# Measuring and modeling diffuse scattering in protein X-ray crystallography

Andrew H. Van Benschoten[1], Ana Gonzalez[2], Aaron S. Brewster[3], Nicholas K. Sauter[3], James S. Fraser[1,*], Michael E. Wall[4,*]

1 - Dept. of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA 11 94158

2 - Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA 94025

3 - Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

4 - Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, 87545

*correspondence: jfraser@fraserlab.com, mewall@lanl.gov

## Abstract:

X-ray diffraction has the potential to provide rich information about the structural dynamics of macromolecules. To realize this potential fully, it will be necessary to measure and model both Bragg scattering, which is currently used to derive macromolecular structures, and diffuse scattering, which reports on correlations in charge density variations. However, due to the extra effort of collecting diffuse data, experimental measurement of diffuse scattering from protein crystals has been limited to a small number of systems. Here, we present three-dimensional measurements of diffuse intensity collected from crystals of the model enzymes cyclophilin A and trypsin. The measurements were obtained from the same X-ray diffraction images as the Bragg data, using best practices for data collection in protein crystallography. To model the underlying dynamics in a practical way that can be used during structure refinement, we tested both Translation-Libration-Screw (TLS) and Liquid-Like Motions (LLM) models of protein motions. Whereas different TLS groupings yielded similar Bragg intensities, the diffuse intensities were clearly different. The agreement of the LLM models with the diffuse data was much stronger than for TLS. These results demonstrate a path to substantially increase the number of diffuse datasets available to the wider biosciences community. These and future datasets will be useful to benchmark computational tools for modeling correlated motions in macromolecules and to improve refinement of dynamics-inspired structural models emerging in X-ray crystallography.

## Significance:

The structural details of protein motions are critical to many biological processes, but they are hidden to conventional biophysical techniques. Diffuse X-ray scattering can reveal these details by measuring the correlated movements between atoms. However, diffuse scattering data collection historically has required extra effort and dedicated experimental protocols. We have measured three-dimensional diffuse intensities in X-ray diffraction from CypA and trypsin crystals using standard crystallographic data collection techniques. Analysis of the resulting data is consistent with the presence of liquid-like motions (LLM) in both crystals. Our results show that using diffuse scattering to model protein motions can become a component of routine crystallographic analysis through the extension of commonplace methods.

## Introduction:

X-ray crystallography can be a key tool for elucidating the structural basis of protein motions that play critical roles in enzymatic reactions, protein-protein interactions and signaling cascades (van den Bedem and Fraser, 2015). X-ray diffraction yields an ensemble-averaged picture of the protein structure: each photon simultaneously probes multiple unit cells that can vary due to internal rearrangements or changes to the crystal lattice. Bragg analysis of X-ray diffraction only yields the mean charge density of the unit cell, however, which fundamentally limits the information that can be obtained about protein dynamics (Clarage and Phillips, 1997; Keen and Goodwin, 2015).

A key limitation inherent in Bragg analysis is that alternative models with different correlations between atomic motions can yield the same mean charge density (Kuzmanic et al., 2011). The traditional approach to modeling atom movement is to assume a single structural model with individual atomic displacement parameters (B factors). Given sufficient data, anisotropic displacement factors can be modeled, yielding directional insights into motions that might cause variations in the crystal. When the data are more limited, Translation-Libration-Screw (TLS) structural refinement, in which motions are described using rigid body segments of the molecule (Schomaker and Trueblood, 1968), has emerged as a tool to model protein domain movements in crystallography (Painter and Merritt, 2005, 2006). However, TLS refinements that vary in the rigid body definitions can predict very different motions while maintaining equivalent agreement to Bragg X-ray diffraction data (Urzhumtsev et al., 2015; Van Benschoten et al., 2015).

Additional sources of information have been used to overcome the inherent limitations of Bragg analysis in identifying collective protein motion. Patterns of steric clashes between alternative local conformations (van den Bedem et al., 2013) or time-averaged ensemble refinement (Burnley et al., 2012) can be used to suggest certain modes of concerted motion. However, the atomistic details of these correlated motions may only be reliably (yet indirectly) identified at high resolution, and time-averaged ensemble refinement is additionally complicated by the use of an underlying TLS model to account for crystal packing variations (Burnley et al., 2012). Alternative methods such as solid-state NMR experiments (Ma et al., 2015) or long time scale molecular dynamics simulations (Janowski et al., 2013; Janowski et al., 2015; Wall et al., 2014b) can be used to probe the structural basis of crystal packing variations and internal protein motions.

Complementary information about internal protein motions also can be obtained in the X-ray crystallography experiment itself by analysis of diffuse scattering. Diffuse scattering arises when deviations away from a perfect crystal cause X-rays to be diffracted away from Bragg reflections. When the deviations are due to crystal vibrations, they can be described using textbook temperature diffuse scattering theory (see, e.g. (James, 1948)). When each unit cell varies independently, the diffuse intensity is proportional to the variance in the unit cell structure factor (Guinier, 1963) which is equivalent to the Fourier transform of the Patterson function of the charge density variations. The approximation of independent unit cells can break down when correlations extend across unit cell boundaries; however, motions with long correlation lengths result in diffuse intensity concentrated in the immediate neighborhood of Bragg peaks.

When analyzing the more broadly distributed diffuse intensity that corresponds to small correlation lengths (Caspar et al., 1988; Clarage et al., 1992; Wall et al., 1997a; Wall et al., 1997b) the contribution of inter-unit cell atom pairs is a small fraction of the total signal, which is therefore dominated by internal protein motions.

Several approaches have been used to connect macromolecular diffuse scattering data to models of protein motion and lattice disorder. Notably, Peter Moore has emphasized the need to validate TLS models using diffuse scattering (Moore, 2009), as has been performed in a limited number of cases (Doucet and Benoit, 1987; Perez et al., 1996; Van Benschoten et al., 2015). Good agreement with the data has previously been observed for liquid-like motions (LLM) models (Caspar et al., 1988; Clarage et al., 1992; Wall et al., 1997a; Wall et al., 1997b). In the LLM model, the atoms in the protein are assumed to move randomly, like in a homogeneous medium; the motions were termed “liquid-like” by Caspar et al (Caspar et al., 1988) because the correlations in the displacements were assumed to fall off exponentially with the distance between atoms. There is also a longstanding interest both in using diffuse scattering to validate improvements in MD simulations and in using MD to derive a structural basis for the protein motions that give rise to diffuse scattering (Clarage et al., 1995; Faure et al., 1994; Héry et al., 1998; Meinhold et al., 2007; Meinhold and Smith, 2005a, b, 2007; Wall et al., 2014b). Recent advances in computing now enable microsecond duration simulations (Wall et al., 2014b) that can overcome past barriers to accurate calculations seen using 10 ns or shorter MD trajectories (Clarage et al., 1995; Meinhold and Smith, 2005a).

