HHS Public Access

Author manuscript

Chem Rev. Author manuscript; available in PMC 2018 June 28.

Published in final edited form as:

Chem Rev. 2017 June 28; 117(12): 7615–7672. doi:10.1021/acs.chemrev.6b00790.

X-ray scattering studies of protein structural dynamics

Steve P. Meisburger[†,‡], **William C. Thomas**[†,‡], **Maxwell B. Watkins**[†,‡], and **Nozomi Ando**[†]

[†]Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States

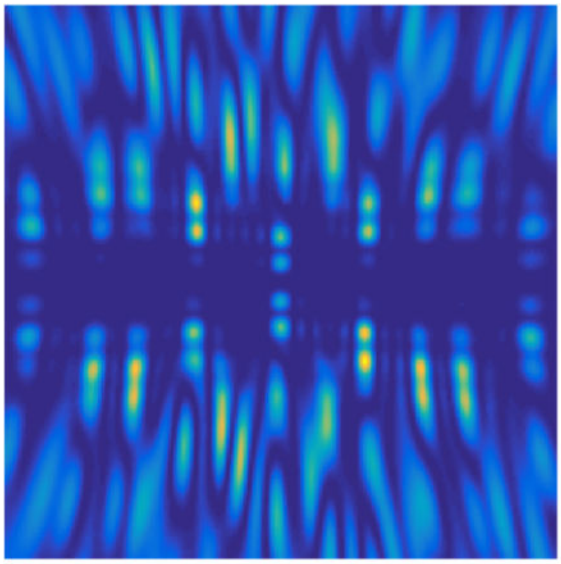
Abstract

X-ray scattering is uniquely suited to the study of disordered systems and thus has the potential to provide insight into dynamic processes where diffraction methods fail. In particular, while X-ray crystallography has been a staple of structural biology for more than half a century and will continue to remain so, a major limitation of this technique has been the lack of dynamic information. Solution X-ray scattering has become an invaluable tool in structural and mechanistic studies of biological macromolecules where large conformational changes are involved. Such systems include allosteric enzymes that play key roles in directing metabolic fluxes of biochemical pathways, as well as large, assembly-line type enzymes that synthesize secondary metabolites with pharmaceutical applications. Furthermore, crystallography has the potential to provide information on protein dynamics via the diffuse scattering patterns that are overlaid with Bragg diffraction. Historically, these patterns have been very difficult to interpret, but recent advances in X-ray detection have led to a renewed interest in diffuse scattering analysis as a way to probe correlated motions. Here, we will review X-ray scattering theory and highlight recent advances in scattering-based investigations of protein solutions and crystals, with a particular focus on complex enzymes.

Graphical Abstract

Correspondence to: Nozomi Ando.

[‡]Contributed equally to this work



1 Introduction

One of the defining challenges of biological chemistry involves untangling the complex mechanisms relating protein structure and function. Although X-ray crystallography remains the technique of choice for obtaining high-resolution snapshots of macromolecules, the next frontier of structural enzymology, and indeed all of structural biology, is to understand the role of dynamics in molecular function. Both diffraction and scattering result from the interference of X-rays elastically scattered from electrons in the sample. The degree of disorder in the sample leads to characteristic patterns; whereas a perfectly ordered lattice produce sharp diffraction spots known as Bragg peaks, the more disordered the sample, the more cloudy and diffuse the resultant scattering becomes (Figure 1). Crystal structures result from analysis of Bragg peak intensities (Figure 1, top), and thus they represent the average atomic positions of the repeating molecule in the lattice. As a result of this limitation, crystal structures are often thought to represent static snapshots. However, proteins are constantly in motion, and such inherently disordered systems are better characterized by scattering techniques than by diffraction (Figure 1, bottom two). In particular, solution X-ray scattering provides a quantitative means to investigate large conformational changes, and the diffuse scattering from crystals provides a unique opportunity for studying the breathing motions of proteins.

Solution scattering methods such as small- and wide-angle X-ray scattering (SAXS/WAXS) have in recent years emerged as powerful approaches for the study of enzymes of biochemical and biomedical importance, particularly in combination with other biophysical techniques. Proteins that are randomly oriented in solution give rise to rotationally symmetric scattering images (Figure 1, bottom) that are integrated to produce one-dimensional scattering profiles. These scattering profiles contain a surprising wealth of structural information, such as radius of gyration, mass, compactness, flexibility, and overall shape. Notably, experimental scattering profiles can be directly compared with theoretical

scattering profiles calculated from high-resolution models.[2] Furthermore, these multivariate data allow for mathematical separation techniques, and thus solution scattering is a particularly powerful tool for investigating dynamic proteins, which exist as mixtures of multiple forms and confound other structural methods.

Diffuse scattering from crystals arises from disorder and thus has the potential to contain information on biochemically important motions (Figure 1, middle). Although protein motions are generally more restricted in a crystal, side-chain and even domain motions can occur.[3–5] In conventional crystallography, such deviations from the average atomic positions are described by the so-called Debye-Waller factor or thermal B-factor. While high B-factors can imply high mobility, they cannot distinguish different modes of motion. However, correlated motions give rise to a faint, textured interference pattern in the background of Bragg diffraction, which is typically ignored in conventional crystallography.[1,5–7] Interpretation of these diffuse patterns has until recently been limited by detector technology and the complexity of the data. As such, this field is still nascent – much like the SAXS field prior to the introduction of data analysis algorithms in the mid-1990's.[2] However, with the advent of pixel-array detectors and hard X-ray free electron lasers (XFELs), there has been a resurgence of interest in this field.

In this review, we will focus on the unique capabilities of X-ray scattering in probing protein structural ensembles over a wide range of length scales. A comprehensive review of X-ray scattering theory will be given in Section 2. As solution scattering has been extensively discussed in other reviews,[8–15] we will devote Section 3 to reviewing studies where this technique has had a major impact on our understanding of enzyme mechanism and regulation. Finally, Section 4 will comprehensively review experimental investigations of protein diffuse scattering, as well as the emergent field known as crystallography beyond Bragg diffraction.

2 X-ray Scattering Theory

2.1 Physics of X-ray Scattering

Macromolecular crystallography (MX), SAXS, and crystal diffuse scattering differ in the nature of the sample, experimental geometry, and obtainable information, but they share the same underlying physics. Owing to historical tradition, the theories of SAXS and MX are usually presented with different notational conventions. Here, we use a common notation ("physicists' notation") to clarify the connections between the three techniques. The first section provides an overview of the physics of X-ray scattering of biological macromolecules. The theory of solution scattering is outlined in the second section, and the theory of diffuse scattering from crystals is given in the final section. Further details can be found in classic texts for the theory of biological SAXS[16–18] and crystallography.[18–20]

2.1.1 Scattering of X-rays by Electrons—Scattering results from the interaction between X-ray photons and electrons in a sample. The quantum mechanical theory for the interaction between a photon and a free electron was first derived by Klein and Nishina in 1929.[21] Depending on the energy of the incident photon, the scattering process may be elastic or inelastic. In structural biology, where the photon energy (typically ~ 6 – 20 keV) is

much less than the rest mass of the electron (511 keV), elastic scattering predominates. For structural studies, the most important consequence of elastic scattering is that phase coherence is maintained between the incident and scattered photons, and hence the scattered waves from electrons in the sample interfere. Scattered photons are recorded on a detector, and information concerning the relative positions of the electrons can be inferred from the interference pattern.

In a typical scattering experiment, a monochromatic X-ray source with wavelength λ is highly collimated, so that on short length scales it resembles a plane wave with wavevector $\mathbf{k} = 2\pi\hat{\mathbf{s}}/\lambda$. A detector pixel in the direction $\hat{\mathbf{s}}'$ records the scattered photons. Since the distance from the sample to the detector is much larger than the distances between electrons in the sample, we are performing what is known as far field diffraction, or Frauenhoffer diffraction. A detector in the far field sees a scattered wave with wavevector $\mathbf{k}' = 2\pi\hat{\mathbf{s}}'/\lambda$ (for elastic scattering, $|\mathbf{k}'| = |\mathbf{k}|$). The wavevector transfer $\mathbf{q}$, or scattering vector,

$$\mathbf{q}=\mathbf{k}'-\mathbf{k}=2\pi\left(\hat{\mathbf{s}}'-\hat{\mathbf{s}}\right)/\lambda, \quad (1)$$

is of central importance in SAXS and crystallography. The magnitude of $\mathbf{q}$ is

$$q \equiv |\mathbf{q}|=4\pi\sin(\theta)/\lambda, \quad (2)$$

where 2θ is the angle between $\hat{\mathbf{s}}$ and $\hat{\mathbf{s}}'$. When a pair of scatterers are separated by a vector, $\mathbf{r}$, their scattered waves have a phase difference at the detector of $\Delta\phi = (\mathbf{k}' - \mathbf{k}) \cdot \mathbf{r} = \mathbf{q}\cdot\mathbf{r}$. The phase difference is expressed using a complex exponential, $e^{i\mathbf{q}\cdot\mathbf{r}}$. A system of free electrons, each with position $\mathbf{r}_n$, produces a scattered wave with an amplitude that is proportional to the form factor,

$$F(\mathbf{q})=\sum_n e^{i\mathbf{q}\cdot\mathbf{r}_n}. \quad (3)$$

The intensity of the interference pattern, the quantity actually recorded by the detector, is proportional to the absolute square of the amplitude. The intensity can be expressed as a differential scattering cross-section ($d\sigma/d\Omega$), defined in terms of experimental quantities as,

$$\frac{d\sigma}{d\Omega} \equiv \frac{J\left(\hat{\mathbf{s}}'\right)}{J_0\Delta\Omega}$$

where J_0 is incident X-ray flux (photons per second per unit area), $\Delta\Omega$ is the solid angle of the detector pixel in the direction $\hat{\mathbf{s}}'$, and $J(\hat{\mathbf{s}}')$ is the number of scattered photons per

second arriving at the detector pixel. Thus defined, the scattered intensity for a system of free electrons is,

$$\frac{d\sigma}{d\Omega}=\left(\frac{d\sigma}{d\Omega}\right)_e |F(\mathbf{q})|^2. \quad (4)$$

where $(d\sigma/d\Omega)_e$ is the differential cross-section of a single electron. Since $(d\sigma/d\Omega)_e$ can be calculated theoretically (see Section A.1), $|F(\mathbf{q})|^2$ can be obtained directly from a scattering measurement.

As electrons in a molecule are bound to atoms, it useful to define the molecular form factor,

$$F(\mathbf{q})=\sum_n f_n(\mathbf{q})\, e^{i\mathbf{q}\cdot\mathbf{r}_n}, \quad (5)$$

where $\mathbf{r}_n$ are the positions of the atoms and $f_n(\mathbf{q})$ are the atomic scattering factors. The atomic scattering factors depend on the shape of the electron density around the atom and the X-ray photon energy $\hbar\omega$. The energy dependence comes from resonant scattering when the photon excites an electronic transition. At X-ray energies used for structure determination, resonant effects may be significant for heavy elements such as the first-row transition metals, but they are insignificant for the light elements which make up the major part of proteins, including carbon, oxygen, and nitrogen. Resonant scattering has important applications such as phase determination in crystallography and contrast variation in SAXS. For the purposes of this review, however, resonant scattering can be neglected (for further discussion of resonant scattering, see Section A.2). Then, the atomic scattering factor is equal to the energy-independent part (Equation 159),

$$f_n(\mathbf{q})=\int \rho_n(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} d^3\mathbf{r}, \quad (6)$$

where $\rho_n(\mathbf{r})$ is the electron density around the atom. In biological SAXS and MX, we also neglect the angular dependence of the atomic electron density, and retain only the radial component, approximating $f(\mathbf{q})$ as $f(q)$. Theoretical values of $f(q)$ have been tabulated for each atom.[22]

As a consequence of Equation 6, the molecular form factor (Equation 5) can be written in terms of the total electron density of the system, $\rho(\mathbf{r}) = \Sigma_n \rho_n(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}_n)$, as,

$$F(\mathbf{q})=\int \rho(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} d^3\mathbf{r}. \quad (7)$$

In other words, the form factor is the Fourier transform of the electron density. Thus, through measurement of the interference pattern of scattered waves, we can learn about the structure ($\rho(\mathbf{r})$).

Autocorrelation of the Electron Density: Although the scattered intensity allows for a measurement of $|F(\mathbf{q})|$, one cannot invert Equation 7 directly to obtain the electron density because the phase of the scattered wave is not recorded. What is actually measured is equivalent to the electron density's autocorrelation function. To derive this relationship, we take the absolute square Equation 7 and apply a change of variables that emphasizes the relative distances between scatterers (first, $\mathbf{u} = \mathbf{r}' - \mathbf{r}$, then swap the labels $\mathbf{r}$ and $\mathbf{u}$),

$$\begin{aligned} |F(\mathbf{q})|^2 &= \int\int \rho(\mathbf{r})\rho(\mathbf{r}')e^{i\mathbf{q}\cdot(\mathbf{r}'-\mathbf{r})}d^3\mathbf{r}d^3\mathbf{r}' \\ &= \int \left(\int \rho(\mathbf{u})\rho(\mathbf{r}+\mathbf{u})\, d^3\mathbf{u}\right)\, e^{i\mathbf{q}\cdot\mathbf{r}}d^3\mathbf{r}. \end{aligned} \quad (8)$$

The integral in parentheses is the autocorrelation of the electron density, $(\rho \star \rho)(\mathbf{r})$, also known as the Patterson function,

$$(\rho\star\rho)(\mathbf{r}) = \int \rho(\mathbf{u})\rho(\mathbf{r}+\mathbf{u})\, d^3\mathbf{u}. \quad (9)$$

Plugging this back into Equation 8, we see that the squared form factor is the Fourier transform of the autocorrelation function

$$|F(\mathbf{q})|^2 = \int (\rho\star\rho)(\mathbf{r})\, e^{i\mathbf{q}\cdot\mathbf{r}}d^3\mathbf{r}, \quad (10)$$

and this relationship can be inverted by taking the inverse Fourier transform,

$$(\rho\star\rho)(\mathbf{r}) = \frac{1}{(2\pi)^3}\int |F(\mathbf{q})|^2 e^{-i\mathbf{q}\cdot\mathbf{r}}d^3\mathbf{q}. \quad (11)$$

Thus, measuring $|F(\mathbf{q})|^2$ provides direct access to the electron density autocorrelation.

2.1.2 Form Factors of Extended Objects—In scattering experiments, one is interested in the microscopic structure of the sample, and not the macroscopic shape of the crystal or solution. Since a macromolecule is typically many orders of magnitude smaller than the sample itself, ignoring the sample shape is equivalent to ignoring the extreme low-angle scattering, which in any case coincides with the un-scattered beam and is therefore difficult to measure. Furthermore, we choose crystals and solutions to be large so that most of the signal comes from molecules in the bulk of the sample, rather than the surface. As a consequence, one can assume mathematically that any particular part of the sample is

equivalent, on average, to any other part. In other words, the sample is assumed to be an isotropic material.

In the previous section, the arrangement of matter was assumed to be static. Here, we introduce the (grand canonical) ensemble average $\langle\cdots\rangle$, which can be viewed as a time average of a dynamically fluctuating system at finite temperature, or, given certain assumptions, as an average over the position in an isotropic, disordered material. We consider a general system where the number of electrons, $N_e \equiv \int_V \rho(\mathbf{r}) d^3\mathbf{r}$, and their positions within the illuminated volume, V, may fluctuate. The ensemble average of the squared form factor (Equation 3) is

$$\begin{aligned}\langle |F(\mathbf{q})|^2 \rangle &= \left\langle \sum_{n=1}^{N_e} \sum_{m=1}^{N_e} e^{i\mathbf{q}\cdot(\mathbf{r}_n - \mathbf{r}_m)} \right\rangle \\ &= \langle N_e \rangle + \left\langle \sum_n \sum_{m \neq n} e^{i\mathbf{q}\cdot(\mathbf{r}_n - \mathbf{r}_m)} \right\rangle \\ &= \langle N_e \rangle \left(1 + \int_V \eta(\mathbf{r})\, e^{i\mathbf{q}\cdot\mathbf{r}} d^3\mathbf{r}\right). \end{aligned} \quad (12)$$

In the last step, we introduced a correlation function $\eta(\mathbf{r})$, equal to the average number density of electrons at a distance $\mathbf{r}$ relative to any reference electron. The correlation function is normalized to

$$\int_V \eta(\mathbf{r})\, d^3\mathbf{r} = \frac{\langle N_e^2 \rangle - \langle N_e \rangle}{\langle N_e \rangle} \quad (13)$$

so that $\langle |F(0)|^2 \rangle = \langle N_e^2 \rangle$. For large V, the integral in Equation 12 is sharply peaked at $\mathbf{q}=\mathbf{0}$, approaching a delta function. Letting $V \to \infty$ and adding and subtracting the average electron density from $\eta(\mathbf{r})$,

$$\langle |F(\mathbf{q})|^2 \rangle = \langle N_e \rangle \left(1 + \int_V \left(\eta(\mathbf{r}) - \frac{\langle N_e \rangle}{V}\right) e^{i\mathbf{q}\cdot\mathbf{r}} d^3\mathbf{r}\right) + (2\pi)^3 \frac{\langle N_e \rangle^2}{V} \delta(\mathbf{q}). \quad (14)$$

The delta function is not measurable because it coincides with the un-scattered beam. The remaining terms correspond to the scattered intensity.

One implication of Equation 14 is that scattering does not depend on the average electron density in the sample. To see this more clearly, we introduce the excess electron density in a given volume V, $\Delta\rho(\mathbf{r}) \equiv \rho(\mathbf{r}) - \langle N_e \rangle / V$, and derive the ensemble averaged autocorrelation function of $\Delta\rho$,

$$\begin{aligned}\langle(\Delta\rho\star\Delta\rho)(\mathbf{r})\rangle &= \int_V \langle(\rho(\mathbf{u})-\langle N_e\rangle/V)(\rho(\mathbf{r}+\mathbf{u})-\langle N_e\rangle/V)\rangle\, d^3\mathbf{u} \\ &= \langle(\rho\star\rho)(\mathbf{r})\rangle - \frac{\langle N_e\rangle^2}{V}.\end{aligned} \quad (15)$$

Note that in going from the first line to the second, we assumed of an isotropic material so that

$$\langle N_e\rangle = \int_V \langle\rho(\mathbf{u})\rangle\, d^3\mathbf{u} = \int_V \langle\rho(\mathbf{r}+\mathbf{u})\rangle\, d^3\mathbf{u}. \quad (16)$$

If the measurable intensity is defined by subtracting the delta function term in Equation 14,

$$I(\mathbf{q}) \equiv \langle|F(\mathbf{q})|^2\rangle - (2\pi)^3 \frac{\langle N_e\rangle^2}{V}\delta(\mathbf{q}), \quad (17)$$

then Equations 10, 15 and 17 can be combined to show that

$$I(\mathbf{q}) = \int \langle(\Delta\rho\star\Delta\rho)(\mathbf{r})\rangle\, e^{i\mathbf{q}\cdot\mathbf{r}} d^3\mathbf{r}. \quad (18)$$

Thus, the scattering from an isotropic material depends only on the fluctuation of the electron density about the average value.

2.1.3 Scattering Invariants—Several statistical properties can be derived from $I(\mathbf{q})$. Here, we give expressions for the forward scattering intensity, $I(\mathbf{0})$, and the total scattering intensity, $\int I(\mathbf{q}) d^3\mathbf{q}$.

Forward Scattering Intensity: The forward scattering is found by combining Equations 13, 14, and 17:

$$\begin{aligned}I(\mathbf{0}) &= \langle N_e\rangle \left(1 + \int_V \left(\eta(\mathbf{r}) - \frac{\langle N_e\rangle}{V}\right) d^3\mathbf{r}\right) \\ &= \langle N_e\rangle \left(1 + \int_V \eta(\mathbf{r})\, d^3\mathbf{r} - \langle N_e\rangle\right) \\ &= \langle(N_e - \langle N_e\rangle)^2\rangle.\end{aligned} \quad (19)$$

Thus, $I(\mathbf{0})$ is equal to the variance of the number of electrons present in the illuminated volume. For single component liquids, such as water, the electron number variance is proportional to the molecule number variance, and therefore $I(\mathbf{0})$ is proportional to the isothermal compressibility of the liquid (see Section A.3).

Total Scattering Intensity: The total scattering is another important invariant of an isotropic system. It can be derived starting from Equation 17,

$$\int I(\mathbf{q}) d^3\mathbf{q} = \int \left(\langle |F(\mathbf{q})|^2 \rangle - (2\pi)^3 \frac{\langle N_e \rangle^2}{V} \delta(\mathbf{q}) \right) d^3\mathbf{q} = \int \langle |F(\mathbf{q})|^2 \rangle \, d^3\mathbf{q} - (2\pi)^3 \frac{\langle N_e \rangle^2}{V}. \quad (20)$$

Since $F(\mathbf{q})$ is a Fourier transform of the electron density, $\rho(\mathbf{r})$, Plancherel's theorem implies

$$\int \langle |F(\mathbf{q})|^2 \rangle \, d^3\mathbf{q} = (2\pi)^3 \int_V \langle \rho^2(\mathbf{r}) \rangle \, d^3\mathbf{r}. \quad (21)$$

Combining Equations 20 and 21,

$$\int I(\mathbf{q}) d^3\mathbf{q} = (2\pi)^3 \left(\int_V \langle \rho^2(\mathbf{r}) \rangle \, d^3\mathbf{r} - \frac{\langle N_e \rangle^2}{V} \right) = (2\pi)^3 \int_V \langle (\rho(\mathbf{r}) - \langle \rho \rangle)^2 \rangle \, d^3\mathbf{r}. \quad (22)$$

Thus, the total scattering is proportional to the mean squared electron density fluctuation of a system.

2.2 Scattering from Proteins in Solution

X-ray scattering from macromolecular solutions can reveal the conformational preferences and dynamic equilibria that are important for biological function. The technique is referred to as SAXS when the low-q region of the scattering is analyzed, and WAXS when the high-q region is analyzed. The distinction between low-q and high-q is operationally defined. SAXS provides information about a molecule's low-resolution shape (envelope), oligomerization state, colloidal stability, and molecular weight. WAXS provides information about the internal structure of a molecule. Both SAXS and WAXS report ensemble-averaged quantities and can be used to study equilibrium mixtures and time-resolved dynamics.

The best-known application of solution scattering is particle shape analysis. This application requires the solution to be dilute and monodisperse. *Monodisperse*, in this context, means that all of the macromolecular particles are identical, at least when viewed at low resolution. *Dilute* has a special meaning in SAXS; it refers to the fact that, at sufficiently low particle concentrations, non-specific colloidal interactions between particles are minimized, causing the SAXS data to resemble the orientationally-averaged intensity of a single particle. Scattering from dilute and monodisperse solutions can be used in two powerful ways. First, the low-resolution shape may be *reconstructed* from the scattering data. Second, the scattering data can be used to assemble or refine high-resolution, atomically-detailed models derived from other techniques, such as MX, cryo-electron microscopy (cryo-EM), nuclear magnetic resonance (NMR), homology modeling, or molecular dynamics (MD). The term SWAXS is sometimes used when high-resolution models are compared with SAXS and WAXS data together.

In Section 2.2.1, we outline the theory of solution scattering applicable to macromolecular ensembles. We begin with the effects of macromolecular solvation and contrast and derive the approximate form of the total scattering, including the interparticle interference. Section 2.2.2 is devoted to the theory of solution scattering in the dilute limit. We derive the relationship between the scattered intensity and particle shape for a monodisperse solution and extend the theory to include multi-component mixtures. Finally, in Section 2.2.3, we provide an overview of methods for collecting and analyzing solution scattering data.

2.2.1 Macromolecular Solutions

Orientational Average: Molecules in solution have a random orientation. First, we derive the squared form factor for a single molecule that samples random displacements **T** and orientations $\boldsymbol{\omega}$, disregarding the contribution of the aqueous solvent for the time being. The instantaneous form factor of the molecule is

$$F(\mathbf{q},\boldsymbol{T},\boldsymbol{\omega})=\sum_{n=1}^{N} f_n(q)\, e^{i\,\mathbf{q}\cdot(\mathscr{R}_{\boldsymbol{\omega}}\mathbf{r}_n+\boldsymbol{T})} \tag{23}$$

where $\mathscr{R}_{\boldsymbol{\omega}}$ is a 3×3 rotation matrix parameterized by the vector $\boldsymbol{\omega}$ (the three Euler angles, for example). This can be rearranged,

$$\begin{aligned} F(\mathbf{q},\boldsymbol{T},\boldsymbol{\omega}) &= e^{i\,\mathbf{q}\cdot\boldsymbol{T}} \sum_{n=1}^{N} f_n(q)\, e^{i(\mathscr{R}_{\boldsymbol{\omega}}^{T}\mathbf{q})\cdot\mathbf{r}_n} \\ &= e^{i\,\mathbf{q}\cdot\boldsymbol{T}} F(\mathscr{R}_{\boldsymbol{\omega}}^{T}\mathbf{q}) \end{aligned} \tag{24}$$

Note that the effect of rotating the particle by $\mathscr{R}_{\boldsymbol{\omega}}$ is equivalent to rotating **q** in the opposite sense (by $\mathscr{R}_{\boldsymbol{\omega}}^{T}$), and the effect of translation is to add an overall phase. Evaluating the squared form factor causes the phase to drop out so that the configurational, or ensemble, average is equivalent the average over the direction of **q**

$$\begin{aligned} \langle |F(\mathbf{q},\boldsymbol{T},\boldsymbol{\omega})|^2 \rangle_{\boldsymbol{T},\boldsymbol{\omega}} &= \left\langle |F(\mathscr{R}_{\boldsymbol{\omega}}^{T}\mathbf{q})|^2 \right\rangle_{\boldsymbol{\omega}} \\ &= \tfrac{1}{4\pi}\textstyle\int d\Omega_{\hat{\mathbf{q}}} |F(\mathbf{q})|^2 \\ &= \tfrac{1}{4\pi}\textstyle\int d\Omega_{\hat{\mathbf{q}}} \sum_{n=1}^{N}\sum_{m=1}^{N} f_n(q) f_m(q)\, e^{i\,\mathbf{q}\cdot(\mathbf{r}_n-\mathbf{r}_m)} \end{aligned} \tag{25}$$

where $d\Omega_{\hat{\mathbf{q}}}$ is the differential solid angle in the $\hat{\mathbf{q}}$ direction. The angular integral can be performed explicitly using the identity,

$$\frac{1}{4\pi}\int d\Omega_{\hat{\mathbf{q}}}\, e^{i\,\mathbf{q}\cdot\mathbf{r}} = \frac{\sin(q\,r)}{q\,r}, \tag{26}$$

which leaves,

$$\langle |F(\mathbf{q},\boldsymbol{T},\boldsymbol{\omega})|^2 \rangle_{\boldsymbol{T},\boldsymbol{\omega}} = \sum_{n=1}^{N}\sum_{m=1}^{N} f_n(q) f_m(q) \frac{\sin(q\, r_{nm})}{q\, r_{nm}} \quad (27)$$

where $r_{nm} = |\mathbf{r}_n - \mathbf{r}_m|$ is the distance between a pair of atoms indexed by n and m. Equation 27 is known as the Debye formula.

Solvation: The Debye formula (Equation 27) applies to a molecule in vacuum. For molecules in solution, it is important to consider the contribution of the aqueous solvent. The total form factor of a solution can be split into two terms, $F_{\text{mol.}}(\mathbf{q})$ and $F_{\text{solv.}}(\mathbf{q})$, for the atoms belonging to the macromolecules and solvent, respectively,

$$F_{\text{soln.}}\ (\mathbf{q}) = F_{\text{mol.}}\ (\mathbf{q}) + F_{\text{solv.}}\ (\mathbf{q}). \quad (28)$$

The term $F_{\text{solv.}}(\mathbf{q})$ differs from the form factor of bulk solvent, $F_{\text{bulk}}(\mathbf{q})$, because the macromolecule both excludes solvent molecules, and perturbs the solvent structure near its surface.

These effects are captured by defining a hydration term

$$\Delta F_{\text{hydr.}}\ (\mathbf{q}) \equiv F_{\text{solv.}}\ (\mathbf{q}) - F_{\text{bulk}}(\mathbf{q}). \quad (29)$$

Then, the ensemble-averaged, squared structure factor of the solution is

$$\begin{aligned} \langle |F_{\text{soln.}}\ (\mathbf{q})|^2 \rangle &= \langle |F_{\text{mol}}(\mathbf{q}) + \Delta F_{\text{hydr.}}\ (\mathbf{q}) + F_{\text{bulk}}(\mathbf{q})|^2 \rangle \\ &= \langle |F_{\text{mol}}(\mathbf{q}) + \Delta F_{\text{hydr.}}\ (\mathbf{q})|^2 \rangle + \langle |F_{\text{bulk}}(\mathbf{q})|^2 \rangle \end{aligned} \quad (30)$$

The scattering contribution of the macromolecules is

$$\tilde{I}(q) = \langle |F_{\text{soln.}}\ (\mathbf{q})|^2 \rangle - \langle |F_{\text{bulk}}(\mathbf{q})|^2 \rangle \quad (31)$$

$$= \langle |F_{\text{mol.}}\ (\mathbf{q}) + \Delta F_{\text{hydr.}}\ (\mathbf{q})|^2 \rangle \quad (32)$$

$\tilde{I}(q)$ is determined experimentally by taking the difference between measurements of two samples: a solution of the macromolecules in aqueous solvent, and the aqueous solvent without macromolecules.

Contrast: Adding solvent causes the forward scattering to deviate from its expected value in vacuum (Equation 27). In dilute solution, the forward scattering is

$$\tilde{I}(0)=N\,\mathrm{MM}^2\,(\Delta\rho_M)^2 \quad (33)$$

where N is the number of particles, MM is the molecular mass of the particle, and $\Delta\rho_M$ is the contrast,

$$\Delta\rho_M=\rho_{M,\mathrm{mol.}}-\rho_{\mathrm{solv.}}\,\bar{v}. \quad (34)$$

where $\rho_{M,\mathrm{mol.}}$ is number of electrons per dry mass of the molecule, $\rho_{\mathrm{solv.}}$ is the electron density of the solvent, and $\bar{v}$ is the partial specific volume of the molecule. In vacuum, the contrast is simply $\rho_{M,\mathrm{mol.}}$. The reduction in scattering can be seen by computing the ratio of the scattering in solution to the scattering in vacuum,

$$\frac{\tilde{I}(0)}{|F(0)|^2}=\left(1-\frac{\rho_{\mathrm{solv.}}\,\bar{v}}{\rho_{M,\mathrm{mol.}}}\right)^2. \quad (35)$$

For a typical protein solution,[23] the contrast is quite low, and the scattering in solution is only about 5% of the scattering in vacuum (Equation 35 with $\rho_{M,\mathrm{prot.}} = 3.22\times10^{23}\ e^-\ \mathrm{g}^{-1}$, $\rho_{\mathrm{solv.}} = 3.34\times10^{23}\ e^-\ \mathrm{cm}^{-3}$, and $\bar{v}\approx 0.7425\ \mathrm{cm}^3\ \mathrm{g}^{-1}$). For nucleic acids, which have significantly higher contrast, the scattering ratio is around 15%.

