Bringing Diffuse X-ray Scattering Into Focus

Michael E. Wall[1], Alexander M. Wolff[2,3], James S. Fraser[2]

[1]Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, 87545
[2] – Graduate Group in Biophysics, University of California San Francisco, San Francisco, CA 94158, USA
[3] – Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94158, USA
Correspondence: mewall@lanl.gov, jfraser@fraserlab.com

Abstract

Now that electron microscopy and micro electron diffraction have entered the arena of high-resolution structure determination, X-ray crystallography is experiencing a renaissance as a method for probing the protein conformational ensemble. The inherent limitations of Bragg analysis, however, which only reveals the mean structure, have given way to a surge in interest in diffuse scattering, which is caused by structure variations. Diffuse scattering is present in all macromolecular crystallography experiments. Recent studies are shedding light on the origins of diffuse scattering in protein crystallography, and provide clues for leveraging diffuse scattering to model protein motions with atomic detail.

Introduction

With over 100,000 X-ray structures deposited in the wwPDB [1], improvements in data processing pipelines, and the advent of completely unattended data collection, it seems hard to imagine that there are any aspects of protein X-ray crystallography that remain to be optimized. However, only half of the X-rays scattered by the crystalline sample are currently being analyzed – those in the Bragg peaks. The weaker, more smoothly varying features in diffraction images, known as diffuse scattering, are largely ignored by current practices. While the analysis of diffuse scattering is an established method in the fields of small molecule crystallography [2] and materials science [3], there are only very few foundational studies of diffuse scattering in macromolecular crystallography [4-17]. However, the relative scarcity of diffuse scattering studies is poised to change as activity in the field has recently increased.

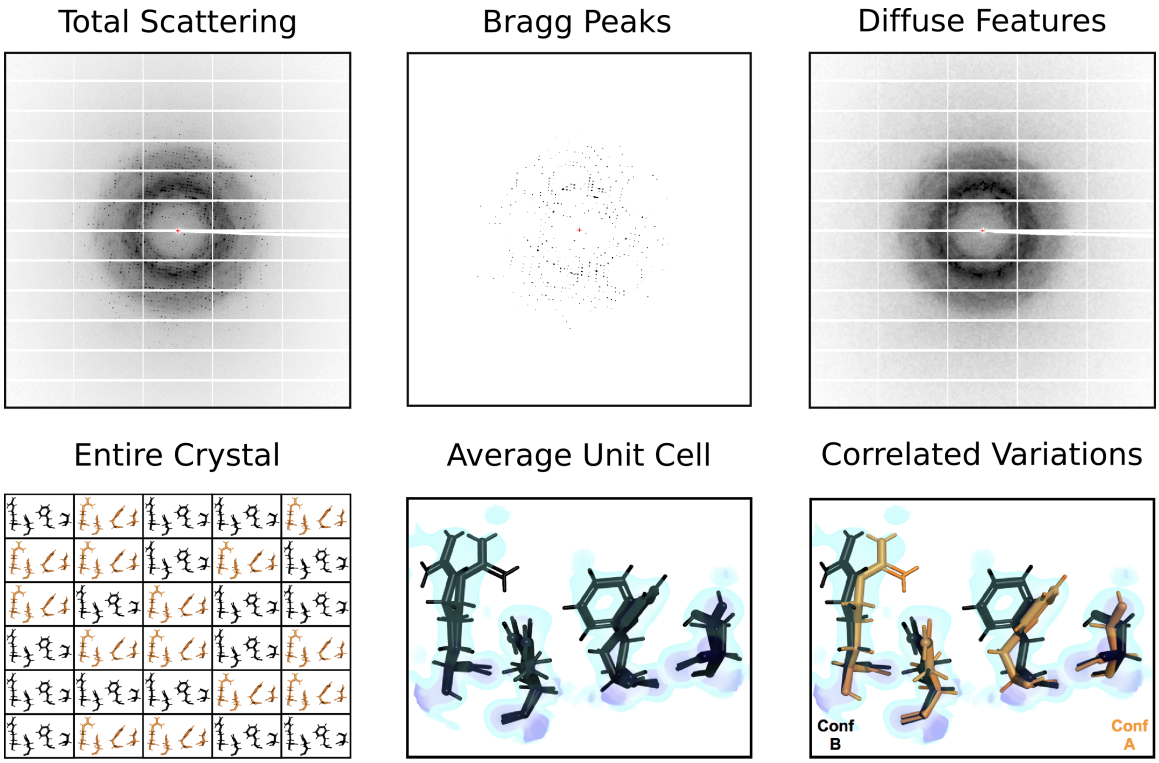

Figure 1. A typical detector image in X-ray crystallography (from [18]) (*upper, left*) records all of the X-rays scattered by a protein crystal during a single exposure. Dark pixels correspond to high X-ray intensities. A cartoon crystal is depicted (*lower, left*) that contains a series of unit cells, with the contents of any given unit cell adopting one of two conformations (the conformations are expected to be more varied in a real protein crystal). Conformation A is shown in orange, while conformation B is shown in black (*lower panel*). During analysis, data are reduced by examining only the Bragg peaks (*upper, middle*), which report on the average charge density within a unit cell (*lower, middle*). The electron density is shown in blue, with areas of especially strong charge highlighted in purple. While multiple conformations may be modelled into the average density, assigning which conformations occur together across residues requires additional information. Current modeling practices use geometric constraints to help classify different alternative conformation groups [19]. The diffuse scattering left behind during data reduction (*upper, right*) is an additional potential source of such information. Diffuse scattering includes an isotropic component that is largely created by solvent scattering, and an anisotropic component that is driven by correlated motions within the crystal. Analyzing this anisotropic signal might help to distinguish networks of residues that move together (*lower, right*).

