IUCrJ

Volume 5 (2018)

Supporting information for article:

Intermolecular correlations are necessary to explain diffuse scattering from protein crystals

Ariana Peck, Frédéric Poitevin and Thomas J. Lane

Supplemental Materials

I. EXTENDED METHODS

The disorder models assessed in this study are described in more detail below. With the exception of the traditional model of liquid-like motions, in which correlations extend throughout the crystal, these models assume that correlations are confined within the boundaries of the asymmetric unit. Because of this assumption, the models of rigid body disorder and liquid-like motions without crystal neighbors share the symmetrized molecular transform, I_{m}, as their basis:

$$I_{\mathrm{m}} = \sum_{\mathrm{asu}} |F_{\mathrm{asu}}(\mathbf{q})|^2 \tag{S1}$$

where the summation is across all asymmetric units in the unit cell. Thus, this subset of models predicts that the diffuse scattering is a blurred image of the symmetrized molecular transform, with the form of the blurring dependent on the nature of the disorder. This incoherent sum of the asymmetric unit intensities is distinct from the coherent sum of scattered intensities across the crystal or, equivalently, the crystal transform, I_{c}:

$$I_{\mathrm{c}} = |S(\mathbf{q})|^2 \left| \sum_{\mathrm{asu}} F_{\mathrm{asu}}(\mathbf{q})^2 \right| \tag{S2}$$

where the Dirac comb, $|S(\mathbf{q})|^2$, is a nonzero constant at integral Miller indices and zero at all other $\mathbf{q}$. This is in contrast to the molecular transform, which is characterized by positive, nonzero intensities throughout reciprocal space. This distinction is important for the two variants of liquid-like motions models considered in this work, as noted below.

What follows is a derivation of a general expression for diffuse scattering in the Gaussian approximation, a category that encompasses the elastic network, rigid body translations, and liquid-like motions models described below. An expression for the total ensemble (or in ergodic systems, time) averaged scattered intensity at some wavevector $\mathbf{q}$ is:

$$\langle I(\mathbf{q}) \rangle = \left\langle \sum_{c,d} e^{-i\mathbf{q}\cdot\mathbf{u}_{cd}} \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} e^{-i\mathbf{q}\cdot\boldsymbol{\delta}_{c_i d_j}} \right\rangle \tag{S3}$$

where f_i is the atomic form factor for atom i, $\mathbf{u}_{cd}$ is the vector between the origins of unit cells c and d, $\mathbf{r}_{ij}$ is the interatomic distance vector between atoms i and j, and $\boldsymbol{\delta}_{c_i d_j}$ is the interatomic displacement vector between atom i in unit cell c and atom j in unit cell d. In the Gaussian approximation, the statistical ensemble of atomic displacements may be described by a pairwise multivariate normal distribution, with zero mean and covariance matrix, $V_{c_i d_j} = \langle \boldsymbol{\delta}_{c_i}^T \boldsymbol{\delta}_{d_j} \rangle \in \mathcal{R}^{3\times 3}$:

$$p(\boldsymbol{\delta}_{c_i}, \boldsymbol{\delta}_{d_j}) \sim \mathrm{MVN}(\mathbf{0}, V_{c_i d_j})$$

Since the average is over pairwise probability distributions, eq. S3 may be rewritten:

$$\begin{aligned} \langle I(\mathbf{q}) \rangle &= \sum_{c,d} e^{-i\mathbf{q}\cdot\mathbf{u}_{cd}} \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} \iint p(\boldsymbol{\delta}_{c_i}, \boldsymbol{\delta}_{d_j}) e^{-i\mathbf{q}\cdot(\boldsymbol{\delta}_{c_i} - \boldsymbol{\delta}_{d_j})} d\boldsymbol{\delta}_{c_i}\, d\boldsymbol{\delta}_{d_j} \\ &= \sum_{c,d} e^{-i\mathbf{q}\cdot\mathbf{u}_{cd}} \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} e^{-\frac{1}{2}\mathbf{q}^T V_{c_i c_i}\mathbf{q} - \frac{1}{2}\mathbf{q}^T V_{d_j d_j}\mathbf{q} + \mathbf{q}^T V_{c_i d_j}\mathbf{q}} \end{aligned} \tag{S4}$$

We further assume that:

1. Correlations between atomic displacements in different unit cells are independent: $\langle \boldsymbol{\delta}_{c_i}^T \boldsymbol{\delta}_{d_j} \rangle = \mathbf{0}$ if $c \neq d$.

2. Atoms in different unit cells behave identically in a statistical fashion: $p(\boldsymbol{\delta}_{ci}) = p(\boldsymbol{\delta}_{di})$ for all i.

With these simplifying assumptions, $V_{c_i d_j} = \mathbf{0}$ if $c \neq d$, and $V_{c_i c_i}$ is identical for all c, such that $V_{c_i c_i} = V_{ii}$ (and similarly $V_{d_j d_j} = V_{jj}$). Eq. S4 can then be split into two terms: one expressing interference between unit cells (where