Inter-particle Interference: The degree to which macromolecules scatter as independent particles depends on their concentration. At high particle densities, non-specific colloidal interactions produce inter-particle interference in the scattering pattern. As a short-hand, we introduce the excess structure factor of particular molecule n,

$$\tilde{F}_n(\mathbf{q})=F_{n,\mathrm{mol.}}(\mathbf{q})+\Delta F_{n,\mathrm{hydr.}}(\mathbf{q}), \quad (36)$$

and its corresponding excess intensity,

$$\tilde{I}_n(\mathbf{q})=\left|\tilde{F}_n(\mathbf{q})\right|^2. \quad (37)$$

Then, the macromolecular component of the total scattering (Equation 31) can be written

$$\tilde{I}(q)=\left\langle \sum_n \tilde{I}_n(\mathbf{q}) \right\rangle + \left\langle \left| \sum_n \sum_{m\neq n} \tilde{F}_{\mathrm{n}}(\mathbf{q}) \tilde{F}^*_{\mathrm{m}}(\mathbf{q}) \right|^2 \right\rangle . \quad (38)$$

Here, the ensemble average $\langle \cdots \rangle$ has contributions from translational and rotational motion $\langle \cdots \rangle_{\mathbf{T},\Omega}$, as well configurations of the particle (macromolecule and surrounding solvent) $\langle \cdots \rangle_p$. Next, we invoke the de-coupling approximation, which assumes that the rotational, translational, and internal degrees of freedom are independent. Then,

$$\tilde{I}(q) \approx \langle N \rangle \left\langle \tilde{I}_1(\mathbf{q}) \right\rangle_{\Omega,p} + \left| \left\langle \tilde{F}_1(\mathbf{q}) \right\rangle_{\Omega,p} \right|^2 \left\langle \sum_n \sum_{n\neq m} e^{i\mathbf{q}\cdot(\boldsymbol{T}_n - \boldsymbol{T}_m)} \right\rangle_T . \quad (39)$$

The average of the translational phase factors can be expressed as using a two-particle correlation function, $g(r)$ (see Section A.3), and the intensity can be written as a product of two terms, as follows,

$$\tilde{I}(q)=\langle N \rangle \left\langle \tilde{I}_1(\mathbf{q}) \right\rangle_{\Omega,p} S(q). \quad (40)$$

The first factor is the scattering from N non-interacting particles. The second factor, $S(q)$, or the structure factor, captures the inter-particle interactions,

$$S(q)=1-n_p H(q) \int_0^\infty 4\pi r^2 (1-g(r)) \frac{\sin(qr)}{qr} dr, \quad (41)$$

where $n_p = \langle N \rangle / V$ is the particle density and $H(q)$ is a shape factor,

$$H(q) \equiv \left| \left\langle \tilde{F}_1(\mathbf{q}) \right\rangle_{\Omega,p} \right|^2 / \left\langle \tilde{I}_1(\mathbf{q}) \right\rangle_{\Omega,p} . \quad (42)$$

At $q = 0$, the structure factor is proportional to the osmotic compressibility,[24] equal to the inverse derivative of the osmotic pressure, Π, with respect to the particle density, n_p

$$S(0)=k_B T \left(\frac{\partial \Pi}{\partial n_p} \right)^{-1} \quad (43)$$

At low concentration, the osmotic pressure can be expanded as a power series

$$\frac{\Pi}{k_B T} = n_p + B_2 n_p^2 + B_3 n_p^3 + \dots \quad (44)$$

where B_n are the virial coefficients. Combining Equations 43 and 44,

$$\frac{1}{S(0)} = 1 + 2B_2 n_p + \dots \quad (45)$$

The second virial coefficient B_2 depends on the potential of mean force for pairs of interacting particles, $V(r)$

$$B_2 = 2\pi \int_0^\infty \left(1 - e^{-V(r)/k_B T}\right) r^2 dr \quad (46)$$

In semi-dilute solutions, where only the second term in the virial expansion is significant, the structure factor at $q = 0$ can be understood as a measure of pair-wise interparticle interactions.[14] A net repulsive interaction between particles produces a positive B_2 and $S(0) < 1$, leading to a "downturn" in the scattering profile at low-q. Conversely, attractive interactions have negative B_2 and $S(0) > 1$, giving the appearance of an "upturn" at low-q.

2.2.2 Shape Information—In a dilute solution, the structure factor (Equation 41) approaches unity, making the total intensity (Equation 40) proportional to the intensity of a single particle, $\tilde{I}_1(q)$. Thus, measurement of the scattering from a dilute solution can be used to obtain shape information. In this section, we derive the relationships between particle shape and $\tilde{I}_1(q)$.

Pair Distance Distribution: The pair distance distribution function of a particle, $P(r)$, is equal to the probability of observing a pair of electrons at a distance r. The relationship between scattering and $P(r)$ can be derived by taking the continuum limit of the Debye formula (Equation 27),

$$I(q) = 4\pi \int_0^{d_{\max}} P(r) \frac{\sin(q\,r)}{q\,r} dr \quad (47)$$

The integral terminates at the maximum dimension of the particle, $d_{\max}$, because $P(r) = 0$ for $r > d_{\max}$. The integral is a Fourier-Bessel transform whose inverse is

$$P(r)=\frac{r^2}{2\pi^2}\int_0^\infty I(q)\frac{\sin(q\,r)}{q\,r}q^2dq. \quad (48)$$

In most cases, the particle shape may be grasped more intuitively by transforming the scattering data into real space. For example, the value of r where $P(r)$ goes to zero can be used to estimate the maximum particle dimension, $d_{\max}$. In addition, $P(r)$ is the distribution of pairwise distances, it can be used to calculate the radius of gyration

$$R_g^2=\frac{\int r^2P(r)\,dr}{2\int P(r)\,dr}. \quad (49)$$

Information Content: The amount of unique information contained in a SAXS measurement is limited. To see why this is true, note that $P(r)/r = 0$ at $r = 0$ and at $r = d_{\max}$. Therefore, in the interval $[0, d_{\max}]$ it can be represented by a Fourier sine series

$$P(r)/r=\sum_{n=1}b_n\sin(n\pi r/d_{\max}) \quad (50)$$

where b_n are the Fourier coefficients,

$$b_n=\frac{2}{d_{\max}}\int_0^{d_{\max}}P(r)\frac{\sin\,(n\pi r/d_{\max})}{r}dr. \quad (51)$$

Since the sine functions in Equation 50 are an orthogonal basis, each coefficient b_n represents a unique piece of information about $P(r)$. Note that the right-hand side of Equation 51 resembles the Fourier-Bessel transform of $P(r)$ in Equation 47 evaluated at

$$q_n=\frac{n\pi}{d_{\max}}. \quad (52)$$

Thus, the n^{th} coefficient in the Fourier series (Equation 51) is related to the intensity measured at q_n,

$$b_n=\frac{1}{2\pi d_{\max}}I(q_n)q_n. \quad (53)$$

In other words, the *unique information* in $I(q)$ is contained in a discrete set of samples with a spacing of $\pi/d_{\max}$, which is essentially Shannon's criterion from information theory.[25,26]

Shannon's criterion has several implications for SAXS. In a SAXS measurement, the lowest scattering angles coincide with the unscattered beam and are experimentally inaccessible. If the minimum q-value which can be measured using a given experimental setup is $q_{\min}$, Shannon's criterion says that the maximum dimension which can be measured is $d_{\max} = \pi/q_{\min}$. Similarly, the SAXS measurement does not extend to $q = \infty$, but instead terminates at a maximum scattering angle, with a corresponding value $q_{\max}$. The number of samples contained in the measurement interval, ($q_{\min}$, $q_{\max}$), is a rough estimate of the information content of the measurement,

$$n_S = (q_{\max} - q_{\min})\, d_{\max}/\pi \quad (54)$$

where n_S is sometimes called the number of *Shannon channels*. Finally, Shannon's criterion implies that one does not need to measure $I(q)$ more finely than $\pi/d_{\max}$. However, in practice, $d_{\max}$ is not known before the experiment is performed. Therefore, it is usually necessary to *oversample* $I(q)$; the extra information gained by oversampling corresponds to length scales $r > d_{\max}$, so that one can determine $d_{\max}$ by observing where $P(r)$ approaches zero (see Section 2.2.3).

Guinier's Law: The shape of $\tilde{I}(q)$ at low angles can be found by replacing $\sin(q\,r)/(q\,r)$, in the right-hand side of Equation 47, by its Taylor series expansion

$$\begin{aligned}\tilde{I}(q) &= 4\pi \sum_{k=0}^{\infty} \frac{(-1)^k}{(1+2k)!} q^{2k} \int r^{2k} P(r)\, dr \\ &= 4\pi \left(\int P(r)\, dr - \frac{1}{6} q^2 \int r^2 P(r)\, dr + \mathcal{O}(q^4) \right)\end{aligned} \quad (55)$$

Substituting in 49,

$$\tilde{I}(q) \approx \tilde{I}(0) \left(1 - \frac{1}{3} q^2 R_g^2 + \mathcal{O}(q^4) \right). \quad (56)$$

The expression in parentheses is identical to the Taylor series expansion of a Gaussian (to third order in q),

$$\tilde{I}(q) \approx \tilde{I}(0) \exp\left(-q^2 R_g^2/3 \right) \quad (57)$$

Equation 57 is known as Guinier's law.

Homogeneous Approximation and Porod's Law: Small-angle scattering can be defined as the part of $I(q)$ which depends on the molecule's low-resolution shape. The concept of shape is made more concrete by defining a smooth envelope where the excess electron density is uniform inside, and zero outside.

$$\rho(\mathbf{r})-\langle\rho\rangle=\begin{cases}\rho_p-\rho_s & \text{inside}\\ 0 & \text{outside}\end{cases} \quad (58)$$

This model is known as the homogeneous particle approximation, and is extremely important in the theory of SAXS. A number of shape parameters may be derived from this simplified model.

The equivalent of the total scattering in SAXS is Porod's invariant,

$$Q \equiv \int_0^\infty I(q)q^2dq. \quad (59)$$

Since $I(\mathbf{q})$ of a solution depends on the magnitude of $\mathbf{q}$, but not its direction, Q is equivalent to the total scattering (Equation 22) written in spherical coordinates,

$$\begin{aligned}\int_0^\infty I(q)q^2dq&=\tfrac{1}{4\pi}\int I(\mathbf{q})d^3\mathbf{q}\\ &=2\pi^2\int_V\langle(\rho(\mathbf{r})-\langle\rho\rangle)^2\rangle\,d^3\mathbf{r}\end{aligned} \quad (60)$$

In the homogeneous approximation, Q may be calculated exactly by substituting Equation 58 for the electron density in Equation 60,

$$Q=2\pi^2\,\langle N\rangle\,V_p\,(\rho_p-\rho_s)^2 \quad (61)$$

where V_p is the volume enclosed by the envelope (sometimes called the Porod volume). Similarly, the forward scattering, $I(0)$, can be derived in the homogeneous approximation,

$$I(0)=\langle N\rangle\,V_p^2\,(\rho_p-\rho_s)^2. \quad (62)$$

Since $Q\propto V_p$ and $I(0)\propto V_p^2$, the ratio between Q and $I(0)$ can be used to find V_p,

$$V_p = 2\pi^2 \frac{I(0)}{Q} \quad (63)$$

The envelope's surface area can also be obtained from $I(q)$, although the derivation is somewhat more involved. In the homogeneous approximation, the pair distance distribution $P(r)$ is related to the correlation function $\gamma(r)$ by $P(r) = \gamma(r)\, r^2$. Assuming that the envelope defines a compact, smooth particle, it can be shown that $\gamma(r)$ has continuous first and second derivatives in the interval $(0, d_{\max})$.[17] The intensity is found by substituting $\gamma(r)\, r^2$ for $P(r)$ in Equation 47 and integrating by parts three times:

$$\begin{aligned} \tfrac{1}{4\pi} I(q) = & -2\gamma'(0)\, q^{-4} + D\gamma'(d_{\max}) \sin(q\, d_{\max})\, q^{-3} \\ & + (2\gamma'(d_{\max}) + D\gamma''(d_{\max})) \cos(q\, d_{\max})\, q^{-4} \\ & - q^{-4} \int_0^{d_{\max}} \cos(qr) \left[\gamma(r)\, r\right]^{(3)} dr. \end{aligned} \quad (64)$$

As $q \to \infty$, all of the terms in $I(q)$ decay rapidly, and the leading non-oscillating trend is given by the first term, $-2\gamma'(0)\, q^{-4}$. Furthermore, $\gamma'(0)$ can be shown to be depend on the particle's surface area, S_p,[17] as

$$\gamma'(0) = -\frac{\langle N \rangle (\rho_p - \rho_s)^2}{4} S_p. \quad (65)$$

Therefore, the intensity at high-q tends to decay as

$$\lim_{q \to \infty} I(q) \approx \frac{2\pi \langle N \rangle (\rho_p - \rho_s)^2 S_p}{q^4}. \quad (66)$$

Equation 66 is known as Porod's law, and S_p is the Porod surface. Equations 66 and 61 can be combined to give the surface-to-volume ratio

$$\frac{S_p}{V_p} \approx \frac{\pi}{Q} \left(\lim_{q \to \infty} q^4 I(q) \right). \quad (67)$$

If SAXS is defined as the part of $I(q)$ which can be described using the homogeneous approximation, then WAXS is the remainder. The q^{-4} decay implied by Porod's law eventually breaks down in the WAXS region. There, the scattering is dominated by the internal density fluctuations and rough surface features that were neglected in the homogeneous approximation.

Mixtures: The theory for the scattering of single-component solutions can be generalized for more complex mixtures. In the dilute limit, the total scattering for a mixture of M types of particle, each with concentration $\langle N_m \rangle / V$ and excess scattering intensity $\tilde{I}_{1,m}(q)$, is

$$I(q) = \sum_{m=1}^{M} \langle N_m \rangle \, \tilde{I}_{1,m}(q) \tag{68}$$

If a normalized intensity $P_m(q) \equiv \tilde{I}_{1,m}(q)/\tilde{I}_{1,m}(0)$ is used instead, the total intensity can be written in terms of the contrast parameters for each component

$$I(q) = N_{\mathrm{A}} \sum_{m=1}^{M} c_m \mathrm{MM}_m \left(\Delta\rho_{M,m}\right)^2 P_m(q) \tag{69}$$

where N_A is Avogadro's number, c_m is the concentration (mass per unit volume), MM_m is the molecular mass, and $\Delta\rho_{M,m}$ is the contrast (Equation 34).

Commonly, all of the components in a given mixture are proteins with similar contrasts, $\Delta\rho_{M,m} \approx \Delta\rho_M$. If the total protein concentration is c, and the mass fraction of each component is defined $x_m \equiv c_m/c$, the intensity can be written

$$I(q) \approx N_{\mathrm{A}} \left(\Delta\rho_M\right)^2 c \sum_{m=1}^{M} x_m \mathrm{MM}_m P_m(q). \tag{70}$$

In the sum, each term is weighted by the molecular mass. There are two important consequences. First, components in a mixture with higher molecular mass will tend to dominate the scattering. This makes SAXS rather sensitive to the presence of aggretated material. Second, the scattering at low angles reports changes in molecular mass due to protein-protein interactions.

2.2.3 Interpretation of Data—There have been several articles published recently that detail current best practices for SAXS, including how to prepare samples, evaluate data quality, and utilize modern analysis software.[8,10,12] Those who wish to perform a SAXS experiment should consult these excellent sources. Our aim in this section is give enough background to make sense of the methods section of a SAXS paper.

Data Collection: Modern biological SAXS setups are highly adaptable, providing specialized X-ray beams, detectors, and sample delivery methods to suit the needs of the experiment. Looking beyond these differences, however, the basic geometry of the scattering experiment is nearly universal. A monochromatic X-ray beam is collimated and defined using slits to minimize background scattering at low angles. The beam impinges on a liquid sample held in a cell with X-ray transparent windows. Most of the beam passes through the

sample or is absorbed, and a small fraction of the X-rays are scattered. One or more position-sensitive detectors is placed at an appropriate distance from the sample to capture the q-range of interest. To protect the detector, the transmitted beam is usually blocked by a beamstop, which may also contain a sensing element to measure the beam intensity. In order to measure the lowest scattering angles, the beamstop placed as close as possible to the detector, and the intervening space between the sample and beamstop is either evacuated or filled with helium to reduce the chance that X-rays will be scattered by air. The two-dimensional scattering pattern recorded on the detector is integrated to obtain a one-dimensional scattering curve (the intensities recorded in rings of constant $|\mathbf{q}|$ are averaged together).

The SAXS profile $\tilde{I}(q)$ (Equation 31) is obtained experimentally by measuring the scattering from two solutions: the *sample* containing the macromolecules, and its *matching buffer* (a dialysis buffer or running buffer from the final step of purification). Both measurements are usually performed in the same sample cell, so that instrumental artifacts will cancel when the two measurements are subtracted. Several measurements with different macromolecular concentrations are usually performed in order to verify that the dilute limit has been reached, or to analyze the inter-particle interference.

The scattered intensity is usually measured on an arbitrary scale, and for many applications it is not necessary to measure the absolute intensity. The intensity may be placed on an absolute scale with the aid of a standard. Liquid water is often used for this purpose, because its X-ray absorption is similar to that of the sample, it is readily available, and its forward scattering on an absolute scale can be calculated accurately (Equation 173).[27] Alternately, a macromolecule standard may be used.[23]

Diagnostic Plots and Scattering Moments: The first step in analyzing SAXS data is to visualize $\tilde{I}(q)$. Specialized plots are commonly used, including the *Guinier plot* and the *Kratky plot*. In a Guinier plot, the natural logarithm of the intensity is plotted as a function of q^2, converting the Gaussian form of Guinier's law (Equation 57) to a straight line,

$$\ln\tilde{I}(q)=\ln\tilde{I}(0)-q^2R_g^2/3 \quad (71)$$

This plot is a useful diagnostic because nonlinearity at low-q can signal interparticle interference, sample purity issues, or failure of background subtraction. When a straight line fit is possible, Guinier's law allows the scattering profile to be extrapolated to $q = 0$ to derive $\tilde{I}(0)$ (this part of the profile is blocked by the beamstop). In addition, the radius of gyration can be estimated from the slope of the line.

A Kratky plot is used to check whether the SAXS profile resembles that of a compact, homogeneous particle. Porod's law (Equation 66) states that $I(q)$ of a compact homogeneous particle with a smooth surface should decay as q^{-4} for large q. The axes of a Kratky plot are $q^2 \times I(q)$ vs. q, which converts the scattering profile to a bell-shaped curve if Porod's law is obeyed. The Kratky plot also emphasizes mid-q features, and is often used in publications

for displaying subtle changes in the scattering (caused by conformational changes, for example). Since the area under a Kratky plot is Porod's invariant (Equation 59), a bell-shaped curve means that this quantity can be determined from the data. When the Kratky plot does not decay to zero, this can be evidence that the molecule is partly unfolded or otherwise not compact. A Kratky plot that does not decay to zero may also signal poor buffer matching.[12]

The molecular mass can be estimated by one of several methods. If the intensity is calibrated on an absolute scale, $I(0)$ can be converted directly into molecular mass provided the molecule's concentration and contrast are known.[23,27] An alternate method, which does not rely on absolute calibration or knowledge of the protein concentration, is to find the Porod volume V_p using the ratio of $I(0)$ and Porod's invariant (Equation 63), which can be combined with an estimate of the density to determine molecular mass. An approximate rule-of-thumb for proteins is MM ~ $V_p/(1.7\text{Å}^3\ \text{Da}^{-1})$.[28] An heuristic alternative to V_p, based on the integral of $q \times I(q)$ instead of $q^2 \times I(q)$, has also been used to estimate molecular mass.[29] It should be noted that none of these methods is consistently accurate; the expected error in molecular mass from SAXS is ~ 10%.[23,29]

Statistical Inference and the Pair Distance Distribution: The first step in determining a particle's shape from $I(q)$ is to calculate the pair distance distribution function $P(r)$. $P(r)$ is the inverse Fourier-Bessel transform of $I(q)$ (Equation 48). In this transform, $I(q)$ is integrated from $q = 0$ to infinity, however $I(q)$ is measured only over a limited q–range. Instead of calculating $P(r)$ using a Fourier-Bessel transform of the data, an *indirect Fourier transform* (IFT) is performed.[30] In the IFT, $P(r)$ is expressed using a set of parameters, which are fit to the data by performing the transform in the *forward* direction (Equation 47). The IFT is underdetermined, so regularization methods must be applied. The classic regularization method for SAXS is to minimize the second derivative of $P(r)$, essentially requiring it to be as smooth as the data allow.[30] The popular IFT program GNOM adds additional criteria to determine the regularization parameter robustly.[31] More recently, Bayesian regularization has been applied to the IFT.[32] Bayesian methods in general give a rigorous way to infer unknown quantities based on limited information. Using a Bayesian IFT, one can infer $d_{\max}$ and other parameters, as well as their experimental uncertainties.[33]

Multi-scale Modeling: Solution scattering is often used to test the validity of structures obtained by other techniques. This application requires methods to calculate the scattering curve given an atomic model. The main difficulty lies in accounting for hydration effects, and the solutions depend on the level of detail required. A number of software programs have been developed for this purpose.[2,34–43] Here, we outline the general strategies for hydration modeling, and explain in greater detail the two programs which are widely used for SAXS profile calculations: CRYSOL[2] and FoXS.[41]

Before describing these popular programs, however, we first emphasize that the most rigorous treatment of hydration effects is to use explicit-solvent molecular dynamics (MD).[34,35] In this approach, two simulations are performed: one of the macromolecule in aqueous solvent, and one of the aqueous solvent alone,[35] and the SAXS experiment is simulated using Equation 31. Of course, MD is computationally expensive, and so other

methods have been developed with the goal of achieving similar accuracy without the computational cost.[36–40] For example, an integral equation theory with exact pair potentials known as 3D-RISM has been shown to provide a similar accuracy to MD with no free parameters.[40] The use of an accurate hydration model is most critical for the WAXS regime.

For SAXS data, simpler hydration models can provide reasonable accuracy.[2,41–43] CRYSOL was one of the first programs to account for the scattering of the hydration layer around the protein. CRYSOL models this layer as a shell of constant density and thickness that follows the outer surface of the molecule. In addition, the excluded solvent associated with each atom is modeled using a spherical Gaussian[44,45]

$$g_j(q) = V_j \exp\left(-\frac{1}{4\pi} q^2 V_j^{2/3}\right) \quad (72)$$

where V_j is the excluded volume of each atom. The intensity is calculated through Equation 32 with the aid of a spherical harmonic expansion (SHE). Since multipole components are orthogonal, cross-terms cancel in the orientational average, giving a simple expression for the intensity[2]

$$\begin{aligned} I(q) &= \langle |A(q) - \rho_0 C(q) + \Delta\rho B(q)|^2 \rangle_\Omega \\ &= \sum_{l=0}^{L} \sum_{m=-l}^{l} |A_{lm}(q) - \rho_0 C_{lm}(q) + \Delta\rho B_{lm}(q)|^2 \end{aligned} \quad (73)$$

where ρ_0 is the solvent electron density; $\Delta\rho$ is the hydration layer contrast; $A(q)$, $B(q)$, and $C(q)$ are the form factors of the atoms, hydration layer volume, and excluded volume, respectively; and $A_{lm}(q)$, $B_{lm}(q)$, and $C_{lm}(q)$ are their spherical harmonic components. The SHE in Equation 73 is truncated at a value L, which allows for a compromise between resolution and computational efficiency.

FoXS is another popular program designed for rapid SAXS profile calculations. Like CRYSOL, FoXS uses an approximate model for the excluded volume (Equation 72) and hydration layer.[41] Where FoXS differs from CRYSOL is in its use of the Debye formula (Equation 27) instead of the SHE, which means that the intensity calculation is exact. The main trade-off is that the computational cost of the Debye formula scales as $\sim \mathcal{O}(N^2)$ (where N is the number of atoms), while the SHE scales as $\sim \mathcal{O}(L^2 N)$. Thus, for very large molecules, SHE is more efficient than the Debye formula.

Both FoXS and CRYSOL have hydration parameters that can be refined by minimizing the χ^2 between the measured and calculated intensities,

$$\chi^2=\frac{1}{N}\sum_{n=1}^{N}\left(\frac{I_{\text{meas.}}(q_n)-c\,I_{\text{calc.}}(q_n)}{\Delta I_{\text{meas.}}(q_n)}\right)^2, \quad (74)$$

where c is a scale factor and $\Delta I_{\text{meas.}}(q)$ is the experimental uncertainty.

Flexible Fitting: It is often the case that atomic resolution structures are available only for individual proteins or domains, and one wishes to determine how they are oriented relative to each other in a complex, or how they move during a conformational change. The program SASREF[46] performs rigid-body optimization using simulated annealing, with optional distance restraints. When flexible linkers are involved, the program CORAL[28] generates these restraints automatically, and includes the linker residues in the SAXS calculation. Molecular dynamics may also be used to sample conformational space.[47]

Analysis of Mixtures: A second application of SAXS is to measure equilibrium mixtures of molecules. The dilute limit is usually required, so that each component adds linearly to the scattering (Equation 68). When the scattering profile of each pure component is known in advance, or can be calculated from atomic models, the relative concentrations of each component can be determined by least-squares fitting.

Mixture analysis is particularly important in methods with varying populations, such as equilibrium titrations and time-resolved data. When many scattering profiles are available, linear algebra methods may be used to determine how components co-vary, without prior knowledge of the structures. A commonly used technique is singular value decomposition (SVD)

$$U\,S\,V^T=\text{svd}(D) \quad (75)$$

where D is the data matrix with scattering profiles arranged side-by-side in columns, S is a diagonal matrix with *singular values* in decreasing order along the diagonal, and U and V are matrices whose columns are the left and right *singular vectors*, respectively. Since errors are unequally distributed as a function of q, the best practice is to pre-weight the columns of D by the average experimental uncertainty, so that each element of D has an expected variance of ~ 1 due to statistical noise.

The SVD can be analyzed to determine the minimum number of components in the data when the number of scattering profiles is significantly larger than the number of components. In this case, most of the singular values will be small, and the corresponding columns of U and V contain only noise, and can be ignored. The number of remaining singular values is equal to the number of co-varying components in the dataset, n. The first n columns of U contain the scattering profiles of those components, and the first n columns of V are their concentrations. It is important to note that n is a lower bound on the number of actual components; a component whose scattering is well-approximated by a linear

combination of the scattering from other components cannot be distinguished by SVD. An additional limitation of SVD is that U is an orthonormal basis, and so it will not correspond to physically meaningful scattering profiles. It is usually necessary to *rotate* the basis, transforming $U \rightarrow U'$ and $V \rightarrow V'$. This rotation can be performed automatically given certain criteria. For example, U' and V' should be positive, since they correspond to concentrations and scattering intensities. More complex constraints can be applied to the shapes of V' depending on the type of experiment performed. For example, the columns of V' could be modeled as a two-state transition or a chromatographic peak.[48] In addition, one can require that the columns of U' satisfy Guinier's law (Equation 57).[49]

Analysis of Ensembles: SAXS can provide structural information about molecules that are intrinsically flexible and non-compact. The scattering from a flexible molecule is effectively that of a mixture (Equation 68) where each component is a member of the equilibrium statistical ensemble. The first step in ensemble modeling is to generate a library of candidate structures, for example using molecular dynamics or Monte-Carlo simulations. Once a library is available, the scattering curve from each model can be predicted. Next, a program must choose subsets of the library that best fit the scattering data, minimizing Equation 74. In one approach, taken by the program EOM,[50,51] a genetic algorithm chooses subsets of the library with a fixed number of members. In contrast, the program MES[47] searches for the smallest ensemble that can fit the data.

Shape Reconstruction: SAXS data are often used to reconstruct the molecule's shape directly. Initially, low-order spherical harmonics were used to define the molecular envelope (Equation 58), and their coefficients were refined.[52] Low-order spherical harmonics, however, are limited in their ability to represent realistic molecular shapes. To solve this problem, the program DAMMIN[53] was developed to optimize a collection of connected, close-packed spherical beads. The bead method is able to uniquely reconstruct many simple shapes.[54] Today, bead models are widely used shape reconstructions from SAXS data. However, since bead models approximate a uniform density, they are only accurate in the SAXS regime. The program GASBOR[55] uses protein-like chains instead of beads, and can therefore be used for *ab-initio* modeling of WAXS data.

2.3 Diffuse Scattering from Protein Crystals

2.3.1 Crystal Diffraction—Crystals are characterized by a repeating structural unit, known as the unit cell. The unit cells are arranged in a 3D lattice with each cell having an origin

$$\mathbf{R_n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \quad (76)$$

where $\mathbf{a}_j$ are the lattice vectors, $\mathbf{n} = [n_1\ n_2\ n_3]$, and n_j are integers. The form factor of a crystal is derived by writing the coordinates of each atom relative to the origin of the unit cell to which it belongs (Equation 5 with $\mathbf{r}_{\mathbf{n},j} \rightarrow \mathbf{R_n} + \mathbf{r}_j$);

$$F_{\text{crystal}}(\mathbf{q}) = \sum_{\mathbf{n}} \sum_{\substack{\text{atom } j \\ \text{in cell } \mathbf{n}}} f_j(q)\, e^{i\mathbf{q}\cdot(\mathbf{R}_n + \mathbf{r}_j)} = \sum_{\mathbf{n}} e^{i\mathbf{q}\cdot\mathbf{R}_n} \sum_{\substack{\text{atom } j \\ \text{in cell } \mathbf{n}}} f_j(q)\, e^{i\mathbf{q}\cdot\mathbf{r}_j} = \sum_{\mathbf{n}} e^{i\mathbf{q}\cdot\mathbf{R}_n} F_{\mathbf{n}}(\mathbf{q}), \quad (77)$$

where $F_{\mathbf{n}}(\mathbf{q})$ is the form factor of unit cell **n**, calculated relative to its origin, $\mathbf{R}_{\mathbf{n}}$. The square of the whole-crystal form factor is

$$|F_{\text{crystal}}(\mathbf{q})|^2 = \sum_{\mathbf{n}} \sum_{\mathbf{m}} F_{\mathbf{n}}(\mathbf{q}) F^*_{\mathbf{m}}(\mathbf{q})\, e^{i\mathbf{q}\cdot(\mathbf{R}_{\mathbf{n}} - \mathbf{R}_{\mathbf{m}})}. \quad (78)$$

In a perfect crystal, all of the unit cells are identical ($F_{\mathbf{n}}(\mathbf{q}) = F(\mathbf{q})$), allowing $|F(\mathbf{q})|^2$ to be factored out. In a real crystal, the unit cells differ from one another. By defining the average form factor,

$$F_{\text{av.}}(\mathbf{q}) = \frac{1}{N} \sum_{\mathbf{m}} F_{\mathbf{m}}(\mathbf{q}), \quad (79)$$

where N is the number of unit cells, Equation 78 can separated into two terms,

$$|F_{\text{crystal}}(\mathbf{q})|^2 = I_{\text{Bragg}}(\mathbf{q}) + I_{\text{D}}(\mathbf{q}). \quad (80)$$

The first term, $I_{\text{Bragg}}(\mathbf{q})$, is the intensity of a perfect crystal, where $F(\mathbf{q})$ is replaced by the average, $F_{\text{av.}}(\mathbf{q})$:

$$I_{\text{Bragg}}(\mathbf{q}) = |F_{\text{av.}}(\mathbf{q})|^2 \left| \sum_{\mathbf{n}} e^{i\mathbf{q}\cdot\mathbf{R}_{\mathbf{n}}} \right|^2. \quad (81)$$

The second term, $I_{\text{D}}(\mathbf{q})$, is the diffuse scattering, which arises from the deviations of the form factors from their average values:

$$I_{\text{D}}(\mathbf{q}) = \sum_{\mathbf{n}} \sum_{\mathbf{m}} \left(F_{\mathbf{n}}(\mathbf{q}) F^*_{\mathbf{m}}(\mathbf{q}) - |F_{\text{av.}}(\mathbf{q})|^2 \right) e^{i\mathbf{q}\cdot(\mathbf{R}_{\mathbf{n}} - \mathbf{R}_{\mathbf{m}})} \quad (82)$$

Equation 81 is the basis of macromolecular crystallography, and further properties of this equation are derived below. The diffuse term will be discussed in Section 2.3.