A small group of researchers (including MEW and JSF) met in 2014 to discuss the challenges and opportunities of investigating macromolecular diffuse scattering [20]. Our attention was drawn to several key developments in the field of macromolecular crystallography that motivated and enabled assessment of the diffuse signal. First, structural models were reaching a plateau in quality. The origin of this plateau and the "R-factor gap" is likely due to the underlying inadequacies of the structural models refined against crystallographic data [21]. These inadequacies can only be overcome if we can improve the modeling of conformational heterogeneity (especially in data collected at room temperature [22]), solvation, and lattice imperfections that break the assumptions of "perfect crystals" used in data reduction and refinement. Second, new

detectors were enabling collection of data with lower noise, higher dynamic range, and highly localized signal. Third, new light sources were emerging with very bright, micro-focused beams (e.g. X-ray free-electron lasers). Collectively, these factors made us optimistic that diffuse scattering data both was needed and could be measured accurately enough to improve structural modeling. In early 2017 we met again to discuss the progress of the field with respect to each of these challenges identified in 2014 [23]. Below, we summarize this progress. While there have been exciting developments in recent years, there are still major challenges ahead. For example, whether diffuse scattering data can be leveraged for resolution extension and crystallographic phasing, as recently claimed [24], still requires additional examination and application to more than one system. Additional remaining challenges include modeling atomic motions in protein crystals using diffuse scattering data with accuracy comparable to the Bragg analysis, as well as utilizing these models of protein motions to distinguish between competing biochemical mechanisms.

Data collection

Extraction of diffuse scattering data from conventional protein crystallography experiments is becoming straightforward thanks to the increased accessibility of photon-counting pixel array detectors (PADs, e.g. Pilatus detectors). These detectors have greater dynamic range and do not suffer from "blooming" overloads that obscured diffuse signals near Bragg peaks on conventional charge-coupled device (CCD) detectors. (An early CCD detector was programmed to drain excess charge away from overflowing pixels to enable measurement of diffuse scattering data [17,25]; however, this feature was not implemented in commercial detectors). Additionally PADs have enabled new collection strategies, such as fine phi angle scans, that facilitate analysis of Bragg peaks and diffuse features from the same set of images [18]. A second major advance is the measurement of diffuse scattering using an X-ray free-electron laser (XFEL) in a serial femtosecond crystallography (SFX) experiment [24]. Using an XFEL enables collection of radiation-damage-free room temperature data, as well the potential to examine time-resolved changes in the diffuse scattering signal.

Despite these advances in collection of diffuse scattering data, minimizing background scattering remains the most important obstacle to collecting high quality data. While it is possible to remove some background scattering during data processing, the cleanest separation requires one to remove scattering extraneous to the crystal during the experiment. Factors to consider during collection of single crystal datasets include the thickness and orientation of the loop (for relevant mounting schemes), the volume of liquid surrounding the crystal, and the amount of airspace between the crystal and the detector. Background air scatter can be also reduced by a Helium path between sample and detector. Collection of SFX data adds additional complexity, as the injection stream and crystal size will vary. Ayyer et al [24] addressed this challenge by selecting only the frames with the strongest diffuse scattering signal, in which the crystal was expected to span the width of the jet. As the landscape of sample delivery devices for SFX and conventional crystallography continues to evolve, mounted sample delivery on materials

such as graphene [26] provides a promising route for minimization of background scattering.

Data integration

Early studies of protein diffuse scattering focused on explaining features in individual diffraction images. The introduction of methods for three-dimensional diffuse data integration enabled quantitative validation of models of correlated motions [17]. Several approaches to 3D data integration now have been implemented [24,25,27-29]. These approaches differ in several key ways: (1) the scaling of intensities when merging the data; (2) the handling of intensities in the neighborhood of the Bragg peak; and (3) the strategy for sampling of reciprocal space. In the *Lunus* software for diffuse scattering (https://github.com/mewall/lunus) we have chosen:

(1) To use the diffuse intensity itself to scale the diffuse data (as opposed to using the Bragg peaks, as in Ref. [28]). This choice avoids artifacts due to differences in the way the Bragg and diffuse scattering vary with radiation damage.

(2) To ignore or filter intensity values in regions where the variations are sharper than the 3D grid that will hold the integrated data. This includes masking intensities too close to a Bragg peak, and kernel-based image processing to remove Bragg peaks from diffraction images. These steps avoid the mixing of signal associated with sharp features into the longer wavelength signals (a potential issue with integration scheme in Ref. [28]), contaminating the large-scale diffuse features. The sharply varying features (e.g. streaks) are an important component of the signal; however, to avoid artifacts in analysis, they should be measured on a grid that is fine enough to resolve them. If the grid is too coarse to resolve them, they should be removed and handled separately from the large-scale features.