$V_{c_i d_j} = \mathbf{0}$), and one expressing interference *within* repeats of a single cell:

$$\langle I(\mathbf{q})\rangle = \sum_{c,d\neq c} e^{-i\mathbf{q}\cdot\mathbf{u}_{cd}} \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} e^{-\frac{1}{2}\mathbf{q}^T V_{ii}\mathbf{q} - \frac{1}{2}\mathbf{q}^T V_{jj}\mathbf{q}} + N\sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} e^{-\frac{1}{2}\mathbf{q}^T V_{ii}\mathbf{q} - \frac{1}{2}\mathbf{q}^T V_{jj}\mathbf{q} + \mathbf{q}^T V_{ij}\mathbf{q}} \tag{S5}$$

$$= \sum_{c,d} e^{-i\mathbf{q}\cdot\mathbf{u}_{cd}} \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} e^{-\frac{1}{2}\mathbf{q}^T V_{ii}\mathbf{q} - \frac{1}{2}\mathbf{q}^T V_{jj}\mathbf{q}} + N\sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} e^{-\frac{1}{2}\mathbf{q}^T V_{ii}\mathbf{q} - \frac{1}{2}\mathbf{q}^T V_{jj}\mathbf{q}} \left[e^{\mathbf{q}^T V_{ij}\mathbf{q}} - 1\right] \tag{S6}$$

where N is the number of unit cells. The first term is recognizable as the expression corresponding to Bragg diffraction:

$$I(\mathbf{q})_{\text{Bragg}} = \sum_{c,d} e^{-i\mathbf{q}\cdot\mathbf{u}_{cd}} \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} e^{-\frac{1}{2}\mathbf{q}^T V_{ii}\mathbf{q} - \frac{1}{2}\mathbf{q}^T V_{jj}\mathbf{q}} \tag{S7}$$

$$= \left| \left(\sum_c e^{-i\mathbf{q}\cdot\mathbf{u}_c}\right) \left(\sum_i f_i e^{-i\mathbf{q}\cdot\mathbf{r}_i} e^{-\frac{1}{2}\mathbf{q}^T V_{ii}\mathbf{q}}\right) \right|^2 \tag{S8}$$

$$= |S(\mathbf{q})|^2 \, |F(\mathbf{q})|^2 \tag{S9}$$

where atomic form factors are scaled by anisotropic Debye-Waller factors. $|S(\mathbf{q})|^2$ becomes a Dirac comb as the number of unit cells grows, showing this scattering is localized to discrete regions of $\mathbf{q}$. We consider the remaining scattering $\langle I(\mathbf{q})\rangle - I(\mathbf{q})_{\text{Bragg}}$ to be the diffuse scattering intensity:

$$I_{\text{diffuse}}(\mathbf{q}) = N\sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} e^{-\frac{1}{2}\mathbf{q}^T V_{ii}\mathbf{q} - \frac{1}{2}\mathbf{q}^T V_{jj}\mathbf{q}} \left[e^{\mathbf{q}^T V_{ij}\mathbf{q}} - 1\right] \tag{S10}$$

There are two notable features of the diffuse scattering. First, lacking the lattice transform $|S(\mathbf{q})|^2$ it is *not* localized in reciprocal space. Second, it is non-trivial only if there are correlated displacements between atoms, i.e. when $V_{ij} \neq \mathbf{0}$.

Elastic network model

This model makes the assumptions outlined above in deriving eq. S10, with the further restriction that correlations be confined within the boundaries of the asymmetric unit. This additional assumption renders this model more biologically interpretable. For each system, the covariance matrix, V_{ij}, was determined from an elastic network model of the ordered atoms in the asymmetric unit. Specifically, the normal modes of each system were generated based on the protein's topology in torsion angle space, with a uniform spring constant for all atom pairs within a certain distance in this internal coordinate space [S6]. The first ten normal modes were then summed to generate C_{ij}, the isotropic correlation coefficient between the displacements of asymmetric unit atoms i and j. Entries in this correlation matrix were converted to covariances using the following formula:

$$V_{ij} = C_{ij}\sqrt{\langle\delta_i^2\rangle\langle\delta_j^2\rangle} \tag{S11}$$

where the mean-square atomic displacements, $\langle\delta_i^2\rangle$, are related to the isotropic B factors by: $B_i = 8\pi^2\langle\delta_i^2\rangle$. Thus, the amplitudes of motions described by the covariance matrix are consistent with the refined Bragg models. Diffuse scattering maps were predicted from these covariance matrices using eq. S10.

Rigid body translational disorder

The model of rigid body translational disorder is a special case of diffuse scattering in the Gaussian approximation (eq. S10) that further assumes that all atoms in the asymmetric unit are displaced as a rigid structural unit. The

displacement covariance between all atom pairs is thus identical and assuming isotropic translations can be described by a scalar, σ^2. The expression for the diffuse scattering intensity is:

$$\begin{aligned} I_{\text{diffuse}}(\mathbf{q}) &= N \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} \left[1 - e^{-q^2\sigma^2}\right] \\ &= N \left[1 - e^{-q^2\sigma^2}\right] I_{\text{m}} \end{aligned} \tag{S12}$$

where N is the number of unit cells and I_m is the symmetrized molecular transform (eq. S1). This expression has previously been derived in Refs. [S2] and [S3]. For each system, the molecular transform was computed from the refined Bragg coordinates, excluding solvent, hydrogen, and crystallographically-unresolved atoms (which were assumed to exhibit uncorrelated disordered behavior and thus contribute to radially symmetric rather than anisotropic diffuse scattering). Best fit values of σ were determined by scanning over this parameter to maximize the CC between the experimental and predicted maps.

Liquid-like motions

The liquid-like motions model is a specific case of Gaussian translational disorder in which correlated motions between atoms decay with interatomic distance [S4, S5]. The expression for the diffuse intensity predicted by this model has previously been derived by making use of the Patterson [S5]; here, we provide a derivation based on the scattered intensity in reciprocal space. This model assumes the following:

1. A global isotropic displacement parameter: $V_{c_i c_i} = \sigma^2$ for all c and i.
2. Interatomic covariances are isotropic and depend on interatomic distance: $V_{c_i d_j} = \sigma^2 \Gamma(\mathbf{r}_{c_i d_j})$, where the kernel Γ describes the correlation length.