Structure Factor of the Lattice: Equation 81 is the product of the square of the average form factor and a term which depends only on the crystal lattice:

$$I_{\text{Bragg}}(\mathbf{q}) = |F_{\text{av.}}\ (\mathbf{q})|^2 |S_N(\mathbf{q})|^2, \quad (83)$$

where $S_N(\mathbf{q})$ is the lattice structure factor for N unit cells, defined as

$$S_N(\mathbf{q}) \equiv \sum_{\mathbf{n}} e^{i\mathbf{q}\cdot\mathbf{R_n}}. \quad (84)$$

At certain values of $\mathbf{q}$, the terms add in phase with $e^{i\mathbf{q}\cdot\mathbf{R_n}} = 1$, implying

$$\mathbf{q}\cdot\mathbf{R_n} = 2\pi \times \text{integer}. \quad (85)$$

The values of $\mathbf{q}$ where this occurs can be found by introducing a reciprocal basis. The reciprocal basis vectors $\mathbf{a}_j^*$ have the property that

$$\mathbf{a}_i \cdot \mathbf{a}_j^* = 2\pi\, \delta_{ij}, \quad (86)$$

where δ_{ij} is the Kronecker delta. The vectors $\mathbf{a}_j^*$ satisfying Equation 86 are

$$\mathbf{a}_1^* = 2\pi \frac{\mathbf{a}_2 \times \mathbf{a}_3}{v_c},\ \mathbf{a}_2^* = 2\pi \frac{\mathbf{a}_3 \times \mathbf{a}_1}{v_c},\ \mathbf{a}_3^* = 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_2}{v_c}, \quad (87)$$

where v_c is the unit cell volume, $v_c = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)$. The vectors $\mathbf{a}_j^*$ define a reciprocal lattice with nodes

$$\mathbf{G_h} = h_1\mathbf{a}_1^* + h_2\mathbf{a}_2^* + h_3\mathbf{a}_3^*, \quad (88)$$

where $\mathbf{h} = [h_1\ h_2\ h_3]$, and h_j are integers (the Miller indices h, k, and l). The dot product of a reciprocal lattice vector and a crystal lattice vector is found from Equations 76, 86, and 88,

$$\mathbf{G_h} \cdot \mathbf{R_n} = 2\pi\,(h_1 n_1 + h_2 n_2 + h_3 n_3). \quad (89)$$

Comparing Equations 89 and 85, it is clear that $e^{i\mathbf{q}\cdot\mathbf{R_n}} = 1$ when $\mathbf{q} = \mathbf{G_h}$. In other words, scattered waves constructively interfere when the crystal is oriented so that a particular value $\mathbf{q}$ coincides with a node of the reciprocal lattice. This statement is equivalent to Bragg's law.

The lattice structure factor depends on the shape of the crystal. For example, if the crystal consists of an array of $N = N_1 \times N_2 \times N_3$ unit cells, the square of the structure factor is

$$|S_{N_1 N_2 N_3}(\mathbf{q})|^2 = \left| \prod_{j=1}^{3} \sum_{n_j=1}^{N_j} e^{i\, n_j (\mathbf{q}\cdot\mathbf{a}_j)} \right|^2 \quad (90)$$

The sum is a geometric series, which can be evaluated explicitly,

$$|S_{N_1 N_2 N_3}(\mathbf{q})|^2 = \prod_{j=1}^{3} \left(\frac{\sin\left(N_j\, \mathbf{q}\cdot\mathbf{a}_j/2\right)}{\sin\left(\mathbf{q}\cdot\mathbf{a}_j/2\right)} \right)^2 \quad (91)$$

This function is sharply peaked at nodes of the reciprocal lattice and contains minor peaks, or fringes, between the nodes. Equation 91 has been used to describe the continuous diffraction of nanocrystals.[56] When the number of unit cells is large, $|S_N(\mathbf{q})|^2$ approaches a series of delta functions centered on the reciprocal lattice nodes:

$$\lim_{N\to\infty} |S_N(\mathbf{q})|^2 = N\, v_c^* \sum_{\mathbf{h}} \delta(\mathbf{q} - \mathbf{G_h}), \quad (92)$$

where v_c^* is the volume of a unit cell of the reciprocal lattice. Thus, the intensity for an ideal macroscopic crystal is

$$I_{\text{Bragg}}(\mathbf{q}) = |F_{\text{av.}}\ (\mathbf{q})|^2 N\, v_c^* \sum_{\mathbf{h}} \delta(\mathbf{q} - \mathbf{G_h}) \text{ as } N \to \infty. \quad (93)$$

Equation 93 implies that Bragg diffraction allows for the measurement of $|F_{\text{av.}}(\mathbf{q})|^2$ at nodes in the reciprocal lattice $\mathbf{G_h}$.

Integrated Intensity of a Bragg Peak: In practice, macroscopic protein crystals have lattice defects that prevent the entire crystal from diffracting coherently according to Equation 93. However, Equation 93 still describes the diffraction for the parts of the crystal that do scatter coherently. In the kinematic theory of diffraction, the macroscopic crystal is approximated as

a collection of coherently scattering crystallites with slightly different orientations, known as mosaic blocks. It is also assumed that the crystal is “ideally mosaic”, meaning that no two mosaic blocks are perfectly aligned. Then, the diffraction from the entire crystal is the sum of $I_{\text{Bragg}}(\mathbf{q})$ for each mosaic block. The misalignment of mosaic blocks causes the Bragg peaks to be less sharp than expected for a perfect crystal. However, the integrated intensity of a Bragg peak associated with the reciprocal lattice node $\mathbf{G_h}$ is the same as that of a coherently-diffracting crystal (Equation 93).

The exact expression for the integrated intensity depends on the details of the measurement. Typically, photon counts are accumulated on a detector while the crystal is rotated continuously around a fixed axis. If, for a given Bragg peak, the rotation axis is normal to the scattering plane defined by $\hat{\mathbf{s}}$ and $\hat{\mathbf{s}}'$, the integrated intensity of the peak is

$$I_{\mathbf{h}} = |F_{\text{av.}}\ (\mathbf{G_h})|^2 N \frac{\lambda^3}{v_c} \frac{1}{\sin 2\theta}. \quad (94)$$

This equation is derived from Equation 93 by integrating $\delta(\mathbf{q}–\mathbf{G_h})$ over the direction of the diffracted beam and the angle of the crystal (see Ref. 18). The geometric factor $(\sin 2\theta)^{-1}$ is known as the Lorentz factor, L. When the rotation axis is not normal to the scattering plane, the Lorentz factor must be modified; however, it only depends on the geometry of the experiment, $L = L(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$.

2.3.2 General Theory of Diffuse Scattering

Guinier’s Equation: In Section 2.3.1, we saw that the total intensity of an imperfect crystal can be separated into two terms: the Bragg term and the diffuse term (Equation 80). The derivation assumed that the arrangement of atoms in each unit cell is fixed. Here, we generalize the derivation by allowing the distribution of matter in the crystal to sample from a statistical ensemble. If the (grand canonical) ensemble average is denoted by $\langle \cdots \rangle$, the diffuse scattering (Equation 82) is

$$I_{\text{D}}(\mathbf{q}) = \sum_{\mathbf{n}} \sum_{\mathbf{m}} \left(\langle F_{\text{n}}(\mathbf{q}) F^*_{\mathbf{m}}(\mathbf{q}) \rangle - |\langle F_{\text{cell}}(\mathbf{q}) \rangle|^2 \right) e^{i\mathbf{q}\cdot(\mathbf{R_n}-\mathbf{R_m})} \quad (95)$$

$$= \langle |F_{\text{crystal}}\ (\mathbf{q})|^2 \rangle - |\langle F_{\text{crystal}}\ (\mathbf{q}) \rangle|^2. \quad (96)$$

Equation 96 is known as Guinier’s equation.[57] It can be rearranged as

$$I_{\text{D}}(\mathbf{q}) = \langle |F_{\text{crystal}}\ (\mathbf{q}) - \langle F_{\text{crystal}}\ (\mathbf{q}) \rangle|^2 \rangle. \quad (97)$$

Since $F(\mathbf{q})$ is the Fourier transform of the electron density, the difference $F(\mathbf{q}) - \langle F(\mathbf{q}) \rangle$ appearing in Equation 97 can be interpreted as an electron density difference between that of the real crystal and that of a perfect crystal with each unit cell's density replaced by the ensemble average.

Independent Unit Cells: An important special case of Equation 95 is where the arrangement of atoms in one unit cell is statistically independent of its neighbors, which implies

$$\langle F_{\mathbf{m}}(\mathbf{q})\, F^{*}_{\mathbf{n}}(\mathbf{q})\rangle = \begin{cases} \langle |F_{\text{cell}}(\mathbf{q})|^2 \rangle & \mathbf{m}=\mathbf{n} \\ |\langle F_{\text{cell}}(\mathbf{q})\rangle|^2 & \mathbf{m} \neq \mathbf{n}. \end{cases} \quad (98)$$

In this case, all terms with $\mathbf{m} \neq \mathbf{n}$ in Equation 95 will be zero, and the diffuse scattering is

$$I_{\text{D}}(\mathbf{q}) = N \left(\langle |F_{\text{cell}}(\mathbf{q})|^2 \rangle - |\langle F_{\text{cell}}(\mathbf{q})\rangle|^2 \right) \quad (99)$$

One important property of Equation 99 is that the lattice sum completely disappears. Thus, when correlations do not cross the unit cell boundary, the diffuse scattering pattern is spread throughout reciprocal space. The pattern is cloudy in appearance and smoothly varying; as a consequence, it is relatively straightforward to separate diffuse scattering of this type from the Bragg peaks.

2.3.3 Harmonic Approximation—Disorder in a crystal that is caused by atomic motion is known as displacement disorder. In macromolecular crystallography, it is common to approximate displacement disorder using a three-dimensional Gaussian distribution for each atom,

$$p(\mathbf{u}_j) = \frac{1}{\sqrt{(2\pi)^3 |\boldsymbol{U}_j|}} e^{-\frac{1}{2}\mathbf{u}_j^T \boldsymbol{U}_j^{-1} \mathbf{u}_j}, \quad (100)$$

where $\mathbf{u}_j = \mathbf{r}_j - \langle \mathbf{r}_j \rangle$ is the displacement of atom j from its average position. Here, the notational convention is that all vectors are column vectors, so that $\mathbf{u}^T\mathbf{u}$ is a scalar and $\mathbf{u}\mathbf{u}^T$ is a 3 × 3 matrix. $\boldsymbol{U}_j$ is the variance-covariance matrix of atomic displacements,

$$\boldsymbol{U}_j = \langle \mathbf{u}_j\, \mathbf{u}_j^T \rangle. \quad (101)$$

The mean squared displacement is $\langle u_j^2 \rangle = \text{Tr}(\boldsymbol{U}_j)/3$. When refining a crystal structure, $\boldsymbol{U}_j$ may be determined along with the coordinate for each atom. In the context of

crystallographic refinement, the 6 unique elements of the 3 × 3 symmetric matrix $\boldsymbol{U}_j$ are known as the *anisotropic displacement parameter*s (ADPs), and the mean squared displacement is usually given in terms of the *isotropic B-factor*, $B_j=8\pi^2\langle u_j^2\rangle$.

If the atoms in a crystal undergo thermally excited motions in a harmonic potential, their positions obey a Gaussian distribution, and Equation 100 is exact. However, protein dynamics at ambient temperature are characterized by diffusion on a complex free energy landscape,[58] and hence the harmonic model does not apply, except perhaps at local minima. Even so, the Gaussian distribution implied by the harmonic model can be used to approximate the true distribution of atomic positions. The reason Equation 100 is useful is that it simplifies the ensemble average of phase factors,

$$\begin{aligned}\langle e^{i\mathbf{q}\cdot\mathbf{r}}\rangle &= \int e^{i\mathbf{q}\cdot\mathbf{r}}p(\mathbf{r}-\langle\mathbf{r}\rangle)d^3\mathbf{r}\\ &= e^{i\mathbf{q}\cdot\langle\mathbf{r}\rangle}\int e^{i\mathbf{q}\cdot\mathbf{u}}p(\mathbf{u})d^3\mathbf{u}\\ &= e^{i\mathbf{q}\cdot\langle\mathbf{r}\rangle}e^{-\frac{1}{2}\mathbf{q}^T\boldsymbol{U}\mathbf{q}}\end{aligned} \tag{102}$$

Comparing Equations 101 and 102, the harmonic approximation implies

$$\langle e^{i\mathbf{q}\cdot\mathbf{u}}\rangle = e^{-\frac{1}{2}\mathbf{q}^T\langle\mathbf{u}\mathbf{u}^T\rangle\mathbf{q}}. \tag{103}$$

Thus, the ensemble average of the form factor (Equation 77) is

$$\begin{aligned}\langle F_{\text{crystal}}(\mathbf{q})\rangle &= \sum_{\mathbf{n}j} e^{i\,\mathbf{q}\cdot\mathbf{R}_\mathbf{n}} f_j(q)\ \langle e^{i\,\mathbf{q}\cdot\mathbf{r}_{\mathbf{n}j}}\rangle\\ &= \sum_{\mathbf{n}j} e^{i\,\mathbf{q}\cdot\mathbf{R}_\mathbf{n}} f_j(q)\,T_j(\mathbf{q})\,e^{i\,\mathbf{q}\cdot\langle\mathbf{r}_j\rangle},\end{aligned} \tag{104}$$

where the index $\mathbf{n}j$ is shorthand for atom j in unit cell $\mathbf{n}$, and

$$T_j(\mathbf{q}) = e^{-\frac{1}{2}\mathbf{q}^T\boldsymbol{U}_j\mathbf{q}} \tag{105}$$

is the *Debye-Waller factor* for atom j. Equations 81 and 104 give the Bragg diffraction in the harmonic approximation,

$$\begin{aligned}I_{\text{Bragg}}(\mathbf{q}) &= |\langle F_{\text{crystal}}(\mathbf{q})\rangle|^2\\ &= \left|\sum_{\mathbf{n}} e^{i\,\mathbf{q}\cdot\mathbf{R}_\mathbf{n}}\right|^2 \left|\sum_j f_j(q)\,T_j(\mathbf{q})\,e^{i\,\mathbf{q}\cdot\langle\mathbf{r}_j\rangle}\right|^2\end{aligned} \tag{106}$$

To calculate the diffuse scattering from Guinier's equation (Equation 96) in the harmonic approximation, the ensemble average of the squared form factor is needed,

$$\langle |F_{\text{crystal}}(\mathbf{q})|^2 \rangle = \sum_{\mathbf{n}j} \sum_{\mathbf{m}k} f_j(q) f_k^*(q) e^{i\mathbf{q}\cdot(\mathbf{R_n}-\mathbf{R_m})} \langle e^{i\,\mathbf{q}\cdot(\mathbf{r}_{\mathbf{n}j}-\mathbf{r}_{\mathbf{m}k})} \rangle. \tag{107}$$

To calculate the average phase factors in Equation 107 using the harmonic approximation, we apply Equation 103,

$$\begin{aligned} \langle e^{i\,\mathbf{q}\cdot(\mathbf{r}_{\mathbf{n}j}-\mathbf{r}_{\mathbf{m}k})} \rangle &= e^{i\,\mathbf{q}\cdot(\langle \mathbf{r}_j \rangle - \langle \mathbf{r}_k \rangle)} e^{-\frac{1}{2}\mathbf{q}^T \langle (\mathbf{u}_{\mathbf{n}j}-\mathbf{u}_{\mathbf{m}k})(\mathbf{u}_{\mathbf{n}j}-\mathbf{u}_{\mathbf{m}k})^T \rangle \mathbf{q}} \\ &= e^{i\,\mathbf{q}\cdot(\langle \mathbf{r}_j \rangle - \langle \mathbf{r}_k \rangle)} T_j(\mathbf{q}) T_k(\mathbf{q})\, e^{\mathbf{q}^T \boldsymbol{V}_{\mathbf{n}j\mathbf{m}k} \mathbf{q}} \end{aligned} \tag{108}$$

where $\boldsymbol{V}_{\mathbf{n}j\mathbf{m}k} = \langle \mathbf{u}_{\mathbf{n}j}\, \mathbf{u}_{\mathbf{m}k}{}^T \rangle$ is a 3 × 3 matrix giving the displacement covariances for each pair of atoms in the crystal. Finally, combining Equations 95, 104, 107, and 108, the diffuse scattering becomes in the harmonic approximation,

$$I_D(\mathbf{q}) = \sum_{\mathbf{n}j} \sum_{\mathbf{m}k} f_j(q) f_k^*(q) e^{i\mathbf{q}\cdot(\mathbf{R_n}-\mathbf{R_m})} e^{i\,\mathbf{q}\cdot(\langle \mathbf{r}_j \rangle - \langle \mathbf{r}_k \rangle)} \times T_j(\mathbf{q}) T_k(\mathbf{q}) \left\{ e^{\mathbf{q}^T \boldsymbol{V}_{\mathbf{n}j\mathbf{m}k} \mathbf{q}} - 1 \right\}. \tag{109}$$

Equation 109 is the starting point for most of the diffuse scattering models in the literature.

Many of the factors appearing in Equation 109 can be determined by solving the structure using the Bragg diffraction (Equation 106), including the crystal lattice $\mathbf{R_n}$, the average atomic coordinates, $\langle \mathbf{r}_j \rangle$, and the displacement parameters determining T_j (isotropic B-factors and/or ADPs). The single exception is the factor $\{e^{\mathbf{q}^T \boldsymbol{V}_{\mathbf{n}j\mathbf{m}k} \mathbf{q}} - 1\}$, which is unique to diffuse scattering. *This term is nonzero for any pair of atoms whose displacements are correlated.*

When the displacements are small compared to the resolution, the exponential in Equation 109 can be replaced by its Taylor expansion,

$$e^{\mathbf{q}^T \boldsymbol{V}_{\mathbf{n}j\mathbf{m}k} \mathbf{q}} = 1 + \mathbf{q}^T \boldsymbol{V}_{\mathbf{n}j\mathbf{m}k} \mathbf{q} + \frac{1}{2} \left(\mathbf{q}^T \boldsymbol{V}_{\mathbf{n}j\mathbf{m}k} \mathbf{q} \right)^2 + \ldots, \tag{110}$$

allowing the total intensity (Equation 80) to be written

$$I_{\text{Total}}(\mathbf{q}) = I_{\text{Bragg}}(\mathbf{q}) + I_{D1}(\mathbf{q}) + I_{D2}(\mathbf{q}) + \ldots \tag{111}$$

Here, $I_{Di}(\mathbf{q})$ is the *ith-order diffuse scattering*. The first order diffuse scattering is then equal to Equation 109 where only the first two terms in the Taylor expansion are retained,

$$I_{D1}(\mathbf{q})=\sum_{\mathbf{n}j}\sum_{\mathbf{m}k} f_j(q)\, f_k^*(q)\, e^{i\mathbf{q}\cdot(\mathbf{R_n}-\mathbf{R_m})} e^{i\,\mathbf{q}\cdot(\langle\mathbf{r}_j\rangle-\langle\mathbf{r}_k\rangle)} \times T_j(\mathbf{q})T_k(\mathbf{q})\,\left\{\mathbf{q}^T\boldsymbol{V}_{\mathbf{n}j\mathbf{m}k}\mathbf{q}\right\} \quad (112)$$

Because of its mathematical convenience, Equation 112 is often the starting point for simplified analytical models.

The goal of collecting diffuse scattering data is to determine how motions are correlated in the crystal, information which is embodied by the variance-covariance tensor $\boldsymbol{V}_{\mathbf{n}j\mathbf{m}k}$. However, it is impossible to determine this tensor directly from the data because it has an enormous number of elements: nine for every *pair* of atoms in the crystal. Thus, diffuse scattering must be interpreted with the aid of a dynamical model.

2.3.4 Models of Correlated Motion—In the following sections, we give several approximate forms of $\boldsymbol{V}_{\mathbf{n}j\mathbf{m}k}$ that have been used to interpret diffuse scattering.

Independent Unit Cells: When the diffuse scattering pattern is cloudy in appearance and not strongly correlated with the nodes of the reciprocal lattice, it is then assumed that the motions within a unit cell are independent of those in neighboring cells (Equation 98). Under this assumption, the covariance matrix is zero except when a pair of atoms belongs to the same unit cell,

$$\boldsymbol{V}_{\mathbf{n}j\mathbf{m}k}=\delta_{\mathbf{nm}}\boldsymbol{V}_{jk}, \quad (113)$$

where $\delta_{\mathbf{nm}}$ is the Kronecker delta. Combining Equations 109 and 113, the diffuse scattering for independent unit cells in the harmonic approximation is

$$I_D(\mathbf{q})=N\sum_j\sum_k f_j(q)f_k^*(q)\, e^{i\,\mathbf{q}\cdot(\langle\mathbf{r}_j\rangle-\langle\mathbf{r}_k\rangle)} T_j(\mathbf{q})T_k(\mathbf{q})\,\left\{e^{\mathbf{q}^T\boldsymbol{V}_{ij}\mathbf{q}}-1\right\} \quad (114)$$

Independent Atoms: Similarly, the covariance matrix for independently vibrating atoms is

$$\boldsymbol{V}_{\mathbf{n}j\mathbf{m}k}=\delta_{\mathbf{nm}}\delta_{jk}U_j, \quad (115)$$

which can be combined with Equation 109 for the diffuse scattering,

$$I_D(\mathbf{q}) = N\sum_j |f_j(q)|^2 T_j^2(\mathbf{q}) \left\{ e^{\mathbf{q}^T \boldsymbol{U}_j \mathbf{q}} - 1 \right\} = N\sum_j |f_j(q)|^2 \{1 - T_j^2(\mathbf{q})\}. \quad (116)$$

Although this model is quite simple, it underscores the point that all protein crystals with non-zero B-factors must generate some amount of diffuse scattering, even if correlated motions are insignificant. Intuitively, this must be the case because the total amount of scattering for a collection of atoms is constant; when B-factors decrease the intensity of Bragg peaks, these extra photons have to go somewhere else. This independent-atom contribution (which may be pulled out of the sum in Equation 109) is present in all diffuse scattering patterns as a smoothly varying background.

Independent Rigid Groups: In general, we do not expect atomic motion to be completely independent. Atoms interact with their neighbors through covalent bonds, electrostatic forces, and short-ranged effects such as hydrogen bonds and van derWaals interactions. When interactions within a group of atoms are strong, correlated motions can be modeled assuming the atoms move together as a rigid body. For example, the atoms in a protein side chain might be assumed to move in this way. Even an entire domain or unit cell might undergo small collective rotations and translations in the crystal.

The theory of rigid-body motion is important in conventional crystallography, where it is sometimes applied during structure refinement. In that context, it is known as *TLS refinement* (T, L, and S stand for translation, libration, and screw motion), which was first introduced by Schomaker and Trueblood in 1968.[59] Crystallographic refinement must invert Equation 106 using a limited number of observations – an under-constrained problem that is prone to over-fitting. Each atom's ADP has 6 parameters, and unless the resolution of the data is very good, there will not be enough information in the data to determine the ADPs. TLS refinement solves this problem by imposing a rigid-body model to increase the data-to-parameter ratio. A single TLS group has 21 vibrational parameters, of which 20 can be determined from Bragg data, so as long as the number of atoms per group is greater than 3, the data-to-parameter ratio increases. TLS refinement is often applied in protein crystallography,[60] where it is common to define entire domains as TLS groups,[61] making it possible to refine ADPs even when the resolution is poor.

Briefly, we derive the TLS model of rigid-body motion. For an atom belonging to a rigid body, the fluctuations in position of the j^{th} atom are

$$\mathbf{u}_j = \mathbf{t} + \mathscr{R}_{\boldsymbol{\lambda}} \bar{\mathbf{r}}_j, \quad (117)$$

where **t** is a translation vector for the group, $\mathscr{R}_{\boldsymbol{\lambda}}$ is a rotation matrix for the group, and $\bar{\mathbf{r}}_j$ is the position of atom *j* when the group is stationary (this may be slightly different from $\langle \mathbf{r}_j \rangle$, defined earlier). The rotations may be parameterized by a vector $\boldsymbol{\lambda}$ pointing along the

instantaneous rotation axis with a magnitude equal to the rotation angle. For small rotations, Equation 117 can be written

$$\mathbf{u}_j \approx \mathbf{t}+\boldsymbol{\lambda}+\bar{\mathbf{r}}_j=\mathbf{t}+[\bar{\mathbf{r}}_j]_\times\boldsymbol{\lambda}, \quad (118)$$

where the cross product matrix $[\mathbf{r}]_\times$ for a vector $\mathbf{r} = [x, y, z]$ is

$$[\mathbf{r}]_\times \equiv \begin{bmatrix} 0 & z & -y \\ -z & 0 & x \\ y & -x & 0 \end{bmatrix}. \quad (119)$$

If atoms *j* and *k* belong to the same group, their displacement covariance matrix can be written using the covariances of the translation vector, **t** and the rotation vector, $\boldsymbol{\lambda}$, according to Equation 118,

$$\begin{aligned} \boldsymbol{V}_{jk} &= \left\langle \mathbf{u}_j\, \mathbf{u}_k^{\mathrm{T}} \right\rangle \\ &= \left\langle \mathbf{t}\,\mathbf{t}^{\mathrm{T}} \right\rangle + \left\langle \mathbf{t}\,\boldsymbol{\lambda}^{\mathrm{T}} \right\rangle [\bar{\mathbf{r}}_k]_\times^{\mathrm{T}} + [\bar{\mathbf{r}}_j]_\times \left\langle \boldsymbol{\lambda}\,\boldsymbol{t}^{\mathrm{T}} \right\rangle + [\bar{\mathbf{r}}_j]_\times \left\langle \boldsymbol{\lambda}\,\boldsymbol{\lambda}^{\mathrm{T}} \right\rangle [\bar{r}_k]_\times^{\mathrm{T}} \end{aligned} \quad (120)$$

To simplify the notation, three matrices are defined; $\boldsymbol{T} \equiv \langle \mathbf{t}\,\mathbf{t}^{\mathrm{T}} \rangle$, $\boldsymbol{L} \equiv \langle \boldsymbol{\lambda}\boldsymbol{\lambda}^{\mathrm{T}} \rangle$, and $\boldsymbol{S} \equiv \langle \boldsymbol{\lambda}\boldsymbol{t}^{\mathrm{T}} \rangle$. For a pair of atoms belonging to a TLS group labeled by the index α, Equation 120 is

$$\begin{aligned} \boldsymbol{V}_{jk}^{(\alpha)} &= \boldsymbol{T}_\alpha + \boldsymbol{S}_\alpha^{\mathrm{T}} [\bar{\mathbf{r}}_k]_\times^{\mathrm{T}} + [\bar{\mathbf{r}}_j]_\times \boldsymbol{S}_\alpha + [\bar{\mathbf{r}}_j]_\times \boldsymbol{L}_\alpha [\bar{\mathbf{r}}_k]_\times^{\mathrm{T}} \\ &= \begin{bmatrix} \boldsymbol{I}_3 & [\bar{\mathbf{r}}_j]_\times \end{bmatrix} \sum_\alpha \begin{bmatrix} \boldsymbol{I}_3 & [\bar{\mathbf{r}}_k]_\times \end{bmatrix}^{\mathrm{T}} \end{aligned} \quad (121)$$

where $\boldsymbol{I}_3$ is a 3 × 3 identity matrix, and $\boldsymbol{\Sigma}_\alpha$ is a 6 × 6 positive semidefinite matrix,

$$\sum_\alpha \equiv \begin{bmatrix} \boldsymbol{T}_\alpha & \boldsymbol{S}_\alpha^{\mathrm{T}} \\ \boldsymbol{S}_\alpha & \boldsymbol{L}_\alpha \end{bmatrix}. \quad (122)$$

In TLS refinement, the Debye-Waller factor (Equation 105) is replaced by

$$T_j(\mathbf{q}) = e^{-\frac{1}{2}\mathbf{q}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{I}_3 & [\bar{\mathbf{r}}_j]_\times \end{bmatrix} \sum_\alpha \begin{bmatrix} \boldsymbol{I}_3 & [\bar{\mathbf{r}}_j]_\times \end{bmatrix}^{\mathrm{T}} \mathbf{q}} \quad (123)$$

and the elements of $\boldsymbol{T}$, $\boldsymbol{L}$, and $\boldsymbol{S}$ are optimized to fit the Bragg data as closely as possible.

Rigid-body motion gives rise to diffuse scattering. In general, it is necessary to account for the correlation between motions of atoms *j* and *k* when they belong to separate groups. When atoms *j* and *k* are in groups α and β, respectively, these correlations are determined by matrices such as $\langle \boldsymbol{\lambda}_\alpha \mathbf{t}_\beta^T \rangle$.[62] The simplest assumption is that the groups are independent,[5] so that $\boldsymbol{V}_{jk} = 0$ for this atom pair. It follows that the diffuse scattering from the rigid groups is additive,

$$I_D(\mathbf{q}) = N \sum_\alpha I_D^{(\alpha)}(\mathbf{q}) \quad (124)$$

where $I_D^{(\alpha)}(\mathbf{q})$ is the diffuse scattering calculated for atoms belonging to group α (Equation 114),

$$I_D^{(\alpha)}(\mathbf{q}) = \sum_j \sum_k f_j(q) f_k^*(q)\, e^{i\,\mathbf{q}\cdot(\langle \mathbf{r}_j \rangle - \langle \mathbf{r}_k \rangle)} T_j(\mathbf{q}) T_k(\mathbf{q}) \times \left\{ e^{\mathbf{q}^T \begin{bmatrix} \boldsymbol{I}_3 & [\bar{\mathbf{r}}_j]_\times \end{bmatrix} \sum_\alpha \begin{bmatrix} \boldsymbol{I}_3 & [\bar{\mathbf{r}}_k]_\times \end{bmatrix}^T \mathbf{q}} - 1 \right\} \quad (125)$$

Equation 125 takes a particularly simple form when the the motion is purely translational, because $\boldsymbol{V}_{jk}^{(\alpha)} = \boldsymbol{T}_\alpha$ is the same for every atom in the group. Then, the diffuse scattering from that group is

$$I_D^{(\alpha)}(\mathbf{q}) = \left\{ e^{\mathbf{q}^T \boldsymbol{T}_\alpha \mathbf{q}} - 1 \right\} |\langle F_\alpha(\mathbf{q}) \rangle|^2, \quad (126)$$

where $\langle F_\alpha(\mathbf{q}) \rangle$ is the average form factor of the group (including the Debye-Waller factors). Since in this case the Debye-Waller factors are the same for all atoms in the group, $T_j(\mathbf{q}) = \exp(-\frac{1}{2}\mathbf{q}^T \boldsymbol{T}_\alpha \mathbf{q})$, they can be factored out of the average form factor,

$$I_D^{(\alpha)}(\mathbf{q}) = \left\{ 1 - e^{-\mathbf{q}^T \boldsymbol{T}_\alpha \mathbf{q}} \right\} |F_\alpha(\mathbf{q})|^2. \quad (127)$$

Liquid-like Motion: One of the first diffuse scattering models developed for protein crystals was the liquid-like motion model, proposed by Caspar *et al.* in their pioneering study of crystalline insulin[3] and later generalized.[63,64] The three fundamental assumptions of the model are:

1. Atomic displacements can be described by the harmonic approximation, with a variance-covariance tensor, $\boldsymbol{V}_{\mathbf{n}j\mathbf{m}k}$, that is a function of the vector between the atoms *j* and *k*,

$$\Delta\mathbf{r}_{\mathbf{n}j\mathbf{m}k} = \langle \mathbf{r}_{\mathbf{n}j} \rangle - \langle \mathbf{r}_{\mathbf{m}k} \rangle , \quad (128)$$

   where $\langle \mathbf{r}_{\mathbf{n}j} \rangle = \mathbf{R}_{\mathbf{n}} + \langle \mathbf{r}_j \rangle$ is the average position of atom *j*.

