit cells predominate. To reveal the correlations from protein dynamics, one option is to simulate the diffuse Patterson from the lattice dynamics model (Fig. 3D, bottom center) and subtract it from the experimental Patterson. However, this must be done very carefully, as any errors in the model could produce spurious signals. A safer approach is to simulate internal motions riding on top of lattice motions (Fig. 3D, bottom right) and to check whether adding internal motions improves agreement in this central region of the Patterson. Importantly, model-data agreement should be checked with both datasets, Bragg and diffuse. Thus, the combination of internal and lattice motions should agree with the B-factors.

Fitting a dynamical model to diffuse scattering and B-factors simultaneously is a long-standing goal of this field[81] that has yet to be realized in a robust way. However, it is generally possible to derive several realistic candidate models of protein dynamics from the B-factors corrected for

lattice disorder and rank them according to agreement with the diffuse Patterson. Such an approach has the additional advantage of illustrating how well the data discriminate between candidates. Although there is not yet an unbiased indicator to prevent overfitting that is comparable to R_{work}/R_{free} in conventional crystallographic refinement, we can employ the dynamical equivalent of the omit map by testing whether model refinement with certain motions of interest suppressed worsens the model-data agreement[54].

As in conventional crystallography, multiple metrics should be used for assessing model accuracy. In using model-data correlations (CCs), it is important to note that the intense halos will dominate the statistics. One way to put lattice and internal motion on a more equal footing is to focus on the central part of the Patterson only for computing CCs[54]. Although CCs are useful, they are normalized by the variance in each resolution shell, and thus it is important to verify that the model correctly predicts the signal strength (variance) vs. resolution[68]. It should be noted that diffuse scattering patterns tend to correlate with the molecular transform, and therefore even a very crude model can easily obtain CCs of ~0.5[49]. Substantially better agreement is needed to establish the accuracy of a model.

Challenges and Opportunities for the Future

Over the past several decades, the importance of correlated motions in protein function has become increasingly apparent. However, as with all areas of protein science, the information we can obtain by examining the properties of one specific protein sequence is limited. In truth, all proteins have evolutionarily related counterparts, which may display similar yet divergent behaviors. Thus, an emerging trend is to use evolution as an additional dimension to understand protein functions by comparing sequences that share the same evolutionary origin[82,83]. In the

context of enzyme catalysis, the utility of this approach is exemplified by studies on DHFR, where coupled networks of residues were identified by mixed quantum/classical MD simulations of the hydride transfer reaction and correlated with sequence conservation across multiple organisms[33]. In a later study[39], a curated MSA of DHFR was constructed in order to identify rare events that demark changes from one invariant sequence motif to another. The dynamic pattern of the active-site Met20 loop from one species was engineered into that of another by substituting divergent sites with evolutionary significance identified in this way[39]. Statistical coupling analysis (SCA)[38] of an MSA of DHFR also revealed coevolving residues that form a contiguous network in the structure and show strong correlation with residues identified by NMR to be involved in millisecond dynamics during the catalytic cycle[35].

In addition to comparing extant sequences from different organisms, sequences resurrected by ancestral sequence reconstruction can be compared with extant homologs to track the evolutionary trajectory of correlated motions and their role in protein function[84]. For example, Zou et al.[85] utilized MD simulations to compare the dynamics of a β-lactamase specific for penicillin degradation with its ancestral counterparts, which show promiscuity but no substantial structural differences. They were able to identify regions with altered dynamics and key residues that contribute to the difference, signifying the importance of dynamics in the tradeoff between activity and specificity in the evolution of β-lactamase.

Conformational dynamics are also thought to be a key contributor to protein evolvability[86,87], defined as the ability of proteins to adopt new functions through mutations. According to this view, protein fluctuations sample minor conformations that serve as precursors for new functions. To test such intriguing proposals, directed evolution of novel enzymes is of particular interest, where sequences from different iterations may be compared to elucidate the relationship

between correlated motions and changes in traits[83,88,89]. As an example, variants of artificial retro-aldolases that were produced by directed evolution[90] were later examined by MD[89], and a population shift toward catalytically competent arrangement of active-site residues was observed along the evolutionary pathway, which interestingly, also included distal mutations. With a novel algorithm, residues exhibiting correlated motions were also inferred from the MD trajectory which further rationalized the conformational conversion. Recently, directed evolution of a bifunctional ancestor enzyme was demonstrated using a library of mutants with altered backbone dynamics generated by transposon-based random insertions or deletions (indels)[91].