With these assumptions, eq. S4 (which still permits correlations between unit cells) can be rewritten:

$$I(\mathbf{q}) = \sum_{c,d} \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{c_i d_j}} e^{-q^2\sigma^2} e^{q^2\sigma^2\Gamma(\mathbf{r}_{c_i d_j})} \tag{S13}$$

If we additionally assume that $q^2\sigma^2\Gamma(\mathbf{r}_{c_i d_j})$ is small, then we can perform Taylor expansion on the exponential $\exp\{q^2\sigma^2\Gamma(\mathbf{r}_{c_i d_j})\} \approx 1 + q^2\sigma^2\Gamma(\mathbf{r}_{c_i d_j})$. Although the exclusion of higher-order terms renders this model less valid at high resolution, this assumption was generally reasonable for the values of q and σ considered here. In the case of CypA, for instance, inclusion of the second-order term affected the CC between the predicted and experimental maps by less than 0.01 (data not shown). Eq. S13 can then be further simplified:

$$\begin{aligned} I(\mathbf{q}) &= e^{-q^2\sigma^2} \sum_{c,d} \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{c_i d_j}} \left[1 + q^2\sigma^2\Gamma(\mathbf{r}_{c_i d_j})\right] \\ &= e^{-q^2\sigma^2} \sum_{c,d} \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{c_i d_j}} + q^2\sigma^2 e^{-q^2\sigma^2} \sum_{c,d} \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{c_i d_j}} \Gamma(\mathbf{r}_{c_i d_j}) \\ &= e^{-q^2\sigma^2} I_c + q^2\sigma^2 e^{-q^2\sigma^2} \sum_{c,d} \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{c_i d_j}} \Gamma(\mathbf{r}_{c_i d_j}) \end{aligned} \tag{S14}$$

The first term corresponds to the Bragg intensity scaled by a global Debye-Waller factor, while the remaining scattering in the second term corresponds to the diffuse intensity predicted by the liquid-like motions model. This term can be simplified by defining a new kernel function:

$$s(\mathbf{q}, \mathbf{r}) = \sum_{c,d} \sum_{i,j} f_i f_j \delta(\mathbf{r} - \mathbf{r}_{c_i d_j}) \tag{S15}$$

where $\delta(\cdot)$ denotes the Dirac delta function (rather than a displacement). Using this kernel, the second term in eq. S14 can be rewritten:

$$\begin{aligned} I_{\text{diffuse}}(\mathbf{q}) &= q^2\sigma^2 e^{-q^2\sigma^2} \int s(\mathbf{q}, \mathbf{r}) \Gamma(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} \, d\mathbf{r} \\ &= q^2\sigma^2 \, e^{-q^2\sigma^2} \frac{1}{16\pi^6} \left[\left(\int s(\mathbf{q}, \mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} \, d\mathbf{r} \right) * \left(\int \Gamma(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} \, d\mathbf{r} \right) \right] \end{aligned} \tag{S16}$$

where the second step takes advantage of the Fourier convolution theorem, with $*$ denoting convolution. Then:

$$\begin{aligned}\int s(\mathbf{q},\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}\,d\mathbf{r} &= \int \sum_{c,d}\sum_{i,j} f_i f_j \delta(\mathbf{r}-\mathbf{r}_{c_i d_j})e^{i\mathbf{q}\cdot\mathbf{r}}\,d\mathbf{r} \\ &= \sum_{c,d}\sum_{i,j} f_i f_j e^{i\mathbf{q}\cdot\mathbf{r}_{c_i d_j}} = I_c \end{aligned} \tag{S17}$$

where I_c is the scattered intensity of the coherently diffracting volume, in this case the crystal transform. The diffuse scattering predicted by the liquid-like motions model is thus a convolution between this intensity function and the Fourier transform of the correlation kernel, Γ:

$$I_{\text{diffuse}}(\mathbf{q}) = q^2\sigma^2\, e^{-q^2\sigma^2}\,\frac{1}{16\pi^6}\left[I_{\text{c}} * \tilde{\Gamma}(\mathbf{q})\right]. \tag{S18}$$

where $\tilde{\Gamma}(\mathbf{q}) = \int \Gamma(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}\,d\mathbf{r}$. We call this model, where correlated disorder to extends between neighboring asymmetric units and across unit cell boundaries, the model of liquid-like motions "with neighbors".

If, however, we initially assume that correlations are confined within the boundaries of the asymmetric unit ($\Gamma(r_{c_i d_j}) = 0$ if atoms c_i and d_j are not members of the same asymmetric unit), then we can re-formulate the liquid-like motions model as a function of the molecular, rather than crystal, transform. To see this, start again with the second (diffuse) term in eq. S14,

$$I_{\text{diffuse}}(\mathbf{q}) = q^2\sigma^2 e^{-q^2\sigma^2}\sum_{c,d}\sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{c_i d_j}}\Gamma(\mathbf{r}_{c_i d_j})$$

Under our approximation, many terms in this sum are zero. Specifically, any where atoms c_i and d_j are not members of the same asymmetric unit. Dropping these terms, we obtain

$$I_{\text{diffuse}}(\mathbf{q}) = q^2\sigma^2 e^{-q^2\sigma^2} N \sum_{\text{asu}}\sum_{i,j\in\text{asu}} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}}\Gamma(\mathbf{r}_{ij}) \tag{S19}$$

where, again, $\sum_{asu}$ indicates a single summation over all unique copies of the asymmetric unit within a single unit cell replica. Note the sum across unit cells produces the scale factor N. No asymmetric unit spans more than one unit cell.