2. $\boldsymbol{V}_{\mathbf{n}j\mathbf{m}k}$ decays to zero as the length of $\Delta\mathbf{r}_{\mathbf{n}j\mathbf{m}k}$ increases.

3. Each atom's displacements are isotropic, so that $\boldsymbol{U}_j = \left\langle u_j^2 \right\rangle$

With these assumptions, the functional form of $\boldsymbol{V}_{\mathbf{n}j\mathbf{m}k}$ can be defined as,

$$\boldsymbol{V}_{\mathbf{n}j\mathbf{m}k} = \left(\left\langle u_j^2 \right\rangle \left\langle u_k^2 \right\rangle \right)^{1/2} \Gamma(\Delta\mathbf{r}_{\mathbf{n}j\mathbf{m}k}), \quad (129)$$

where $\Gamma(\Delta\mathbf{r})$ is the correlation function, which equals 1 when $\Delta\mathbf{r} = 0$. The diffuse scattering for the general liquid-like model is

$$I_D(\mathbf{q}) = \sum_{\mathbf{n}j}\sum_{\mathbf{m}k} f_j(q) f_k^*(q) e^{i\mathbf{q}\cdot\Delta\mathbf{r}_{\mathbf{n}j\mathbf{m}k}} T_j(\mathbf{q}) T_k(\mathbf{q}) \left\{ e^{q^2\left(\langle u_j^2\rangle\langle u_k^2\rangle\right)^{1/2}\Gamma(\Delta\mathbf{r}_{\mathbf{n}j\mathbf{m}k})} - 1 \right\} \quad (130)$$

In diffuse scattering studies to-date, various approximations to Equation 130 have been applied. Assuming that all atoms have the same displacements, $\left\langle u_j^2 \right\rangle = \langle u^2 \rangle$, Caspar et al.[3] applied the first-order scattering approximation (Equation 112) to Equation 130 to obtain the approximation

$$I_{D1}(\mathbf{q}) = q^2 \left\langle u^2 \right\rangle e^{-q^2\langle u^2\rangle} \sum_{\mathbf{n}j}\sum_{\mathbf{m}k} f_j(q) f_k^*(q) e^{i\mathbf{q}\cdot\Delta\mathbf{r}_{\mathbf{n}j\mathbf{m}k}} \Gamma(\Delta\mathbf{r}_{\mathbf{n}j\mathbf{m}k}). \quad (131)$$

Equation 131 can be written in a more intuitive form if the atoms in the crystal are assumed to be point-like, so that the unperturbed electron density can be approximated as a sum of delta functions, $\rho_0(\mathbf{r}) \approx \Sigma_{\mathbf{n}j} Z_j \delta(\mathbf{r} - \mathbf{r}_{\mathbf{n}j})$, which implies $f_j(q) \approx Z_j$. With this approximation, the double sum in Equation 131 can be written,

$$\begin{aligned}&\sum_{\mathbf{n}j}\sum_{\mathbf{m}k} Z_j Z_k e^{i\mathbf{q}\cdot\Delta\mathbf{r}_{\mathbf{n}j\mathbf{m}k}}\Gamma(\Delta\mathbf{r}_{\mathbf{n}j\mathbf{m}k})\\ =&\sum_{\mathbf{n}j}\sum_{\mathbf{m}k}\int\int Z_j\delta(\mathbf{r}-\overline{\mathbf{r}}_{nj})Z_k\delta(\mathbf{r}'-\overline{\mathbf{r}}_{mk})\\ &\times\Gamma(\mathbf{r}-\mathbf{r}')\,e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')}d^3\mathbf{r}d^3\mathbf{r}'\\ =&\int\int\rho_0(\mathbf{r})\rho_0(\mathbf{r}')\Gamma(\mathbf{r}-\mathbf{r}')\,e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')}d^3\mathbf{r}d^3\mathbf{r}'\\ =&\int\left(\int\rho_0(\mathbf{u}+\mathbf{r})\rho_0(\mathbf{u})d^3\mathbf{u}\right)\Gamma(\mathbf{r})\,e^{i\mathbf{q}\cdot\mathbf{r}}d^3\mathbf{r}\\ =&\mathscr{F}\{(\rho_0\star\rho_0)(\mathbf{r})\,\Gamma(\mathbf{r})\},\end{aligned}\tag{132}$$

where $(\rho_0 \star \rho_0)(\mathbf{r})$ is the Patterson function of the unperturbed crystal with point-like atoms, and $\mathscr{F}\{\cdots\}$ denotes the Fourier transform. According to Equation 132, the correlation function acts as an envelope of the Patterson function, suppressing features at large distances. Applying the convolution theorem to Equation 132, we obtain

$$\begin{aligned}\mathscr{F}\{(\rho_0\star\rho_0)(\mathbf{r})\,\Gamma(\mathbf{r})\}&=\tfrac{1}{(2\pi)^3}\mathscr{F}\{(\rho_0\star\rho_0)(\mathbf{r})\}*\mathscr{F}\{\Gamma(\mathbf{r})\}\\ &=\tfrac{1}{(2\pi)^3}|F_0(\mathbf{q})|^2*\hat{\Gamma}(\mathbf{q})\end{aligned}\tag{133}$$

where $\hat{\Gamma}(\mathbf{q})$ is the Fourier transform of the correlation function, $\Gamma(\mathbf{r})$. Equation 131 for the diffuse scattering is

$$I_{D1}(\mathbf{q})=\frac{1}{(2\pi)^3}q^2\left\langle u^2\right\rangle e^{-q^2\langle u^2\rangle}|F_0(\mathbf{q})|^2*\hat{\Gamma}(\mathbf{q})\tag{134}$$

Note that the structure factor of the unperturbed crystal can be written as a sum of delta functions (Equation 92), as follows

$$|F_0(\mathbf{q})|^2=N\,v_c^*|F_{\text{cell}}(\mathbf{q})|^2\sum_{\mathbf{h}}\delta(\mathbf{q}-\mathbf{G}_\mathbf{h}).\tag{135}$$

Then, Equation 134 can alternately be expressed,

$$I_{D1}(\mathbf{q})=\frac{N}{v_c}q^2\langle u^2\rangle e^{-q^2\langle u^2\rangle}\sum_{\mathbf{h}}|F_{\text{cell}}(\mathbf{G}_\mathbf{h})|^2\hat{\Gamma}(\mathbf{q}-\mathbf{G}_\mathbf{h}).\tag{136}$$

Equations 134 and 136 are the two most common representations of the liquid-like model.

The correlation function $\Gamma(\mathbf{r})$ is usually modeled as a sum of exponentials, characterized by correlation lengths γ_i,

$$\Gamma(\mathbf{r})=\sum_l e^{-\gamma_l^{-1}|\mathbf{r}|} \quad (137)$$

The Fourier transform of Equation 137 is

$$\hat{\Gamma}(\mathbf{q})=\sum_l \frac{8\pi\gamma_l^3}{(1+q^2\gamma_l^2)^2}. \quad (138)$$

Anisotropic correlation lengths can also be modeled,[64]

$$\Gamma(\mathbf{r})=\sum_l e^{-|\mathbf{\Lambda}_l^{-1}\mathbf{r}|} \quad (139)$$

where $\boldsymbol{\lambda}_l$ is a 3 × 3 matrix that generalizes the correlation length in three dimensions (its eigenvalues being the correlation lengths along 3 orthogonal directions). The Fourier transform of Equation 139 is[64]

$$\hat{\Gamma}(\mathbf{q})=\sum_l \frac{8\pi \det(\mathbf{\Lambda}_l)}{\left(1+|\mathbf{\Lambda}_l\mathbf{q}|^2\right)^2}. \quad (140)$$

Normal Modes: The normal modes model is a logical extension of the harmonic approximation of disorder in a crystal. The protein atoms are assumed to vibrate collectively around their equilibrium positions in a harmonic potential. The instantaneous displacements of all N atoms in the crystal are represented by a vector $\mathbf{u}(t)$ of length $3N$. The general solution to this physics problem is a sum of $3N$ normal modes, $\mathbf{v}_\alpha$, with time-dependent coefficients, $c_\alpha(t)$, such that

$$\mathbf{u}(t)=\mathbf{M}^{-1/2}\sum_{\alpha=1}^{3N}\mathbf{v}_\alpha c_\alpha(t) \quad (141)$$

where $\mathbf{M}$ is matrix with atomic masses along the diagonal. The displacement covariance matrix is then

$$\boldsymbol{V}=\left\langle \mathbf{u}\mathbf{u}^T\right\rangle=\mathbf{M}^{-1}\sum_\alpha\sum_\beta \mathbf{v}_\alpha\left\langle c_\alpha c_\beta\right\rangle \mathbf{v}_\beta^T. \quad (142)$$

Averaged over time in thermal equilibrium, the normal modes are uncorrelated, so $\langle c_\alpha c_\beta \rangle = 0$ unless $\alpha = \beta$. Invoking the equipartition theorem of classical statistical mechanics, the matrix elements are

$$\langle c_\alpha c_\beta \rangle = \delta_{\alpha\beta} \frac{k_B T}{\omega_\alpha^2}, \quad (143)$$

where ω_α is the vibrational frequency of normal mode α. Thus, equation 142 can be written,

$$\boldsymbol{V} = \left\langle \mathbf{u}\mathbf{u}^T \right\rangle = \mathbf{M}^{-1} \sum_\alpha \mathbf{v}_\alpha \frac{k_B T}{\omega_\alpha^2} \mathbf{v}_\alpha^T \quad (144)$$

Equation 144 shows that the normal modes are eigenvectors of the mass-weighted covariance matrix with eigenvalues $k_B T/\omega_\alpha^2$.

In the next section, the thermally excited normal modes (Equation 144) are used to derive the lattice vibration model of diffuse scattering, and the frequencies in this case are physically meaningful: they relate to the propagation of acoustic waves in the crystal through the dispersion relation. However, as discussed in Section 2.3.3, Equation 144 is not a good model for the collective motions of proteins at room temperature, since protein dynamics are generally not harmonic.[58] Instead, the normal mode model is used as an orthonormal basis in which to represent the true conformational ensemble. Viewed in this way, the normal modes are useful as a convenient representation of the *important conformational subspace* of $\boldsymbol{V}$.[65] Thus, the true eigenvectors of $\boldsymbol{V}$ are not the normal modes, but they may be approximated by linear combinations of the normal modes derived from a harmonic potential. Assuming the eigenvectors of $\boldsymbol{V}$ are linear combinations of the first M normal modes, the displacement covariance matrix (Equation 142) is equivalent to

$$\boldsymbol{V} \approx \sum_{\alpha=1}^{M} \sum_{\beta=1}^{M} \mathbf{v}_\alpha \langle \sigma_\alpha \sigma_\beta \rangle \mathbf{v}_\beta^T, \quad (145)$$

where $\langle \sigma_\alpha \sigma_\beta \rangle$ is an $M \times M$ matrix of correlation coefficients, which need not be diagonal. Equation 145 was first proposed by Mizuguchi et al. [63] as a way to co-refine Bragg diffraction and diffuse scattering data; the intensities are calculated by combining Equation 145 with Equations 106 and 114, and the coefficients $\langle \sigma_\alpha \sigma_\beta \rangle$ are varied to improve the fit. Much like TLS refinement, normal mode refinement reduces the number of degrees of freedom in the ADPs when M is small. Compared with the independent rigid-body and LLM models for diffuse scattering, Equation 145 has the advantage that arbitrarily complex correlations may be included.

Lattice Vibrations: The small-scale features in diffuse scattering patterns, which correspond to long-ranged correlations in the crystal, can be predicted using lattice vibration models. In these models, internal protein degrees of freedom are typically neglected, and only unit cell translational motions are considered. Then, the covariances are the same for pairs of atoms belonging to the same unit cell, $V_{\mathbf{njmk}} = V_{\mathbf{nm}}$. The motions are also assumed to be small, so that the first-order diffuse scattering formula can be used,

$$I_{\mathrm{D1}}(\mathbf{q}) = |\langle F_{\mathrm{cell}}(\mathbf{q})\rangle|^2 \sum_{\mathbf{n}}\sum_{\mathbf{m}} \mathbf{q}^T V_{\mathbf{nm}} \mathbf{q}\, e^{i\mathbf{q}\cdot(\mathbf{R_n}-\mathbf{R_m})}. \quad (146)$$

A normal mode vibrational model is applied to calculate the covariance matrix, $V_{\mathbf{nm}}$, using Equation 144,

$$I_{\mathrm{D1}}(\mathbf{q}) = |\langle F_{\mathrm{cell}}(\mathbf{q})\rangle|^2 \sum_{\alpha} \frac{k_B T}{M_\alpha \omega_\alpha^2} \left| \sum_{\mathbf{n}} \mathbf{q}\cdot \mathbf{v}_{\mathbf{n}\alpha} e^{i\mathbf{q}\cdot\mathbf{R}_n} \right|^2 \quad (147)$$

where $\mathbf{v}_{\mathbf{n}\alpha}$ is the component of normal mode α associated with the displacement of the atoms in unit cell **n**. The normal modes in this case are longitudinal and transverse displacement waves in the crystal,

$$\mathbf{v}_{\mathbf{n}\alpha} = \mathbf{1}_\alpha e^{-i\mathbf{Q}_\alpha\cdot\mathbf{R_n}}, \quad (148)$$

where $\mathbf{Q}_\alpha$ is the wavevector of mode α, and $\mathbf{1}_\alpha$ is a unit vector along the polarization direction of the wave (perpendicular to $\mathbf{Q}_\alpha$ for transverse modes, and parallel to $\mathbf{Q}_\alpha$ for longitudinal modes). The frequency ω_α is a function of $\mathbf{Q}_\alpha$ according to the dispersion relation for the crystal. Using Equation 148 in Equation 147, the diffuse scattering is

$$I_{\mathrm{D1}}(\mathbf{q}) = |\langle F_{\mathrm{cell}}(\mathbf{q})\rangle|^2 k_B T \sum_{\alpha} \frac{(\mathbf{q}\cdot\mathbf{1}_\alpha)^2}{M_\alpha \omega_\alpha^2} \left| \sum_{\mathbf{n}} e^{i(\mathbf{q}-\mathbf{Q}_\alpha)\cdot\mathbf{R_n}} \right|^2 \quad (149)$$

$$= N\, v_c^* k_B T \sum_{\alpha} \frac{(\mathbf{q}\cdot\mathbf{1}_\alpha)^2}{M_\alpha \omega_\alpha^2} |\langle F_{\mathrm{cell}}(\mathbf{q})\rangle|^2 \delta(\mathbf{q}-\mathbf{Q}_\alpha-\mathbf{G_h}) \quad (150)$$

where Equation 150 was derived from Equation 149 by replacing the lattice sum by a delta function (see Equation 92).

When **q** is close to one of the reciprocal lattice nodes $\mathbf{G_h}$, long-wavelength vibrations are the dominant contributors to the diffuse scattering (modes with small $\mathbf{Q}^{\alpha}$). Equation 150 with $\mathbf{Q}_{\alpha} \ll \mathbf{q}$ is approximately[66]

$$I_{D1}(\mathbf{q}\,\text{near}\,\mathbf{G_h}) \approx N\, v_c^* q^2 |\langle F_{\text{cell}}(\mathbf{G_h})\rangle|^2 k_B T \sum_{\alpha} \frac{(\mathbf{1_h}\cdot\mathbf{1}_{\alpha})^2}{M_{\alpha}\omega_{\alpha}^2} \quad (151)$$

where $\mathbf{1_h}$ is a unit vector in the direction of the lattice node $\mathbf{G_h}$. For long-wavelength vibrations, a linear dispersion relation is expected with $\omega \approx vQ$, where v is the speed of sound. Therefore, Equation 151 describes a halo that is is sharply peaked underneath the Bragg reflection and decays as Q^{-2}, a well-known result for thermal diffuse scattering. Equation 151 with $\omega_{\alpha} \propto |\mathbf{q} - \mathbf{G_h}|$ is

$$I_{D1}(\mathbf{q}\,\text{near}\,\mathbf{G_h}) \approx\propto q^2 I_{\text{Bragg}}(\mathbf{G_h}) \frac{1}{|\mathbf{q}-\mathbf{G_h}|^2} \quad (152)$$

It is interesting to note the similarity between Equation 152 for lattice dynamics and Equation 136 for liquid-like motions. If Γ is chosen with a correlation length that is much longer than the unit cell dimension, the intensity near a Bragg peak decays as $\hat{\Gamma}(\mathbf{q}\ \text{near}\ \mathbf{G_h}) \approx\propto |\mathbf{q}\text{-}\mathbf{G_h}|^{-4}$ according to Equation 138. This decay is much steeper than would be expected for vibrational modes with a linear dispersion relation (Equation 152). Alternate functional forms for Γ have been devised with the property that $\hat{\Gamma}(\mathbf{q}\ \text{near}\ \mathbf{G_h}) \approx\propto |\mathbf{q}\text{-}\mathbf{G_h}|^{-2}$, so that diffuse scattering from lattice vibrations can be modeled approximately within the LLM formalism.[64]

Elastic Deformation: The characteristic Q^{-2} decay of the diffuse scattering near the Bragg peaks has been observed in diffraction images of RNase crystals, and the scattering was attributed to acoustic vibrations.[67] This interpretation was later confirmed by lattice dynamics simulations.[68] However, Bragg peak halos can also result from static disorder.[69] In fact, the Q^{-2} decay that characterizes acoustic vibrational scattering is not unique to that model; a Q^{-2} decay can also be caused by lattice deformation around crystal defects, known as Huang scattering.[69,70] The importance of Huang scattering for protein crystals has not yet been determined. However, we note that protein crystals tend to have a high defect density.[71] Thus, we urge caution when attributing Bragg peak halos to acoustic motions.

2.3.5 Interpretation of Data

Data Collection: Diffuse scattering measurements to date have been collected using setups designed for MX with minor modifications. In the basic MX setup, the crystal is held in a loop or capillary and placed in a well-collimated and monochromatic X-ray beam. An area detector is placed at an appropriate distance from the crystal in order to record the highest resolution reflections, while ensuring that neighboring Bragg peaks do not overlap. A beamstop is placed between the crystal and the detector. The position of the beamstop is

chosen to be as close as possible to the crystal in order to reduce scattering from air, but far enough away so that it does not block the low-angle reflections. During the X-ray exposure, the crystal is rotated continuously around a fixed axis. The optimal rotation range per X-ray exposure has been estimated to be approximately half the crystal's mosaicity when modern detectors are used.[72]

In standard MX experiments, it is common for the crystal to be cryo-cooled to temperatures of ~ 100 K. The formation of ice is prevented by cooling rapidly, often with the aid of cryoprotectants introduced during crystal growth or after harvesting. Cryo-cooling mitigates radiation damage effects and facilitates the storage, transportation, and handling of crystals.[73] Rapid cryo-cooling traps a dynamic molecule in a glass-like state, arresting large-scale motion. Nonetheless, diffuse scattering can be observed from cryo-cooled crystals. [66,74] However, significant effects of cryo-cooling on conformational equilibria have been observed.[75,76] In addition, differential contraction during cooling places strain on the crystal lattice, and can increase the apparent mosaicity of the crystal.[77] When the mosaicity is large, it becomes difficult to separate Bragg peaks from the diffuse scattering. Diffuse scattering from cooling-induced lattice strain may also be significant. For these reasons, it makes sense to measure diffuse scattering at ambient temperatures when possible.

Before modern cryo-cooling techniques were developed, ambient temperature data collection was the norm. Today, ambient temperature data collection has seen a resurgence in popularity as crystallography is increasingly used to obtain dynamic information.[76,78] In addition, the development of X-ray free electron lasers (XFELs) has enabled damage-free data collection without cryo-cooling; when the crystal is illuminated by an intense femtosecond pulse from an XFEL, diffraction occurs before the onset of radiation damage, a phenomenon known as *diffraction before destruction*.[79]

MX beamlines are designed for detecting Bragg peaks, which are recorded along with the crystal's diffuse scattering as well as background scattering from the solvent, the sample holder, and air in the beam path. When integrating MX data, the scattering near each peak is subtracted, and therefore the physical origin of the background is not a concern. In diffuse scattering, however, it is important to completely understand and mitigate background scattering when possible. Modifications to the setup and data collection protocols include minimizing the amount of solvent around the crystal, choosing a low-background sample environment such as a helium tunnel,[80] and measuring and subtracting the background scattering that remains.[6]

The X-ray detector is a key component in a diffuse scattering measurement. The ideal detector would have zero noise, high efficiency, uniform response, high dynamic range, and a sharp point-spread function that decays rapidly to zero. Pixel-array detectors (PADs), developed recently, are ideal detectors for both MX and diffuse scattering.[81] Older technologies, such as charge-coupled devices (CCDs), image places (IPs), and X-ray film, have significant drawbacks. However, under certain conditions they can all been used successfully for diffuse scattering measurements. For example, although commercially-available CCDs suffer from an artifact known as "blooming" [80,82] – where excess charge bleeds into pixels surrounding intense Bragg peaks, overwhelming the diffuse signal –

diffuse scattering has been measured using a CCD that was modified to reduce blooming at the expense of dynamic range.[82]

Data Processing and Visualization: A handful of software packages have been developed for diffuse scattering data processing. Lunus[82–84] is the most comprehensive software package for protein diffuse scattering that is currently available. Custom software is also commonly used, as well as routines developed for materials science[80] and XFEL data processing.[85]

Qualitative features in this diffuse pattern may reveal the types of disorder that are present. Therefore, visual inspection of the data is an important first step. However, diffuse patterns are not always obvious in diffraction images. Historically, very long exposures were used, compared with standard MX, to make the diffuse pattern clearly visible. With modern detectors, overexposing each frame is no longer necessary, however some amount of image processing or averaging may be required to observe the signal above the background and noise. For example, a *mode filter* can be applied to each diffraction image, where the intensity of each pixel is replaced by the most common value (the mode) among all pixels within the moving window.[82,83]

Every pixel on the detector can be mapped into the 3D reciprocal space of the crystal by indexing the Bragg peaks using standard MX software. Those pixels which fall within an appropriate range of a reciprocal lattice node are masked out, so that the remaining pixels contain only non-Bragg scattering. At this stage, one usually subtracts the instrumental background and corrects for geometric effects such as solid angle and polarization (Equation 157). The corrected intensity values are proportional to the diffuse scattering of the crystal (Equation 96) plus isotropic sources of background such as solvent scattering and Compton scattering. The isotropic component of the diffuse scattering and the background can be subtracted by taking the average scattering in annular rings of constant $|\mathbf{q}|$. What remains is the anisotropic component of the diffuse scattering, which is most relevant to correlated molecular motion.

Once the rotation images are processed to isolate the diffuse scattering, the next step is to merge the data in 3D. Several strategies for scaling, merging, and symmetry averaging have been developed (discussed in Section 4.3.1). Software packages for merging rotation data include Lunus[82,84] and XCAVATE.[80,86] Software for merging XFEL data in 3D has also been described.[85] The 3D diffuse maps can be visualized using 2D slices through reciprocal space,[85,86] Mercator projections of spherical surfaces of constant $|\mathbf{q}|$,[1,64,83] and 3D isosurface renderings.[1,84,87]

Modeling Diffuse Scattering: To interpret diffuse scattering data quantitatively, two approaches have been used: *predictive modeling*, and *model fitting*.

In the predictive modeling approach, a dynamical model is used to simulate the diffuse scattering through Equation 96, 109, or 112. Typically, two or more predictive models are compared with the data to determine which model gives the best agreement. This approach is suitable for models based on first-principles, such as molecular dynamics simulation, or

models which are fully-constrained by the Bragg data, such as uncorrelated rigid-body motions from TLS refinement. Given a structural ensemble represented by a set of Protein Data Bank (PDB) atomic coordinate files, the program *phenix.diffuse*[74] can be used to calculate 3D diffuse scattering maps using Guinier's equation (Equation 96).

In the model fitting approach, the covariance matrix $\boldsymbol{V}_{\mathbf{n}j\mathbf{n}k}$ is approximated analytically using a small number of free parameters, and Equation 109 or 112 is fit to the data. The LLM model (Equation 134) is most often applied in this context, since it can be calculated efficiently using Fourier-based convolution (such as *fftlt* in Lunus[84]). Depending on the form of the correlation function, isotropic internal motions (Equation 138) or anisotropic lattice-coupled motions (Equation 140) can be fit to the data. Effective normal modes (Equation 145) have been proposed as an alternative to LLM for fitting more complex correlated motions.[63]

3 Solution X-ray Scattering

3.1 Allosteric Enzymes

Biology is a complex interplay of many competing processes, and thus, it is no surprise that the activities of countless enzymes are heavily regulated. A particularly important form of regulation is protein allostery. Much like a thermostat that controls the temperature of a room, allosteric enzymes are able to sense changes in the environment by coupling a signal, such as the binding of a ligand at one intramolecular site, to a change in activity at the catalytic site. The resulting change in activity may be mediated by structural changes, as we will discuss here, or they may involve changes in fluctuations, as will be discussed in Section 4.1. Allosteric enzymes play a major role in controlling the fluxes of metabolites and are thus often found at key junctions in biochemical pathways. These remarkable enzymes are also interesting from an evolutionary perspective, as certain structural motifs and regulatory strategies recur across different proteins and organisms.

In many of these enzymes, allosteric regulation involves large shifts in quaternary and tertiary structure upon changes in activity, and these gross conformational changes are perfect targets for investigations by SAXS. For example, these enzymes often exist as mixtures of species that are challenging to separate and structurally characterize. SAXS is a powerful method for quantitatively deconvoluting such dynamic equilibria and for determining changes in oligomerization state or subunit stoichiometry.[88] In this section, we will discuss three examples of allostery with particular attention given to the role that SAXS has played in their characterization. In the first, we will examine common evolutionary strategies used in the regulation of aromatic amino acid biosynthesis and degradation. We will then discuss aspartate transcarbamoylase, an important protein in the pyrimidine biosynthesis pathway and a classic example of protein allostery that has been particularly well characterized by SAXS. Finally, we will discuss recent insights into the complex allosteric regulation of ribonucleotide reductases, the family of enzymes responsible for synthesizing the building blocks of DNA.

3.1.1 Allosteric Regulation of Aromatic Amino Acids—The production and derivatization of aromatic amino acids (AAAs) are major biochemical processes subject to

allosteric control. In addition to being essential for protein biosynthesis in all living cells, the AAAs, L-tryptophan (Trp), L-phenylalanine (Phe), and L-tyrosine (Tyr), are precursors for various natural products in plants and microorganisms as well as neurotransmitters in animals.[89,90] The pathways responsible for AAA metabolism involve many enzymatic steps and multiple branching points, making regulation essential for maintaining the correct intracellular levels of each AAA. To that end, a variety of allosteric mechanisms have evolved in different organisms to control these pathways. Despite incredible diversity in the details of this regulation, however, certain strategies recur throughout these pathways. Here, we will discuss SAXS studies of allosteric enzymes involved in the biosynthesis of AAAs in plants and microorganisms as well as those involved in the digestion of AAAs in animals. Particular attention is given to a common regulatory motif known as the ACT domain.

The ACT domain was first observed in *E. coli* 3-phosphoglycerate synthase (PGDH) as a ligand-binding domain with a βαββαβ topology that is responsible for allosteric inhibition by the amino acid serine.[91] A bioinformatics study based on iterative PSI_BLAST searches later revealed the prevalence of this motif in amino acid metabolism, leading to the classification of this family as the “ACT domain” after three of the members: aspartate kinase, chorismate mutase, and TyrA or prephenate dehydrogenase,[92] where the latter two enzymes are directly involved in AAA biosynthesis. ACT domains and other ACT-like domains have since been found in a host of other enzymes often linked to the N- or C-terminus of a separate catalytic domain.[93] Despite their generally low sequence identity, ACT domains are identifiable by topology and most importantly, their tendency to self-associate upon ligand binding. This interaction between ACT domains often leads to major structural rearrangements and is therefore key to eliciting allosteric responses.[94] In recent years, SAXS has emerged as a particularly useful tool for characterizing these conformational changes.

3.1.1.1 Biosynthesis of Aromatic Amino Acids: The biosynthesis of the AAAs (Phe, Tyr, and Trp) begins with the shikimate pathway. Used by bacteria, fungi, and plants, this pathway is responsible for the production of chorismate (Figure 2, blue pathway), the common precursor for the AAAs and other aromatic compounds. Because this pathway is not found in animals, it is a potential target for the development of antimicrobial agents and herbicides.[95,96] The first committed step of the shikimate pathway is the condensation of two primary metabolites from glycolysis and the pentose phosphate pathway, phosphoenolpyruvate (PEP) and D-erythrose 4-phosphate (E4P), to form 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (DAHP). The enzyme responsible for catalyzing this reaction and thereby directing the carbon flux into the shikimate pathway is DAHP synthase. Once chorismate is produced as the final product of the shikimate pathway, the enzymes anthranilate synthase (AS) and chorismate mutase (CM) respectively act as branching points for the Trp and Phe/Tyr biosynthetic pathways (Figure 2, orange and green pathways). The downstream dehydratase and dehydrogenase enzymes then act as branching points for Phe and Tyr biosynthesis (Figure 2, orange pathway). Many of these branch-point enzymes are under allosteric control by the AAAs (Figure 2, black dotted lines).

Currently, the AAA biosynthetic enzyme that is best characterized by SAXS is DAHP synthase. DAHP synthases are often tightly regulated as they serve as entry points into the

shikimate pathway from primary metabolism, where the main mechanism for the control of carbon flow is feedback inhibition of the enzyme by the three end-product AAAs.[97] All DAHP synthases have a conserved $(\beta/\alpha)^8$ barrel core structure, which on its own is unregulated. Depending on the host organism and isoform, the enzyme can have additional structural elements that allow members of this family of enzymes to employ various regulatory mechanisms.

Most notably, Parker and colleagues have investigated DAHP synthases from various species by combining X-ray crystallography with SAXS data generated by in-line size exclusion chromatography (SEC). SEC-SAXS allows separation of components in solution by size prior to exposure to X-rays. In this way, the common problem of aggregation prevalent in many protein solutions is obviated, as high order species should elute earlier than the species of interest. Moreover, SEC-SAXS allows for a more robust background subtraction, as buffer exchange occurs directly on the column. As will be discussed in Section 3.1.1.2, SEC-SAXS can also generate data that are amenable to mathematical deconvolution methods even in cases where multiple species cannot be resolved chromatographically. In the studies by Parker and colleagues discussed here, SEC-SAXS is used to generate high quality data that is then directly compared with theoretical scattering from crystal structures.