(3) To sample at even subdivisions of integer Miller indices. Off-lattice sampling strategies are valid (as used in Refs. [24,27]), but on-lattice strategies enable a direct connection to models that involve crystalline transforms (as opposed to molecular transforms), which are the most relevant models for protein crystallography.

Recent algorithmic improvements have led to scalable, parallelized methods for real-time processing of single-crystal synchrotron data. These improvements aim to keep pace with real-time analysis of Bragg data at high frame rates, such as those expected at LCLS-II and euXFEL. Initial tests mapped staphylococcal nuclease diffuse data onto a fine-grained reciprocal lattice, using two samples per Miller index [30]. This implementation of the *Lunus* software is capable of processing thousands of diffraction images within a few minutes on a computing cluster.

In addition to improving the scalability of diffuse scattering data processing, we have also developed methods to make analysis of diffuse data push-button. Inspired by the user-friendly environment provided by software for analyzing Bragg peaks, such as xia2 [31], we aimed to reduce the barrier for crystallographers to analyze the diffuse signal in their

data. The resulting pipeline, *Sematura,* is openly available on GitHub (https://github.com/fraser-lab/diffuse_scattering). To ensure portability the project was built upon the CCTBX framework [32], with future work focusing on moving *Sematura* directly into the CCTBX package for ease of access.

Model building and refinement

Liquid-like motions. After early experiments on tropomyosin [14], the liquid-like motions (LLM) model became a key tool in interpreting diffuse features in diffraction images [4,6]. In the LLM model, the crystal is treated as a soft material. All atoms are assumed to exhibit statistically identical normally distributed displacements about their mean position. The correlation between atom displacements is a decreasing function of the distance between the atoms, usually an exponential decay. A LLM was used to interpret early 3D diffuse data sets, refined using a correlation coefficient as a target function [16,17]. Successful refinement of a LLM model was used to demonstrate the successful extraction of diffuse datasets from Bragg diffraction experiments collected on Pilatus detectors [18]. Peck et al. [28] recently found the ability of the LLM to capture correlations across unit cell boundaries was essential for modeling the diffuse signal in several 3D datasets. This result is intuitive, as the nearest neighbors of an atom are often found in symmetry related molecules. Overall the LLM model has proven to be a simple means of capturing the data with a straightforward interpretation, and therefore remains an important first approach to analysis of protein diffuse X-ray scattering.

Normal mode analysis and elastic network models. Beyond the LLM model, normal mode analysis (NMA) of elastic network models (ENMs) can provide insights into the soft modes of protein dynamics in more detail, helping to reveal mechanisms that bridge protein structure and function [33]. In an ENM, the atoms of the crystal structure are connected by springs, and the resulting network is coupled to a thermal bath. NMA then yields the covariance matrix of atom displacements. The diagonal elements of the covariance matrix correspond to the crystallographic B factors, which come from the Bragg analysis through the crystal structure model. Riccardi et al. [34] showed how to renormalize the entire covariance matrix using the crystallographic B factors. Importantly, this allows one to scale an ENM in a manner consistent with traditional crystallographic metrics. Despite this strength, different ENMs can match the same Bragg data equally well. This happens when different covariance matrices have the same diagonal elements following renormalization, even though the off-diagonal elements differ. Thus, as with Translation-Libration-Screw refinement [35], the Bragg data alone cannot be used to distinguish between similar ENMs, and also cannot be used to refine ENMs. Diffuse scattering could help differentiate between these ENMs because the off-diagonal elements directly influence the diffuse signal. Thus, there is an opportunity for carefully measured diffuse data to be used in refinement of ENM models, and subsequent refinement of protein structural models.

Indeed, many key elements needed for refinement of normal modes models using diffuse scattering already have been demonstrated. Cloudy diffuse features in X-ray diffraction

from lysozyme crystals resemble the diffuse scattering predicted from simulations of normal modes models [9,12]. Similarly, sharper diffuse features in the neighborhood of Bragg peaks in ribonuclease crystals can be captured by lattice normal modes [36] Different varieties of ENMs for staphylococcal nuclease give rise to distinct diffuse scattering patterns, even when renormalized using the crystallographic B factors [34].

Three-dimensional diffuse scattering data from trypsin and proline isomerase (CypA) recently were modeled using ENMs [18]. The agreement was substantial, considering that the models were not refined. On the other hand, Peck et al. [28] found a low agreement between ENM models and diffuse data. How much can refinement improve the agreement of an ENM model? Here we provide an example. In our example, the asymmetric unit of PDB ID 4WOR was expanded to the P1 unit cell, and an ENM was constructed as in Ref. [18]. The spring force constants between C-alpha atoms were computed as $e^{-r_{ij}/\lambda}$, where r_{ij} is the closest distance between atoms i and j, either in the same unit cell or in neighboring unit cells of the crystal structure. All atoms on the same residue as the C-alpha were assumed to move rigidly as a unit. The initial value λ = 10.5 Å yielded a linear correlation of 0.07 with the anisotropic component of the diffuse data, as computed in Ref. [18]. Powell minimization using the *scipy.optimize.minimize* method was used to refine the value of λ, using the negative correlation as a target. The final correlation was 0.54 for a value λ = 0.157 Å – a substantial improvement, but one that indicates that the direct interactions are essentially limited to nearest neighbors. Simulated diffuse intensity in diffraction images calculated using the model vs. the data show similarities in cloudy diffuse features (Fig. 2). Key strategies for improving the model are: extending from a Calpha network to an all-atom network; using crystalline normal modes that extend beyond a single unit cell (prior studies used the Born von Karman method to compute these modes [34,36], but did not fully include the resulting modes in the thermal diffuse scattering calculation [37]); and allowing spring constants to deviate locally from the exponential behavior. Optimizing this type of model has applications beyond diffuse scattering validation and model refinement, as structures derived from normal modes analysis of network models have been useful for providing alternative starting points for molecular replacement [38] and have recently been used in an exciting local refinement procedure in cryo electron microscopy [39].