Now we follow the same tactic as above. Define:

$$s_{\text{asu}}(\mathbf{q},\mathbf{r}) = \sum_{i,j\in\text{asu}} f_i f_j \delta(\mathbf{r}-\mathbf{r}_{ij}) \tag{S20}$$

and re-write eq S19

$$\begin{aligned} I_{\text{diffuse}}(\mathbf{q}) &= q^2\sigma^2 e^{-q^2\sigma^2} N \sum_{\text{asu}} \int s_{\text{asu}}(\mathbf{q},\mathbf{r})\,\Gamma(\mathbf{r})\,e^{-i\mathbf{q}\cdot\mathbf{r}}\,d\mathbf{r} \\ &= q^2\sigma^2 e^{-q^2\sigma^2}\frac{N}{16\pi^6}\sum_{\text{asu}}\left[\left(\int s_{\text{asu}}(\mathbf{q},\mathbf{r})\,e^{i\mathbf{q}\cdot\mathbf{r}}\,d\mathbf{r}\right) * \left(\int \Gamma(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}\,d\mathbf{r}\right)\right] \\ &= q^2\sigma^2 e^{-q^2\sigma^2}\frac{N}{16\pi^6}\left[\left(\sum_{\text{asu}}\sum_{i,j\in\text{asu}} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}}\right) * \left(\int \Gamma(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}\,d\mathbf{r}\right)\right] \end{aligned} \tag{S21}$$

where in the last step we have used the fact that convolution, as a linear operation, distributes. Recalling

$$\sum_{\text{asu}}\sum_{i,j\in\text{asu}} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} = I_m$$

we obtain

$$I_{\text{diffuse}}(\mathbf{q}) = q^2\sigma^2\, e^{-q^2\sigma^2}\,\frac{N}{16\pi^6}\left[I_{\text{m}} * \tilde{\Gamma}(\mathbf{q})\right]. \tag{S22}$$

The lack of coherence between unit cells results in loss of the Dirac comb from the intensity function, and the addition of a factor that scales the expression by the number of unit cells, N. We refer to this as the ASU-confined liquid-like motions model.

For both models, we employed a previously-described form of the kernel, in which covariances decay exponentially as a function of interatomic distance [S4, S5]:

$$\Gamma(\mathbf{q}) = \frac{8\pi\gamma^3}{(1+q^2\gamma^2)^2} \tag{S23}$$

Prior studies on the liquid-like motions model with neighbors analyzed experimental maps that sampled the observed diffuse scattering intensity at integral Miller indices only [S4]. In this study, we compare the predicted signal to experimental maps that sample the diffuse signal at fractional Miller indices, which enables us to assess long range correlations that extend beyond a unit cell. For both liquid-like motions models, best fit values of the isotropic displacement parameter and correlation length were determined by a grid search to maximize the CC between the experimental and predicted maps. As an example, convergence of these parameters for the internally-disordered model is shown in Fig. S6B.

Rigid body rotational disorder

The simplest case of rigid body rotational disorder was evaluated, in which there is no preferred axis of rotation and rotation angles are sampled from a normal distribution. Rotations were additionally assumed to be independent and uncorrelated between asymmetric units. Diffuse scattering maps were predicted from an ensemble of rotated molecules using Guinier's equation [S1]:

$$I_{\text{diffuse}}(\mathbf{q}) \propto \sum_{\text{asu}} \left[\langle |F_{\text{asu, n}}(\mathbf{q})|^2 \rangle - \langle |F_{\text{asu, n}}(\mathbf{q})| \rangle^2 \right]. \tag{S24}$$

where $F_{\text{asu, n}}(\mathbf{q})$ represents the asymmetric unit transform for the nth ensemble member and $\langle ... \rangle$ indicates the time or ensemble average (which these data cannot distinguish between). In order to focus on the diffuse scattering predicted solely by rotational disorder, the asymmetric unit transforms were not scaled by Debye-Waller factors, which account for translational disorder effects. For each map, the best fit standard deviation of the rotational distribution was determined by scanning over values of this parameter to maximize the CC between the experimental and predicted maps (Fig. S6A). Convergence of the ensembles was determined by ensuring that the CCs with the experimental map were within 0.01 for independent 'trajectories' generated with the same rotation parameter.

Ensemble models

Many types of protein disorder involve transitions between discrete states rather than along a continuum of alternate conformations. In real space, this disorder can be modeled as an ensemble of representative "snapshots" of distinct protein configurations. The refined Bragg coordinates of the CypA and AP crystal structures suggested the existence of specific ensembles for these systems. Here we chose the simplest representation of each ensemble: a two-state model, with each state represented by a single, probability-weighted conformation. Probability weights were derived from the refined occupancy values in the Bragg model, and each state was assumed to exhibit uncorrelated atomic disorder that could be adequately described by isotropic B-factors. Diffuse scattering maps were predicted from each two-state ensemble using Guinier's equation (eq. S24) [S1] and scaled by the Debye-Waller factor, $e^{-q^2\sigma^2}$, with the global atomic displacement factor σ computed from the Bragg Wilson B factor.