In one study, SEC-SAXS was utilized to investigate the type Iα DAHP synthase from *Neisseria meningitides*. Type Iα DAHP synthases contain the core structure common to all DAHP synthases and additionally have an N-terminal extension consisting of a β strand and two α-helices as well as an insertion of a two-stranded β-sheet.[98] In *Escherichia coli*, there are three isoforms, each of which is sensitive to one of the three AAA end products. By contrast, the *N. meningitides* genome encodes for a single type Iα DAHP synthase, which is sensitive to all three AAAs, with Phe having the most potent inhibitory effect.[99] The identity of a residue in the inhibitor-binding site plays an important role in determining the inhibitor selectivity. In the *N. meningitides* enzyme, mutating this residue, Ser213, to Gly (S213G) effectively changes the inhibitor selectivity from Phe to Tyr. Interestingly, the SAXS profiles were nearly superimposable for both the wild-type protein and the S213G mutant in the absence and presence of inhibitors, indicating that inhibition is not accompanied by conformational changes.[99] The mechanism of allostery was thus proposed to involve a change in protein dynamics rather than structure.

By contrast, large conformational changes were observed in the type Iβ DAHP synthase from *Thermotoga maritima* (*Tma*DAHP synthase). *Tma*DAHP synthase is predominantly inhibited by Tyr and contains an N-terminal ACT domain flexibly linked to the catalytic core. Crystal structures from multiple groups establish the role of the ACT domain in both stabilizing and regulating the enzyme.[100,101] A crystal structure of an “open” homotetramer was obtained in the absence of Tyr, where the ACT domains are not interacting and the entrances to catalytic sites are accessible[100] (Figure 3A, inset). Removal of the ACT domains generated a dimeric form that was active yet completely unregulated by Tyr.[101] Consistent with this result, co-crystallization of the full-length *Tma*DAHP synthase with Tyr led to a “closed” structure, in which two molecules of Tyr are bound at the interface of two ACT domains interacting like clasped hands, with the β-sheets arranged face-to-face (Figure 3B, inset). Allosteric inhibition by Tyr thus entails an overall conformational change in

which a pair of ACT domains dimerize across each side of the tetrameric assembly, obstructing entrance to the catalytic sites.[101] The binding of ligands at the interface of multiple ACT domains is a hallmark of this domain's unique form of allosteric control.[94,101] The formation of such a ligand-bound state requires only two opposing ACT domains that are flexibly linked to the enzyme, and thus this regulatory element is optimal for modular addition to a variety of proteins.

Using SEC-SAXS,[101] Cross *et al.* further demonstrated that the crystallographically observed structural changes also occur in solution. First using CRYSOL,[2] they showed that the experimental profile of *Tma*DAHP synthase obtained in the absence of Tyr (Figure 3A, black) was in remarkable agreement with the theoretical profile of the open structure (Figure 3A, blue), whereas the theoretical profile of the closed structure showed a poor fit (Figure 3A, orange). Conversely, the closed structure showed a much better fit to the experimental profile obtained in the presence of Tyr than the open structure (Figure 3B). Although the single-component fits to the experimental profiles were convincing, Cross *et al.* further estimated the fraction of each population in solution by fitting linear combinations of the theoretical profiles of the open and closed structures to the experimental profiles. Using the program OLIGOMER,[102] they found that the scattering obtained in the absence of Tyr was consistent with ~90% of the protein adopting the open form and ~10% adopting the closed form. Upon addition of Tyr, the scattering becomes consistent with ~80% of the protein in the closed, inactive form. Although the usage of multi-component fits should be justified with additional evidence for the number of species in solution, such a quantitative analysis exemplifies one of the advantages of SAXS over other structural methods. This study also demonstrated that SAXS can differentiate between the open and closed states of this enzyme, motivating further studies that would be difficult with crystallography alone.

Parker and colleagues later adapted their SAXS-based approach to a number of *Tma*DAHP synthase constructs of interest. In one study, several mutants of *Tma*DAHP synthase were biochemically assayed, and their structural responses to the inhibitors Tyr and Phe were characterized via SEC-SAXS and OLIGOMER analysis.[102] In this way, Cross and Parker were able to differentiate mutations that weakened inhibition by favoring the open state versus those that altered the binding affinity of inhibitors.[103] In a particularly interesting study, Cross *et al.* generated a chimeric DAHP synthase by combining the catalytic domain of the normally unregulated *Pyrococcus furiosus* enzyme with the ACT domain of the *T. maritima* enzyme (Figure 4, inset). Not only was this chimeric protein inhibited by Tyr like the wild-type *Tma*DAHP synthase, but a similar structural response was observed by crystallography and SAXS in the presence of inhibitor (Figure 4). Thus, it was concluded that without any altering of the catalytic domain, transfer of an ACT domain was sufficient to induce allosteric control on a normally unregulated protein. This finding provides further evidence for the modularity and adaptability of the ACT domain.[104]

Like DAHP synthase, other enzymes in AAA biosynthesis are tightly regulated (Figure 2, dotted) and worth investigation via SAXS. Some of these have already been studied via SAXS.[105,106] However, for many of these enzymes, these investigations have only just begun, and further exploration with more advanced SAXS techniques is likely to be fruitful in the coming years. A particularly promising area is in the investigation of protein-protein

interactions in the pathway, such as those of DAHP synthase and chorismate mutase (CM). In a recent example, complex interplay was reported between CM and DAHP synthase of *Geobacillus sp.* In this species, DAHP synthase exists as a fusion protein with an N-terminal CM domain and is allosterically inhibited by prephenate, the product of CM.[107] SAXS studies revealed that a tighter association between the dimeric CM domains and the tetrameric DAHP synthase domains forms upon binding of prephenate, thus disrupting DAHP synthase function by occluding the active site.[107] Namely, Guinier analysis showed a subtle decrease in the radius of gyration R_g from 36.5 ± 0.4 Å to 34.8 ± 0.6 Å in the presence of prephenate. Evidence for compaction was also seen in the Kratky plots (Figure 5, blue to orange). Here, upon binding of prephenate, the main peak shifts to the right, indicative of a decrease in molecule size.

Another notable example of unique protein-protein interactions in AAA biosynthesis is the interaction between type II DAHP synthase of *Mycobacterium tuberculosis* (*Mtu*DAHP synthase) and the cytoplasmic chorismate mutase (*Mtu*CM).[108–110] *Mtu*DAHP synthase is unusual in that it is inhibited synergistically by combinations of AAAs, yet no structural changes are observed by SAXS, suggesting that inhibition of this enzyme may involve a change in dynamics.[108,109] Furthermore, *Mtu*DAHP synthase acts as an allosteric activator for *Mtu*CM, which on its own is a poor catalyst. To determine if inhibition of *Mtu*DAHP synthase can alter its interaction with *Mtu*CM, Munack *et al.* examined the enzyme complex in the presence of AAAs using multiple techniques, including SAXS and SEC-coupled multi-angle light scattering (SEC-MALS). Interestingly, addition of Phe and Tyr led to the dissociation of *Mtu*CM from *Mtu*DAHP synthase and subsequent deactivation of both enzymes. [109] This result makes particular sense when one considers that chorismate mutase is the enzyme that acts on the end product of the shikimate pathway and is located at the branch point between the biosynthesis of Trp and Tyr/Phe (Figure 2). Notably, the dissociation of the complex was observed via SEC-MALS but not SAXS, which was performed on an equimolar 20 μM sample of *Mtu*DAHP synthase and *Mtu*CM. Specifically, the scattering from this sample closely resembled the theoretical scattering calculated from the crystal structure of the heterooctameric complex. The authors attributed the apparent lack of dissociation in the SAXS sample to the high protein concentration, which favors complexation even after addition of inhibitors. However, it is important to note that *Mtu*CM is only 20 kDa, an order of magnitude smaller than the 207 kDa *Mtu*DAHP synthase, and thus detection of dissociation is challenging with one experiment performed at a single protein concentration. Indeed, SEC-SAXS can be helpful in similar situations by enabling study at lower concentrations or by separating complexes from individual subunits.

3.1.1.2 Aromatic Amino Acid Hydroxylases: The three AAAs serve as important precursors to neurotransmitters in animals. However, because animals lack the AAA biosynthetic enzymes of the shikimate pathway, Phe and Trp must be obtained from diet, while Tyr can be obtained as a byproduct of Phe catabolism.[90,111] A family of pterin-dependent hydroxylase enzymes controls the levels of the three AAAs by directing them towards the synthesis of neurotransmitters, such as dopamine and seratonin (Figure 6). Interestingly, all eukaryotic enzymes from this family contain ACT domains. In particular, a regulatory ACT domain is found in mammalian phenylalanine hydroxylase (PheH), a non-

heme Fe-dependent enzyme that uses tetrahydrobiopterin to convert Phe to Tyr. The ACT domain has long been known to be required for allosteric activation of the enzyme by excess levels of Phe.[112] Proper regulation of PheH activity is essential for maintaining safe levels of Phe in the blood, and mutations in the gene encoding PheH cause the genetic disorder phenylketonuria. Because of its medical importance, PheH has been studied extensively using the tools of structural biology.[113] Yet full-length crystal structures have been notoriously difficult to obtain, likely because of the protein's flexible multi-domain architecture. Thus, SAXS has played a critical role in revealing the mechanism of allosteric activation.

Recently, Arturo *et al.* succeeded in determining the first crystal structure of the full-length rat PheH, revealing a homotetramer in an open but inactive state[114] (Figure 7). SAXS was then used to probe for any conformational changes that accompany allosteric activation. In the absence of Phe, the SAXS profile for rat PheH resembles the theoretical profile based the crystal structure, where missing residues were modeled in CORAL.[28] At saturating Phe concentration, however, the scattering changed significantly,[114] providing evidence for a major conformational change upon activation. This change in shape corresponded to a redistribution of major distances in the pair distance distribution function, whereas the maximum dimension of the enzyme remained unchanged.

In the crystal structure of the inactive PheH, the ACT domains are far apart from one another, and no obvious binding sites for Phe are observed[114] (Figure 7). A similar crystal structure of the full-length PheH was obtained by Meisburger *et al.*[115] Based on structures of other enzymes with ACT domains, it was proposed that allosteric activation would involve dimerization of these domains.[116] Notably, several key residues are conserved between the ACT domains of PheH and prephenate dehydratase (PDT), an enzyme from the shikimate pathway (Figure 2).[117] In PDT, Phe acts as an allosteric inhibitor, and the ACT domains form a side-by-side dimer with two molecules of Phe bound at the interface.[118] The same arrangement was recently observed for the ACT-domain fragment of PheH bound with Phe by Zhang *et al.* using NMR[119] and subsequently by Patel *et al.* using crystallography.[120] In the latter study, SAXS experiments were also performed to determine the effects of Phe. Although the SAXS data show that the truncated construct is prone to heavy aggregation, it was concluded that the dimeric form seen in the crystal structure is preferentially stabilized by Phe,[120] in accord with prior NMR and analytical ultracentrifugation studies.[119,121,122]

The question of allosteric activation in full-length rat PheH was approached by Meisburger *et al.* with advanced SAXS methods.[115] Two types of SAXS experiments were performed. First, a Phe titration established that changes in the SAXS profile occurred in concert with the known activation thresholds for the enzyme (Figure 8A). Here, singular value decomposition (SVD) showed that activation was consistent with a cooperative, two-state transition (Figure 8B). Importantly, a similar cooperative transition was observed even when an active-site mutation (R270K) was introduced to prevent Phe-binding in the active site. Together with other biophysical data, these SAXS titration experiments demonstrated that allosteric activation involves cooperative Phe binding to a site other than the active site. Secondly, SEC-SAXS was performed in place of the basic flow-cell setup, both in the

absence of Phe and with Phe in saturating amounts. In both cases, SEC alone was not sufficient to separate the tetramer from aggregates.

To obtain the scattering profile of tetrameric PheH, Meisburger *et al.* adapted an SVD-based technique previously used in the field of chemometrics to separate overlapping absorption spectra (Figure 9A–B).[123] In general, separation by SVD is possible because the SAXS profiles of components in solution add linearly, and the combined dataset can be treated as a matrix that can then be decomposed into scattering components that co-vary. This decomposition, however, is not unique, and the resulting scattering profiles may be linearly combined to yield an equally valid solution. Fortunately, chromatography data has the property that components elute sequentially, and in favorable cases, this is sufficient to determine the components uniquely. Taking advantage of this property, Meisburger *et al.* used Evolving Factor Analysis (EFA) to confirm that the broad peak obtained from SEC-SAXS involved the sequential elution of two, overlapping peaks, which could then be separated computationally.

With the aggregation-free SAXS profiles determined by EFA, structural modeling was performed in SASREF[46] using the maximum number of spherical harmonics in order to accurately model the wide-angle region of the scattering. The EFA-derived SAXS profile obtained in the absence of Phe was well described by the predicted scattering from a structural model derived from the inactive PheH crystal structures, in which the dimer-of-dimer interface was slightly more expanded in solution. In contrast, the EFA-derived SAXS profile of the Phe-bound state instead was consistent with a model in which the ACT domains were rotated to interface with their diagonal partners in the tetramer crystal structure, as seen in structures of the isolated domains[119,120] (9C). In this conformation, the N-terminal tail of PheH that directly precedes the ACT sequence cannot occlude access to the active site as observed in the crystal structures of the inactive state.[114,115,124]

Together with crystallography,[114,115,120,124] NMR,[119] and other biophysical methods,[119,121,122,125] SAXS has provided direct evidence for a structural change accompanying allosteric activation in PheH,[114,115] while also informing our understanding of the transition between states.[115] The degree of cooperativity observed in the Phe titration experiments[115] suggested that the transition between an inactive tetramer and active structure required no dissociation, in contrast to a previously suggested model in which the tetramer must dissociate into dimers upon activation.[116] This position is supported by the expanded conformation noted in solution compared to the restricted crystal structure,[115] which would lessen the steric clash thought to be an issue for the rotation of the four ACT domains during activation.

3.1.2 Aspartate Carbamoyltransferase—Discovered over 50 years ago, *E. coli* aspartate transcarbamoylase (ATCase) – also known as aspartate carbamoyltransferase – serves as a paradigm of enzymatic characterization that has drawn heavily on SAXS techniques.[15,126] ATCase catalyzes the first step of the biosynthetic pathway of pyrimidines, the carbamylation of aspartate by carbamoyl phosphate (CP) to produce carbamoyl aspartate and inorganic phosphate (Figure 10). The enzyme first provoked interest due to its complex regulation involving both its amino acid substrate and the eventual nucleotide products of the

pyrimidine biosynthetic pathway.[127] The enzyme preferentially binds CP first and displays cooperative binding of aspartate upon initial binding of CP.[128] ATCase is further regulated by the purine nucleotide adenosine triphosphate (ATP), which acts as an allosteric activator, and by the downstream pyrimidine products, cytidine triphosphate (CTP) and uridine triphosphate (UTP), which act in tandem as allosteric inhibitors (Figure 10). Structurally, regulation of this enzyme involves a twisting motion that is readily resolved by SAXS.

The regulation of ATCase has been described in the formalism of the MWC or *concerted* mechanism put forth by Monod, Wyman, and Changeux.[129] Two major structural states of ATCase have been described, where allosteric inhibitors favor a less active "tense" or T-state, and activators favor the more active "relaxed" or R-state. In this model, ligand binding to any catalytic subunit shifts the equilibrium to favor the R-state over the T-state, and all subunits cooperatively transition between the two states with no intermediates. Numerous investigations have been made over the past several decades to elucidate the details of ATCase regulation, and SAXS has been at the forefront of this effort because of its unique suitability for probing the structural transitions of ATCase.

3.1.2.1 Major Conformations of ATCase: Seminal biophysical studies on the quaternary structure of *E. coli* ATCase showed that the enzyme is a heterododecamer consisting of two trimers of catalytic subunits and three dimers of regulatory subunits.[130–132] ATCase is predominantly in the T-state in the absence of ligands or in the presence of allosteric inhibitors, whereas binding of substrates shifts the equilibrium of ATCase to the R-state. The R-state can also be stabilized by the ligand N-phosphonacety-L-aspartate (PALA), a bisubstrate analog that mimics the reaction tetrahedral intermediate and thus acts as a potent competitive inhibitor for the natural substrates. An early SAXS study showed that the addition of PALA to ATCase leads to a subtle increase in R_g from 45.9 ± 0.5 Å to 48.4 ± 1.0 Å that is accompanied by a significant shift in the peak position of the pair distribution function,[133] indicative of a "swelling" in the overall shape of the enzyme.

Crystallographic studies by Lipscomb, Kantrowitz, and colleagues[126,134,135] have shown that the transition from the T- to R-state is indeed accompanied by an opening in structure. Overall, ATCase consists of two catalytic trimers, which are stacked such that the active sites face the interior, and three regulatory dimers, which act as posts connecting the trimers. Whereas the CTP-bound structure (representing the T-state) is relatively closed with limited access to the active sites (Figure 11A),[128,136] the PALA-bound structure (representing the R-state) is open and displays a 12° rotation of one catalytic trimer relative to the other trimer, concomitant with a rotation of each of the three regulatory dimers about their two-fold axes by 15°[137] (Figure 11B). The opening of the structure caused by this twisting motion leads to an overall elongation of about 11 Å along the molecular three-fold axis.[126]

Consistent with the dramatic subunit rearrangements observed crystallographically, the scattering profiles of ATCase undergo significant changes in shape in the mid- to high-q region with the addition of PALA.[133] However, the changes in scattering could not be easily interpreted until the development of CRYSOL by Dmitri Svergun and colleagues in the mid-1990's, which had enormous impact on the growth of SAXS as a structural biology technique.[2] Interestingly, ATCase served as one of its first test cases. Using coordinates from

a T-state structure of ATCase, Svergun *et al.* showed that the experimental scattering from ligand-free enzyme can be successfully simulated by accounting for contributions from a hydration layer.[2] The R-state was examined in a subsequent study. Unexpectedly, the positions of the subsidiary maxima and minima observed in the experimental scattering of the PALA-bound ATCase could not be matched by the simulated scattering of the crystal structure, even with the addition of a hydration layer in the calculation.[138] It was thus proposed that PALA-bound ATCase is slightly more open in solution than depicted in the crystal structure and that crystal-packing forces not present in solution prevent the heterododecamer from adopting a more open conformation. Recent explicit solvent MD simulations of ATCase are consistent with an R-state that is more open than is seen in the PALA-bound crystal structure. [139] It is for this reason that in the SAXS literature, the PALA-bound crystal structure is often referred to as the R_{cryst}-state, to distinguish from the solution R-state.

3.1.2.2 Evidence for a Concerted Mechanism: Along with biochemical studies, SAXS played an important role in demonstrating that ATCase undergoes a cooperative allosteric transition between the two states. Many advanced SAXS studies came from Vachette and colleagues. In one notable study, Fetler *et al.* performed a SAXS experiment in which ATCase was titrated with increasing concentrations of PALA, leading to a shift from the T- to R-state. By generating a set of curves at various points in the transition, several important observations could be made. First, clear iso-scattering points were visible in the data, suggestive of a two-state transition (Figure 12A). Furthermore, the transition reached saturation at substoichiometric amounts of PALA, indicative of cooperative binding (Figure 12B). Moreover, SVD analysis could be performed in a manner similar to that described in Section 3.1.1.2 for the PheH titration study by Meisburger *et al.* Using this model-independent method, it was confirmed that the titration of ATCase with PALA could be explained with only two states, implying a single, concerted transition between the T- and R-states.[140,141]

Evidence for a preexisting structural equilibrium was demonstrated in a later study by Fetler *et al.* Here, a mutant ATCase was investigated in which Asp236 was mutated to Ala (D236A) in the catalytic chains of the enzyme, weakening a catalytic-regulatory interface.[142] The overall effect of this mutation was to destabilize the T-state, and as a result, the [T]/[R] ratio was shifted to heavily favor the R-state at room temperature in the mutant, even in the absence of ligand.[143] SAXS measurements at increasing temperatures revealed the [T]/[R] ratio to be temperature dependent in a reversible manner, thereby indicating thermodynamic equilibrium between the two states.

In another approach, Kantrowitz and colleagues combined SAXS, crystallography, and advanced protein preparation methods to provide evidence for a concerted mechanism of cooperativity. The active-site residue Arg105 was mutated to Ala (R105A) to disrupt binding of substrate or the substrate analog PALA. Co-crystallization of the mutant enzyme in the presence of excess PALA led to a crystal structure of the T-state with no density for PALA, thus demonstrating that this mutation prevents binding of this ligand, which in turn prevents the transition from the T- to R-state. A hybrid enzyme was then constructed consisting of six wild-type regulatory domains, one wild-type catalytic domain, and five mutant catalytic

domains. SAXS comparison of solutions with and without PALA revealed that while an ATCase enzyme with only mutant catalytic domains remained in the T-state (Figure 13B), the hybrid enzyme with a single wild-type catalytic domain can undergo a transition to the R-state (Figure 13C).[144] Although there remains some debate over whether the same conclusion can be made for the wild-type holoenzyme,[145,146] this elegant experiment by Macol, *et al.* showed direct evidence of a concerted transition, where only a single PALA molecule was necessary to shift the [T]/[R] equilibrium heavily to the R state.

3.1.2.3 Effect of Allosteric Nucleotide Effectors: SAXS has also played a major role in probing the effects of the allosteric effectors CTP, UTP, and ATP on the T- to R-state transition of ATCase. Based on the concerted model of cooperativity, it was predicted that an allosteric inhibitor of the enzyme would act by shifting the structural equilibrium toward the T-state, while an activator would shift it toward the R-state. In their 1995 study, Fetler *et al.* performed a PALA titration experiment in the presence of the pathway product CTP at saturating concentrations.[140] Importantly, SVD analysis indicated a two-state transition and that CTP shifts the saturation point to higher [PALA], indicating that CTP acts an inhibitor. A later study by Cockrell *et al.* demonstrated the importance of Mg^{2+} as a cation for the nucleotide effectors.[147] In this work, it was shown that although CTP on its own can inhibit the enzyme, activity is restored with addition of physiological levels of Mg^{2+}. Instead, the highest degree of inhibition was observed with the combination of the pathway products CTP and UTP in the presence of Mg^{2+} (Figure 10). In support of this result, Cockrell *et al.* presented a crystal structure of the ATCase in a PALA-bound R-state conformation with CTP and UTP preferentially binding to two distinct sites in the regulatory domain, with the phosphate moieties coordinating a central Mg^{2+} ion.

In comparison, debates over the structural effects of the allosteric activator ATP have been more challenging to resolve. Again, Fetler *et al.* used SVD analysis to show that PALA titration in the presence of saturating ATP leads to a two-state transition. Surprisingly, however, no shift in the transition point was observed.[140] In a second study, Fetler and Vachette revisited the effect of ATP, this time in the presence of Mg^{2+}.[148] Interestingly, addition of saturating levels of Mg^{2+}-ATP to PALA-bound ATCase led to a subtle change in the scattering curve. Concomitant with a slight increase in the R_g, the subsidiary maxima and minima move slightly to lower q (Figure 14, orange to red). However, the scattering curve could not be explained as a linear combination of the profiles of the unliganded and PALA-bound enzyme. The subtle change in the SAXS profile was thus interpreted as Mg^{2+}-ATP causing a further expansion of the molecule along the three-fold axis than with PALA alone.[148] Essentially, Fetler *et al.* argued that ATP activates ATCase by inducing a new R′-state that is structurally distinct from the PALA-bound R-state.

The effects of nucleotide effectors were also examined by crystallography. In a study by Cockrell *et al.*, crystals of ATCase co-crystallized with PALA were soaked with nucleotides. [147] The resultant crystal structures show no major differences in quaternary structure, thus seemingly at odds with an ATP-induced structural change as proposed by Fetler *et al.* Instead, these structures show that upon binding of inhibitors, the N-terminus of each regulatory domain interacts with its own chain, whereas the binding of activators instead causes the N-termini to interact with the adjacent regulatory chain, thus forming a

more stable interface. Cockrell *et al.* thus argued that activation by ATP and Mg^{2+} and deactivation by CTP, UTP, and Mg^{2+} are a result of the respective stabilization and destabilization of the R-state. In the same study, Cockrell *et al.* considered the scattering contribution of ligands to explain the changes in the SAXS profiles that Fetler and Vachette observed with Mg^{2+}-ATP. Using their crystal structures, the theoretical scattering of the R-state was calculated in FoXS.[149] When the ATP and Mg^{2+} are removed from the calculation, Cockrell *et al.* found that the theoretical profile of the ATP-bound R-state structure is superimposable to that of the nucleotide-free R-state structure, as anticipated by the lack of major differences in their crystal structures. From this observation, it was suggested that the subtle changes in scattering observed by Fetler and Vachette[148] were solely due to the scattering of the nucleotide, rather than a relaxation of the R-state in solution. However, no experimental scattering was presented in this study.

Indeed, as shown by Svergun's seminal CRYSOL work with ATCase,[138] the solution scattering of R-state has proved to be difficult to simulate from crystal structures, likely due to a more open structure in solution.[139] Thus, a purely crystallographic argument may not take into account the added flexibility observable only via solution scattering. With the crystal structures from Cockrell *et al.*[147] and SAXS profiles from Fetler *et al.*[148] now available, this question can be revisited. Again, we find that the experimental scattering of the unliganded T-state (Figure 14, green) can be fit well with the predicted scattering of the crystal structure of ligand-free ATCase (Figure 14, gray). As noted by Cockrell *et al.*, the predicted scattering of the PALA-bound structure with Mg^{2+}-ATP (Figure 14, black dotted) is slightly greater than that of the structure with only PALA bound (Figure 14, black solid).[147] However, neither of the predicted curves describes the experimentally obtained scattering of PALA-bound ATCase in solution with (Figure 14, red) or without (Figure 14, orange) Mg^{2+}-ATP.

The debate sparked by these studies gives a pointed example of the surprisingly high-resolution information available in SAXS profiles but also reveals the difficulty that can arise in interpretation (Figure 14). Future study on this topic combining both experimental SAXS and crystallographic data may prove particularly fruitful in settling the debate. Collectively, however, these studies have already provided remarkable insight into the local and global structural changes accompanying of the allosteric transitions of ATCase.

3.1.2.4 Transition Kinetics: Although steady-state SAXS can provide insight into thermodynamically stable states, transient intermediates cannot be detected. Time-resolved SAXS can be a valuable probe of such intermediates. A number of time-resolved studies enabled by Tsuruta and colleagues have shed light into the structural kinetics of ATCase. In a set of early experiments, the T- to R-state transition was initiated by rapidly mixing ATCase with substrates[150] or substrate analogs[151] utilizing a specialized stopped-flow cell with a deadtime of 10 ms.[152] Because the transition was found to be too rapid, the reaction mixture was held at −5 to −11 °C with 30% ethylene glycol, and the scattering was sampled with 0.1–0.2 sec exposure times.[151,152] The resultant scattering profiles were described sufficiently well by linear combinations of T- and R-state scattering profiles, and thus structurally distinct intermediates were not detected.[151] Unfortunately, it was also known

that simple alcohols,[153] and by extension ethylene glycol, can inhibit ATCase by stabilizing the T-state.

With the advent of 3rd generation synchrotron sources, higher X-ray flux allowed for shorter exposure times and in turn, enabled observation of transitions previously considered too rapid. The transition of ATCase between T- and R-state was thus revisited by West *et al.*, once again utilizing a specialized stopped-flow mixer but this time without the presence of ethylene glycol and at temperatures ranging from 5 to 5 °C.[154] Using a fast CCD detector, the scattering was sampled at 19-ms time intervals. Scattering profiles were then integrated over the mid-q region sensitive to the T- to R-state transition, and rates were determined by exponential fits to the resultant time courses. These fits suggested a biphasic nature of the transition in certain cases; however, the presence of multiple rates was interpreted as consequence of a multiple ligation states. Interestingly, the ATP was shown to increase the rate of the structural transition, whereas CTP and UTP slowed the transition. Furthermore, by using the natural substrates CP and aspartate, West *et al.* showed that the allosteric transition is not the rate-limiting step of enzyme catalysis (Figure 15), thus demonstrating how time-resolved SAXS can provide information that cannot be easily obtained from biochemical assays alone.

3.1.3 Ribonucleotide Reductases—Ribonucleotide reductases (RNRs) are an essential family of enzymes that convert ribonucleotides to deoxyribonucleotides, the building blocks of DNA, via a unique, radical-based mechanism.[155] No other enzyme performs this process essential to DNA synthesis and repair, and thus RNRs are of vital importance to all forms of life. Their ubiquity in nature and requirement for cell replication make these enzymes prime targets for both antibiotic and anticancer therapies.[156–158] Because RNRs are located at a critical junction point in nucleotide metabolism (Figure 16), these enzymes are tightly regulated both transcriptionally and at the level of cofactor assembly. Moreover, the allosteric regulation of RNRs is among the most complex examples in biochemistry, enabling these enzymes to sense and maintain the correct levels of nucleotides in the cell needed for genetic stability.[159] Due to their complex mechanisms and importance in medicine, RNRs have captivated biochemists for over half a century.

Although RNRs share a common evolutionary lineage, they display diversity in quaternary structures and mechanisms of radical generation. RNRs are categorized into three major classes based on the metal-containing cofactors used to generate the catalytically essential radical species.[155] Class III RNRs utilize *S*-adenosylmethionine bound to a [4Fe-4S] cluster and function only under anaerobic conditions, whereas class II RNRs are adenosylcobalamin-dependent. The most recently evolved class I RNRs utilize a dinuclear metal cofactor in an oxygen-dependent manner and are further subcategorized by the identity of the metal.[160] In all RNRs, once the active-site radical is generated, nucleotide reduction proceeds through a conserved mechanism.[161,162] Although the catalytic mechanism is conserved, accumulating evidence shows that allosteric regulation of activity in RNRs can involve dramatic changes in quaternary structure in an organism-dependent manner.[88,163–168] SAXS has played a particularly important role in characterizing the transient active complex and the large conformational changes involved in the activity

regulation of class Ia RNRs. This review will focus on this subclass of enzymes, as they are the best characterized both biochemically and structurally.

3.1.3.1 Activity Regulation of *E. coli* RNR: The class I RNRs are particularly interesting as they require two separate proteins: an α subunit, which contains the active site, and a smaller β subunit, which harbors the metal center that generates the initial radical (Figure 17A–B). This radical is then transferred from the metallocofactor in β to a conserved cysteine in the active site in the α subunit, generating a thiyl radical that initiates reduction of the ribonucleoside diphosphate substrate (NDP, where N is any of the four bases). Turnover thus requires the interaction of the two subunits. Class I RNRs are further subclassified by metal identity. The class Ia enzymes utilize a diferric tyrosyl radical cofactor and are the predominant RNRs found in eukaryotes and certain aerobic bacteria.