Despite the fact that diffuse scattering analysis is relatively well developed in small-molecule crystallography (Welberry, 2004) and materials science (Keen and Goodwin, 2015), it has been underutilized in protein crystallography. There are relatively few examples of diffuse data analyzed using individual diffraction images from protein crystallography experiments, including studies of tropomyosin (Chacko and Phillips, 1992; Phillips et al., 1980), 6-phosphogluconate dehydrogenase (Helliwell et al., 1986), yeast initiator tRNA (Kolatkar et al., 1994), insulin (Caspar et al., 1988), lysozyme (Clarage et al., 1992; Doucet and Benoit, 1987; Faure et al., 1994; Mizuguchi et al., 1994; Perez et al., 1996), myoglobin (Clarage et al., 1995), Gag protein (Welberry et al., 2011), and the 70s ribosome subunit (Polikanov and Moore, 2015). Moreover, there are an even smaller number of examples involving complete three-dimensional diffuse data sets; these include studies of staphylococcal nuclease (Wall et al., 1997b), and calmodulin (Wall et al., 1997a).

To exploit the increased information that is potentially available from diffuse scattering, there is a pressing need to increase the number of proteins for which complete three-dimensional diffuse datasets have been experimentally measured. Conventional data collection procedures use oscillation exposures to estimate the full Bragg intensities. In contrast, the complete three-dimensional datasets measured by Wall *et al.* (Wall et al., 1997a; Wall et al., 1997b) used specialized methods for integrating three-dimensional diffuse data from still diffraction images. Similar methods now can be generalized and applied to other systems using modern beamlines and X-ray detectors. In particular, the recent commercial development of pixel-array detectors (PADs), which possess tight point-spread functions and single-photon sensitivity (Gruner, 2012),

have created new opportunities for measuring diffuse scattering as a routine tool in protein crystallography experiments using more conventional data collection protocols.

Here, we present diffuse scattering datasets for the human proline isomerase cyclophilin A (CypA) and the bovine serine protease trypsin. These datasets substantially increase the amount of experimental three-dimensional diffuse scattering data available to the macromolecular crystallography community, providing a necessary foundation for further advancement of the field (Wall et al., 2014a). To assess the potential for routine collection of diffuse datasets in crystallography, rather than expending a great deal of effort in optimizing the diffuse data and collecting still images (Wall et al., 1997a; Wall et al., 1997b), we used oscillation images obtained using best practices for high-quality Bragg data collection. The resulting datasets are of sufficient quality that the diffuse scattering can discriminate among alternative TLS refinements (Van Benschoten et al., 2015) and Liquid-Like Motion (LLM) models (Caspar et al., 1988; Clarage et al., 1992). Our results demonstrate that diffuse intensity can, and should, be measured in a typical X-ray crystallography experiment and indicate that diffuse X-ray scattering can be applied broadly as a tool to understand the conformational dynamics of macromolecules.

# Results:

## Experimental diffuse data show crystallographic symmetry

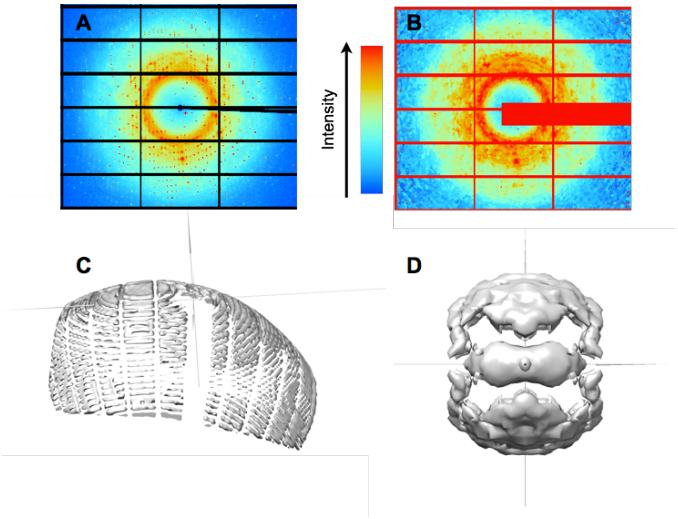


**Figure 1.** Steps in diffuse data integration. (A) Raw CypA diffraction images are processed (B) to remove Bragg peaks and enable direct comparisons of pixel values to models. (C) Pixels in diffraction images are mapped to reciprocal space and values of diffuse intensity are accumulated on a three-dimensional lattice; each diffraction image produces measurements of diffuse intensity on the surface of an Ewald sphere. (D) The data from individual images is combined and symmetrized to yield a complete dataset (isosurface at a value of 65 photon counts in the total intensity, before subtracting the isotropic component).

The symmetrized anisotropic diffuse datasets processed by LUNUS (**Figure 1**) are shown in **Figure 2A** (CypA) and **Figure 2D** (trypsin) and are available as supplementary material. The CypA dataset is 98% complete to a resolution of 1.4 Å, while the trypsin map is 95% complete to 1.25 Å resolution. We used the Friedel symmetry and Laue group symmetry to quantify the level of crystallographic symmetry in each anisotropic map. To evaluate the degree Friedel symmetry, we averaged intensities between Friedel pairs to create a symmetrized map $I_{Friedel}$ and calculated the Pearson Correlation Coefficient (PCC) between the symmetrized and unsymmetrized data to obtain the statistic $CC_{Friedel}$. For CypA and trypsin, $CC_{Friedel}$ = 0.90 and 0.95 respectively, demonstrating that diffuse intensities obey Friedel symmetry. To assess the degree of Laue group symmetry, we averaged P222-related reflections (the Laue symmetry corresponding to the P 21 21 21 space group of both CypA and trypsin crystals) to produce the symmetrized intensities, $I_{P222}$. The linear correlation $CC_{Sym}$ was then computed between the symmetrized and unsymmetrized intensities. The correlations were substantial for both CypA ($CC_{Sym}$ = 0.70) and trypsin ($CC_{Sym}$ = 0.69). Thus, our data are consistent with the diffuse intensity following the Bragg peak symmetry. The trypsin data were integrated using one degree oscillation frames, while the CypA data were integrated using 0.5 degree oscillation frames. The comparable degree of symmetry in the CypA and trypsin data suggests that the measurement of diffuse intensity is robust with respect to this difference in data collection.