II. EXTENDED FIGURES

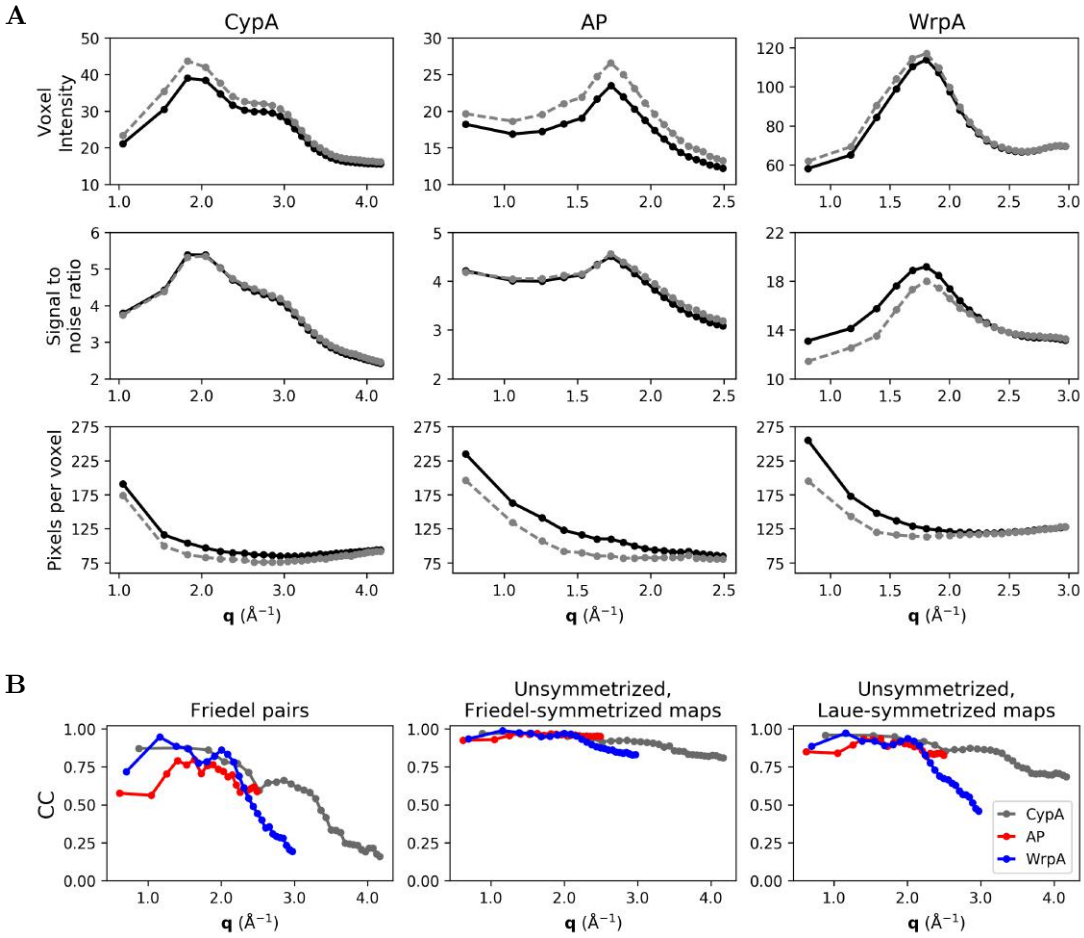

FIG. S1. **Diffuse scattering map statistics by resolution shell.** (A) The median voxel intensity, signal to noise ratio, and number of pixels per voxel are shown across resolution bins for the indicated experimental map. The signal to noise ratio for each voxel was estimated as $\langle I \rangle / \sigma(I)$ for the set of pixel intensities binned into each voxel. The solid and dashed lines correspond to the overall values and the values for voxels centered on integral Miller indices, respectively. (B) Correlation coefficients between Friedel pairs (left), the indicated map before and after averaging Friedel pairs (center), and the indicated map before and after averaging Laue-symmetric voxels (right).