Another hallmark of the class Ia RNRs is the prevalent usage of two different types of allosteric sites in the α subunit. Substrate preference is coupled to the binding of a downstream nucleotide at an allosteric site known as the specificity site (Figure 17A), thus ensuring that deoxyribonucleotide pools are balanced. For example, in the class Ia RNR from *E. coli*, binding of the downstream product dTTP leads to preferred binding of the substrate GDP and disfavors the binding of a different substrate UDP, which is converted to dTTP later in its metabolic pathway (Figure 16). In many class Ia RNRs, the α subunit additionally contains at least one copy of an N-terminal “ATP cone” motif, which can house a second allosteric site, known as the activity site[174] (Figure 17A, orange domains). Binding of ATP to this site increases the rate of reduction, whereas its deoxy form (dATP) acts as an inhibitor at this site. Together, the specificity and activity sites allow RNRs to sense the intracellular balance of ribonucleotides to deoxyribonucleotides and maintain the proper levels of DNA precursors.[159]

While no single RNR is truly representative of the diverse RNR family, the class Ia RNR from *E. coli* was one of the first discovered[175] and is the best studied. Many biophysical studies have focused on elucidating the active and inhibited structures of this enzyme. Early analytical ultracentrifugation (AUC) studies by Thelander *et al.* first revealed the existence of two subunits, the aforementioned α and β, and revealed that these subunits were dimeric in solution.[176] Thelander *et al.* further found that in the presence of the specificity effector dTTP, the dimers came together in 1:1 stoichiometry. These early AUC experiments were also first to show that saturating amounts of the allosteric inhibitor dATP led to higher order oligomers.[176,177] However, the exact nature of these inactive oligomers remained unknown for several decades. Furthermore, while these early studies implied the formation of an active $\alpha_2\beta_2$ complex, the RNR subunits proved challenging to capture together by crystallography. Originally thought to be the result of a weak subunit interaction (0.4 μM),[178] it is now known to be a consequence of the highly dynamic nature of the enzyme, in which multiple quaternary structures are interconverting.[88,167,168,172] In contrast to crystallography, however, SAXS is particularly suited to the study of transient complexes in solution. Together with other biophysical techniques, SAXS studies have helped determine the mechanisms of activation and inhibition of the class Ia *E. coli* RNR.

In a recent study, Ando *et al.* first used AUC to demonstrate that *E. coli* RNR exists as a mixture of multiple oligomerization states under activating conditions.[88] Addition of dATP, by contrast, led to the stabilization of a single state corresponding to a large oligomer with a mass roughly twice that of an $\alpha_2\beta_2$ complex. To structurally characterize this large oligomer, two types of titration experiments were performed utilizing SAXS. First, the allosteric inhibitor dATP was titrated to an equimolar mixture of the subunits under conditions where RNR is active. The SAXS profiles showed a significant change in shape as well as an increase in the forward scattering intensity $I(0)$. SVD analysis and mass estimation by $I(0)$ indicated that the structural change accompanying allosteric inhibition corresponds to a two-state process in which the oligomer doubles in mass. The corresponding Kratky plots from this experiment show a single dominant peak indicative of a compact structure change into a "double-humped" curve at saturating dATP that is characteristic of a highly non-globular species (Figure 18A).

Ando *et al.* complemented this experiment with a subunit titration done in the presence of saturating concentrations of dATP. In this experiment, R_g was monitored at increasing concentrations of α, and the R_g was found to reach a maximum at equimolar amounts of the subunits (Figure 18B), supporting the formation of a 1:1 complex. The titration also revealed that the subunits on their own remained dimeric over a wide range of protein concentrations and that the inactive complex required both dimeric subunits to form. Together with single-particle electron microscopy (EM) data and a low-resolution crystal structure of the inhibited complex, it was determined that dATP induces the formation of a ring-shaped $\alpha_4\beta_4$ complex.[88] In this structure, each N-terminal ATP-cone domain of an α_2 forms an interface with a lobe of a β_2 (Figure 17D). In this structure, the β_2 subunit is displaced away from the active site of the α_2 subunit, thus inhibiting the enzyme by preventing radical transfer. In this study, SAXS was utilized to probe physiologically relevant concentrations of protein and dATP, providing convincing evidence that the $\alpha_4\beta_4$ ring is not just an artifact of crystallization or high concentrations of protein.[88]

The two-state transition observed in the dATP titration was consistent with an $\alpha_2\beta_2$ complex interconverting with the $\alpha_4\beta_4$ ring. By performing a two-component fit to this dataset using different models of the $\alpha_2\beta_2$ complex, Ando *et al.* showed that the active form of *E. coli* RNR is similar to a previously proposed model in which the two subunits are docked along their symmetry axes[171,179] (Figure 17C). In this so-called "docking model" conformation,[171] the active sites of the α_2 dimer are brought in close proximity to the lobes of the β_2 dimer, minimizing the radical transfer distance between the metallocofactor and the active-site cysteine. In a later study, an unnatural NH_2Y_{730} was incorporated in the α subunit to trap the enzyme during turnover by stabilizing the radical at this residue rather than passing it along the radical transfer pathway.[172] This trapped mutant was found via a combination of SAXS and single-particle EM to form the $\alpha_2\beta_2$ previously observed in the interconverting mixture of wild-type RNR (Figure 19A). Based on activity, this compact tetramer is thought to be the active form of *E. coli* RNR that had proven elusive to crystallization. A time-dependent SAXS experiment further showed that this trapped mutant maintains the active $\alpha_2\beta_2$ form even under strongly inhibiting conditions where the $\alpha_4\beta_4$ ring is favored in the wild-type enzyme (Figure 19B). This result thus led to the remarkable

conclusion that *E. coli* RNR can sense the presence of an on-pathway radical and remain locked in the compact $\alpha_2\beta_2$ conformation during turnover.

Further study of *E. coli* RNR revealed an even more intriguing structure: concatenated $\alpha_4\beta_4$ rings. Co-crystallization of this enzyme with either the allosteric inhibitor, dATP, or with the mechanism-based inhibitor, gemcitabine (F_2CDP), both led to lattices with an octamer in a single asymmetric unit. Visualization of two asymmetric units shows that the two octamers are in fact two $\alpha_4\beta_4$ rings interlocked, like linked chains (Figure 17E). SAXS experiments were once again performed to explain this result. Importantly, SAXS showed that at physiologically relevant protein concentrations, F_2CDP-inhibited RNR is largely $\alpha_2\beta_2$ and that the formation of $\alpha_4\beta_4$ ring was due to the higher protein concentrations needed for crystallization. This result is consistent with that the fact turnover, and hence, an active conformation, is required for mechanism-based inhibition. Furthermore, SAXS was used to show that concatenation of the $\alpha_4\beta_4$ rings increased with titration of the crystallization precipitant. The crystallization precipitant contained 25% polyethylene glycol, a known crowding agent that can contribute to closer packing of the complexes. While it was still possible that cellular conditions lead to similar crowding, the SAXS titration experiment pointed to the interlocked structure being a result of crystallization conditions.[167]

3.1.3.2 Activity Regulation of Human RNR: Considering the structural homology of the subunits, one would expect all class Ia RNRs to follow a similar mechanism of dATP-induced inhibition. Interestingly, however, studies of yeast, murine, and human RNRs all suggest that eukaryotic class Ia RNRs utilize a different mechanism than that used by *E. coli* RNR.[163–165,180] In a recent study, a low-resolution crystal structure revealed that human RNR forms an α_6 ring in the presence of dATP with the active site lining the inner hole (Figure 17F). As in *E. coli* RNR, the subunit interfaces are formed by the ATP-cone domains. SAXS titration experiments confirmed that this α_6 ring forms in solution in the presence of dATP. In this state, radical transfer cannot easily occur as the ring is too small to accommodate a β_2 subunit. Curiously, however, SAXS revealed that a similar α_6 also forms in the presence of the allosteric activator ATP. While structurally similar to the dATP-induced α_6, SAXS and EM showed that the ATP-induced α_6 can interconvert with a higher-order filament-like structure at high ATP concentrations. Furthermore, upon titration of β_2 into the reaction mixture containing ATP, the scattering features characteristic of a ring, such as the subsidiary maximum at $q \sim 0.07$ Å, were lost. By contrast, the scattering features of the ring were retained when β_2 was titrated into the reaction mixture containing dATP. Thus, SAXS titration experiments revealed that while human α forms a hexameric ring with either inhibitor or activator, the hexamer formed in the presence of dATP is more stable and resists conversion to active structures.[168]

3.1.3.3 Multiple ATP-cone domains: In the two different inhibition mechanisms illuminated above, the N-terminal ATP-cone domain is responsible for binding of both ATP and dATP as well as initiating the conformational change that leads to inhibited quaternary structures. This domain, commonly characterized by four conserved residues in the N-terminal strands and first helix that form favorable contacts with the nucleotide effector, is an evolutionarily mobile regulatory domain found in all three classes of RNR.[174,181]

However, mounting evidence points to repeated gain and loss of the ATP cone throughout the evolutionary history of RNR, leading to the apparent variety of allosteric mechanisms observed today. A recent structure of the class I RNR from *Pseudomonas aeruginosa* in the presence of inhibiting dATP revealed an α_4 ring composed of two symmetry-related dimers[173] (Figure 17G). As in the α_6 ring seen in human RNR, this ring occludes access of the β subunit to the active site. To determine whether this unusual ring forms in solution, SAXS was performed on the α subunit in the presence and absence of the nucleotides dATP and dGTP. The scattering profiles obtained with dATP or dGTP differed significantly from that obtained with no nucleotides present, where the dATP-induced scattering aligned closely to the curve predicted for the crystal structure of the α_4. What is most notable about this structure is that the *P. aeruginosa* RNR has two ATP-cones rather than one. Interestingly, the crystal structure reveals that one cone domain binds two dATP molecules while the other is nonfunctional. The non-functional domain acts solely as a spacer between the catalytic core and the functional N-terminal ATP cone, enabling the dATP-bound domains to reach across the tetramer interface and form a compact α_4 ring structure. Removal of the functional ATP-cone leads to a scattering profile that is more consistent with an α_2. The non-functional domain is thought to have lost function due to evolutionary atrophy under conditions in which regulation by dATP was unnecessary. It is further thought that when regulation again became beneficial later in the enzyme's history, a new ATP cone was incorporated rather than reversion of the original one to a functional state.[173]

In all class I RNRs examined thus far by SAXS, inhibition by dATP leads to a quaternary structure in which radical transfer is sterically prevented. As class I RNRs can have up to three ATP-cone domains,[174] it is likely that further study of activity regulation in these enzymes will involve quaternary structures not yet seen. In a notable example, strong inhibition by dATP has been reported for the class Ib RNR from *Bacillus subtilis*, which lacks the ATP-cone motif.[182] How dATP may alter the structure of this enzyme is thus of particular interest. SAXS will undoubtedly play a major role in these investigations.

3.1.4 Conclusions—In this section, the use of SAXS as a primary and complementary means toward the study of allosteric enzymes has been examined in detail. In all three case studies explored, complex intra- or inter-enzyme interactions occur on length scales readily observed via SAXS, and thus solution scattering techniques are ideal for biophysical characterization. The results of these studies have considerably broadened our understanding of allosteric mechanisms and informed the design of future biophysical studies. However, we have also noted the difficulties that can arise in interpretation of SAXS studies of complex enzyme systems, whether due to problems of aggregation or discrepancies between crystallographic and solution data. While interpretation will continue to require attention to detail and awareness of potential pitfalls, the case studies discussed in this section exemplify the power of the technique for deconvoluting the intricate allosteric mechanisms that have evolved in nature. In the next section, we will turn to SAXS-based investigations of another type of enzyme system: large, multi-modular proteins.

3.2 Flexible Multi-Modular Enzymes

Perhaps one of the most potent uses of SAXS is in investigating the architecture and motions of multi-modular proteins. Of particular interest recently are large, multifunctional enzymes, such as polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) that produce secondary metabolites in an assembly-line fashion.[183–185] Secondary metabolites are molecules that are not essential to the normal functions of an organism. Instead, they often serve to further the organism's competitiveness in its environment, often by interacting with other organisms. Because of their unique roles in nature, these complex molecules have served as an important source of antibiotics such as penicillin, erythromycin, and vancomycin, among others.[183] Polyketides and non-ribosomal peptides are particularly fascinating as they are constructed one monomer unit at a time, in a manner reminiscent to the way in which proteins are made linearly from amino acids by the ribosome. In addition to polymerization, the biosynthesis of these complex secondary metabolites involves extensive modifications of the monomer units, including racemization, formylation, cyclization and many others. Importantly, unlike the biosynthesis of polypeptides, the "blueprint" for the product comes not from the genetic code but instead from the architecture of the biosynthetic enzymes. PKSs and NRPSs are thus extraordinarily large,[185] consisting of multiple modules, each responsible for the installation and modification of one monomer unit of the final product. The modules themselves contain multiple functional domains, and hence, major domain motions are expected in order to shuttle the covalently tethered substrate from one active site to the next.

Although these megasynthases have great potential for bioengineering applications, their size, complexity, and inherent flexibility have made them difficult to study by traditional structural methods. Thus far, crystallographic investigations of these enzymes have largely focused on gaining mechanistic insight into the reactions catalyzed by individual domains. By comparison, visualization of multiple domains and modules working in concert has been very challenging. To date, only a handful of high-resolution structures of intact modules have been solved via crystallography or cryo-EM,[186–188] and none have been solved for any megasynthase in its entirety. With crystal structures of many of the prototypical domains now available, the time is ripe for the investigation of multiple domains and modules working together. Here, we will discuss recent advances in the use of SAXS to examine PKSs and NRPSs with particular attention given to investigations of full module architectures.

3.2.1 Polyketide Synthases—Found in bacteria, fungi, and plants, polyketide synthases (PKSs) are responsible for the biosynthesis of small molecules derived from acyl-coenzyme A (CoA) precursors, such as the well-known antibiotics, erythromycin A and tetracycline. PKSs are broadly classified into three types based on domain organization and usage.[189] Here, we will focus on the type I PKSs for which much of the structural work has been done thus far. The type I enzymes are large multifunctional enzymes, which are organized into individual operational units called modules, each responsible for the incorporation of an acyl monomer unit in the final product. The polyketide products are assembled in an analogous way to fatty acids, whose molecular assembly machinery, fatty acid synthases (FASs), are thought to be evolutionarily related to PKSs.[190] For type I PKSs, a basic elongation module

contains three distinct domains: a ketosynthase (KS) domain, an acyltransferase (AT) domain, and an acyl-carrier protein (ACP), also called a thiolation (T) domain. The apo-ACP is post-translationally modified to form the holo-ACP with a 4′-phosphopantetheine (Ppant) "arm," which tethers the growing proto-product via a thioester bond. Additionally, many PKS modules contain tailoring domains, which are not necessary for the extension of the polyketide chain but instead introduce important structural modifications to the product.[183,190,191]

The assembly-line strategy of PKSs can be generalized into three steps: initiation, elongation and termination. During initiation, the first acyl unit is covalently attached onto the first ACP domain typically by the initiation module's AT domain. This initiated chain can then be transferred to the first elongation module. Type I elongation modules contain the same basic KS-AT-ACP architecture, with the elongation process proceeding in a relatively straightforward fashion (Figure 20). First, the AT domain of the elongation module binds a new acyl unit and acts to catalyze its transfer to the ACP domain. Next, the KS domain of that module accepts the growing chain from the previous module and subsequently catalyzes a carbon-carbon bond-forming reaction, resulting in the addition of the next acyl unit. The growing chain can then be transferred to the next elongation module or be released by a thioesterase (TE) domain, where chain termination can involve cyclization.[190]

While the individual domains of PKSs can now be readily classified by bioinformatics methods and appear to be relatively conserved, how the architecture of the domains within PKS modules yields the incredible diversity of their products remains poorly understood. [183,190] In particular, prediction of the end product based on structure and bioinformatics has been largely unsuccessful. In recent years, SAXS has emerged as a useful complementary tool to high-resolution structural data in many studies of these enzymes.

3.2.1.1 Erythromycin Biosynthesis: 6-Deoxyerythronolide B synthase (DEBS) has served as the prototypical PKS since the genes encoding this enzyme were first characterized in the early 1990's.[190,191] DEBS is responsible for the biosynthesis of 6-deoxyerythronolide B, a macrocycle, which acts as the primary precursor to erythromycin A, one of the most widely deployed and well-known antibiotics. The DEBS pathway involves the incorporation of a propionyl-CoA initiation unit and 6 methylmalonyl-CoA elongation units into the cyclic end product, using 6 equivalents of an NADPH cofactor (Figure 21).[191,192] DEBS is organized as a 2 MDa $\alpha_2\beta_2\gamma_2$ homodimeric assembly of three multi-modular proteins, DEBS1, DEBS2, and DEBS3 (Figure 21A). Previous studies on DEBS3 indicated a high degree of flexibility between modules, as well as additional flexibility between the individual domains.[193–195] While crystal structures representative of every component of DEBS had been solved,[196–200] their relative orientations were not known. In a major undertaking, Khosla and colleagues used crystallographic data and SAXS to build rigid-body models depicting the relative domain orientations within a minimal PKS construct as well as the relative orientations of modules within DEBS3.[192]

In the aforementioned study, Edwards *et al.* first set out to determine whether known crystal structures of DEBS fragments were consistent with their solution scattering. Four different fragments were examined by SEC-SAXS, and several hundred frames were collected during

each SEC-SAXS run. To determine sample homogeneity, the R_g values determined from Guinier analysis and the peak position of the Kratky plots were plotted as a function of frame number. Solution samples of the TE domain, the KR domain from module 1 (KR1), and the holo-ACP domain from module 3 (ACP3) were shown to be monodisperse by this method. Furthermore, the resulting scattering profiles showed excellent agreement with the theoretical scattering from high resolution structures (Figure 22). Because there is no crystal structure of ACP3, the SAXS data for this domain were fit to the crystal structure of the ACP domain from module 2 lacking the Ppant arm (apo-ACP2). In these fits, terminal regions missing from the crystal structures were modeled in CORAL.[28] Importantly, the TE domain was shown to form a homodimer (Figure 22, red curve), and the KS-AT didomain from module 3 (KSAT3) eluted as a mixture, with the higher oligomer species resembling the dimeric crystal structure (Figure 22, green curve). These domains were later shown to form the homodimer interface of the DEBS3 module. Together, these results indicated that the solution states of these DEBS fragments are well represented by existing high-resolution structures, thus justifying their use as rigid bodies in modelling DEBS modules.

One of the challenges of modeling flexible proteins is that they are likely to sample multiple conformations in solution. In certain cases, rigid-body modeling may not be appropriate, and instead, ensemble modeling may be a better approach. A number of criteria have been proposed to investigate the overall flexibility of a protein from SAXS data.[201] Edwards *et al.* concluded from several of these criteria, namely baseline convergence of the Kratky high-q region, a q^{-4} dependence in the mid-q region, and an R_g value much less than would be expected from an unfolded protein of similar size, that both the minimal PKS construct and DEBS3 were largely conformationally constrained and could be reasonably approximated by rigid-body modeling.

Following this analysis, overall architecture determination from SEC-SAXS data was performed on a catalytically competent, minimal PKS construct consisting of module 3 from DEBS2 linked to a C-terminal TE domain (Figure 21B, middle).[202] Multiple rounds of rigid-body modeling in CORAL were performed with two-fold symmetry imposed and then clustered in DAMCLUST.[28] Two major clusters with similar architectures emerged from this analysis that were highly consistent with the SAXS data. A representative structure from the best-fitting cluster is shown in Figure 23A and consists of a ring-like structure, where the TE, KR and ACP domains lie on top of the KS-AT didomain. To support their modeling, Edwards *et al.* also repeated this analysis for fragments of the minimal PKS, namely module 3 on its own (Figure 23B) and a module 3 construct lacking the ACP domain. The domain arrangements in the resultant rigid-body models showed remarkable agreement with that of the minimal PKS construct (Figure 23A–B).

Likewise, SEC-SAXS data were obtained on the homodimeric DEBS3 and rigid-body modeling was performed with two-fold symmetry imposed. Components for modules 5 and 6 and the TE dimer were oriented collinearly in the starting model based on the elongated appearance of an *ab initio* shape reconstruction performed in DAMMIN. Two major clusters of rigid-body models emerged from this analysis, where the individual domain arrangements within each module (Figure 24) were similar to those observed for the single-module minimal PKS construct (Figure 23A). The primary difference between the two clusters was

the angle between the two modules, suggestive of flexibility in this region. Importantly, the overall architecture of DEBS3 observed in both clusters is consistent with a previous model based on limited proteolysis and AUC experiments.[193] Corroborating data are in many cases essential to increase the robustness of SAXS data analysis, as SAXS data can be very easy to misinterpret or over-analyze.

Interestingly, the SAXS-derived models suggest that, even though the ACP domains must undergo fairly large motions to interact with the various other domains, the relative orientation of the other domains allows for these ACP motions without major changes to the overall module architecture. Though it should be noted that these studies were done in the absence of substrate, rigid-body modeling suggests that the architecture is at least relatively consistent with a catalytic state, given that the ACP domains of both module 3 and DEBS3 modules can be positioned near each active site without changing the overall architecture of either. Ongoing biochemical studies[203] are anticipated to shed further light on the influence of protein-protein interactions in governing domain motions during catalysis.

In addition to the SAXS models described above, medium-resolution structures of a full type I PKS module in multiple catalytic states have been solved by cryo-EM.[186,188] These studies by Skiniotis and colleagues show the structure of module 5 from the pikromycin synthesis pathway, PikAIII, which is closely related to DEBS. The EM-derived model of PikAIII module 5 reported by Dutta *et al.* (Figure 23C) displays a surprisingly different architecture than the SAXS-derived model for DEBS module 3 (Figure 23B).[188] Most notably, the KS-AT didomain adopts a different arrangement in the cryo-EM model, leading to an arched structure where the AT domains fold down from a central KS dimer. In an accompanying study, Whicher *et al.* were also able to follow the localization of the ACP domain to each different catalytic domain during substrate turnover.[186] These EM structures suggest that the movement of the ACP domain does not require changes to the overall architecture of the module, consistent with what was proposed in the DEBS model.

The different conformations of the KS-AT didomain seen in the EM model of the PikAIII module[188] and the SAXS model of the DEBS module[192] has raised questions on data interpretation. Specifically, it has been noted that the SAXS-based rigid-body modeling of the DEBS module was done with the KS and AT domains as a single rigid body (Figure 23B), which would necessarily lead to a different model than that observed by EM.[191] Although the PikAIII model was also constructed by fitting the EM map (EM Databank entry 5647) with crystal structures of DEBS components, the KS and AT domains were treated as separate rigid bodies[188] (Figure 23C). Using CRYSOL,[2] it is possible to determine if the overall architecture of the EM model can better explain the SAXS data obtained by Edwards *et al.* for the DEBS module 3. By plotting the fits in Kratky representation to emphasize differences in the mid-q region, we find that the domain arrangement of the SAXS model (Figure 23D, blue) yields a better fit to the pronounced secondary peak in the data (Figure 23D, black) than the EM model (Figure 23D, pink). Thus, the discrepancy between the EM and SAXS models may reflect an actual difference between the PikAIII and DEBS modules in question. It is noted that the SEC-SAXS data on the DEBS module 3 indicated a monomer-dimer mixture, which can be problematic for data interpretation. However, the set of scattering profiles that were averaged for rigid-body modeling appeared

to represent a single species, as determined by the peak positions of the Kratky plots.[192] The use of EFA (described in Section 3.1.1.2) may be useful in a future study to further confirm sample homogeneity.

3.2.1.2 Virginiamycin Biosynthesis: Although the majority of PKS studies thus far have focused on *cis*-AT PKSs like DEBS, efforts to characterize more recently identified systems, such as type I PKSs with *trans*-AT architectures, are ongoing. Unlike the *cis*-AT enzymes, which contain AT domains within each module, *trans*-AT PKSs utilize a standalone AT domain that interacts iteratively with each module.[191] Other notable features of these *trans*-AT enzymes include the involvement of *trans*-acting enzymes. One such *trans*-AT system is the biosynthetic pathway of the antibiotic virginiamycin, which consists of 7 PKS modules and, interestingly, 2 NRPS modules organized into three proteins: VirA (modules 1–5), VirFG (modules 6 and 7), and VirH (module 9).

Recently, Weissman and colleagues undertook a SAXS investigation of module 5 from VirA, which contains an N-terminal KS domain, a 298-residue linker region, two sequential ACP domains, and a putative dimerization (DD) motif at the C-terminus.[204] This module is known to interact with multiple different catalytic partners, including the upstream and downstream domains of modules 4 and 6, the standalone AT VirI, the phosphopantetheinyl transferase VirK, the proofreading TE VirJ, and the VirC-VirE β-methylation cassette, and thus, serves as a good model system for examining protein-protein interactions within PKSs. Unlike the better characterized DEBS, few high-resolution structures of the *trans*-AT PKS domains exist. Hence, the authors took a ‘dissect and build’ approach, where fragments of VirA of varying length were studied by combining homology modeling, NMR, and SAXS.

Notably, VirA module 5 contains a long 298-residue linker region between the KS and ACP domains. In a *cis*-AT PKS elongation module, this region would contain the so-called KS-AT linker, which adopts an αβ fold, the AT domain, and another region known as the *post-AT linker*, which wraps back to interact with the KS domain. Based on sequence alignments, Weissman and colleagues assigned three regions to the 298-residue linker, with the first two regions having sequence homology to the DEBS KS-AT linker and the post-AT linker, and the third 159 residue region thus referred to as the *post-post-AT linker*. As such, a construct consisting of KS and the two following linker regions was examined by SEC-SAXS. In the absence of structures from VirA, dimeric homology models were generated for this construct using other PKS structures, including that of DEBS, as templates. The theoretical scattering of these homology models showed surprisingly good agreement with SEC-SAXS data, a significant result given that DEBS is a *cis*-AT PKS while VirA is a *trans*-AT PKS. Additionally, the linker region appeared to have some degree of conformational flexibility in solution. Improved fits to the data were obtained with rigid-body refinement in SASREF, allowing for flexibility between the KS domain and the KS-AT linker. It was thus concluded that the KS-AT linker can rotate relative to the KS domain, a hypothesis that may have importance for the docking of a *trans*-acting AT protein.

Other results from this study indicate a highly extended structure for the full module. First, a construct consisting only of the tandem ACP domains resulted in a pair-distance distribution function with a bimodal shape and a relatively large D_{max}, indicative of an extended

dumbbell-shaped molecule. Additional constructs containing the KS domains, the intervening linker region, and the tandem ACP domains resulted in pair-distance distribution functions with a strong peak at ~40 Å and a tail that extends to large D_{max} values, consistent with a central KS-domain dimer and extended arms. Here, the post-post-AT linker was hypothesized to be similar to intrinsically unfolded proteins, while still adopting a compact conformation. The extendedness of the arms protruding from the KS dimer was instead attributed to the tandem ACP-domains. Interestingly, the presence of the putative C-terminal DD motif did not lead to the closing of the structure, and it was thus suggested that the natural dimerization partner for this motif may be the N-terminal region of the next module in VirEG.

Overall, the SAXS experiments in this particular study yielded insight into the foldedness of the various linker regions between domains as well as the nature of interactions between PKS domains. Moreover, the unexpected architecture of the full module construct highlights the limitations of our predictive abilities and the importance of directly studying the protein-protein interactions between PKS domains and modules. SAXS, especially in conjunction with other techniques like AUC and NMR, can be an invaluable tool for studying these interactions. However, much of the structural modeling in this study was done in real space, in which rigid-bodies were fitted into *ab initio* shape reconstructions. Future studies will benefit from ensemble minimization approaches, where the flexible regions are modeled computationally, and the resultant ensembles are fitted to SAXS data in reciprocal space.

3.2.1.3 Modifying Regions of Polyketide Synthases: In addition to providing insight into the structural organization of PKS condensing domains (KS-AT-ACP),[192,204] SAXS has been used to examine the architecture of the so-called modifying regions of PKSs, which can consist of ketoreductase (KR), dehydratase (DH), and enolreductase (ER) domains. In one recent example, Maier and colleagues used crystallography and SAXS to investigate the mycocerosic acid synthase (MAS), a PKS that iteratively elongates starter fatty acids with methyl-malonyl-CoA units to produce key components of the mycobacterial cell envelope.[205] MAS contains a complete set of modifying domains, which is flexibly linked to the condensing region. A crystal structure of the modifying region revealed a closed dimer in which the DH and ER domains form the dimer interface. SAXS was used to show that this closed domain architecture not only exists in solution but is also consistent with scattering data obtained from modifying regions of two other PKSs.

3.2.2 Non-ribosomal Peptide Synthetases—Solution scattering experiments have also been performed on members of another class of modular assembly-line proteins, non-ribosomal peptide synthetases (NRPSs). NRPSs function in a conceptually similar way to type I PKSs, in that they employ a modular architecture where each module contains multiple domains that perform specific reactions. However, instead of acyl-CoA building blocks, NRPSs use amino acids to build their products. In fact, the variety of precursor amino acids is astounding. While ribosomal peptides are comprised of just the 20 proteogenic amino acids, NRPSs have been found to incorporate upwards of 500 different non-canonical amino acids into their products.[206]

Like PKSs, NRPSs contain several core domains that are necessary for chain extension and are present in every module. Those domains are the thiolation or peptidyl carrier protein (T or PCP) domain, the adenylation (A) domain, and the condensation (C) domain. Additionally, the final module of an NRPS will contain a thioesterase termination (TE) domain, which will release the product. The basic synthetic cycle of a minimal NRPS domain is shown in Figure 25. First, a conserved serine residue on the T domain is thiolated by the attachment of a Ppant arm from CoA. Then, following adenylation of the amino acid building block by the A domain using ATP, the amino acid is covalently linked via a thioester bond to the T domain. The C domain then catalyzes a peptide bond-forming reaction between two T-domain-bound amino acids, resulting in the extension of the growing peptide chain. When the product is complete, it is then either released or cyclized and released by the TE domain.[185,206] Again, like PKSs, many NRPSs contain modifying domains that perform reactions such as formylation (F), reduction (R), or epimerization (E).

3.2.2.1 Linear Gramicidin Biosynthesis: Structural characterization of NRPS modules has lagged behind compared to FASs or PKSs. The first crystal structure of a full-length NRPS module was solved in 2008 and depicted the final module involved in the biosynthesis of surfactin (SrfA–C), with C, A, PCP, and TE domains.[207] Due to extensive interactions between the C and the A domains, this structure suggested that these two domains may form a single “catalytic platform” around which the PCP domain can simply move in order to access the C and A domain active sites. More recent whole-module crystal structures indicate that additional domain motions may be necessary to move the intermediate product between the different domains during synthesis.[187,208] By using mechanism-based inhibitors, Drake *et al.* solved three different crystal structures that represented a four-stage catalytic cycle of a minimal NRPS module, corroborated by single-particle EM data.[208] In the other study, Reimer *et al.* presented four different crystal structures of LgrA, the initiation module of the linear gramicidin synthesis pathway.[187] In addition to the minimal set of NRPS domains, this module contains a modifying formylation (F) domain. In a manner reminiscent to that seen in SrfA-C, the modifying F domain in LrgA makes extensive interactions with the A domain (Figure 26A, orange and blue) forming a platform. These crystal structures additionally reveal large domain motions undertaken by both the PCP (Figure 26A, green) and the so-called A_{sub} domain, the C-terminal portion of the A domain (Figure 26A, light blue).[187]

To verify the presence of an extended F-A platform, SAXS was performed on the full-length module in solution. The theoretical scattering curves calculated from structures representing the thiolation and formylation states agree well with the experimental data, particularly at low-q (Figure 26B). Furthermore, the authors modeled flexibility in the structure via the Ensemble Optimization Method (EOM), where a large pool of conformations is generated, and a minimal subset of the ensemble is selected by fitting to experimental SAXS data.[51] In this case, EOM yielded a minimal ensemble with ~ 60% of the conformations similar to the formylation-state crystal structure and the remainder being similar to the thiolation-state structure. Overall, the SAXS analysis is consistent with the crystallographic module structure, in which the F and A domains form a stable catalytic platform for the smaller and highly mobile PCP and A_{sub} domains.