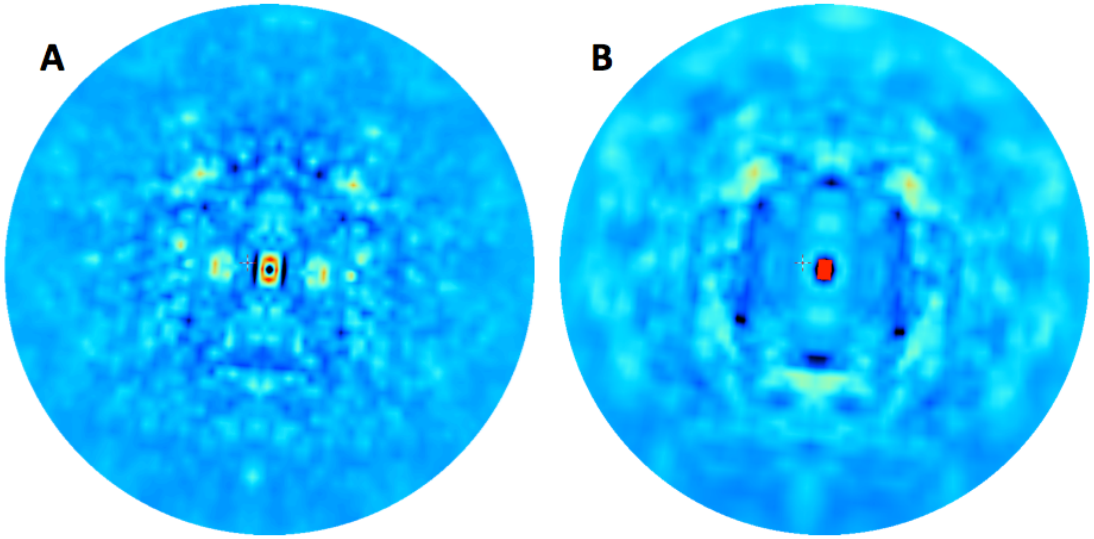

Figure 2. Comparison of simulated diffuse intensity in diffraction images computed from (A) a refined ENM of staphylococcal nuclease and (B) experimental data from Ref. [17].

Ensemble refinement. A great promise of diffuse scattering is the potential to inform ensemble or multiconformer models of protein structures (Figure 1). As for TLS and ENM models, diffuse signal might be able differentiate between ensembles resulting in the same average structures. Even if information about atomistic conformations remains out of reach, the signal could potentially be leveraged to improve ensemble models derived from time-averaged refinement using the scheme by Gros and colleagues [40]. Currently, this procedure operates on the rationale that large scale deviations can be modelled using a TLS model, and the residual local deviations are then sampled by a molecular dynamics simulation with a time-averaged difference electron density term. Our work has revealed that diffuse scattering calculated from TLS models of disorder do not match the measured diffuse signal, however, indicating that TLS is a poor descriptor of the disorder within a crystal [18]. Given the improvements seen when including neighboring unit cells in LLM models [28], the disorder of the crystal environment might be better accounted for by a coarse-grained model of intramolecular motion using a NMA model refined against the diffuse scattering signal. Once large-scale disorder is accounted for by NMA, local anharmonic deviations from the modes can be explored using MD simulations restrained by the X-ray data. As diffuse analysis becomes more sensitive, the selection of the final representative ensemble also can be optimized against the diffuse data. This selection step could supplement the current practice of selecting ensemble that matches the final rolling Rfree value.

Molecular dynamics simulations

In addition to refining models of protein motions, diffuse scattering can be used to validate MD simulations [7,9,41-45]. Early efforts were hindered by the use of 10 ns or shorter simulation durations [7,42], which lacked sufficient sampling for the calculations. Microsecond duration simulations of protein crystals are now becoming routine [30,45-47]. For staphylococcal nuclease, microsecond simulations overcome the sampling

limitations for diffuse scattering calculations, while providing insight into ligand binding and catalysis [45].