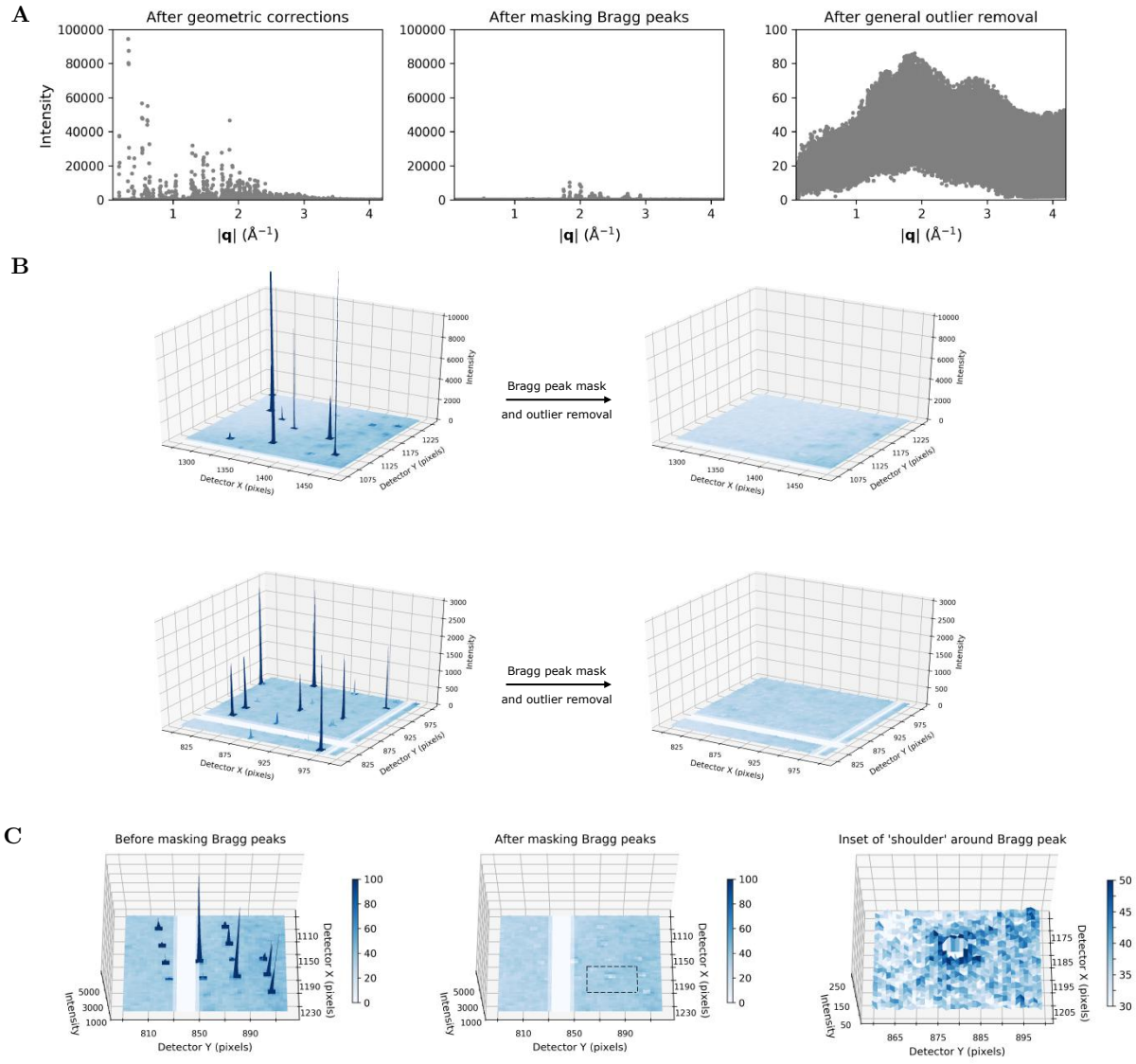

FIG. S2. **Elimination of Bragg peaks.** (A) Intensity distributions for a representative diffraction image from the CypA dataset after applying geometric corrections, masking Bragg peaks, and general outlier removal from left to right. (B) Surface plots of a 200 x 200 pixel region of the image at 5.7 (top) and 2.3 Å (bottom) resolution before and after removal of Bragg peaks and outliers. (C) Representative surface plots as in (B), except tilted. The rightmost panel is shows the diffuse 'shoulder around the masked Bragg peak boxed in the center panel.

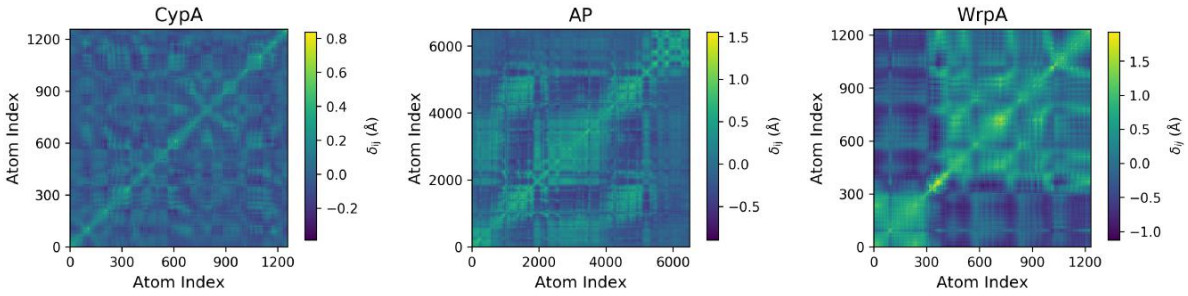

FIG. S3. **Covariance matrices generated from elastic network models.** The correlation matrix of interatomic Gaussian displacements was predicted for each system by an elastic network model. The covariance matrix was then computed from this correlation matrix and normalized by the refined B factors to ensure consistency with the Bragg data.