3.2.2.2 Ligand-binding to NRPS Domains: In addition to the work discussed previously, SAXS has been used to examine individual NRPS domains. For instance, in two studies from Gokhale and colleagues, SAXS was used to probe conformational changes associated with cofactor binding to NRPS R domains.[209,210] In the first study, the authors observed decreases in the R_g, and D_{max} in response to NADPH binding to an R domain from a mycobacterial NRPS,[209] suggestive of a compaction. Additionally, a later study using solution NMR and SAXS showed that R domains from two different mycobacterial NRPSs proceed by a random bi-bi reaction mechanism, unlike other similar short-chain dehydrogenase reductases. The SAXS data taken here again indicate a compaction in response to NADPH binding as well as a more subtle closing in response to binding of the other co-factor, lauroyl-CoA.[210] These findings are corroborated by their solution NMR data, which reveal clear changes in chemical shifts in response to the NADPH binding and more subtle changes upon lauroyl-CoA binding.

3.2.3 Conclusions—The studies highlighted in this section showcase the potential of SAXS as a technique to structurally characterize large multi-modular enzymes that are challenging to crystallize. In particular, PKSs and NRPSs are prime examples of how solution scattering studies can help elucidate the chemical mechanisms of multifunctional enzymes, as well as provide critical insight into their domain architecture, flexibility, and inter- or intra-protein interactions. [187,192,204] At the same time, solution scattering studies of these megasynthases are still in a nascent stage. These flexible systems present unique challenges, and interpretation of SAXS data from these systems remains difficult without high-resolution structures of the components. To go beyond modeling the overall architecture and begin investigating domain motions within these megasynthases, mathematical deconvolution analyses and advanced forms of the solution scattering techniques, such as time-resolved SAXS, will likely be important. The methods used to study PKSs, NRPSs, and other flexible systems not discussed in this review[211–214] are constantly improving,[215,216] and the use of SAXS in conjunction with other solution techniques such as NMR and FRET is becoming increasingly common. Future investigations of dynamic enzymes with SAXS in combination with other techniques will continue to improve our understanding of the forces that govern how these extended systems function and will ultimately inform our ability to reliably manipulate and engineer them for a wide range of applications.

4 Diffuse Scattering

4.1 Introduction

One of the greatest achievements of X-ray crystallography thus far has been in relating how the spatial arrangement of atoms in a protein gives rise to its activity. For example, crystallography was instrumental in our understanding of how protein activity can be allosterically regulated by ligand-induced changes to structure.[217,218] However, a protein is a network of many residues, and in many cases, we do not fully understand how distant sites within a protein communicate. Clearly, protein motions are required to transduce signals from one site to another, and in fact, it is known that distant sites can communicate without an overall change in structure, in an entropically driven phenomenon known as “dynamic

allostery".[219,220] To understand how allosteric sites couple to the active site of enzyme or how multiple active sites coordinate in a multifunctional enzyme, it is critical that we obtain information beyond the average conformation. Although Bragg diffraction can provide insight into whether deviations from the average conformation can occur, it cannot provide information on how those motions are coupled. This information exists elsewhere in crystal diffractions images, in the so-called diffuse scattering background. By combining Bragg diffraction and diffuse scattering, the future of crystallography has the potential to yield insight into the correlated motions that govern protein catalysis and allostery. Here, we will review what has been done and what remains to be resolved in this relatively young field.

Crystal lattices give rise to sharp Bragg peaks in diffraction images, and crystal structures of the repeating molecule are derived from the intensities of these peaks. Conversely, scattering of X-rays away from the Bragg peaks occurs when molecules in a crystal are displaced from their average positions, and the resultant scattering pattern reveals how these displacements are correlated. In materials science, diffuse scattering arising from crystalline samples is commonly used to characterize correlated disorder resulting from diverse phenomena, from packing defects to electronic structure.[221] The potential for protein diffuse scattering was first explored in the late 1980s and early 1990s with pioneering studies from the labs of Donald L.D. Caspar on lysozyme[222] and insulin[3] (Figure 27), Jean-Pierre Benoit and Jean Doucet on DNA[223] and lysozyme,[7,224,225] and George N. Phillips Jr. on cardiac tropomyosin[226,227] and yeast initiator tRNA[228] (Figure 28). These studies showed that diffuse scattering measurement from biological crystals is feasible, documented the variety of patterns which can be observed, and established modeling strategies — such as liquid-like motions (LLM), normal modes, and rigid-body motions — that are still used today. These early studies are described in several reviews from the period.[6,229–231]

With the recent development of pixel-array detectors, biological diffuse scattering is gaining attention once again.[81] In Section 4.2, we will discuss evidence of motions that can occur in crystals, and we specifically highlight pioneering studies on the diffuse scattering of lysozyme, an illuminating and somewhat controversial test case that was revisited several times as the field developed. In Section 4.3, we review the next phase in diffuse scattering research, which began with work of Michael Wall and Sol Gruner in the development of a custom CCD-based X-ray detector and methods to measure and compare three-dimensional diffuse scattering maps with theoretical models.[83] Diffuse scattering is also highly sensitive to all types of disorder, whether or not the disorder has a dynamic origin. Studies of such correlated disorder in protein crystals are reviewed in Section 4.4. Finally, in Section 4.5 we will discuss open questions and future goals for the field of biological diffuse scattering.

4.2 Pioneering Studies

4.2.1 Protein Motion in Crystals—Constraints imposed by the crystal lattice might be expected to suppress protein dynamics. However, the dynamic motions important for protein function can often occur in the crystal. Importantly, many dynamic enzymes retain activity when crystallized.[232] In this section, we review the evidence – from Bragg diffraction, Mössbauer spectroscopy, NMR, and MD simulation – that protein crystals are dynamic.

Although Bragg diffraction is normally used to determine atomic coordinates, one of the parameters determined along with the position of each atom is a measure of how blurry the atom's electron density appears. Commonly, the blurriness is represented by a single number, the root mean square (RMS) displacement, u, of each atom, which appears in the thermal B-factor, $B = 4\pi^2 \langle u^2 \rangle$ (see Section 2.3.3). B-factors tend to vary among atoms in a protein, and although this can arise from various effects, it is often interpreted in terms of molecular motion.[233,234]

Crystallographic refinement schemes incorporate different assumptions about the nature of correlated displacements in order to reduce the number of degrees of freedom in the model, a particularly important consideration when refining low-resolution data. Rigid-body, or TLS (for Translation, Libration, Screw),[59] refinement is the most commonly used method of this type (see Section 2.3.4). TLS has been used to infer collective motions of macromolecules including lysozyme,[235] calmodulin,[236,237] GroEL[238] and 70S ribosome.[239] An alternative to TLS is to parameterize atomic displacements using low-frequency normal modes for the purposes of crystallographic refinement, which affords a more direct dynamical interpretation. Normal mode refinement has been applied in a handful of cases.[65,240–242] However, several authors have cautioned against interpreting the TLS or normal mode models as evidence of collective motions, because the Bragg data does not depend on the covariances of atomic displacements ($\boldsymbol{V}_{\mathbf{n}j\mathbf{m}k}$ in Section 2.3).[5,229] In other words, models describing different collective motions may fit Bragg data equally well.

Although B-factors are often interpreted as indications of molecular motions, they also contain contributions from lattice-coupled motions. Protein crystal lattices have been shown to support acoustic phonons, or displacement waves.[243] The relative contributions of lattice-coupled motion and internal motion have been determined for myoglobin crystals, using narrow-bandwidth Mössbauer radiation to separate the elastic and inelastic components of a 3.63Å Bragg reflection.[244] At room temperature, lattice-coupled motion on a < 100 ns time-scale of $\langle u^2 \rangle = 0.02 Å^2$ accounted for only a small fraction of the total displacement, $\langle u^2 \rangle = 0.16 Å^2$, given by the crystallographic B-factor. In a different study of myoglobin crystals, Mössbauer spectroscopy was used to monitor the RMS displacement of heme ^{57}Fe as a function of temperature.[245] At low temperature, the RMS displacement rose gradually, as expected for lattice-coupled motions. However, above ~ 200 K, the RMS displacement rose sharply in an apparent dynamical transition.[245] These results suggest that relatively fast and short-range correlated motions dominate in protein crystals at room temperature.[229]

Protein motions in crystals have also been measured by solid-state NMR spectroscopy. In crystals of ^{2}H-labeled lysozyme and ribonuclease, the peptide planes were found to rotate stochastically by 12–15 degrees at room temperature.[246] More recently, magic-angle spinning (MAS) NMR has been used to characterize the slow rotational dynamics in three crystal forms of [^{2}H, ^{13}C, ^{15}N]-labeled ubiquitin.[247] The spin relaxation rate constant, ^{15}N $R_{1\rho}$, was shown to be sensitive to the slow reorientation of peptide planes and was interpreted in terms of an overall rocking motion of several degrees with a time-scale of 0.1 – 100 μs. This interpretation of the NMR data was also supported by X-ray diffraction and molecular dynamics (MD) simulations.

MD simulations and crystallography have had a synergistic relationship. For example, MD simulations of protein crystals have been used to better understand X-ray crystallographic structures, including the effect of the crystal lattice on protein structure and dynamics[248,249] and the contribution of disorder to the so-called “R-factor gap” in crystallographic refinement statistics.[250] Conversely, crystallographic data have also been use to evaluate MD-simulated force fields.[251–256] Most recently, a 3 × 2 × 2 supercell of triclinic lysozyme was simulated using a microsecond-scale all-atom molecular dynamics simulation.[256] In this study, a comparison of four force fields revealed that the more modern force fields did a better job reproducing the experimentally determined atomic positions and B-factors. However, the most accurate simulations still show room for improvement.

4.2.2 Lysozyme—Lysozyme was the first enzyme to have its three-dimensional structure determined by X-ray crystallography[257] and has since served as an important model system for investigating the role of dynamics in enzyme function. In nature, lysozyme is involved in fighting bacterial infections by degrading the protective peptidoglycan cell wall of gram-positive bacteria. Structures of lysozyme depicting small sugars bound to the active site have led to the now-classic model of catalysis, in which enzymes catalyze a reaction by stabilizing the intermediate.[258] These structures revealed a “hinge-bending” motion of the two lobes of the active-site cleft. Understanding lysozyme’s hinge-bending motion became an early goal for the nascent field of molecular mechanics.[259] Shortly thereafter, it was realized that lysozyme’s internal dynamics might be revealed by X-ray crystallography from an analysis of B-factors variations among different regions of the protein.[233]

As discussed above, analysis of an enzyme’s internal dynamics is complicated by the fact that atomic displacements in a crystal are compounded by external motions, such as translations and rotations of the entire molecule within the lattice. In the context of small molecule crystallography, Schomaker and Trueblood derived the functional form of the B-factors that would result from such rigid-body motions in terms of three tensors, T (translation), L (libration), and S (screw), which describe the covariances of rotation and translation.[59] This so-called TLS model was first applied to protein crystallography by Sternberg et al. who showed that the B-factors of tetragonal lysozyme crystals could be described largely by TLS motions of pairs of molecules in the unit cell.[235] However, whether global rigid-body displacements were solely responsible for the variation in atomic B-factors was impossible to prove from Bragg diffraction data alone.

To tease apart internal and external motions, researchers examined the patterns of diffuse scattering in X-ray images of lysozyme crystals. Between 1987 and 1998, several groups measured the diffuse scattering of triclinic, tetragonal, and orthorhombic lysozyme crystals and developed a variety of dynamical models to analyze the diffuse patterns. The first diffuse scattering analysis of lysozyme was performed by Doucet and Benoit on the orthorhombic crystal form.[224] Diffraction images recorded on X-ray film showed diffuse streaks between Bragg reflections along the reciprocal lattice planes perpendicular to $\mathbf{a}^*$ and $\mathbf{c}^*$, indicating that correlated motions occur along the **a** and **c** directions in the crystal (Figure 29A). Examining the packing relationships in the crystal lattice, they found that “rows” of symmetry-related molecules could be identified along each of the crystalline axes. Intramolecular contacts along the **a** and **c** directions were extensive, while contacts along the

b direction were relatively minor, explaining the strong anisotropy of the diffuse streaks. Finally, they simulated the diffuse scattering produced by small correlated displacements of pairs of molecules along **a** and **c** and were able to reproduce the streaked features they observed (Figure 29B).

Although the work of Doucet and Benoit supported a picture of a protein crystal as an elastic solid composed of relatively rigid protein units, this view was thoroughly challenged by subsequent studies. Clarage et al. collected diffuse scattering images from tetragonal and triclinic lysozyme crystals, which displayed striking cloudy patterns that were not confined to reciprocal lattice planes[222] (Figure 30). Similarly cloudy patterns were observed from insulin crystals several years earlier, which were interpreted using a liquid-like correlation function that decays exponentially with distance.[3] A key advantage of this liquid-like motion (LLM) model is that only two parameters need to be optimized: the RMS amplitude and the correlation length. Clarage et al. found that the cloudy diffuse patterns in both tetragonal and triclinic lysozyme crystals could be modeled using a correlation length of 6±1 Å, which is similar to the value obtained for insulin[3] and compares roughly with the size of an amino acid residue. The RMS displacements for the best-fit model were 0.49Å (tetragonal) and 0.33Å (triclinic). An LLM model was also used to describe the halos surrounding the Bragg peaks, which are sometimes referred to as thermal diffuse scattering, or TDS. For both crystal forms, the optimal LLM parameters for these halos were a correlation length of 50±5Å and an RMS displacement of 0.11 Å. Clarage et al. estimate that internal motions accounted for 95% of the motion in tetragonal lysozyme and 90% of the motion in triclinic lysozyme.

However, the conclusion of Clarage et al. that the RMS displacements derive mainly from internal motions relies on the LLM model, whose key assumptions are perhaps overly simplistic. The LLM model assumes that correlated motions are always co-linear, neglecting the possibility of more complex motions such as rigid-body rotations. In addition, LLM assumes that all pairs of atoms in the protein obey the same pair correlation function, independent of their position. Both assumptions are clearly at odds with the distribution of B-factors in tetragonal lysozyme, which vary significantly throughout the protein and show evidence of rigid-body rotations.[233,235]

In a later study, Pérez et al. collected new diffuse scattering data for tetragonal lysozyme (Figure 31A) and simulated the diffuse scattering (Figure 31B) that would result from independent translations and rotations of each copy of protein about its center of mass, of which there are 8 per unit cell. Although neither isotropic translations (iso-T) nor rotations (iso-L) accounted for the cloudy diffuse pattern on their own, an equal mixture of independent rotation and translation (iso-TL) achieved satisfactory, qualitative agreement. Additionally, the iso-TL motions reproduced most of the trends in the measured B-factors. Thus, Pérez et al. argued that since rigid-body motions can account for both diffuse scattering and Bragg diffraction, a rigid-body interpretation of tetragonal lysozyme dynamics should be preferred over that of the LLM model.[7]

Since the interpretation of diffuse scattering can be ambiguous, it is most powerful when used to test the predictions of self-contained, realistic models. Lysozyme was the first

system where diffuse scattering was applied in this way. Faure et al. tested the predictions of normal mode (NM) and molecular dynamics (MD) simulations using the cloudy diffuse scattering from orthorhombic lysozyme, which became visible once image plate detectors replaced X-ray film[225](Figure 32A). Both simulation methods reproduced some of the diffuse features, although the NM simulation was found to agree slightly better (Figure 32B–C). Interestingly, the hinge-bending motion was among the 15 low-frequency modes used to calculate the diffuse scattering. It is important to note that these simulations were rather crude by current standards; neither simulation accounted for the crystal lattice or solvent. More rigorous MD simulations were performed later by Héry et al.; diffuse scattering calculated from a 1 ns all-atom MD simulation of an orthorhombic lysozyme unit cell showed fair agreement with the data of Faure et al..[260] It is important to note, however, that none of these studies used a quantitative figure of merit, and so statements of relative agreement are difficult to judge.

Summary: These early studies on lysozyme demonstrated the feasibility of measuring diffuse scattering from macromolecules and established that the scattering derives from correlated motions, which cannot be measured by conventional Bragg diffraction. Although in some cases it was possible to interpret diffuse scattering in terms of internal dynamic motions, the evidence for ignoring external motions was by no means conclusive. In particular, diffuse scattering from tetragonal lysozyme could be explained equally well by the seemingly opposite LLM and TLS models.[7,222] Interpretation in these early studies was limited to qualitative visual comparisons, largely because methods to separate Bragg peaks and diffuse scattering had not yet been developed. This limitation meant that accurate reciprocal space values were not available, and by extension, the quality of fit between between experiment and theory could not be evaluated rigorously.[260]

Although the X-ray equipment and computational methods utilized are crude by current standards, early diffuse scattering work remains highly relevant. Diffuse scattering was upheld as a check on the increasingly common use of Bragg data to infer correlated motions.[229,261] This use of Bragg data has dramatically increased in popularity and sophistication, and remains a strong motivation for diffuse scattering studies.[4,5] Dynamical models developed to explain diffuse patterns -- including LLM, TLS, NM, and MD -- remain in use. Finally, the observation that the same diffuse patterns can sometimes be explained with very different models has led to a debate about the proper interpretation of diffuse scattering that has yet to be resolved.

4.3 Diffuse Maps for Dynamic Enzymes

4.3.1 Three Dimensional Diffuse Scattering Maps—In early diffuse scattering studies, analysis was performed on single diffraction images which were recorded on X-ray film or image plates. The next phase in diffuse scattering research began when methods were developed by Michael Wall, while a student in Sol Gruner's lab, for measuring three-dimensional diffuse scattering maps using a custom CCD-based X-ray detector that was optimized for these measurements.[83] The maps allowed for quantitative comparisons to be made between experiment and theory. Subsequent 3D mapping experiments focused on using the diffuse maps to discriminate between alternate models of protein dynamics.

Briefly, we describe the method for mapping diffuse scattering throughout reciprocal space.[1,82,83] A series of X-ray diffraction patterns is acquired for many different crystal orientations, as in standard crystallography. Each image is then processed to filter out the Bragg reflections using a mode filter, a technique previously used to "de-star" telescope images, where the value at each pixel is replaced by the mode (most common value) within a small patch centered on that pixel. Then, the intensity of the scattering around each Bragg peak is averaged, and this value is taken to represent the intensity of the scattering underneath the Bragg reflection (at integer values of *h*, *k*, and *l*). After this initial step, the diffuse intensities are stored in a table and processed in a similar way to Bragg data. Importantly, processing of diffuse scattering in this way allows for comparisons of experimental data to model predictions, such that meaningful R-factors or correlation coefficients can be calculated.

Complete 3D diffuse maps are available for several proteins. Notably, Staphylococcal nuclease[1] and calmodulin[64] were measured using a specialized CCD detector.[82] For nearly two decades, these were the only datasets available, likely because commercially-available CCDs suffered from blooming artifacts, which causes intense Bragg peaks to bleed into surrounding pixels. Recently, interest in protein diffuse scattering has revived,[4] in part because pixel array detectors (PADs) are now available. These detectors have a negligible point-spread function,[262] making it possible to measure weak diffuse signals and strong Bragg peaks simultaneously. [81] Recently, complete 3D diffuse maps were reported for two new proteins: bovine trypsin[84] and human cyclophilin A (Cyp A).[84] These 3D diffuse maps have been compared quantitatively to several dynamical models, including LLM,[1,64,84] MD,[87,263] TLS,[74,84] and elastic networks.[84]

4.3.2 Staphylococcal Nuclease—Staphylococcal nuclease (SNase) is an extracellular, Ca^{2+}-dependent enzyme that catalyzes hydrolysis of the phosphodiester bond in RNA and DNA. SNase has been an important model system for studying protein folding and stability, beginning with the work of Anfinsen,[264] and more recently with the aid of high-pressure structural techniques such as pressure-jump SAXS[265] and high-pressure NMR.[266] The crystal structure of the enzyme bound to Ca^{2+} and the inhibitor thymidine 3′, 5′-bisphosphate (pdTp) was first reported in 1971[267] and eventually led to a complete description of the catalytic mechanism.[268]

Diffuse scattering measurements of SNase crystals were reported by Wall et al. in 1997.[1] The data were collected from room temperature crystals sealed in thin glass capillaries (Figure 33A). The diffraction images were processed as described in Section 4.3.1 to obtain the diffuse scattering at integer multiples of the reciprocal lattice vectors. The spherically symmetric component of the scattering, which included contributions from air and solvent, was subtracted from the map. The data quality was then assessed by calculating the linear correlation and R-factors for symmetry-related regions prior to averaging. The resulting 3D maps were visualized in 3D as isosurfaces (Figure 33C), or in 2D by mercator projections of spherical shells (Figure 33D).

Wall et al. considered various models of motion, including rigid-body displacements,[83] and found that an LLM model with a correlation length of 10Å and RMS displacement of 0.36Å

provided the best fit with the data[1](Figure 33D). The correlation between this model and the data was 0.595 – less than the experimental precision (C_{sym} = 0.93) – suggesting that better models could further improve the fit. Molecular dynamics is an obvious candidate. However, MD simulations of diffuse scattering require a significant amount of time to converge.[269] A 10 ns simulation of the SNase unit cell achieved good agreement with the isotropic component of the diffuse scattering,[263] which depends only on very short-range correlations. However, the anisotropic component of the diffuse scattering requires the full positional covariance matrix, which had not yet converged by the end of the 10 ns simulation. With more recent advances in computing, a 1.1 μs simulation of the SNase unit cell was achieved[87] (Figure 34A). Again, the correlation of the simulation with the data was excellent for the isotropic component of the scattering (Figure 34B). However, agreement with the anisotropic component was limited (Figure 34C–F), with an overall correlation coefficient of 0.40. These results highlight the need to improve the accuracy of the MD model and computational power, and the potential role that diffuse scattering can play in validation of force fields.

4.3.3 Calmodulin—Calmodulin (CaM) is a eukaryotic Ca^{2+}-binding protein that interacts with a wide range of protein targets, conferring Ca^{2+} sensitivity for regulation and signaling. CaM has a dumbbell-like architecture, with a flexible linker joining N- and C-terminal domains, or “lobes”. The lobes are similar in sequence and overall fold; each includes a pair of helix-turn-helix motifs (EF-hands) that bind two Ca^{2+} ions with micromolar affinity. Ca^{2+} binding induces a conformational change in the lobes, exposing a hydrophobic patch to solvent.[270] This patch is buried again when CaM binds its target. Because the linker between the N- and C-terminal lobes is flexible, CaM can interact with its targets in a variety of conformations, from compact to extended.[271]

In neurons, CaM confers Ca^{2+} sensitivity to CaM-dependent protein kinase II (CaMKII). The dynamics of Ca^{2+}-dependent activation of CaMKII are thought to be responsible for the long-term potentiation of synaptic connections, which is important for learning and memory. In the brain, CaM activates CaMKII by binding a short autoregulatory peptide, which permits auto-phosphorylation by CaMKII to promote activity in the absence of Ca^{2+}.[272] The structure of bovine brain Ca^{2+}-CaM bound to the autoinhibitory peptide of CaMKII (alpha isoform) was determined by X-ray crystallography,[273] and CaM was shown to bind the peptide in a compact conformation. Electron density was missing for residues 74–83 of the linker, suggesting that this region is highly flexible.

To investigate the motions of CaM in the crystal, Wall et al. performed multi-conformer refinement of the Bragg data and analyzed the diffuse scattering.[64] The diffuse pattern showed weak cloudy features and intense streaks (Figure 35A). These features were then separated by sampling the cloudy pattern in the region of systematically absent reflections of the $C222_1$ space group, where streaks did not extend. The cloudy pattern was then mapped in 3D and fit using an isotropic LLM model, with RMS displacement of 0.38Å and correlation length of 4.8 ± 1.0 Å, which is somewhat shorter than the typical correlation length of 6 – 10Å observed for crystals of SNase, lysozyme, and insulin.

In this study, Wall et al. also mapped the *streaked* diffuse patterns at high resolution in 3D[64] (Figure 35C), the first such map for a protein crystal. Computer memory limitations restricted the map to a cube in reciprocal space with 8 reciprocal cells on each side. The data were fit using an anisotropic form of the LLM model (Figure 35B,D). The optimal parameter values of the model indicated 0.4Å motion with a correlation length of 135 Å, oriented along a diagonal direction in the **a**–**b** plane that aligns with a continuous end-to-end packing direction in the crystal. Interestingly, the multi-conformer refinement of the Bragg data also showed a preference for displacements in this direction. The authors suggested that the short correlation length in the perpendicular direction was partly caused by the linker, which is too flexible to make rigid lattice contacts.

4.3.4 Glycerophosphodiesterase—Glycerophosphodiesterase (GpdQ) is an example an enzyme whose collective motions had been inferred from TLS refinement of Bragg data, and later tested using diffuse scattering. GpdQ employs a binuclear metal center to catalyze the degradation of glycerophosphodiesters as part of the phospholipid remodeling and synthesis pathways. GpdQ is unique among known metallo-phosphodiesterases in that it displays broad substrate specificity,[274] which has made this enzyme the subject of bioengineering efforts for the bioremediation of organophosphate pesticides, including EA 2192, a highly toxic degradation product of the nerve agent VX.[275,276]

The first crystal structure of Zn-bound GpdQ from *Enterobacter aerogenes*, solved at 2.92Å resolution,[277] depicts a domain-swapped dimer in the asymmetric unit, which is cross-linked by an inter-chain disulfide bond. In this structure, the physiologically relevant hexameric form is generated from symmetry-related dimers about the 3-fold crystallographic axis. The domain-swapped “cap” domain sits over the active site and is suggested to play a role in determining substrate specificity in this enzyme family. Interestingly, the active site cleft formed by the “cap” was too narrow to accommodate the leaving group of the particular substrate used in activity assays, suggesting that breathing motions were required as part of the catalytic cycle.[277] In support of the presence of such motions, the strong diffuse scattering in the diffraction images was noted[278] (Figure 36A). Additionally, TLS refinements with 1–6 groups per dimer improved the R-factors significantly.[277]

To model the collective motions of GpdQ, Jackson et al. calculated the normal mode spectrum for an elastic network.[277] When dimeric GpdQ was modeled, the low frequency modes did not result in cleft opening. However, when hexameric GpdQ was modeled, the lowest frequency mode showed a rocking motion about the hexamer interface, which opened the active site cleft. Later, directed evolution experiments showed that mutations at the “cap” domain interface, including the cysteine involved in cross-linking, led to increased activity toward an unnatural substrate,[279] which is consistent with the hypothesis that dynamics of the cap domain are important in substrate discrimination. However, the only direct evidence for the existence of collective motions of GpdQ comes from the crystallographic B-factors derived from TLS refinement.

It has been strongly argued that the existence of biochemically relevant collective motions should not be inferred from TLS parameters, unless they correctly predict the diffuse scattering present alongside the Bragg data used in refinement.[5] Unfortunately, the mapping

of TLS matrices to diffuse scattering is not straightforward. As discussed earlier in Section 4.2.2, the diffuse scattering of tetragonal lysozyme has been modeled previously using isotropic T and L matrices,[7] but methods making use of the full T, L, and S matrices to model diffuse scattering were not described until recently.[280] Given a PDB file containing a structural model refined using TLS, the algorithm generates a representative ensemble.[280] This ensemble can then be used to predict the diffuse scattering with Guinier's formula (Equation 96).[74]

To compare the diffuse scattering of GpdQ with predictions from TLS refinement, Van Benschoten et al. first processed individual X-ray images to remove the Bragg peaks (Figure 36B). Three different TLS refinements were performed with 1–6 groups per asymmetric unit, treating either the dimer (1 group), the monomer (2 groups), or each subdomain (6 groups) as separate rigid bodies (Figure 36C–E). The isotropic component of the simulated and measured diffuse scattering were then compared directly. Remarkably, none of the TLS-derived ensembles showed significant agreement with the experimental curves (Figure 36F). The underlying reason is not yet clear. It may be that the dynamics of GpdQ are not well-described by rigid-body motion. However, we note that rigid body motion alone is unlikely to account for the isotropic component of diffuse scattering without also modeling the non-rigid thermal motion of protein and solvent. In addition, it is also worth noting that the determination of TLS parameters from Bragg data is highly ambiguous, making it rather unlikely that the results correspond to a physically meaningful model.[280] For example, the GpdQ ensembles generated from the 6-group TLS refinement are clearly unphysical.[280] This study highlights the gap between what currently can be inferred from TLS parameters, and what is actually needed to predict diffuse scattering profiles accurately.

4.3.5 Cyclophilin A—The dynamics of human cyclophilin A (CypA) have been the subject of a number of investigations performed with X-ray diffraction, diffuse scattering, and other biophysical techniques. CypA is a peptidyl prolyl cis-trans isomerase, which catalyzes the interconversion of the cis and trans isomers of peptide bonds involving a proline. Due to the high-energy barrier of ~ 20 kcal/mol, this isomerization occurs on the timescale of seconds in solution and can therefore limit the rate of protein folding.[281] The roles of CypA in the cell are diverse and include important cell signaling functions. The drug cyclosporin A, a cyclic non-ribosomal peptide, induces an immunosuppressive response by forming a ternary complex with CypA and calcineurin, inhibiting the phosphatase activity of calcineurin and blocking its ability to regulate the transcription of cytokine genes.[282,283] CypA also modulates HIV-1 infection, and the enzyme from the host is packaged within the viral capsid.[284,285]

The equilibrium ensemble of CypA has been visualized independently by Fraser and colleagues using X-ray crystallography. Fraser et al. first presented multi-conformer fitting to the low-level electron density,[78] and later this technique was combined with filtering of steric clashes to infer networks of correlated motion.[286] These studies motivated a subsequent diffuse scattering study,[84] which aimed to validate these models and characterize more subtly correlated motions in CypA.[287] In particular, diffuse scattering data were reported for crystals of substrate-free CypA, in space group $P2_12_12_1$ with one protein per asymmetric unit.[84] X-ray images were recorded using a pixel-array detector with the crystal

at 273 K using a standard crystallography setup. The diffuse intensity was extracted and symmetry-averaged (Figure 37A–B) as in the previous 3D diffuse mapping studies of SNase and calmodulin[64,82,83] and reported alongside diffuse scattering maps for $P2_12_12_1$ crystals of bovine trypsin in complex with the inhibitor benzamidine.

The 3D diffuse maps of CypA over a resolution range of 31.2–1.45Å were then compared with three classes of models: TLS ensemble models, LLM, and the normal modes of an elastic network (EN). TLS models with either 1 or 8 groups per protein were refined against the Bragg data. None of the diffuse maps predicted by TLS showed significant agreement with the experimental maps. The isotropic LLM model gave an improved fit to the data (correlation of 0.518, Figure 37C,F) with a correlation length $\gamma = 7.1$Å and RMS displacement $\sigma = 0.38$ Å. An EN derived from the CypA unit cell with periodic boundary conditions was introduced as a third model. Agreement with the normal modes of the EN was less compelling (correlation of 0.41, Figure 37E,F); however, unlike the LLM model, the EN was able to account for the crystallographic B-factors.[84]

Although the LLM model exhibited the highest global correlation with the diffuse scattering of CypA and trypsin,[84] the global correlation was lower than for previously measured proteins SNase[1] and calmodulin.[64] Several issues may be at play. The first is that the LLM model is expected to be less accurate at high resolution, since it is a first-order expansion of the diffuse intensity.[63] It thus follows that the model performs more poorly when describing the 1.45Å CypA and 1.46Å trypsin maps than describing the 2.5Å SNase and 2.1Å calmodulin maps. Another issue may be the differing qualities of the 3D diffuse maps. For the CypA and trypsin datasets, which were collected using a standard crystallography setup, the correlation coefficient for symmetry-related measurements[84] was significantly lower than that obtained for SNase using a specialized setup.[64,83] These results highlight that obtaining the highest quality data is the first important step in diffuse scattering analysis.