The agreement of the total diffuse intensity with MD simulations is high for staphylococcal nuclease [43,45], yielding a linear correlation of 0.94 for a microsecond simulation [20]. Agreement with the 10-fold weaker anisotropic component is lower [30,45], but is more sensitive to the details of the simulation, creating opportunities for increasing the accuracy of MD models. For example, expanding the staphylococcal nuclease model from a single periodic unit cell to a 2x2x2 supercell increased the correlation with the anisotropic component from 0.42 to 0.68 for a microsecond simulation [30]. The agreement with the MD is tantalizingly close to what is expected for an initial molecular replacement model in the Bragg analysis, suggesting that the combination of MD simulations and diffuse scattering might soon yield experimentally validated atomic details of protein motions. In addition, recent solid state NMR (ssNMR) experiments combined with crystalline protein simulations [48-50] create opportunities for joint validation of MD simulations using crystallography and NMR.

Phasing and resolution extension – paradigm shift or confirmation bias?

In a high-profile publication, the Chapman and Fromme groups integrated the first three-dimensional diffuse scattering dataset from a serial femtosecond protein crystallography experiment at an X-ray free electron laser [24]. Their analysis focused on the potential for phasing and resolution extension of a charge density map of photosystem II (PSII). The method, based on the difference-map algorithm [51], depends critically on the assumption that the diffuse signal is proportional to the molecular transform of the unit to be resolved. In this respect, the work is closely related to that of Stroud and Agard [52] and Makowski [53] on phasing using continuous diffraction data. Despite the widespread importance of short-scale correlated motions in explaining diffuse scattering across several systems [4,6,16,17], Ayyer and colleagues claimed that lateral contacts between PSII layers in the crystal caused independent rigid-body translations of the dimer, with longer-scale correlated motions contributing to the diffuse signal.

However, despite the promise this breakthrough analysis of SFX diffuse scattering heralds [24], many questions remain unaddressed:

- Bragg spots are visible in the 4.5-3.5 Å range in Fig. 2 of Ref. [24]. The intensities for improving the charge density therefore include contributions from Bragg peaks. What effect do the Bragg peaks have on phasing and resolution extension in the 4.5-3.5 Å range? Depositing even the 2848 raw diffraction images selected for this analysis from the total dataset of 25,585 in a repository such as the SBGrid Databank [54] or CXIdb [55] could help to distinguish the role of Bragg peak contamination in this dataset.

- Electron density maps can be improved using random data beyond the resolution of the Bragg data – the *free lunch* effect [56]. How does the improvement in the

PSII map compare to what would be obtained by using randomized data in the 4.5-3.3 Å range? The R-factors in the extended resolution range reported in PDB 5E79 are very high (over 50%) and several bins have Rfree < Rwork. How would this compare to pseudo-crystallographic refinement [57] of the expected average intensity in these bins?

- The support envelope in the difference-map algorithm is generated from a blurred charge density, but the boundary between the accepted and rejected region is sharp. The application of the support mask therefore enables the low-resolution molecular envelope to the bias high-resolution features of the calculated map. How important is this bias in determining the final charge density? Would seeding the model with a spherical support yield similar improvement in the charge density, using a shrink-wrap method [58]?

- Omit maps [59] can be prepared using the diffuse data by setting the charge density of the model to zero in some region and computing a 2FoFc map using the experimental amplitudes. How robust are the improved features of the charge density in Ref. [24] to omit map analysis? Most experimental phasing experiments abandon or seriously down-weight the experimental information as soon as the model quality allows. It is unclear when model phases are being used in their approach and how heavily they are weighted in refinement by MLHL-refinement methods used here.

- Does a rigid-body translation model describe the data better than the models reviewed here (LLM, ENM, MD)? In particular, the most accurate models of diffuse scattering so far – the LLM and supercell MD – include couplings across unit cell boundaries. Peck et al. [28] point out that the diffuse scattering for models with couplings across unit cell boundaries is theoretically derived from the crystal transform (sampled only at Miller indices), not the molecular transform (sampled everywhere in reciprocal space). Diffuse scattering from protein crystals may therefore be missing the key information between the Bragg peaks required for phasing.

- The assumption of rigid-body translations used in Ref. [24] additionally requires that the data used for phasing and resolution extension are measured in a regime $s^2\sigma^2 >> 1$, where s is the scattering vector length and σ is the amplitude of motion for the rigid-body translation model. What is the actual value of σ for the best-fit rigid-body translation model to the PSII diffuse data? If the requirement is for the intensity values of the Bragg peaks to be comparable to the continuous diffraction at a resolution of 4.5 Å (according to Ref. [24], a factor of 100 weaker), the condition is $e^{-s^2\sigma^2} = 0.01$, yielding σ = 9.7 Å, or a B factor of about 7,400 Å^2. If a factor of 10 attenuation is required, the condition yields sigma = 6.8 Å, or a B factor of about 3,680 Å^2. For attenuation by e, σ = 4.5 Å, and B = 1,600 Å^2. All of these B factors are much larger than the Wilson B factor in Extended Data Table 1 of Ref. [24] (191.6 Å^2). The authors' own assumptions lead to σ values that are much larger than the Wilson B factors (which in turn are

presumably much larger than those in the 4.5 $Å^2$ structural model). The conclusion that rigid-body motions of this magnitude are responsible for the diffuse scattering from PSII crystals seems to run counter to the fact that the crystals even diffract to 4.5 Å.