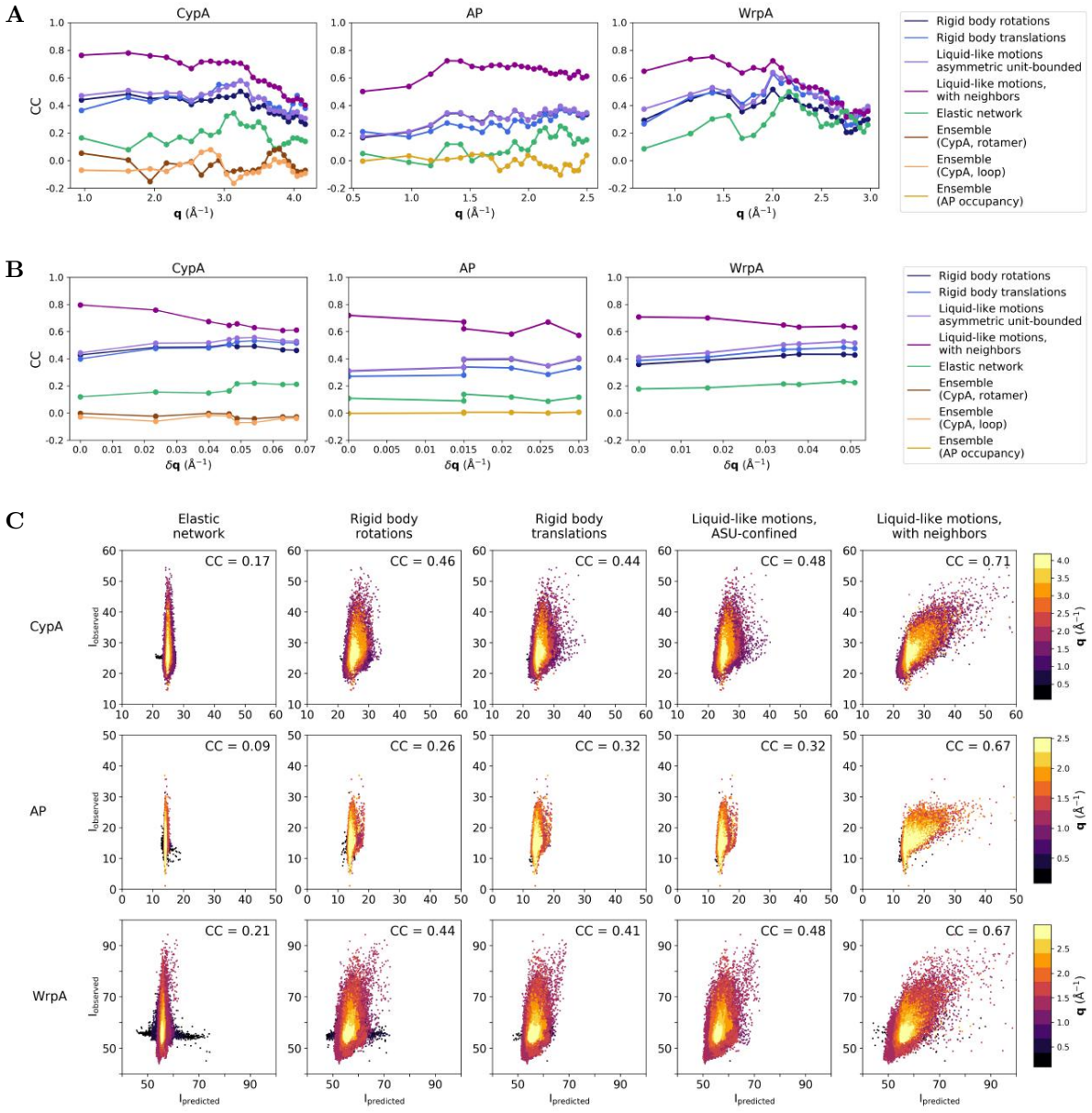

FIG. S4. **Correlation coefficients by resolution shell.** The multiplicity-weighted correlation coefficient between the experimental and predicted map for the indicated models is plotted as a function of (A) resolution shell and (B) distance from reciprocal lattice sites. (C) Scatter plots of the model-predicted versus observed intensities, with points colored by resolution; the overall correlation coefficient is noted in the upper right corner of each plot.