As evident from the examples above, the emerging interest in evolution of correlated motions relies heavily on the synergy between computational and experimental methods. We see several opportunities for advancing these studies with total scattering analysis. Perhaps the most critical area is in improving the accuracy of MD. With the recent advance in measurement, it has become clear that all-atom MD is not yet able to accurately predict correlated motions implied by diffuse scattering[54]. A major contributor to this discrepancy is the fact that the predictive ability of MD (RMSD of ~0.4 Å for triclinic lysozyme[92]) cannot match the coordinate precision of Bragg diffraction (~0.03 Å for PDB 6o2h[93]), which has a cascading effect on the ability of MD to predict the diffuse signal[54]. In the future, methods to restrain or otherwise improve MD using total scattering data will be of great importance, both for the accuracy of the simulation itself and for gaining atomistic insight into correlated networks in proteins. Integration of multiple experimental methods will also lead to a deeper understanding of the information contained in different types of data. Of particular interest is combining total scattering analysis with solid-state NMR, which can both be performed on protein crystals[94]. Finally, the quest to understand the link between evolutionary sequence correlations and functional dynamics represents a grand

challenge in molecular biophysics (Fig. 4). From sequence analysis, it is clear that selection pressure drives certain groups of residues to coevolve and that these groups can highlight areas of functional importance in a structure[95]. In the case of the PDZ domain, a deep double-mutational library of multiple homologs was experimentally characterized to show that the SCA matrix couplings can have thermodynamic interpretation[95]. However, a general connection between co-evolving residues and *motion* has not been established. Thus, direct comparisons between theory and dynamic experiments will play a crucial role in in gaining a precise understanding of how correlated motions can be predicted from sequence.

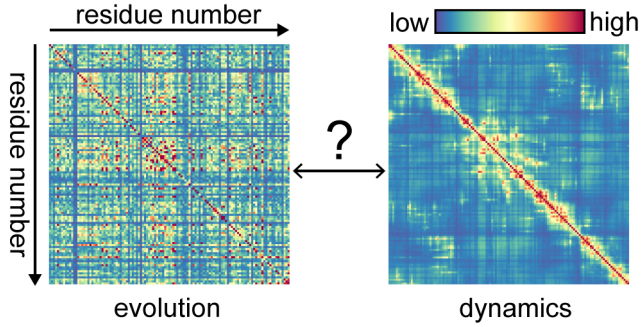

Figure 4. Evolutionary and dynamic perspectives on residue-residue correlations. (Left) Evolutionary correlation according to SCA[41] applied to an MSA of a hydrolase family that contains lysozyme (Pfam PF00062). (Right) Displacement correlations[96] in lysozyme according to an ENM derived from total scattering analysis (Fig. 3). Establishing the connection between these two perspectives is necessary to fully understand protein function and allostery.

To bring total scattering analysis to a wider audience, several areas of development are of high priority. The first is to get around the issue of crystal size for room-temperature studies. Although we recommend maximizing signal to noise with large (>100 μm) crystals, inevitably, we will need to work with smaller ones in a serial fashion, i.e. by collecting one or a few frames

of images from many crystals. We see no fundamental reason why such experiments cannot be done at synchrotrons or X-FELs today, however accurate measurement of total scattering has special requirements for the X-ray detector, properties of the beam such as divergence and energy bandwidth, and sample environment. The primary requirements for a detector are that photons are detected directly (i.e. via a semiconductor), that dynamic range is sufficient to measure Bragg peaks and diffuse scattering simultaneously, and that any variability in pixel response is well-characterized. We must also consider how to minimize background scattering and computationally correct for sources that can't be eliminated. One promising experimental approach is to eschew X-ray windows and instead use humidified helium gas to prevent crystal dehydration[97]. Second is the issue of data processing software. As with all structural techniques, the availability of user-friendly software packages will be important for bringing total scattering analysis to a wide audience. This is contingent on standardization in the field, and we hope that our recommendations for data collection, processing, and validation will accelerate this process. Finally, we foresee a potential for machine-learning methods in the future, either in data processing or in interpretation of total scattering data[98]. For example, it would be of great interest to use machine-learning to classify signals that are from different types of motion. Towards such a goal, we as a community must first produce the learning data, namely, a collection of total scattering datasets that are fully understood.