Integrating diffuse scattering with Bragg diffraction to improve crystallographic models is a major goal in the field [16,29,60]. The work by Ayyer, Chapman, and colleagues [24] claims to mark a paradigm shift toward solving structures with poorly diffracting crystals. While this is an exciting idea that many crystallographers would like to see realized, as illustrated in the examples above, the paper raises many questions that must be answered to determine how robust, feasible, and general this method might become.

Future perspective:

The massive investment in structural genomics in the 2000s dramatically increased the robustness of X-ray crystallography data collection, processing, and refinement. Although diffuse scattering remained relatively unstudied during that time, it is now poised to piggy-back on these technological improvements and standardizations. As attention shifts toward electron microscopy, ascendant as a go-to method for determining novel macromolecular structures, and with electron crystallography (microED) making a comeback, it is reasonable to ask: why study the origins of diffuse X-ray scattering? First, despite being present in all macromolecular diffraction patterns, the origins of diffuse scattering in protein crystallography remain mysterious. Whether it is due to long-range [24] or short-range disorder [18,28,30], diffuse scattering can be potentially informative for structural modeling. There are additional parallels between diffuse scattering and the multiple "dynamic" scattering of electrons that are currently being ignored (intentionally and surprisingly without much consequence) in microED studies. Second, the types of conformational heterogeneity that can be validated and, potentially, refined against diffuse scattering data can guide us to define better structural models. As the structural biology toolkit expands, X-ray scattering, including diffuse scattering, still provides unique capabilities to probe conformational ensembles over many length scales, as captured in a recent review by Meisburger et al. [61]. Ultimately, the better models of concerted motions will have far ranging impact beyond the average structure that is accessible using conventional X-ray crystallography and cryo-electron microscopy data, yielding a deeper understanding of biochemical mechanism.

1. Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S: **Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive**. *Methods Mol Biol* 2017, **1607**:627-641.
2. Welberry TR: *Diffuse X-Ray Scattering and Models of Disorder*. Oxford: Oxford University Press; 2004.
3. Keen DA, Goodwin AL: **The crystallography of correlated disorder**. *Nature* 2015, **521**:303-309.
4. Caspar DL, Clarage J, Salunke DM, Clarage M: **Liquid-like movements in crystalline insulin**. *Nature* 1988, **332**:659-662.

5. Chacko S, Phillips GN, Jr.: **Diffuse X-ray scattering from tropomyosin crystals**. *Biophys J* 1992, **61**:1256-1266.
6. Clarage JB, Clarage MS, Phillips WC, Sweet RM, Caspar DL: **Correlations of atomic movements in lysozyme crystals**. *Proteins* 1992, **12**:145-157.
7. Clarage JB, Romo T, Andrews BK, Pettitt BM, Phillips GN, Jr.: **A sampling problem in molecular dynamics simulations of macromolecules**. *Proceedings of the National Academy of Sciences of the United States of America* 1995, **92**:3288-3292.
8. Doucet J, Benoit JP: **Molecular dynamics studied by analysis of the X-ray diffuse scattering from lysozyme crystals**. *Nature* 1987, **325**:643-646.
9. Faure P, Micu A, Pérahia D, Doucet J, Smith JC, Benoit JP: **Correlated intramolecular motions and diffuse X-ray scattering in lysozyme**. *Nat Struct Biol* 1994, **1**:124-128.
10. Helliwell JR, Glover ID, Jones A, Pantos E, Moss DS: **Protein dynamics - use of computer-graphics and protein crystal diffuse-scattering recorded with synchrotron X-radiation**. *Biochem Soc Transact* 1986, **14**:653-655.
11. Kolatkar AR, Clarage JB, Phillips GN, Jr.: **Analysis of diffuse scattering from yeast initiator tRNA crystals**. *Acta Crystallogr D* 1994, **50**:210-218.
12. Mizuguchi K, Kidera A, Gō N: **Collective motions in proteins investigated by X-ray diffuse scattering**. *Proteins* 1994, **18**:34-48.
13. Perez J, Faure P, Benoit JP: **Molecular rigid-body displacements in a tetragonal lysozyme crystal confirmed by X-ray diffuse scattering**. *Acta Crystallogr D Biol Crystallogr* 1996, **52**:722-729.
14. Phillips GN, Jr., Fillers JP, Cohen C: **Motions of tropomyosin. Crystal as metaphor**. *Biophys J* 1980, **32**:485-502.
15. Polikanov YS, Moore PB: **Acoustic vibrations contribute to the diffuse scatter produced by ribosome crystals**. *Acta Crystallogr D Biol Crystallogr* 2015, **71**:2021-2031.

* Careful analysis of the diffuse scattering present in 70S ribosome crystals revealed that lattice vibrations may explain a significant portion of the diffuse signal, highlighting the importance of models that account for correlated variations across unit cell boundaries.

16. Wall ME, Clarage JB, Phillips GN: **Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering**. *Structure* 1997, **5**:1599-1612.
17. Wall ME, Ealick SE, Gruner SM: **Three-dimensional diffuse X-ray scattering from crystals of *Staphylococcal* nuclease**. *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**:6180-6184.
18. Van Benschoten AH, Liu L, Gonzalez A, Brewster AS, Sauter NK, Fraser JS, Wall ME: **Measuring and modeling diffuse scattering in protein X-ray crystallography**. *Proc Natl Acad Sci U S A* 2016, **113**:4069-4074.