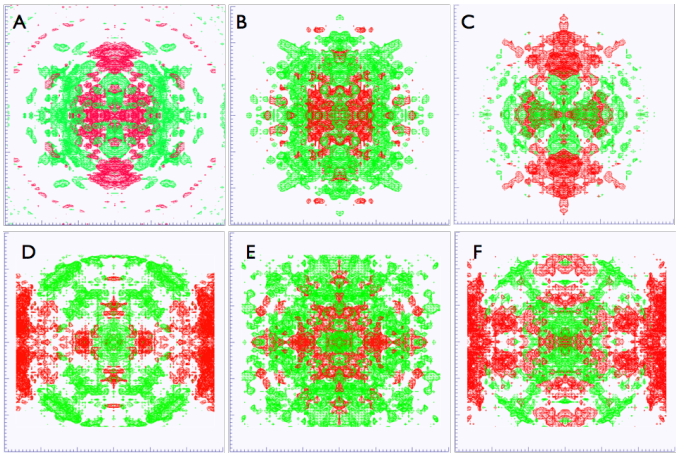


**Figure 2.** Visualization of anisotropic diffuse intensities. (A) CypA experimental data with isosurfaces shown using wireframes at a level of 2 photon counts in the resolution range 4.16 Å – 2.97 Å. Positive intensity is rendered in green, negative in red. (B) Isosurfaces for diffuse scattering predicted by the CypA LLM model. (C) Residual diffuse scattering (experimental data (A) minus LLM (B). (D) Trypsin experimental data with isosurfaces shown using wireframes at a level of 3 photon counts in the resolution range 4.53 Å – 3.26. (E) Isosurfaces for diffuse scattering predicted by the Trypsin LLM model. (F) Residual diffuse scattering (experimental data (D) minus LLM (E).

## TLS models yield low correlation with diffuse scattering data

To investigate how well TLS models agree with the molecular motions in the CypA crystal, we compared the experimental diffuse data to intensities calculated from three alternative TLS models: *phenix*, *tlsmd* and *whole molecule* (**Figure 3A-D**). Although all three models predict

different motions, the R-factors are very similar: R,R-free = 16.4%,18.1% for the *whole molecule* and *Phenix* models; and 16.2%,18.1% for the TLSMD model. The correlations between the calculated diffuse intensity for these models and the anisotropic experimental data are low: 0.03 for the *phenix* model; 0.04 for the *TLSMD* model; and 0.14 for the *whole molecule* model. In addition, the pairwise correlations of the calculated diffuse intensities are low: 0.066 for *whole molecule/TLSMD*; 0.116 for *whole molecule/Phenix*; and 0.220 for *Phenix/TLSMD*.

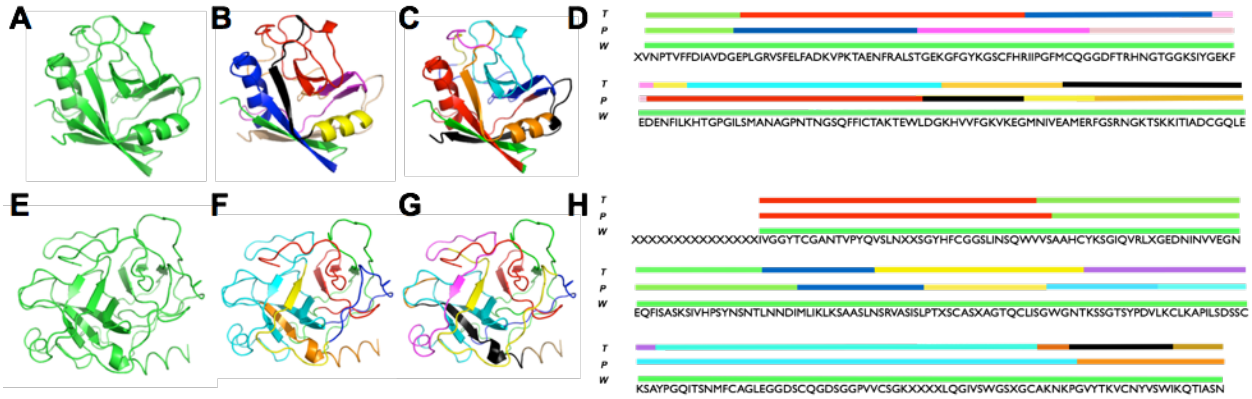


**Figure 3**. Rigid body domain definitions used for TLS models. CypA and Trypsin TLS groups shown on the tertiary structure for whole molecule (A, E), Phenix (B, F), and TLSMD (C, G) and shown on the primary sequence (D, H).

Like CypA, the three trypsin TLS models (**Figure 3E-H**) yielded very similar R,R-free values: 15.1%,16.7% for the *whole molecule* model; 15.3%,16.6% for the Phenix model; and 15.2%,16.6% for the TLSMD model. Correlations between the calculated and experimental diffuse intensities are again low: 0.02 for the *Phenix* and *TLSMD* models, and 0.08 for the *Whole molecule* model. Comparisons of the calculated anisotropic diffuse intensity show that the *Whole molecule* motion is dissimilar to both the *Phenix* and *TLSMD* predictions (PCC = 0.03 and 0.05, respectively). In contrast, the *Phenix* and *TLSMD* models yield much more similar diffuse intensities (PCC = 0.515). The relatively high correlation between these models is consistent with the similarity in the TLS groups (**Figure 3F-H)**.

There are several possible explanations for the low correlation between the TLS model and diffuse data for CypA and trypsin. First, TLS domain groupings other than those identified here might yield higher agreement with the data. Second, the method used for generating ensembles (Van Benschoten, 2015) assumes that TLS domains vary independently; it is possible that accounting for correlations among the domains would more accurately describe the variations. Lastly, similar to the rigid body motions model of Doucet & Benoit (Doucet and Benoit, 1987), the correlations among TLS domains might lead to substantial correlations across unit cell boundaries, which would produce small scale diffuse features in the immediate neighborhood of Bragg peaks. The data integration methods used here cannot resolve these features, as the measurements are mapped to a Bragg lattice. Methods to integrate the small-scale features in protein crystallography onto a finer three-dimensional reciprocal space grid do exist (Wall et al., 1997a) and might be used to address this last possibility in the future. In any case, the low correlation of TLS models with the diffuse intensity for CypA and trypsin suggests that the

variations in the protein crystal might not be best explained by motions of relatively large, rigid domains, and instead might involve motions that are correlated on a shorter length scale than accounted for by these models.