These are exciting times to be a structural biologist. With technical breakthroughs in cryo-EM and X-ray diffraction experiments, we have an unprecedented array of tools to address virtually any structural question. But perhaps more importantly, these methods give a new window into the correlated motions of proteins. We anticipate that in the coming decade, total scattering will provide a much-needed bridge between dynamics on short time scales from NMR and MD with

highly precise structural measurements required for insight into chemical mechanism. Especially when combined with bioinformatic approaches to evolution, structural biology may finally answer the question that started the field: how does allostery work?

AUTHOR INFORMATION

Corresponding Author

*nozomi.ando@cornell.edu

Author Contributions

All authors contributed intellectually and collaborated closely to write the manuscript. Elastic networking modeling of CAP and diffraction simulations were performed by SPM. Statistical coupling analysis of the hydrolase family was performed by DX. ‡These authors contributed equally.

Funding Sources

This work was supported by National Science Foundation (NSF) grant MCB-1942668 and National Institutes of Health, National Institute of General Medical Sciences (NIH/NIGMS) grant GM124847.

ACKNOWLEDGMENT

The authors are grateful for Sol Gruner (Cornell) for his feedback and helpful discussions.

ABBREVIATIONS

EM, electron microscopy; XFEL, X-ray free electron laser; MSA, multiple sequence alignment; ENM, elastic network model; MD, molecular dynamics; CAP, catabolite activator protein; DHFR, dihydrofolate reductase; NMR, nuclear magnetic resonance; TLS, translation libration

screw-axis; CC, correlation coefficient; PDB, Protein Data Bank; RMSD, root mean squared distance; SCA, statistical coupling analysis.

REFERENCES

(1) Kühlbrandt, W. (2014) The resolution revolution. *Science 343*, 1443–1444.

(2) Reich, E. S. (2013) Ultimate upgrade for US synchrotron. *Nature 501*, 148–149.

(3) Hand, E. (2009) X-ray free-electron lasers fire up. *Nature 461*, 708–709.

(4) Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., Hunter, M. S., Schulz, J., Deponte, D. P., Weierstall, U., Doak, R. B., Maia, F. R. N. C., Martin, A. V., Schlichting, I., Lomb, L., Coppola, N., Shoeman, R. L., Epp, S. W., Hartmann, R., Rolles, D., Rudenko, A., Foucar, L., Kimmel, N., Weidenspointner, G., Holl, P., Liang, M., Barthelmess, M., Caleman, C., Boutet, S., Bogan, M. J., Krzywinski, J., Bostedt, C., Bajt, S., Gumprecht, L., Rudek, B., Erk, B., Schmidt, C., Hömke, A., Reich, C., Pietschner, D., Ströder, L., Hauser, G., Gorke, H., Ullrich, J., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kühnel, K. U., Messerschmidt, M., Bozek, J. D., Hau-Riege, S. P., Frank, M., Hampton, C. Y., Sierra, R. G., Starodub, D., Williams, G. J., Hajdu, J., Timneanu, N., Seibert, M. M., Andreasson, J., Rocker, A., Jönsson, O., Svenda, M., Stern, S., Nass, K., Andritschke, R., Schröter, C. D., Krasniqi, F., Bott, M., Schmidt, K. E., Wang, X., Grotjohann, I., Holton, J. M., Barends, T. R. M., Neutze, R., Marchesini, S., Fromme, R., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Andersson, I., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B., and Spence, J. C. H. (2011) Femtosecond X-ray protein nanocrystallography. *Nature 470*, 73–78.

(5) Stellato, F., Oberthür, D., Liang, M., Bean, R., Gati, C., Yefanov, O., Barty, A., Burkhardt, A.,

Fischer, P., Galli, L., Kirian, R. A., Meyer, J., Panneerselvam, S., Yoon, C. H., Chervinskii, F., Speller, E., White, T. A., Betzel, C., Meents, A., and Chapman, H. N. (2014) Room-temperature macromolecular serial crystallography using synchrotron radiation. *IUCrJ 1*, 204–212.

(6) Callaway, E. (2020) "It will change everything": DeepMind's AI makes gigantic leap in solving protein structures. *Nature 588*, 203–204.

(7) Rossmann, M. G. (2017) John C. Kendrew and His Times. *J. Mol. Biol. 429*, 2601–2602.