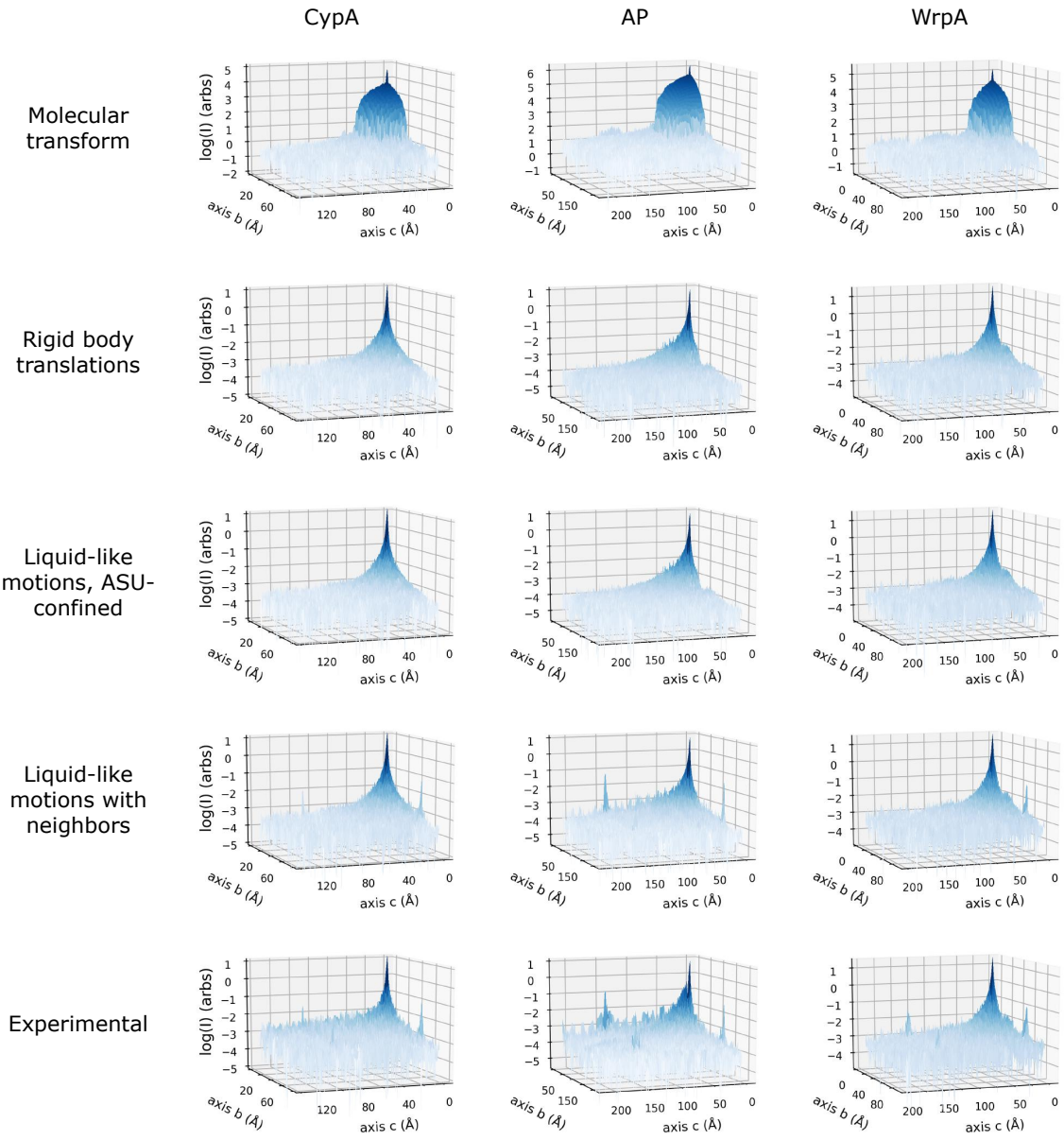

FIG. S5. **Autocorrelation functions of the predicted and experimental maps.** The 3D autocorrelation function of the indicated map was computed by Fourier methods; shown are quadrants from the projection along the crystallographic a axis. The low intensity peaks near the boundaries of the map are consistent with the unit cell dimensions along the relevant crystallographic axis. Note that difference vectors in autocorrelation space overestimate the shape of the real space object by a factor of two, but this effect is balanced by only viewing a quadrant of the map (and considering distance to the origin, rather than to the symmetric peak).

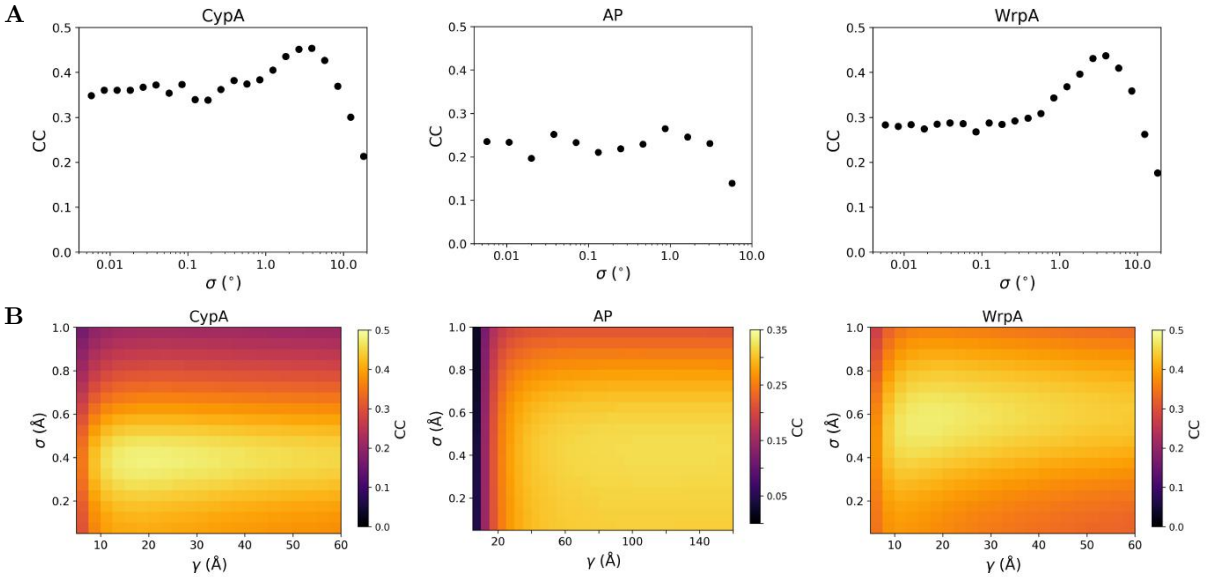

FIG. S6. **Convergence of disorder parameters for the rigid body rotational disorder and liquid-like motions models.** (A) Scans over the standard deviation of the rotational distribution, σ, were performed to fit the rigid body rotational disorder model to each experimental map. (B) Grid scans over the global atomic displacement factor, σ, and the correlation length, γ, parameters of the liquid-like motions model. The color indicates the overall correlation coefficient between the experimental and predicted maps.

[S1] A. Guinier, *X-ray diffraction in crystals, imperfect crystals, and amorphous bodies* (W. H. Freeman, San Francisco, 1963).
[S2] P. B. Moore, Structure **17**, 1307 (2009)
[S3] K. Ayyer, O. M. Yefanov, D. Oberthur, S. RoyChowdhury, L. Galli, V. Mariani, S. Basu, J. Coe, C. E. Conrad, R. Fromme, A. Schaffer, K. Dorner, D. James, C. Kupitz, M. Metz, G. Nelson, P. L. Xavier, K. R. Beyerlein, M. Schmidt, I. Sarrou, J. C. Spence, U. Weierstall, T. A. White, J. H. Yang, Y. Zhao, M. Liang, A. Aquila, M. S. Hunter, J. S. Robinson, J. E. Koglin, S. Boutet, P. Fromme, A. Barty, and H. N. Chapman, Nature **530**, 202 (2016).
[S4] M. E. Wall, J. B. Clarage, and G. N. Phillips, Structure **5**, 1599 (1997).
[S5] J. B. Clarage, M. S. Clarage, W. C. Phillips, R. M. Sweet, and D. L. Caspar, Proteins **12**, 145 (1992).
[S6] J. K. Bray, D. R. Weiss, and M. Levitt, Biophys. J. **101**, 2966 (2011).