## Liquid-like motions models yield substantial correlation with diffuse scattering data

One model that accounts for short-range correlations is Liquid-Like Motions (LLM) (Caspar et al., 1988; Clarage et al., 1992). The LLM model assumes that atomic displacements are uncorrelated between different unit cells, but are correlated within the unit cells. The correlation in the displacements is assumed to decay exponentially as $f(x) = e^{-x/\gamma}$, where x is the separation of the atoms, and **γ** is the length scale of the correlation. The displacements of all atoms are assigned a standard deviation of **σ**. The LLM model previously has been refined against three-dimensional diffuse intensities obtained from crystalline staphylococcal nuclease (Wall et al., 1997b) and calmodulin (Wall et al., 1997a), yielding insights into correlated motions.

We refined isotropic LLM models of motions in CypA and trypsin against the experimental diffuse intensities (**Figure 2B, E and Methods**). The CypA model was refined using data in the resolution range 31.2 Å – 1.45 Å, and the trypsin model using 68 Å – 1.46 Å data. For CypA, the refinement yielded **γ** = 7.1 Å and **σ** = 0.38 Å with a correlation of 0.518 between the calculated and experimental anisotropic intensities. The highest correlation between data and experiment occurs in the range 3.67 Å – 3.28 Å, where the value is 0.74 **(Figure 4A)**. For the trypsin dataset, the refinement yielded **γ** = 8.35 Å and **σ** = 0.32Å with a correlation of 0.44, which is lower than for CypA. The peak value is 0.72 in the resolution range 4.53 Å -4.00 Å (**Figure 4B)**.

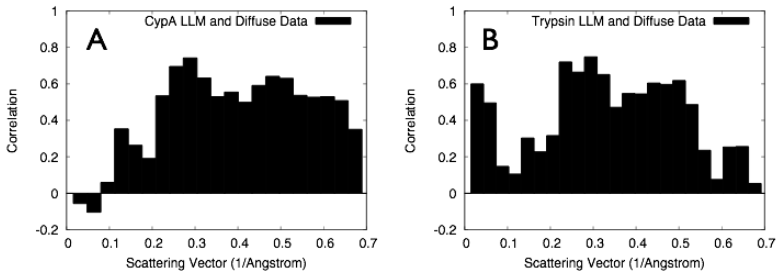


**Figure 4.** Correlations between LLM models and diffuse data computed by resolution shell for (A) CypA and (B) Trypsin.

The substantial correlation of the LLM model with the diffuse data for CypA and trypsin indicates that the variations in the protein crystal can be approximately described using a model of the protein as a soft, homogeneous medium. The model implies that the motions of atoms separated by more than 7-8 Å are relatively independent, and that atoms that are closer to each other move in a more concerted way.

## Discussion:

Diffuse X-ray scattering is a potentially valuable yet little exploited source of information about macromolecular dynamics. Diffuse intensities can double the total number of measured data points in the crystallographic experiment while providing a parallel dataset against which structural dynamical models can be refined or validated. Until now measurement of three-dimensional diffuse scattering data only has been pursued in dedicated efforts requiring extra still diffraction images and substantial optimization of experimental design. The present collection of two new datasets obtained using oscillation images using best current practices in room-temperature protein crystallography (Fraser et al., 2011), and the use of the data in evaluating TLS and LLM models, illustrates the potential for using diffuse scattering to increase understanding of protein structure variations in any X-ray crystallography experiment, and represents a significant step towards moving diffuse scattering analysis into the mainstream of structural biology.

Diffuse data obtained for CypA and trypsin can distinguish between the TLS and LLM models of motions. However, the agreement with the data is somewhat lower than in previous LLM models of three-dimensional diffuse scattering (Wall et al., 1997a; Wall et al., 1997b). In this study, the correlation of the LLM model with the data was 0.518 in the range 31.2 Å – 1.45 Å for CypA, and 0.44 in the range 68 Å – 1.46 Å for trypsin; in comparison, the correlation was 0.595 in the range 10 Å – 2.5 Å for staphylococcal nuclease (Wall et al., 1997b) and 0.55 in the range 7.5 Å – 2.1 Å for calmodulin (Wall et al., 1997a). Some possible explanations for the lower agreement for CypA and trypsin include: the use of higher resolution data in the present studies; that LLM might be a better description of motions in staphylococcal nuclease and calmodulin than in CypA and trypsin; and that the measurements might have been more accurate in the past experiments, as the data collection was tailored for diffuse scattering. The apparent alignment of the residual intensity distribution with the unit cell axes (**Figures 2C, 2F**) also suggests that an anisotropic LLM model might be more appropriate than an isotropic LLM model for CypA and trypsin.

The agreement of the LLM models with three-dimensional experimental diffuse data across multiple systems warrants further consideration for using diffuse scattering in model refinement and validation. A key finding is that the agreement of the LLM models with the diffuse data is higher than the TLS models, which currently are used widely in protein crystallography. Interestingly, the 7-8 Å length scale of the correlations is comparable to the size of the TLS domains; however, compared to the sharp domains of the TLS model, the exponential form of the correlations indicates that there is a smooth spatial transition between the correlated and uncorrelated atoms in the LLM. The smooth transition might be key to the increased agreement of the LLM with the diffuse data compared to the rigidly defined regions of the TLS model.

Overall, the three-dimensional diffuse scattering data obtained here for CypA and trypsin, and previously for staphylococcal nuclease (Wall et al., 1997b) and calmodulin (Wall et al., 1997a) suggest that the protein structure varies more like a soft material than like a collection of independent rigid domains. A normal modes model would be similarly soft, suggesting that

normal modes should be further investigated for analysis of protein motions using three-dimensional diffuse datasets. Indeed, reasonable qualitative agreement already has been seen using normal modes to model diffuse intensity in individual diffraction images (Faure et al., 1994; Mizuguchi et al., 1994), and, like TLS refinement, normal modes refinement methods have been developed for Bragg analysis (Gniewek et al., 2012; Kidera et al., 1994; Lu and Ma, 2008). An important consideration in developing these new refinement methods is to maintain a key advantage of TLS refinement at lower resolutions: the introduction of relatively few new parameters for refinement. This requirement also would be satisfied by the LLM, which has a low computational cost and general applicability, making it a promising model for integrating diffuse scattering into crystallographic model building and refinement (Wall et al., 2014a).