(8) Muirhead, H., and Perutz, M. F. (1963) Structure Of Hæmoglobin: A Three-Dimensional Fourier Synthesis of Reduced Human Haemoglobin at 5.5 Å Resolution. *Nature 199*, 633–638.

(9) Blake, C., Koenig, D., Mair, G., North, A., Phillips, D., and Sarma, V. (1965) The three-dimensional structure of hen eggwhite lysozyme. *Nature 206*, 757–761.

(10) Johnson, L. N., and Phillips, D. C. (1965) Structure of Some Crystalline Lysozyme-Inhibitor Complexes Determined by X-Ray Analysis At 6 Å Resolution. *Nature 206*, 761–763.

(11) Kupitz, C., Basu, S., Grotjohann, I., Fromme, R., Zatsepin, N. A., Rendek, K. N., Hunter, M. S., Shoeman, R. L., White, T. A., Wang, D., James, D., Yang, J. H., Cobb, D. E., Reeder, B., Sierra, R. G., Liu, H., Barty, A., Aquila, A. L., Deponte, D., Kirian, R. A., Bari, S., Bergkamp, J. J., Beyerlein, K. R., Bogan, M. J., Caleman, C., Chao, T. C., Conrad, C. E., Davis, K. M., Fleckenstein, H., Galli, L., Hau-Riege, S. P., Kassemeyer, S., Laksmono, H., Liang, M., Lomb, L., Marchesini, S., Martin, A. V., Messerschmidt, M., Milathianaki, D., Nass, K., Ros, A., Roy-Chowdhury, S., Schmidt, K., Seibert, M., Steinbrener, J., Stellato, F., Yan, L., Yoon, C., Moore, T. A., Moore, A. L., Pushkar, Y., Williams, G. J., Boutet, S., Doak, R. B., Weierstall, U., Frank, M., Chapman, H. N., Spence, J. C. H., and Fromme, P. (2014) Serial time-resolved crystallography

of photosystem II using a femtosecond X-ray laser. *Nature 513*, 261–265.

(12) Beyerlein, K. R., Dierksmeyer, D., Mariani, V., Kuhn, M., Sarrou, I., Ottaviano, A., Awel, S., Knoska, J., Fuglerud, S., Jönsson, O., Stern, S., Wiedorn, M. O., Yefanov, O., Adriano, L., Bean, R., Burkhardt, A., Fischer, P., Heymann, M., Horke, D. A., Jungnickel, K. E. J., Kovaleva, E., Lorbeer, O., Metz, M., Meyer, J., Morgan, A., Pande, K., Panneerselvam, S., Seuring, C., Tolstikova, A., Lieske, J., Aplin, S., Roessle, M., White, T. A., Chapman, H. N., Meents, A., and Oberthuer, D. (2017) Mix-and-diffuse serial synchrotron crystallography. *IUCrJ 4*, 769–777.

(13) Hekstra, D. R., White, K. I., Socolich, M. A., Henning, R. W., Šrajer, V., and Ranganathan, R. (2016) Electric-field-stimulated protein mechanics. *Nature 540*, 400–405.

(14) Tom Burnley, B., Afonine, P. V., Adams, P. D., and Gros, P. (2012) Modelling dynamics in protein crystal structures by ensemble refinement. *Elife 2012*, 1–29.

(15) Fraser, J. S., Van Den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N., and Alber, T. (2011) Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A. 108*, 16247–16252.

(16) Lang, P. T., Holton, J. M., Fraser, J. S., and Alber, T. (2014) Protein structural ensembles are revealed by redefining X-ray electron density noise. *Proc. Natl. Acad. Sci. U. S. A. 111*, 237–242.

(17) Scheres, S. H. W. (2016) Chapter Six - Processing of Structurally Heterogeneous Cryo-EM Data in RELION, in *The Resolution Revolution: Recent Advances In cryoEM*, pp 125–157. Academic Press.

(18) Punjani, A., and Fleet, D. J. (2021) 3D variability analysis: Resolving continuous flexibility

and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol. 213*, 107702.

(19) Frank, J., and Ourmazd, A. (2016) Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods 100*, 61–67.

(20) Moscovich, A., Halevi, A., Andén, J., and Singer, A. (2020) Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes. *Inverse Probl. 36*.