** Diffuse data is extracted from data collected under optimal conditions for Bragg analysis, revealing that modern PAD detectors and fine phi slicing can make diffuse data widely available. Also, LLM and normal modes models of disorder account for a substantial portion of the diffuse signal isolated from cypa and trypsin.

19. van den Bedem H, Bhabha G, Yang K, Wright PE, Fraser JS: **Automated identification of functional dynamic contact networks from X-ray crystallography**. *Nat Methods* 2013, **10**:896-902.
20. Wall ME, Adams PD, Fraser JS, Sauter NK: **Diffuse X-ray scattering to model protein motions**. *Structure* 2014, **22**:182-184.
21. Holton JM, Classen S, Frankel KA, Tainer JA: **The R-factor gap in macromolecular crystallography: an untapped potential for insights on accurate structures**. *FEBS J* 2014, **281**:4046-4060.
22. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T: **Accessing protein conformational ensembles using room-temperature X-ray crystallography**. *Proc Natl Acad Sci U S A* 2011, **108**:16247-16252.
23. Wall ME, Sweet RM, Ando N, Fraser JS, Phillips GN: **Measurement and Interpretation of Diffuse Scattering in X‐Ray Diffraction for Macromolecular Crystallography**. 2017 (http://www.osti.gov/scitech/biblio/1400134).
24. Ayyer K, Yefanov OM, Oberthur D, Roy-Chowdhury S, Galli L, Mariani V, Basu S, Coe J, Conrad CE, Fromme R, et al.: **Macromolecular diffractive imaging using imperfect crystals**. *Nature* 2016, **530**:202-206.
** This work extends the analysis of diffuse X-ray scattering into the realm of XFELS and serial crystallography, while also advocating for the use of diffuse scattering for phasing and resolution extension. Rigid body translations of the PSII dimer are the proposed source of the diffuse signal.

25. Wall ME: **Diffuse Features in X-Ray Diffraction from Protein Crystals**. In *Physics*. Edited by: Princeton University; 1996. vol Ph.D.]
26. Sui S, Wang Y, Kolewe KW, Srajer V, Henning R, Schiffman JD, Dimitrakopoulos C, Perry SL: **Graphene-based microfluidics for serial crystallography**. *Lab Chip* 2016, **16**:3082-3096.
* Graphene coated microfluidic chips enable collection of diffraction data with very high signal-to-noise, and may provide an alternative to jet based delivery systems for SFX experiments.

27. Estermann MA, Steurer W: **Diffuse scattering data acquisition techniques**. *Phase Transitions* 1998, **67**:165-195.
28. Peck A, Poitevin F, Lane TJ: **Intermolecular correlations are necessary to explain diffuse scattering from protein crystals**. 2017 (https://arxiv.org/abs/1707.05433).
** The authors approach diffuse scattering from a modelling perspective, and rigorously test current disorder models against a series of experimental datasets. Quantitative tests allow them to discern that most disorder models explain a limited portion of the diffuse signal, and that a LLM model is the best option, especially when correlations across unit cell boundaries are included.

29. Wall ME: **Methods and software for diffuse X-ray scattering from protein crystals**. *Methods in molecular biology* 2009, **544**:269-279.
30. Wall ME: **Internal protein motions in molecular dynamics simulations of Bragg and diffuse X-ray scattering**. 2017 (https://doi.org/10.1101/190496).
31. Winter G, Lobley CM, Prince SM: **Decision making in xia2**. *Acta Crystallogr D Biol Crystallogr* 2013, **69**:1260-1273.
32. Grosse-Kunstleve RW, Sauter NK, Moriarty NW, Adams PD: **The Computational Crystallography Toolbox: crystallographic algorithms in a reusable software framework**. *Journal of Applied Crystallography* 2002, **35**:126-136.
33. Haliloglu T, Bahar I: **Adaptability of protein structures to enable functional interactions and evolutionary implications**. *Curr Opin Struct Biol* 2015, **35**:17-23.
34. 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.
35. Van Benschoten AH, Afonine PV, Terwilliger TC, Wall ME, Jackson CJ, Sauter NK, Adams PD, Urzhumtsev A, Fraser JS: **Predicting X-ray diffuse scattering from translation-libration-screw structural ensembles**. *Acta Crystallogr D Biol Crystallogr* 2015, **71**:1657-1667.
* Predicted diffuse scattering patterns differ substantially across different TLS models derived from the same data. This provides an important proof of principle for the use of diffuse scattering in refinement of macromolecular models.