Diffuse scattering also can be used to validate more detailed models of molecular motions than were considered here, including models produced by normal mode refinement (Ni et al., 2009), as mentioned above; ensemble refinement (Burnley et al., 2012); multiconformer modeling performed by discrete (Keedy et al., 2015a; van den Bedem et al., 2009) or continuous (Burling and Brünger, 1994; Kuriyan et al., 1991; Wall et al., 1997a) conformational sampling; and molecular dynamics simulations (Clarage and Phillips, 1994; Clarage et al., 1995; Faure et al., 1994; Héry et al., 1998; Janowski et al., 2013; Janowski et al., 2015; Meinhold et al., 2007; Meinhold and Smith, 2005a, b; Wall et al., 2014b). In particular, molecular dynamics simulations now provide sufficient sampling to yield robust calculations of diffuse intensity (Wall et al., 2014b), and these can be used to consider a myriad of intramolecular motions (e.g., loop openings and side chain flips) (Wilson, 2013) and lattice dynamics. Polikanov and Moore (Polikanov and Moore, 2015) recently have demonstrated the importance of lattice vibrations in explaining experimental diffuse scattering measurements of ribosome crystals, which indicates that models should simultaneously account for correlations that are coupled both within and across unit cell boundaries (Clarage et al., 1992; Wall et al., 1997a); accounting for lattice vibrations more accurately also might yield improved Bragg integration (Wall et al., 2014a). Although the initial successes of dynamics-based models of diffuse scattering indicates that crystal defects can play a secondary role in contributing to the diffuse signal, at least in some cases, consideration of crystal defects might become important to achieve the highest model accuracy and most general applicability of diffuse scattering in crystallography. Additionally, as more X-ray data from both brighter conventional and XFEL light sources, accounting for all sources of Bragg and diffuse scattering will be necessary to model the total scattering needed for innovative phasing applications (Gaffney and Chapman, 2007). In summary, the new datasets presented here demonstrate that diffuse scattering can now be routinely collected and that using these data will help us obtain an increasingly realistic picture of motion in protein crystals, including integrated descriptions of intramolecular motions, lattice vibrations, and crystal defects.

# Methods:

### Protein purification and crystallization

Trypsin crystals were obtained according to the method of Liebschner *et.al* (Liebschner et al., 2013). Lyophilized bovine pancreas trypsin was purchased from Sigma-Aldrich (T1005) and dissolved at a concentration of 30 mg/mL into 30mM HEPES pH 7.0, 5 mg/mL benzamidine and 3mM $CaCl_2$. Crystals were obtained from a solution of 200mM Ammonium sulfate, 100mM Na cacodylate pH 6.5, 20% PEG 8000 and 15% glycerol. CypA was purified and crystallized as previously described (Fraser et al., 2009). Briefly, the protein was concentrated to 60 mg/mL in 20mM HEPES pH 7.5, 100mM NaCl and 500mM TCEP. Trays were set with a precipitant solution of 100mM HEPES pH 7.5, 22% PEG 3350 and 5mM TCEP. Both crystal forms were obtained using the hanging-drop method.

### Crystallographic data collection

Diffraction data were collected on beamline 11-1 at the Stanford Synchrotron Radiation Lightsource (Menlo Park, CA). X-ray diffraction images were obtained using a Dectris PILATUS 6M Pixel Array Detector (PAD). Each dataset was collected from a single crystal at an ambient temperature of 273K. To prevent dehydration, crystals were coated in a thin film of paratone with minimal surrounding mother liquor. For CypA, a single set of 0.5 degree oscillation images were collected and used for both Bragg (Keedy et al., 2015b) and diffuse data processing. A total of 360 images were collected across a 180 degree phi rotation. The Trypsin diffraction data consisted of one degree oscillations across a 135 degree phi rotation; this dataset was similarly used for both Bragg and diffuse data analysis. Both datasets were collected to optimize the Bragg signal, not the diffuse signal. Although not used here, we note that data collection using a PAD with fine phi slicing should be especially well suited for simultaneous collection of Bragg and diffuse data, as it would enable integration of diffuse intensity at a tunable level of detail in reciprocal space.

### Bragg data processing

Bragg diffraction data were processed using XDS and XSCALE (Kabsch, 2010) within the *xia2* software package (Winter et al., 2013). Molecular replacement solutions were found using Phaser (McCoy et al., 2007) within the *Phenix* software suite (Adams et al., 2010). The PDB search models were 4I8G for trypsin, and 2CPL for CypA. Initial structural refinement was performed using *phenix.refine* (Afonine et al., 2012). The strategy included refinement of individual atomic coordinates and water picking. Both the X-ray/atomic displacement parameters and X-ray/stereochemistry weights were optimized. Isotropic B-factors were chosen for the initial structures to allow for non-negligible R-factor optimization by subsequent TLS refinement strategies. All structures were refined for a total of 5 macrocycles. Statistics for these initial crystal structure models are shown in **Table 1**.

396 **Table 1.** Refinement statistics for CypA and trypsin models, before TLS modeling is applied.
397

| | **CypA** | **Trypsin** |
|---|---|---|
| Resolution range | 38.66-1.4 | 23.29-1.25 |
| Space group | P 21 21 21 | P 21 21 21 |
| Unit cell | 42.91, 52.44, 89.12 | 54.81, 58.51, 67.42 |
| Completeness (%) | 98 | 95 |
| $R_{work}$ (%) | 17.88 | 15.9 |
| $R_{free}$ (%) | 19.5 | 17.41 |
| RMS (bonds, Å) | 0.007 | 0.013 |
| RMS (angles, degrees) | 1.16 | 1.61 |
| Ramachandran favored % | 97 | 98 |
| Ramachandran allowed % | 3 | 2 |
| Ramachandran outliers % | 0 | 0 |
| Clashscore | 0.79 | 2.59 |
| Average B-factor, $Å^2$ | 21.42 | 14.57 |
| PDB ID | 5F66 | 5F6M |
| SBGrid Data Grid ID | 68 | 201 |

398
399 **Diffuse data integration**
400
401 An overview of the diffuse data integration process is presented in **Figure 1**. Image processing
402 was performed using the LUNUS collection of diffuse scattering tools (Wall, 2009). Pixels
403 corresponding to the beam stop and image edges were masked using the *punchim* and *windim*
404 methods. To focus on the diffuse intensity, which compared to Bragg peaks has low individual
405 pixel values (while, being more broadly distributed in reciprocal space, having comparable total
406 integrated intensity), pixel values outside of the range 1-10,000 photon counts were masked
407 using *threshim*. The beam polarization was determined by analyzing the first frame to determine
408 the azimuthal intensity profile within a 100 pixel wide annulus about the origin, and by fitting the
409 resulting profile to the theoretical profile (Wall, 1996). Pixel values then were corrected for beam
410 polarization using *polarim*. A solid-angle normalization (*normim*) correction was also applied.