(21) Zhong, E. D., Bepler, T., Berger, B., and Davis, J. H. (2021) CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods 18*, 176–185.

(22) Thirumalai, D., Hyeon, C., Zhuravlev, P. I., and Lorimer, G. H. (2019) Symmetry, Rigidity, and Allosteric Signaling: From Monomeric Proteins to Molecular Machines. *Chem. Rev. 119*, 6788–6821.

(23) Cooper, A., and Dryden, D. T. F. (1984) Allostery without conformational change. *Eur. Biophys. J. 11*, 103–109.

(24) Popovych, N., Sun, S., Ebright, R. H., and Kalodimos, C. G. (2006) Dynamically driven protein allostery. *Nat. Struct. Mol. Biol. 13*, 831–838.

(25) Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J., and Lee, A. L. (2009) Hidden dynamic allostery in a PDZ domain. *Proc. Natl. Acad. Sci. 106*, 18249–18254.

(26) Nagel, Z. D., and Klinman, J. P. (2009) A 21st century revisionist's view at a turning point in enzymology. *Nat. Chem. Biol. 5*, 543–550.

(27) Schwartz, S. D., and Schramm, V. L. (2009) Enzymatic transition states and dynamic motion

in barrier crossing. *Nat. Chem. Biol. 5*, 551–558.

(28) Benkovic, S. J., Hammes, G. G., and Hammes-Schiffer, S. (2008) Free-energy landscape of enzyme catalysis. *Biochemistry 47*, 3317–3321.

(29) Kamerlin, S. C. L., and Warshel, A. (2010) At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis. *Proteins Struct. Funct. Bioinforma. 78*, 1339–1375.

(30) Kohen, A. (2015) Role of dynamics in enzyme catalysis: Substantial versus semantic controversies. *Acc. Chem. Res. 48*, 466–473.

(31) Hanoian, P., Liu, C. T., Hammes-Schiffer, S., and Benkovic, S. (2015) Perspectives on electrostatics and conformational motions in enzyme catalysis. *Acc. Chem. Res. 48*, 482–489.

(32) Warshel, A., and Bora, R. P. (2016) Perspective: Defining and quantifying the role of dynamics in enzyme catalysis. *J. Chem. Phys. 144*.

(33) Agarwal, P. K., Billeter, S. R., Rajagopalan, P. T. R., Benkovic, S. J., and Hammes-Schiffer, S. (2002) Network of coupled promoting motions in enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A. 99*, 2794–2799.

(34) Wang, L., Goodey, N. M., Benkovic, S. J., and Kohen, A. (2006) Coordinated effects of distal mutations on environmentally coupled tunneling in dihydrofolate reductase. *Proc. Natl. Acad. Sci. U. S. A. 103*, 15753–15758.

(35) Boehr, D. D., McElheny, D., Dyson, H. J., and Wrightt, P. E. (2006) The dynamic energy landscape of dihydrofolate reductase catalysis. *Science 313*, 1638–1642.

(36) Bhabha, G., Lee, J., Ekiert, D. C., Gam, J., Wilson, I. a, Dyson, H. J., Benkovic, S. J., and Wright, P. E. (2011) A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis. *Science 332*, 234–238.

(37) Stojković, V., Perissinotti, L. L., Willmer, D., Benkovic, S. J., and Kohen, A. (2012) Effects of the donor-acceptor distance and dynamics on hydride tunneling in the dihydrofolate reductase catalyzed reaction. *J. Am. Chem. Soc. 134*, 1738–1745.

(38) Reynolds, K. A., McLaughlin, R. N., and Ranganathan, R. (2011) Hot spots for allosteric regulation on protein surfaces. *Cell 147*, 1564–1575.

(39) Liu, C. T., Hanoian, P., French, J. B., Pringle, T. H., Hammes-Schiffer, S., and Benkovic, S. J. (2013) Functional significance of evolving protein sequence in dihydrofolate reductase from bacteria to humans. *Proc. Natl. Acad. Sci. U. S. A. 110*, 10159–10164.

(40) Liu, H., and Warshel, A. (2007) The catalytic effect of dihydrofolate reductase and its mutants is determined by reorganization energies. *Biochemistry 46*, 6011–6025.

(41) Rivoire, O., Reynolds, K. A., and Ranganathan, R. (2016) Evolution-Based Functional Decomposition of Proteins. *PLOS Comput. Biol. 12*, 1–26.