36. Meinhold L, Merzel F, Smith JC: **Lattice dynamics of a protein crystal**. *Phys Rev Lett* 2007, **99**:138101.
37. James R: *The Optical Principles of the Diffraction of X-Rays*. London: Bell; 1948.
38. Suhre K, Sanejouand YH: **ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement**. *Nucleic Acids Res* 2004, **32**:W610-614.
39. Schilbach S, Hantsche M, Tegunov D, Dienemann C, Wigge C, Urlaub H, Cramer P: **Structures of transcription pre-initiation complex with TFIIH and Mediator**. *Nature* 2017, **551**:204-209.
40. Burnley BT, Afonine PV, Adams PD, Gros P: **Modelling dynamics in protein crystal structures by ensemble refinement**. *Elife* 2012, **1**:e00311.
41. Héry S, Genest D, Smith JC: **X-ray diffuse scattering and rigid-body motion in crystalline lysozyme probed by molecular dynamics simulation**. *Journal of molecular biology* 1998, **279**:303-319.
42. Meinhold L, Smith JC: **Correlated dynamics determining X-ray diffuse scattering from a crystalline protein revealed by molecular dynamics simulation**. *Phys Rev Lett* 2005, **95**:218103.
43. Meinhold L, Smith JC: **Fluctuations and correlations in crystalline protein dynamics: a simulation analysis of Staphylococcal nuclease**. *Biophys J* 2005, **88**:2554-2563.

44. Meinhold L, Smith JC: **Protein dynamics from X-ray crystallography: anisotropic, global motion in diffuse scattering patterns**. *Proteins* 2007, **66**:941-953.
45. 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 U S A* 2014, **111**:17887-17892.
** In this study, the authors demonstrate the importance of long time-scale simulations to accurately sample a protein's correlated motions. The improvement is most evident when examining the anisotropic portion of the diffuse signal attributed to protein dynamics alone.

46. Janowski PA, Cerutti DS, Holton J, Case DA: **Peptide crystal simulations reveal hidden dynamics**. *J Am Chem Soc* 2013, **135**:7938-7948.
47. Janowski PA, Liu C, Deckman J, Case DA: **Molecular dynamics simulation of triclinic lysozyme in a crystal lattice**. *Protein Sci* 2016, **25**:87-102.
* MD simulations of lysozyme in a crystalline lattice reveal enhanced agreement with structural models derived from Bragg data. Nonetheless, convergence is slow, the lattice becomes disordered, and fluctuations of residues involved in crystal contacts are too high, indicating the need for improved MD force fields.

48. Kurauskas V, Izmailov SA, Rogacheva ON, Hessel A, Ayala I, Woodhouse J, Shilova A, Xue Y, Yuwen T, Coquelle N, et al.: **Slow conformational exchange and overall rocking motion in ubiquitin protein crystals**. *Nat Commun* 2017, **8**:145.
49. Ma P, Xue Y, Coquelle N, Haller JD, Yuwen T, Ayala I, Mikhailovskii O, Willbold D, Colletier JP, Skrynnikov NR, et al.: **Observing the overall rocking motion of a protein in a crystal**. *Nat Commun* 2015, **6**:8361.
50. Mollica L, Baias M, Lewandowski JR, Wylie BJ, Sperling LJ, Rienstra CM, Emsley L, Blackledge M: **Atomic-Resolution Structural Dynamics in Crystalline Proteins from NMR and Molecular Simulation**. *J Phys Chem Lett* 2012, **3**:3657-3662.
51. Elser V, Millane RP: **Reconstruction of an object from its symmetry-averaged diffraction pattern**. *Acta Crystallographica Section A* 2008, **64**:273-279.
52. Stroud RM, Agard DA: **Structure determination of asymmetric membrane profiles using an iterative Fourier method**. *Biophys J* 1979, **25**:495-512.
53. Makowski L: **The Use of Continuous Diffraction Data as a Phase Constraint .1. One-Dimensional Theory**. *Journal of Applied Crystallography* 1981, **14**:160-168.
54. Meyer PA, Socias S, Key J, Ransey E, Tjon EC, Buschiazzo A, Lei M, Botka C, Withrow J, Neau D, et al.: **Data publication with the structural biology data grid supports live analysis**. *Nat Commun* 2016, **7**:10882.
55. Maia FR: **The Coherent X-ray Imaging Data Bank**. *Nat Methods* 2012, **9**:854-855.

56. Caliandro R, Carrozzini B, Cascarano GL, De Caro L, Giacovazzo C, Siliqi D: **Phasing at resolution higher than the experimental resolution**. *Acta Crystallogr D Biol Crystallogr* 2005, **61**:556-565.
57. Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, Stark H: **Structure of the E. coli ribosome-EF-Tu complex at <3 A resolution by Cs-corrected cryo-EM**. *Nature* 2015, **520**:567-570.
58. Marchesini S, He H, Chapman HN, Hau-Riege SP, Noy A, Howells MR, Weierstall U, Spence JCH: **X-ray image reconstruction from a diffraction pattern alone**. *Physical Review B* 2003, **68**.
59. Bhat TN, Cohen GH: **OMITMAP: An electron density map suitable for the examination of errors in a macromolecular model**.
60. Clarage JB, Phillips GN, Jr.: **Analysis of diffuse scattering and relation to molecular motion**. *Methods Enzymol* 1997, **277**:407-432.
61. Meisburger SP, Thomas WC, Watkins MB, Ando N: **X-ray Scattering Studies of Protein Structural Dynamics**. *Chem Rev* 2017, **117**:7615-7672.

** This excellent review thoroughly lays out the connection between diffuse scattering, solution scattering, and crystallography. The assumptions and limitations of various approaches to analyzing diffuse data are clearly explained, and several disorder models are explored using case studies of biochemical interest.