411 Mode filtering was used to remove Bragg peaks from diffraction images. This was accomplished
412 using *modeim*, with a mask width of 20 pixels and a single bin for each photon count increment.
413 These steps produced diffraction images in which pixel values could be directly compared to
414 model diffuse intensities. This procedure is similar to the steps used by Polikanov and Moore
415 (Polikanov and Moore, 2015) to process individual ribosome diffraction images for analysis of
416 diffuse scattering data.
417
418 The Lunus processed frames were used to integrate the diffuse data onto a 3D lattice. The
419 integration was performed using a python script that calls DIALS methods within the
420 Computational Crystallography Toolbox (Parkhurst et al., 2014). The script obtains an indexing
421 solution using the *real_space_grid_search* method and uses the results to map each pixel in
422 each diffraction image to fractional Miller indices *h'k'l'* in reciprocal space. It sums the intensities
423 from pixels in the neighborhood of each integer Miller index *hkl* and tracks the corresponding
424 pixel counts, while ignoring pixels that fall within a ½ x ½ x ½ region about *hkl*. It writes the
425 intensity sums and pixel counts for each frame on a grid, populated on an Ewald sphere that
426 varies according to the crystal orientation for each image (**Figure 1C**). A radial scattering vector
427 intensity profile was calculated for each frame using the Lunus *avgrim* method and was used to
428 scale diffuse frames across the entire dataset. The Lunus *sumlt* and *divlt* methods were used to
429 compute the mean diffuse intensity at each grid point using the scaled sums and pixel counts
430 from all of the frames.
431
432 Because the model diffuse intensities were computed without considering solvent, experimental
433 and model diffuse intensities were compared using just the anisotropic component of the signal,
434 which is primarily due to the protein (Wall et al., 2014b). The Lunus *avgrlt* and *subrflt* routines
435 were applied to subtract the radial average and obtain the anisotropic signal. Signal intensities
436 were then symmetrized using *phenix.reflection_file_converter* to obtain a dataset for comparison
437 to models. Datasets were compared to each other and to models using linear correlations
438 computed using the *phenix.reflection_statistics* tool.
439
440 All images are available on SBGrid Data Grid (https://data.sbgrid.org/dataset/68/ for CypA;
441 https://data.sbgrid.org/dataset/201/ for Trypsin) and the integrated diffuse scattering maps are
442 available as Supplementary Material.
443
444 **TLS structure refinement and diffuse scattering model**
445
446 Three independent TLS refinements were performed for CypA (**Figure 3A-D**). The *Whole*
447 *molecule* selection consists of the entire molecule as a single TLS group. The *Phenix* selection
448 consists of the 8 groups (residues 2-14, 15-41, 42-64, 65-84, 85-122, 123-135, 136-145 and
449 146-165) identified by *phenix.find_tls_groups.* The *TLSMD* selection consists of 8 groups
450 (residues 2-15, 16-55, 56-80, 81-85, 86-91, 92-124, 125-143 and 144-165) identified by the TLS
451 Motion Determination web server (Painter and Merritt, 2005, 2006). All TLS refinement was
452 performed within *phenix.refine* through 5 macrocycles. Aside from the inclusion of TLS
453 refinement, these macrocycles were identical to the initial structure refinement described above.
454

Similarly, for trypsin, we selected Whole *Molecule, Phenix,* and *TLSMD* TLS refinement strategies as described above (**Figure 3E-H**). The *Phenix* selection consists of 7 TLS groups: residues 16-54, 55-103, 104-123, 124-140, 141-155, 156-225 and 226-245. The *TLSMD* selection consists of 9 groups: residues 16-52, 53-98, 99-115, 116-144, 145-171, 172-220, 221-224, 225-237 and 238-245.

Structural ensembles of the CypA and trypsin TLS motions were generated through the *Phenix.tls_as_xyz* method (Urzhumtsev et al., 2015). Each ensemble consisted of 1,000 random samples of the underlying TLS atomic displacement distributions, assuming independent distributions for each domain. Diffuse scattering models were calculated from the TLS ensembles using *Phenix.diffuse* (Van Benschoten et al., 2015). CypA and trypsin models were generated to a final resolution of 1.2 Å and 1.4 Å respectively, to match the resolution of the experimental data.

**Liquid-like motions model**

We computed Liquid-like motions (LLM) models of diffuse scattering using the structures refined prior to the TLS refinements (CypA: PDB 5F66; Trypsin: PDB 5F6M). For both CypA and trypsin, the temperature factors for all atoms were set to zero and squared calculated structure factors $I_0(hkl)$ were computed using the *structure_factors*, *as_intensity_array*, and *expand_to_p1* methods in CCTBX. The Lunus *symlt* method was used to fill in missing values in reciprocal space using the appropriate P222 Laue symmetry.

Given a correlation length $\gamma$ and amplitude of motion $\sigma$, the diffuse intensity predicted by the LLM model was calculated as

$$D(\mathbf{s}) = 4\pi^2 s^2 \sigma^2 e^{-4\pi^2 s^2 \sigma^2} I_0(\mathbf{s}) * \Gamma_\gamma(\mathbf{s})$$
$$\Gamma_\gamma(\mathbf{s}) = \frac{8\pi\gamma^3}{(1 + 4\pi^2 s^2 \gamma^2)} \quad (1)$$

Fourier methods in Lunus (*fftlt*) were used to compute the convolution. The agreement with the data was quantified by computing a linear correlation as a target function, using the anisotropic intensities (**Diffuse data integration**). Optimization of the target with respect to $\gamma$ and $\sigma$ was performed in a python script using *scipy.optimize* (www.scipy.org) with the Powell minimization method.