(42) Ming, D., and Wall, M. E. (2005) Quantifying allosteric effects in proteins. *Proteins Struct. Funct. Genet. 59*, 697–707.

(43) Zheng, W., Brooks, B. R., and Thirumalai, D. (2009) Allosteric transitions in biological nanomachines are described by robust normal modes of elastic networks. *Curr. Protein Pept. Sci. 10*, 128–132.

(44) Atilgan, C., and Atilgan, A. R. (2009) Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Comput. Biol. 5*.

(45) Hertig, S., Latorraca, N. R., and Dror, R. O. (2016) Revealing Atomic-Level Mechanisms of Protein Allostery with Molecular Dynamics Simulations. *PLoS Comput. Biol. 12*, 1–16.

(46) Amor, B. R. C., Schaub, M. T., Yaliraki, S. N., and Barahona, M. (2016) Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. *Nat. Commun. 7*, 12477.

(47) Wang, J., Jain, A., McDonald, L. R., Gambogi, C., Lee, A. L., and Dokholyan, N. V. (2020) Mapping allosteric communications within individual proteins. *Nat. Commun. 11*, 3862.

(48) Kovermann, M., Rogne, P., and Wolf-Watz, M. (2016) Protein dynamics and function from solution state NMR spectroscopy. *Q. Rev. Biophys. 49*, e6.

(49) Meisburger, S. P., Thomas, W. C., Watkins, M. B., and Ando, N. (2017) X-ray Scattering Studies of Protein Structural Dynamics. *Chem. Rev. 117*, 7615–7672.

(50) Torchia, D. A. (2015) NMR studies of dynamic biomolecular conformational ensembles. *Prog. Nucl. Magn. Reson. Spectrosc. 84–85*, 14–32.

(51) Smith, J. L., Hendrickson, W. A., Honzatko, R. B., and Sheriff, S. (1986) Structural Heterogeneity in Protein Crystals. *Biochemistry 25*, 5018–5027.

(52) Rupp, B. (2009) Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology. Garland Science.

(53) Passner, J. M., Schultz, S. C., and Steitz, T. A. (2000) Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 Å resolution. *J. Mol. Biol. 304*, 847–859.

(54) Meisburger, S. P., Case, D. A., and Ando, N. (2020) Diffuse X-ray scattering from correlated motions in a protein crystal. *Nat. Commun. 11*, 1271.

(55) Parthasarathy, S., and Murthy, M. R. N. (1997) Analysis of temperature factor distribution in high-resolution protein structures. *Protein Sci.* 2561–2567.

(56) Carugo, O., and Argos, P. (1998) Accessibility to internal cavities and ligand binding sites monitored by protein crystallographic thermal factors. *Proteins Struct. Funct. Genet. 31*, 201–213.

(57) Poon, B. K., Chen, X., Lu, M., Vyas, N. K., Quiocho, F. A., Wang, Q., and Ma, J. (2007) Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42-Å crystallographic resolution. *Proc. Natl. Acad. Sci. U. S. A. 104*, 7869–7874.

(58) Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. Sect. D Biol. Crystallogr. 62*, 439–450.

(59) Sun, Z., Liu, Q., Qu, G., Feng, Y., and Reetz, M. T. (2019) Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chem. Rev.*

(60) Tirion, M. M. (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett. 77*, 1905–1908.

(61) Bahar, I., Atilgan, A. R., and Erman, B. (1997) Direct evaluation of thermal fluctuations in

proteins using a single-parameter harmonic potential. *Fold. Des. 2*, 173–181.

(62) Riccardi, D., Cui, Q., and Phillips, G. N. (2010) Evaluating elastic network models of crystalline biological molecules with temperature factors, correlated motions, and diffuse X-ray scattering. *Biophys. J. 99*, 2616–2625.

(63) Cerutti, D. S., Freddolino, P. L., Duke, R. E., and Case, D. A. (2010) Simulations of a protein crystal with a high resolution X-ray structure: Evaluation of force fields and water models. *J. Phys. Chem. B 114*, 12811–12824.

(64) Soheilifard, R., Makarov, D. E., and Rodin, G. J. (2008) Critical evaluation of simple network models of protein dynamics and their comparison with crystallographic B-factors. *Phys. Biol. 5*.

(65) Meisburger, S. P., and Ando, N. (2017) Correlated motions from crystallography beyond diffraction. *Acc. Chem. Res. 50*, 580–583.