**Acknowledgments**

We are grateful to the UC Office of the President, Multicampus Research Programs and Initiatives grant MR-15-338599 and the Program for Breakthrough Biomedical Research, which is partially funded by the Sandler Foundation. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of

Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). N.K.S. was supported by NIH grant GM095887. J.S.F. was supported by a Searle Scholar Award from the Kinship Foundation, a Pew Scholar Award from the Pew Charitable Trusts, a Packard Fellowship from the David and Lucile Packard Foundation, NIH OD009180, NIH GM110580, and NSF STC-1231306. M.E.W. was supported by the US Department of Energy under Contract DE-AC52-06NA25396 through the Laboratory-Directed Research and Development Program at Los Alamos National Laboratory. This document is LANL technical release #LA-UR-15-28934.

## References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

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

Burling, F.T., and Brünger, A.T. (1994). Thermal motions and conformational disorder in protein crystal structures: comparison of multi-conformer and time-averaging models. Israeli J Chem *34*, 165-175.

Burnley, B.T., Afonine, P.V., Adams, P.D., and Gros, P. (2012). Modelling dynamics in protein crystal structures by ensemble refinement. Elife *1*, e00311.

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

Chacko, S., and Phillips, G.N., Jr. (1992). Diffuse x-ray scattering from tropomyosin crystals. Biophys J *61*, 1256-1266.

Clarage, J.B., Clarage, M.S., Phillips, W.C., Sweet, R.M., and Caspar, D.L. (1992). Correlations of atomic movements in lysozyme crystals. Proteins *12*, 145-157.

Clarage, J.B., and Phillips, G.N., Jr. (1994). Cross-validation tests of time-averaged molecular dynamics refinements for determination of protein structures by X-ray crystallography. Acta crystallographica Section D, Biological crystallography *50*, 24-36.

Clarage, J.B., and Phillips, G.N., Jr. (1997). Analysis of diffuse scattering and relation to molecular motion. Methods Enzymol *277*, 407-432.

Clarage, J.B., Romo, T., Andrews, B.K., Pettitt, B.M., and Phillips, G.N., Jr. (1995). A sampling problem in molecular dynamics simulations of macromolecules. Proc Natl Acad Sci U S A *92*, 3288-3292.

535 Doucet, J., and Benoit, J.P. (1987). Molecular dynamics studied by analysis of the X-ray diffuse
536 scattering from lysozyme crystals. Nature *325*, 643-646.

537 Faure, P., Micu, A., Perahia, D., Doucet, J., Smith, J.C., and Benoit, J.P. (1994). Correlated
538 intramolecular motions and diffuse X-ray scattering in lysozyme. Nat Struct Biol *1*, 124-128.

539 Fraser, J.S., Clarkson, M.W., Degnan, S.C., Erion, R., Kern, D., and Alber, T. (2009). Hidden
540 alternative structures of proline isomerase essential for catalysis. Nature *462*, 669-673.

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

544 Gaffney, K.J., and Chapman, H.N. (2007). Imaging atomic structure and dynamics with ultrafast
545 x-ray scattering. Science *316*, 1444-1448.

546 Gniewek, P., Kolinski, A., Jernigan, R.L., and Kloczkowski, A. (2012). Elastic network normal
547 modes provide a basis for protein structure refinement. J Chem Phys *136*, 195101.

548 Gruner, S.M. (2012). X-ray imaging detectors. Physics Today *65*, 29-34.

549 Guinier, A. (1963). X-ray diffraction in crystals, imperfect crystals, and amorphous bodies
550 (Courier Dover Publications).

551 Helliwell, J., Glover, I., Jones, A., Pantos, E., and Moss, D. (1986). Protein dynamics: use of
552 computer graphics and protein crystal diffuse scattering recorded with synchrotron X-radiation.
553 Biochemical Society Transactions *14*, 653-655.

554 Héry, S., Genest, D., and Smith, J.C. (1998). X-ray diffuse scattering and rigid-body motion in
555 crystalline lysozyme probed by molecular dynamics simulation. Journal of molecular biology
556 *279*, 303-319.

557 James, R. (1948). The Optical Principles of the Diffraction of X-Rays (London: Bell).

558 Janowski, P.A., Cerutti, D.S., Holton, J., and Case, D.A. (2013). Peptide crystal simulations
559 reveal hidden dynamics. J Am Chem Soc *135*, 7938-7948.

560 Janowski, P.A., Liu, C., Deckman, J., and Case, D.A. (2015). Molecular dynamics simulation of
561 triclinic lysozyme in a crystal lattice. Protein Sci.

562 Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.

563 Keedy, D.A., Fraser, J.S., and van den Bedem, H. (2015a). Exposing Hidden Alternative
564 Backbone Conformations in X-ray Crystallography Using qFit. PLoS Comput Biol *11*, e1004507.

565 Keedy, D.A., Kenner, L.R., Warkentin, M., Woldeyes, R.A., Hopkins, J.B., Thompson, M.C.,
566 Brewster, A.S., Van Benschoten, A.H., Baxter, E.L., Uervirojnangkoorn, M.*, et al.* (2015b).
567 Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL
568 crystallography. Elife *4*.

569 Keen, D.A., and Goodwin, A.L. (2015). The crystallography of correlated disorder. Nature *521*,
570 303-309.

571 Kidera, A., Matsushima, M., and Go, N. (1994). Dynamic structure of human lysozyme derived
572 from X-ray crystallography: normal mode refinement. Biophys Chem *50*, 25-31.

573 Kolatkar, A.R., Clarage, J.B., and Phillips, G.N., Jr. (1994). Analysis of diffuse scattering from
574 yeast initiator tRNA crystals. Acta Crystallogr D Biol Crystallogr *50*, 210-218.

575 Kuriyan, J., Osapay, K., Burley, S.K., Brunger, A.T., Hendrickson, W.A., and Karplus, M. (1991).
576 Exploration of disorder in protein structures by X-ray restrained molecular dynamics. Proteins
577 *10*, 340-358.

578 Kuzmanic, A., Kruschel, D., van Gunsteren, W.F., Pannu, N.S., and Zagrovic, B. (2011).
579 Dynamics may significantly influence the estimation of interatomic distances in biomolecular X-
580 ray structures. J Mol Biol *411*, 286-297.

581 Liebschner, D., Dauter, M., Brzuszkiewicz, A., and Dauter, Z. (2013). On the reproducibility of
582 protein crystal structures: five atomic resolution structures of trypsin. Acta Crystallogr D Biol
583 Crystallogr *69*, 1447-1462.

584 Lu, M., and Ma, J. (2008). A minimalist network model for coarse-grained normal mode analysis
585 and its application to biomolecular x-ray crystallography. Proc Natl Acad Sci U S A *105*, 15358-
586 15363.

587 Ma, P., Xue, Y., Coquelle, N., Haller, J.D., Yuwen, T., Ayala, I., Mikhailovskii, O., Willbold, D.,
588 Colletier, J.P., Skrynnikov, N.R.*, et al.* (2015). Observing the overall rocking motion of a protein
589 in a crystal. Nat Commun *6*, 8361.

590 McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J.
591 (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

592 Meinhold, L., Merzel, F., and Smith, J.C. (2007). Lattice dynamics of a protein crystal. Phys Rev
593 Lett *99*, 138101.

594 Meinhold, L., and Smith, J.C. (2005a). Correlated dynamics determining x-ray diffuse scattering
595 from a crystalline protein revealed by molecular dynamics simulation. Phys Rev Lett *95*,
596 218103.

597 Meinhold, L., and Smith, J.C. (2005b). Fluctuations and correlations in crystalline protein
598 dynamics: a simulation analysis of staphylococcal nuclease. Biophys J *88*, 2554-2563.

599 Meinhold, L., and Smith, J.C. (2007). Protein dynamics from X-ray crystallography: anisotropic,
600 global motion in diffuse scattering patterns. Proteins *66*, 941-953.

601 Mizuguchi, K., Kidera, A., and Go, N. (1994). Collective motions in proteins investigated by X-
602 ray diffuse scattering. Proteins *18*, 34-48.

603 Moore, P.B. (2009). On the relationship between diffraction patterns and motions in
604 macromolecular crystals. Structure *17*, 1307-1315.

605 Ni, F., Poon, B.K., Wang, Q., and Ma, J. (2009). Application of normal-mode refinement to X-ray
606 crystal structures at the lower resolution limit. Acta Crystallogr D Biol Crystallogr *65*, 633-643.

607 Painter, J., and Merritt, E.A. (2005). A molecular viewer for the analysis of TLS rigid-body
608 motion in macromolecules. Acta Crystallogr D Biol Crystallogr *61*, 465-471.

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

611 Parkhurst, J.M., Brewster, A.S., Fuentes-Montero, L., Waterman, D.G., Hattne, J., Ashton, A.W.,
612 Echols, N., Evans, G., Sauter, N.K., and Winter, G. (2014). : the diffraction experiment toolbox.
613 J Appl Crystallogr *47*, 1459-1465.

614 Perez, J., Faure, P., and Benoit, J.P. (1996). Molecular rigid-body displacements in a tetragonal
615 lysozyme crystal confirmed by X-ray diffuse scattering. Acta Crystallogr D Biol Crystallogr *52*,
616 722-729.

617 Phillips, G.N., Jr., Fillers, J.P., and Cohen, C. (1980). Motions of tropomyosin. Crystal as
618 metaphor. Biophys J *32*, 485-502.

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

621 Schomaker, V., and Trueblood, K.N. (1968). On the rigid-body motion of molecules in crystals.
622 Acta Crystallographica Section B *24*, 63-76.

623 Urzhumtsev, A., Afonine, P.V., Van Benschoten, A.H., Fraser, J.S., and Adams, P.D. (2015).
624 From deep TLS validation to ensembles of atomic models built from elemental motions. Acta
625 Crystallogr D Biol Crystallogr *71*, 1668-1683.

626 Van Benschoten, A.H., Afonine, P.V., Terwilliger, T.C., Wall, M.E., Jackson, C.J., Sauter, N.K.,
627 Adams, P.D., Urzhumtsev, A., and Fraser, J.S. (2015). Predicting X-ray diffuse scattering from
628 translation-libration-screw structural ensembles. Acta Crystallogr D Biol Crystallogr *71*, 1657-
629 1667.

630 van den Bedem, H., Bhabha, G., Yang, K., Wright, P.E., and Fraser, J.S. (2013). Automated
631 identification of functional dynamic contact networks from X-ray crystallography. Nat Methods
632 *10*, 896-902.

633 van den Bedem, H., Dhanik, A., Latombe, J.C., and Deacon, A.M. (2009). Modeling discrete
634 heterogeneity in X-ray diffraction data by fitting multi-conformers. Acta Crystallogr D Biol
635 Crystallogr *65*, 1107-1117.

636 van den Bedem, H., and Fraser, J.S. (2015). Integrative, dynamic structural biology at atomic
637 resolution--it's about time. Nat Methods *12*, 307-318.

638 Wall, M.E. (2009). Methods and software for diffuse X-ray scattering from protein crystals.
639 Methods Mol Biol *544*, 269-279.

640 Wall, M.E., Adams, P.D., Fraser, J.S., and Sauter, N.K. (2014a). Diffuse X-ray scattering to
641 model protein motions. Structure *22*, 182-184.

642 Wall, M.E., Clarage, J.B., and Phillips, G.N. (1997a). Motions of calmodulin characterized using
643 both Bragg and diffuse X-ray scattering. Structure *5*, 1599-1612.

644 Wall, M.E., Ealick, S.E., and Gruner, S.M. (1997b). Three-dimensional diffuse x-ray scattering
645 from crystals of Staphylococcal nuclease. Proceedings of the National Academy of Sciences of
646 the United States of America *94*, 6180-6184.

647 Wall, M.E., Van Benschoten, A.H., Sauter, N.K., Adams, P.D., Fraser, J.S., and Terwilliger, T.C.
648 (2014b). Conformational dynamics of a crystalline protein from microsecond-scale molecular
649 dynamics simulations and diffuse X-ray scattering. Proc Natl Acad Sci U S A *111*, 17887-17892.

650 Welberry, T.R. (2004). Diffuse X-Ray Scattering and Models of Disorder (Oxford: Oxford
651 University Press).

652 Welberry, T.R., Heerdegen, A.P., Goldstone, D.C., and Taylor, I.A. (2011). Diffuse scattering
653 resulting from macromolecular frustration. Acta Crystallogr B *67*, 516-524.

654 Wilson, M.A. (2013). Visualizing networks of mobility in proteins. Nat Methods *10*, 835-837.

655 Winter, G., Lobley, C.M., and Prince, S.M. (2013). Decision making in xia2. Acta Crystallogr D
656 Biol Crystallogr *69*, 1260-1273.
657