(66) Peck, A., Poitevin, F., and Lane, T. J. (2018) Intermolecular correlations are necessary to explain diffuse scattering from protein crystals. *IUCrJ 5*, 211–222.

(67) Van Benschoten, A. H., Liu, L., Gonzalez, A., Brewster, A. S., Sauter, N. K., Fraser, J. S., and Wall, M. E. (2016) Measuring and modeling diffuse scattering in protein X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A. 113*, 4069–4074.

(68) Polikanov, Y. S., and Moore, P. B. (2015) Acoustic vibrations contribute to the diffuse scatter produced by ribosome crystals. *Acta Crystallogr. Sect. D Biol. Crystallogr. 71*, 2021–2031.

(69) Kriminski, S., Caylor, C. L., Nonato, M. C., Finkelstein, K. D., and Thorne, R. E. (2002) Flash-cooling and annealing of protein crystals. *Acta Crystallogr. Sect. D Biol. Crystallogr. 58*,

459–471.

(70) Nave, C. (1998) A description of imperfections in protein crystals. *Acta Crystallogr. Sect. D Biol. Crystallogr. 54*, 848–853.

(71) Vahedi-Faridi, A., Lovelace, J., Bellamy, H. D., Snell, E. H., and Borgstahl, G. E. O. (2003) Physical and structural studies on the cryocooling of insulin crystals. *Acta Crystallogr. - Sect. D Biol. Crystallogr. 59*, 2169–2182.

(72) Blake, C., and Phillips, D. C. (1962) Effects of X-irradiaton on single crystals of myoglobin, in *Proceedings of the Symposium on the Biological Effects of Ionising Radiation at the Molecular Level*, pp 183–191. International Atomic Energy Agency, Vienna.

(73) Ravelli, R. B., and Garman, E. F. (2006) Radiation damage in macromolecular cryocrystallography. *Curr. Opin. Struct. Biol. 16*, 624–629.

(74) Ayyer, K., Yefanov, O. M., Oberthür, D., Roy-Chowdhury, S., Galli, L., Mariani, V., Basu, S., Coe, J., Conrad, C. E., Fromme, R., Schaffer, A., Dörner, K., James, D., Kupitz, C., Metz, M., Nelson, G., Xavier, P. L., Beyerlein, K. R., Schmidt, M., Sarrou, I., Spence, J. C. H., Weierstall, U., White, T. A., Yang, J. H., Zhao, Y., Liang, M., Aquila, A., Hunter, M. S., Robinson, J. S., Koglin, J. E., Boutet, S., Fromme, P., Barty, A., and Chapman, H. N. (2016) Macromolecular diffractive imaging using imperfect crystals. *Nature 530*, 202–206.

(75) Bellamy, H. D., Snell, E. H., Lovelace, J., Pokross, M., and Borgstahl, G. E. O. (2000) The high-mosaicity illusion: Revealing the true physical characteristics of macromolecular crystals. *Acta Crystallogr. Sect. D Biol. Crystallogr. 56*, 986–995.

(76) De Sanctis, D., Oscarsson, M., Popov, A., Svensson, O., and Leonard, G. (2016) Facilitating best practices in collecting anomalous scattering data for de novo structure solution at the ESRF Structural Biology Beamlines. *Acta Crystallogr. Sect. D Struct. Biol. 72*, 413–420.

(77) Karplus, P. A., and Diederichs, K. (2012) Linking Crystallographic Model and Data Quality. *Science 336*, 1030–1034.

(78) Meinhold, L., and Smith, J. C. (2005) Correlated dynamics determining X-ray diffuse scattering from a crystalline protein revealed by molecular dynamics simulation. *Phys. Rev. Lett. 95*, 1–4.

(79) Caspar, D. L. D., Clarage, J., Salunke, D. M., and Clarage, M. (1988) Liquid-like movements in crystalline insulin. *Nature 332*, 659–662.

(80) Kurauskas, V., Izmailov, S. A., Rogacheva, O. N., Hessel, A., Ayala, I., Woodhouse, J., Shilova, A., Xue, Y., Yuwen, T., Coquelle, N., Colletier, J. P., Skrynnikov, N. R., and Schanda, P. (2017) Slow conformational exchange and overall rocking motion in ubiquitin protein crystals. *Nat. Commun. 8*, 1–11.

(81) Clarage, J. B., and Phillips, G. (1997) [21] Analysis of diffuse scattering and relation to molecular motion, in *Macromolecular Crystallography Part B*, pp 407–432. Academic Press.

(82) Sikosek, T., and Chan, H. S. (2014) Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface 11*.

(83) Petrovic, D., Risso, V. A., Kamerlin, S. C. L., and Sanchez-Ruiz, J. M. (2018) Conformational dynamics and enzyme evolution. *J. R. Soc. Interface 15*.

(84) Campitelli, P., Modi, T., Kumar, S., and Banu Ozkan, S. (2020) The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annu. Rev. Biophys. 49*, 267–288.

(85) Zou, T., Risso, V. A., Gavira, J. A., Sanchez-Ruiz, J. M., and Ozkan, S. B. (2015) Evolution of Conformational Dynamics Determines the Conversion of a Promiscuous Generalist into a Specialist Enzyme. *Mol. Biol. Evol. 32*, 132–143.

(86) James, L. C., and Tawfik, D. S. (2003) Conformational diversity and protein evolution - A 60-year-old hypothesis revisited. *Trends Biochem. Sci. 28*, 361–368.

(87) Tokuriki, N., and Tawfik, D. S. (2009) Protein Dynamism and Evolvability. *Science 324*, 203–207.

(88) Campbell, E., Kaltenbach, M., Correy, G. J., Carr, P. D., Porebski, B. T., Livingstone, E. K., Afriat-Jurnou, L., Buckle, A. M., Weik, M., Hollfelder, F., Tokuriki, N., and Jackson, C. J. (2016) The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol. 12*, 944–950.

(89) Romero-Rivera, A., Garcia-Borràs, M., and Osuna, S. (2017) Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal. 7*, 8524–8532.

(90) Obexer, R., Godina, A., Garrabou, X., Mittl, P. R. E., Baker, D., Griffiths, A. D., and Hilvert, D. (2017) Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem. 9*, 50–56.

(91) Schenkmayerova, A., Pinto, G. P., Toul, M., Marek, M., Hernychova, L., Planas-Iglesias, J., Daniel Liskova, V., Pluskal, D., Vasina, M., Emond, S., Dörr, M., Chaloupkova, R., Bednar, D.,

Prokop, Z., Hollfelder, F., Bornscheuer, U. T., and Damborsky, J. (2021) Engineering the protein dynamics of an ancestral luciferase. *Nat. Commun. 12*, 3616.

(92) Janowski, P. A., Liu, C., Deckman, J., and Case, D. A. (2016) Molecular dynamics simulation of triclinic lysozyme in a crystal lattice. *Protein Sci. 25*, 87–102.

(93) Gurusaran, M., Shankar, M., Nagarajan, R., Helliwell, J. R., and Sekar, K. (2014) Do we see what we should see? Describing non-covalent interactions in protein structures including precision. *IUCrJ 1*, 74–81.

(94) Schanda, P., and Ernst, M. (2016) Studying dynamics by magic-angle spinning solid-state NMR spectroscopy: Principles and applications to biomolecules. *Prog. Nucl. Magn. Reson. Spectrosc. 96*, 1–46.

(95) Salinas, V. H., and Ranganathan, R. (2018) Coevolution-based inference of amino acid interactions underlying protein function. *Elife 7*, 1–20.

(96) Lange, O. F., and Grubmüller, H. (2006) Generalized correlation for biomolecular dynamics. *Proteins Struct. Funct. Genet. 62*, 1053–1061.

(97) Roedig, P., Duman, R., Sanchez-Weatherby, J., Vartiainen, I., Burkhardt, A., Warmer, M., David, C., Wagner, A., and Meents, A. (2016) Room-temperature macromolecular crystallography using a micro-patterned silicon chip with minimal background scattering. *J. Appl. Crystallogr. 49*, 968–975.

(98) Venderley, J., Matty, M., Krogstad, M., Ruff, J., Pleiss, G., Kishore, V., Mandrus, D., Phelan, D., Poudel, L., Wilson, A. G., Weinberger, K., Upreti, P., Rosenkranz, S., Osborn, R., and Kim,

E.-A. (2020) Harnessing Interpretable and Unsupervised Machine Learning to Address Big Data from Modern X-ray Diffraction. *arXiv*.

TOC Artwork