It remains to be seen whether diffuse scattering from CypA supports the correlated motions inferred from the low-level electron density by the multi-conformer approach. A critical assumption of the multi-conformer approach is that the average electron density reports internal conformational heterogeneity, rather than deviations of the entire molecule from its average position in the lattice. Diffuse scattering of CypA was more consistent with LLM than rigid-body TLS models, lending support to the assumptions inherent in the multi-conformer approach. A more direct test of the contact network analysis could be achieved by calculating the diffuse scattering directly from multi-conformer models, instead of fitting the LLM. Such a calculation has not yet been reported for CypA but could potentially illuminate the extent to which contact network models accurately represent correlations present in the ensemble.[287]

The introduction of EN models for diffuse scattering analysis is a logical step toward the eventual goal of combining Bragg and diffuse data during crystal structure refinement. ENs derived from a unit cell with periodic boundary conditions can give a similar exponential decay of the atom-atom correlation function as the LLM,[288] suggesting that the EN approach could give similar levels of agreement with diffuse scattering data, while addressing critical limitations of the LLM, including interpretability. Whereas the

correlation length of LLMs is only distantly related to the functional motions of interest, ENs allow for visualization of the low frequency normal modes.[289] Normal mode refinement of Bragg data has been applied to several proteins,[65,240–242] and a procedure for simultaneously fitting normal modes to Bragg and diffuse scattering has been described.[63,65] Thus, ENs are a good candidate for a simple model that can be refined to all available X-ray data in order to visualize the breathing motions of proteins.

4.4 Correlated Disorder

4.4.1 Gag Polyprotein—Diffuse scattering is produced by deviations from the strict periodic structure of a perfect crystal and it arises irrespective of whether the disorder has a dynamic origin. Correlated disorder in crystalline materials has been characterized extensively by diffuse scattering.[221] Geometric frustration is one such non-dynamic phenomenon that gives rise to crystal disorder, which results from the mismatch between the local symmetry of the repeating object in the crystal and the symmetry of the lattice. Such frustration is exemplified by the antiferromagnetic Ising model on a 2D triangular lattice. In this system, neighboring spins prefer to align in opposite directions, but they are frustrated on a triangular lattice because they cannot all be anti-aligned with their neighbors (Figure 38). This frustration leads to a situation where the ground-state orientation of the spins is not periodic, but the spins are still correlated with one another over significant distances. The diffuse patterns from these materials show characteristic honeycomb-shaped ridges between the Bragg peaks.[221]

Diffuse scattering resulting from geometric frustration was recently observed from crystals of the N-terminal 154 residues of the gag protein from feline foamy virus (FFV).[86] The protein crystallized in the space group $P6_122$ with one monomer per asymmetric unit. Distinctive diffuse rings were observed around certain Bragg peaks in the diffraction images (Figure 39A). When the diffuse data were mapped at high resolution in 3D (Figure 39B), the rings were found to lie only in certain planes in the reciprocal lattice (l = integer), in a pattern that repeated roughly every 5 planes. 2D slices through reciprocal space resembled diffuse patterns produced by an Ising model on a triangular lattice (Figure 39C–D). In a manner analogous to the Ising model described above, the geometric frustration occurring in the protein lattice can be understood by considering two possible arrangements of molecules in the asymmetric unit, A and B, which prefer to pack as AB or BA rather than AA or BB (Figure 38).

Because of the disordered packing arrangement in the gag FFV crystals, the structure could not be determined from the Bragg data by conventional means. Remarkably, however, initial modeling suggests that the extra information contained in diffuse scattering might be used to solve the structure. Reverse Monte-Carlo (RMC) simulations were able to identify the solvent channel and produce some interpretable electron density, assuming that the arrangements A and B differ by a fractional translation of 0.2 l.[290]

4.4.2 70S Ribosome—The ribosome is a dynamic RNA-protein machine that synthesizes polypeptides from genetic information encoded in mRNA. The prokaryotic (70S) ribosome is composed of two subunits: the small (30S) subunit, which, binds the mRNA, and the large

(50S) subunit, which contains the peptidyl-transferase center. Each amino acid is delivered by a specific tRNA, whose anti-codon sequence matches the mRNA template codon. As an amino acid is added to the polypeptide chain, the tRNAs translocate through three sites in the ribosome (A, P, and E). EM studies revealed that the two subunits rotate relative to one another during translocation in what is described as a "ratchet-like" motion.[291]

TLS refinement has also been used to infer dynamical information from crystal structures of *Thermus thermophilus* 70S ribosomes.[239] Interestingly, the libration axes of the subunits were consistent with the ratcheting motion observed in the earlier EM studies. However, if the rigid-body motions implied by the TLS refinement actually occur in ribosome crystals, diffuse scattering should be present in the diffraction images. In particular, Peter Moore has argued that biochemical conclusions derived from TLS should be supported by an analysis of the diffuse scattering.[5] The diffuse scattering from crystals of the 70S ribosome was thus later revisited by Polikanov and Moore.[66]

It can be especially illuminating to analyze the diffuse scattering from a diffraction image that is oriented with one of the crystal axes parallel to the X-ray beam. The symmetry of the scattering pattern is immediately apparent, and symmetry-related regions of the image can be averaged to improve the signal-to-noise, as was previously demonstrated for myoglobin crystals in the P6 space group.[269] Polikanov and Moore searched through ~ 350, 000 diffraction images from cryo-cooled 70S ribosome crystals – collected for standard structure determination using PAD – to find fortuitously oriented patterns (such as Figure 40A). They processed several of these images in order to extract the textured diffuse pattern. The circularly symmetric background was subtracted, and the image was high-pass filtered to remove absorption artifacts such as the beamstop shadow. Images were also binned to reveal the more subtle modulations on scales greater than the distance between Bragg peaks. Binned and un-binned images are shown in Figure 40B–C.

Several features in these images suggest that the diffuse scattering originates from lattice-coupled motions, or acoustic modes, and not the independent rigid-body motions of the type implied by TLS refinement. First, the binned diffuse intensities display the same *mmm* symmetry as the Bragg peaks (such as Figure 40B). Polikanov and Moore point out that this is not a trivial consequence of $P2_12_12_1$ space group symmetry, because screw axes are not point group operations. In addition, the diffuse intensities correlate with the Bragg intensities, which would be expected for lattice-coupled motion (see Section 2.3.4). Finally, in the un-binned images, ridges of intensity that follow the reciprocal lattice planes are clearly visible (such as Figure 40C). A sinusoidal modulation of the diffuse intensity is predicted by the lattice vibrational mode model (Equation 151) with a particular form of the dispersion relation that was derived assuming interactions between adjacent unit cells dominate.[66] Images simulated using this dispersion relation and Equation 151 strongly resemble the diffuse data (Figure 40D).

Although the lattice vibration model describes general features of the diffuse scattering pattern in these crystals, the model does not account for the magnitude of motion. The B-factors predicted for thermally excited acoustic modes at room temperature are about 100

times too small.[66] Thus, the field awaits further progress in both experiment and theory for a complete understanding of the diffuse scattering produced by these crystals.

The work of Polikanov and Moore highlights the need to carefully consider all potential sources of correlated motion when analyzing diffuse scattering. They provide several simple metrics and guidelines which can be used. First, they plot the standard deviation σ_D of the diffuse scattering in thin annular rings of constant $|\mathbf{q}|$ (after subtracting the shot noise contribution) as a function of scattering angle (such as Figure 40F). This profile is compared with characteristic shapes derived for different models using Wilson statistics. For acoustic modes (Figure 40D, solid line),

$$\sigma_D(q) \propto\approx q^2\langle T\rangle\sum_n f_n(q), \quad (153)$$

where $\langle T\rangle$ is the average Debye-Waller factor (Equation 105). For motion which is uncorrelated between unit cells (Figure 40E, dashed line),

$$\sigma_D(q) \propto\approx (1-\langle T\rangle)\sum_n f_n(q). \quad (154)$$

Notably, the characteristic fall-off at high angles is much steeper for acoustic modes (Equation 153) than for short-ranged, uncorrelated motions (Equation 154). Finally, Polikanov and Moore provide a useful list of “trade marks” that can help identify lattice-coupled motions from diffuse scattering:

1. Spatial correlation between Bragg and diffuse intensities
2. Minima in the diffuse intensity mid-way between adjacent Bragg peaks
3. Steep fall-off of $\sigma_D(q)$ at high scattering angles.

These guidelines will undoubtedly be useful in future diffuse scattering studies.

4.4.3 Photosystem II—Time-resolved crystallography allows for the study of concerted conformational changes in proteins over a large range of timescales. This method lends itself particularly well to studying light-driven reactions that can be synchronized by a fast laser pulse (reviewed in Ref. 292). Of particular interest is the photosynthetic water splitting reaction performed by photosystem II (PSII). A prerequisite for understanding the reaction mechanism of PSII is damage-free structural data for each intermediate. The metal cluster in PSII is known to be especially susceptible to X-ray induced damage[293,294] and has motivated the development of new methods to mitigate damage during data collection. The advent of the XFEL that can deliver femtosecond X-ray pulses has inspired serial crystallography approaches based on the principle of “diffraction before destruction”.[79] PSII has served as a poster child for these new sources.

PSII is a multi-subunit complex found in the thylakoid membranes of higher plants, eukaryotic algae, and cyanobacteria, and is highly conserved (Figure 41). Much of the structural work on PSII has focused on PSII from thermophilic cyanobacteria (reviewed in Ref. 295). In cyanobacteria, PSII is a 700 kDa dimeric complex consisting of two 350 kDa momomers, each of which contains 20 distinct subunits. A pair of transmembrane subunits, D1 and D2, contain four chlorophylls collectively known as P680. In close proximity is the oxygen-evolving center (OEC), a cubane-like, μ-oxo bridged Mn_3CaO_4 cluster with a fourth "dangling" Mn. After one P680 chlorophyll absorbs a photon, the excited electron moves to pheophytin and then to a pair of plastiquininones, denoted QA and QB (QB eventually dissociates). Following electron transfer, P680 is reduced by a tyrosine residue in D1 called YZ, which then oxidizes the OEC. After four such reductions have occurred, the OEC splits two water molecules, yielding four protons and molecular oxygen. With each photon absorbed, PSII advances through a series of five states, $S_0 \rightarrow S_4$, known as the Kok cycle,[296] where water splitting occurs in the transition between the unstable S_4 state and the S_0 state.

The XFEL approach has yielded "damage-free" structures of PSII, including a 1.95Å structure of the dark-stable S_1 state.[297] Since XFELs destroy the sample after a single exposure, complete datasets must be obtained by merging measurements from multiple crystals. In order to obtain structures of PSII in the other S-states, a time-resolved (pump-probe) approach has been developed whereby microcrystals are delivered in a liquid jet that flows sequentially through laser and X-ray beams.[298] The first pump-probe measurements of PSII suffered from low resolution (5–5.5 Å), which made interpretation of conformational changes highly ambiguous.[299,300] Progress in this field is rapid, and these controversies are likely to be resolved once high resolution structures are available. Of interest here, however, is the fact that low resolution diffraction from PSII crystals show substantial diffuse scattering. This observation was made even before the advent of the XFEL and interpreted as evidence of weak crystal packing.[301] Strong diffuse scattering was also seen in XFEL measurements of PSII microcrystals by Ayyer et al. (Figure 42A).[302]

What is the origin of the diffuse scattering from PSII? If it is caused by weak crystal packing interactions, and not internal dynamics, the prediction is that the scattering should resemble the symmetry-averaged molecular transforms of the rigid structural subunits undergoing independent translational motions. The Fourier transform of the diffuse intensity (the diffuse Patterson function, Figure 42D) resembles the autocorrelation function of the PSII dimer (the asymmetric unit, Figure 42E), suggesting that the PSII dimer is the rigid subunit in this case – not the entire unit cell (4 dimers, space group $P2_12_12_1$), or individual domains within the complex.[302] If indeed the PSII dimer is more rigid than implied by the diffraction resolution, the diffuse scattering which extends beyond the resolution of the Bragg peaks should contain high resolution information about its structure.

Ayyer et al. developed a clever approach to extend the resolution of the electron density maps by incorporating information from diffuse scattering.[302] Mathematically, the diffuse scattering from rigid-body translational motion is proportional to the molecular transform (see Equation 127). Previous work on lensless X-ray imaging has shown that knowledge of the continuous molecular transform can be sufficient to solve the phase problem.[303] Ayyer et

al. used a difference-map algorithm[304] to compute electron density maps from the diffuse scattering, which was combined with the low-resolution map from the Bragg data.

After adding the information from the diffuse scattering to the Bragg data for PSII, the reported resolution gains were approximately 1Å (from ~ 4.5Å using Bragg data alone, to ~ 3.5Å combined).[302] While this gain is impressive, it is not entirely clear whether the diffuse scattering deserves all of the credit. Given the complexity of the phasing method, there are many open questions, including the degree to which model bias influences the result. It is also unclear whether the original diffraction resolution was indeed 4.5Å because Bragg peaks can be seen extending beyond this limit in some of the published images. It is well known that serial crystallography suffers from non-isomorphism, with merging statistics that are generally poorer than those of single-crystal datasets. Thus, it is possible that the apparent resolution of the merged dataset, according to correlation between random half data sets, would be lower than the resolution of individual crystals prior to merging. One concern, then, is that some weak Bragg diffraction was mistakenly counted as diffuse scattering. Finally, the method rests on the assumption that the diffuse scattering derives entirely from uncorrelated translations of the PSII dimer. The evidence for this is not conclusive, and other potential contributions, including internal protein motions and lattice-coupled motions, were not examined.

These criticisms, however, do not detract from the essential novelty of the idea – that lenseless imaging techniques can be applied to diffuse scattering under certain circumstances in order to improve structure resolution. The idea is useful only if random rigid-body translations are responsible for diffuse scattering in low-resolution crystals more generally. The evidence from all previous 3D diffuse mapping studies of protein crystals — including SNase,[1] CypA,[84] calmodulin,[64] and Trypsin[84] — suggests the opposite. These crystals showed cloudy diffuse patterns that were not consistent with rigid-body models, and liquid-like or other “soft” models showed better agreement. However, the crystals used in these studies diffracted to high resolution compared with PSII. Clearly, more examples of diffuse scattering from crystals of all types are needed.

4.5 Open Questions and Future Prospects

Diffuse scattering, with its unique sensitivity to correlated motion, promises to reveal previously hidden dynamics that underlie allosteric control of enzyme activity. However, the technique has yet to demonstrate its full potential. Perhaps the most glaring criticism is the fact that diffuse scattering has yet to solve a problem of genuine biochemical interest. This is not for a lack of ingenuity. Indeed, as the examples in this review demonstrate, the seminal efforts to measure and interpret diffuse scattering have generated a multitude of processing schemes, dynamical models, and novel insights. There are, it may be said without exaggeration, as many explanations for diffuse scattering as experimental examples. The evidence so far is that protein crystals exhibit many different types of disorder. Thus, before any information of biochemical interest can be extracted from the data, a critical challenge is to identify which models apply. This problem of classification is largely due to our relative inexperience, which comes from the paucity of experimental or simulated examples. With

the advent of ideal X-ray detectors and powerful computers, however, the types of systematic studies that need to be done are well within reach.

A number of open questions need to be resolved before diffuse scattering can be considered a routine measurement. Some of the most pressing include the following. How does radiation damage affect diffuse scattering? Can diffuse scattering from cryo-cooled crystals be trusted? How isomorphic are protein crystals from the perspective of their scattering? Does scattering from poorly-diffracting crystals contain useful information?

These questions should yield to a systematic experimental approach, and the answers will determine the most useful modes of data collection and the breadth of samples that can be reliably analyzed. For example, current methods for processing diffuse data were developed in the CCD era, and deserve renewed attention, especially given improvements in detector technology and computational power. A wish list might include the following:

1. A computational method to separate diffuse background and Bragg peaks that outperforms the mode filter when applied to PAD data.
2. New metrics, akin to R_g, $I(0)$, or D_{max} employed in SAXS, which allow for a model-independent description of a dataset's general properties.
3. A parametrized model of internal protein motion and lattice disorder which is minimally complex, yet sufficiently general that it can be refined simultaneously to Bragg and diffuse data.

While systematic experiments and methodological improvements are one path forward, we should not discount the possibility of surprise. Indeed, the recent application of diffuse scattering for improving structural resolution[302] was not anticipated, to our knowledge, by prior authors. The biological diffuse scattering field at is current stage has reached a point like the biological SAXS field immediately before the introduction of widely distributed software.[2] With creativity, open-mindedness, and rigorous aproaches, the field is poised to bring crystallography into a new era that extends beyond Bragg diffraction.

5 Concluding Remarks

Visualizing protein in motions remains one of the major frontiers in biology, and recent advances in both X-ray diffraction and scattering have made this goal more achievable.[305] The case studies in solution scattering as well as the extensive overview of diffuse scattering discussed in this review are testament to the increasingly compelling biological problems that can be addressed by X-ray scattering. While this review encompasses a diverse array of technical approaches and protein systems, certain topics have not been covered. Chief among these are time-resolved solution scattering, such as the now classic macromolecule folding studies by Pollack and colleagues performed in microfluidic mixers[306–308] and pump-probe studies of so-called "protein quakes" in photoactivated systems, such as myoglobin,[309,310] performed with picosecond pulses at 3rd generation synchrotrons and femtosecond pulses at XFELs. Important milestones in time-resolved diffraction and scattering have been reviewed previously, and the reader is referred to these excellent reviews.[8,305,311–313] With the recent development of the XFEL, time-resolved solution

scattering studies are once again gaining momentum and will undoubtedly continue to shed light onto ever more interesting biochemical processes.

This review has placed into the spotlight a host of solution scattering experiments performed as either primary or secondary characterization techniques toward the study of complex enzymes. The experimental results enumerated here demonstrate the unique capabilities of this approach and are indicative of the future of the field. In particular, whereas highresolution structures may be determined by crystallography or cryo-EM, solution scattering has proven to be a powerful tool for probing conformational changes under solution conditions that are physiologically relevant. Moreover, solution scattering can offer quantitative and model-independent insight into the nature of the structural transitions, such as the number of species and forms of cooperativity involved. Thus, the future awaits more investigations that take advantage of these unique strengths to elucidate the complex regulatory and catalytic mechanisms of dynamic enzymes. Likewise, as our understanding of protein dynamics grows, it has become increasingly clear that crystallographic snapshots afforded by traditional Bragg diffraction are not sufficient for a complete understanding of enzyme mechanisms. Thus, diffuse scattering, a field long dormant, is experiencing a resurgence.[4] Although bottlenecks remain, the field has gained unprecedented momentum with the recent emergence of new light sources and detection capabilities, and future studies are certain to be interesting regardless of the outcome. In sum, as researchers delve into increasingly complex protein systems, X-ray scattering techniques stands at the forefront of biophysical characterization.

Acknowledgments

The authors are grateful to Emily Parker, Patrice Vachette, Evan Kantrowitz, Catherine Drennan, JoAnne Stubbe, Chaitan Khosla, and Martin Schmeing for sharing their data. The authors are especially indebted to the following people for their critical reading of the manuscript: Susannah Shoemaker (theory and diffuse scattering), Paul Fitzpatrick (aromatic amino acids), Patrice Vachette (ATCase), and Yan Kung (flexible enzymes). We also thank Liat Kugelmass for helping with the typesetting of this manuscript, Andrea Edwards for helpful correspondence, and the rest of the Ando lab for their support. This work was supported by National Health Institutes (NIH) grants F32GM117757 to SPM and GM100008 to NA and startup funds from Princeton University to NA.

References

1. Wall ME, Ealick SE, Gruner SM. Three-dimensional diffuse X-ray scattering from crystals of Staphylococcal nuclease. Proc Natl Acad Sci U S A. 1997; 94:6180–6184. [PubMed: 9177191]
2. Svergun D, Barberato C, Koch MHJ. *CRYSOL* – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr. 1995; 28:768–773.
3. Caspar DLD, Clarage J, Salunke DM, Clarage M. Liquid-like movements in crystalline insulin. Nature. 1988; 332:659–662. [PubMed: 3282173]
4. Wall ME, Adams PD, Fraser JS, Sauter NK. Diffuse X-ray scattering to model protein motions. Structure. 2014; 22:182–184. [PubMed: 24507780]
5. Moore PB. On the relationship between diffraction patterns and motions in macro-molecular crystals. Structure. 2009; 17:1307–1315. [PubMed: 19836331]
6. Clarage J, Phillips G Jr. Analysis of diffuse scattering and relation to molecular motion. Methods Enzymol. 1997; 277:407. [PubMed: 9379926]
7. Pérez J, Faure P, Benoit JP. Molecular rigid-body displacements in a tetragonal lysozyme crystal confirmed by X-ray diffuse scattering. Acta Crystallogr, Sect D: Biol Crystallogr. 1996; 52:722–729. [PubMed: 15299635]

8. Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macro-molecular structures, conformations and assemblies in solution. Q Rev Biophys. 2007; 40:191–285. [PubMed: 18078545]
9. Svergun DI, Koch MH. Small-angle scattering studies of biological macromolecules in solution. Rep Prog Phys. 2003; 66:1735–1782.
10. Jacques DA, Trewhella J. Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls. Protein Sci. 2010; 19:642–657. [PubMed: 20120026]
11. Jacques DA, Guss JM, Svergun DI, Trewhella J. Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution. Acta Crystallogr, Sect D: Biol Crystallogr. 2012; 68:620–626. [PubMed: 22683784]
12. Skou S, Gillilan RE, Ando N. Synchrotron-based small-angle X-ray scattering of proteins in solution. Nat Protoc. 2014; 9:1727–1739. [PubMed: 24967622]
13. Lipfert J, Doniach S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. Annu Rev Biophys Biomol Struct. 2007; 36:307–327. [PubMed: 17284163]
14. Pollack L. SAXS studies of ion–nucleic acid interactions. Annu Rev Biophys. 2011; 40:225–242. [PubMed: 21332357]
15. Koch MH, Vachette P, Svergun DI. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q Rev Biophys. 2003; 36:147–227. [PubMed: 14686102]
16. Glatter, O., Kratky, O. Small-angle X-ray scattering. Academic Press; New York: 1982.
17. Feigin, LA., Svergun, DI. Structure analysis by small-angle X-ray and neutron scattering. Taylor, GW., editor. Springer US; Boston, MA: 1987.
18. Als-Nielsen, J., McMorrow, D. Elements of modern X-ray physics. John Wiley & Sons; 2011.
19. Drenth, J. International Tables for Crystallography Volume F: Crystallography of biological macromolecules. Prince, E., editor. Springer; 2006. p. 45-63.
20. Rupp, B. Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology. Garland Science; 2009.
21. Klein O, Nishina Y. Über die streuung von strahlung durch freie elektronen nach der neuen relativistischen quantendynamik von Dirac. Z Phys. 1929; 52:853–868.
22. Brown, PJ., Fox, AG., Maslen, EN., O'Keefe, MA., Willis, BTM. International Tables for Crystallography Volume C: Mathematical, physical and chemical tables. Prince, E., editor. Springer; 2006. p. 554-595.
23. Mylonas E, Svergun DI. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. J Appl Crystallogr. 2007; 40:s245–s249.
24. Tardieu A, Le Verge A, Malfois M, Bonneté F, Finet S, Riès-Kautt M, Belloni L. Proteins in solution: from X-ray scattering intensities to interaction potentials. J Cryst Growth. 1999; 196:193–203.
25. Moore PB. Small-angle scattering. Information content and error analysis. J Appl Crystallogr. 1980; 13:168–175.
26. Taupin D, Luzzati V. Information content and retrieval in solution scattering studies. I. Degrees of freedom and data reduction. Journal of Applied Crystallography. 1982; 15:289–300.
27. Orthaber D, Bergmann A, Glatter O. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. J Appl Crystallogr. 2000; 33:218–225.
28. Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HDT, Konarev PV, Svergun DI. New developments in the *ATSAS* program package for small-angle scattering data analysis. J Appl Crystallogr. 2012; 45:342–350. [PubMed: 25484842]
29. Rambo RP, Tainer JA. Accurate assessment of mass, models and resolution by small-angle scattering. Nature. 2013; 496:477–481. [PubMed: 23619693]
30. Glatter O. A new method for the evaluation of small-angle scattering data. J Appl Crystallogr. 1977; 10:415–421.
31. Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr. 1992; 25:495–503.

32. Hansen, S. Bayesian Methods in Structural Bioinformatics. In: Hamelryck, T.Mardia, K., Ferkinghoff-Borg, J., editors. Statistics for Biology and Health. Springer; Berlin Heidelberg: 2012. p. 313-342.
33. Hansen S. Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering. J Appl Crystallogr. 2000; 33:1415–1421.
34. Merzel F, Smith JC. SASSIM: a method for calculating small-angle X-ray and neutron scattering and the associated molecular envelope from explicit-atom models of solvated proteins. Acta Crystallogr, Sect D: Biol Crystallogr. 2002; 58:242–249. [PubMed: 11807248]
35. Park S, Bardhan JP, Roux B, Makowski L. Simulated X-ray scattering of protein solutions using explicit-solvent models. J Chem Phys. 2009; 130:134114. [PubMed: 19355724]
36. Bardhan J, Park S, Makowski L. *SoftWAXS*: a computational tool for modeling wide-angle X-ray solution scattering from biomolecules. J Appl Crystallogr. 2009; 42:932–943. [PubMed: 21339902]
37. Grishaev A, Guo L, Irving T, Bax A. Improved fitting of solution X-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. J Am Chem Soc. 2010; 132:15484–15486. [PubMed: 20958032]
38. Poitevin F, Orland H, Doniach S, Koehl P, Delarue M. AquaSAXS: a web server for computation and fitting of SAXS profiles with non-uniformally hydrated atomic models. Nucleic Acids Res. 2011; 39:W184–W189. [PubMed: 21665925]
39. Virtanen JJ, Makowski L, Sosnick TR, Freed KF. Modeling the hydration layer around proteins: applications to small- and wide-angle X-ray scattering. Biophys J. 2011; 101:2061–2069. [PubMed: 22004761]
40. Nguyen HT, Pabit SA, Meisburger SP, Pollack L, Case DA. Accurate small and wide angle x-ray scattering profiles from atomic models of proteins and nucleic acids. J Chem Phys. 2014; 141:22D508.
41. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys J. 2013; 105:962–974. [PubMed: 23972848]
42. Liu H, Morris RJ, Hexemer A, Grandison S, Zwart PH. Computation of small-angle scattering profiles with three-dimensional Zernike polynomials. Acta Crystallogr, Sect A: Found Crystallogr. 2012; 68:278–285.
43. Ravikumar KM, Huang W, Yang S. Fast-SAXS-pro: A unified approach to computing SAXS profiles of DNA, RNA, protein, and their complexes. J Chem Phys. 2013; 138:024112. [PubMed: 23320673]
44. Fraser RDB, MacRae TP, Suzuki E. An improved method for calculating the contribution of solvent to the X-ray diffraction pattern of biological molecules. J Appl Crystallogr. 1978; 11:693–694.
45. Tiede DM, Zhang R, Seifert S. Protein conformations explored by difference high-angle solution X-ray scattering: oxidation state and temperature dependent changes in cytochrome C. Biochemistry. 2002; 41:6605–6614. [PubMed: 12022864]
46. Petoukhov MV, Svergun DI. Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J. 2005; 89:1237–1250. [PubMed: 15923225]
47. Pelikan M, Hura G, Hammel M. Structure and flexibility within proteins as identified through small angle X-ray scattering. Gen Physiol Biophys. 2009; 28:174–189.
48. Maeder, M., Neuhold, Y-M. Practical data analysis in chemistry. Elsevier; 2007.
49. Malaby AW, Chakravarthy S, Irving TC, Kathuria SV, Bilsel O, Lambright DG. Methods for analysis of size-exclusion chromatography–small-angle X-ray scattering and reconstruction of protein scattering. J Appl Crystallogr. 2015; 48:1102–1113. [PubMed: 26306089]
50. Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc. 2007; 129:5656–5664. [PubMed: 17411046]
51. Tria G, Mertens HDT, Kachala M, Svergun DI. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCrJ. 2015; 2:207–217.

52. Svergun DI, Volkov VV, Kozin MB, Stuhrmann HB. New developments in direct shape determination from small-angle scattering. 2. Uniqueness. Acta Crystallogr, Sect A: Found Crystallogr. 1996; 52:419–426.
53. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J. 1999; 76:2879–2886. [PubMed: 10354416]
54. Volkov VV, Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. J Appl Crystallogr. 2003; 36:860–864.
55. Svergun DI, Petoukhov MV, Koch MHJ. Determination of domain structure of proteins from X-ray solution scattering. Biophys J. 2001; 80:2946–2953. [PubMed: 11371467]
56. Kirian RA, Wang X, Weierstall U, Schmidt KE, Spence JCH, Hunter M, Fromme P, White T, Chapman HN, Holton J. Femtosecond protein nanocrystallography—data analysis methods. Opt Express. 2010; 18:5713–5723. [PubMed: 20389587]
57. Guinier, A. X-ray diffraction in crystals, imperfect crystals, and amorphous bodies. Courier Corporation; 1994.
58. Doster W, Cusack S, Petry W. Dynamical transition of myoglobin revealed by inelastic neutron scattering. Nature. 1989; 337:754–756. [PubMed: 2918910]
59. Schomaker V, Trueblood KN. On the rigid-body motion of molecules in crystals. Acta Crystallogr, Sect B: Struct Crystallogr Cryst Chem. 1968; 24:63–76.
60. Winn MD, Isupov MN, Murshudov GN. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr, Sect D: Biol Crystallogr. 2001; 57:122–133. [PubMed: 11134934]
61. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr, Sect D: Biol Crystallogr. 2006; 62:439–450. [PubMed: 16552146]
62. Tickle, I., Moss, DS. Modelling rigid-body thermal motion in macromolecular crystal structure refinement. 1999. http://people.cryst.bbk.ac.uk/~tickle/iucr99/iucrcs99.html
63. Mizuguchi K, Kidera A, Gō N. Collective motions in proteins investigated by X-ray diffuse scattering. Proteins: Struct, Funct Bioinf. 1994; 18:34–48.
64. Wall ME, Clarage JB, Phillips GN Jr. Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering. Structure. 1997; 5:1599–1612. [PubMed: 9438860]
65. Kidera A, Gō N. Normal mode refinement: Crystallographic refinement of protein dynamic structure: I. Theory and test by simulated diffraction data. J Mol Biol. 1992; 225:457–475. [PubMed: 1593630]
66. Polikanov YS, Moore PB. Acoustic vibrations contribute to the diffuse scatter produced by ribosome crystals. Acta Crystallogr, Sect D: Biol Crystallogr. 2015; 71:2021–2031. [PubMed: 26457426]
67. Glover ID, Harris GW, Helliwell JR, Moss DS. The variety of X-ray diffuse scattering from macromolecular crystals and its respective components. Acta Crystallogr, Sect B: Struct Sci. 1991; 47:960–968.
68. Meinhold L, Merzel F, Smith JC. Lattice dynamics of a protein crystal. Phys Rev Lett. 2007; 99:138101. [PubMed: 17930640]
69. Welberry TR, Weber T. One hundred years of diffuse scattering. Crystallogr Rev. 2016; 22:2–78.
70. Huang K. X-ray reflexions from dilute solid solutions. Proc R Soc A. 1947; 190:102–117.
71. Malkin AJ, Kuznetsov YG, McPherson A. Defect structure of macromolecular crystals. J Struct Biol. 1996; 117:124–137.
72. Mueller M, Wang M, Schulze-Briese C. Optimal fine Φ-slicing for single-photon-counting pixel detectors. Acta Crystallogr, Sect D: Biol Crystallogr. 2012; 68:42–56. [PubMed: 22194332]
73. Pflugrath JW. Practical macromolecular cryocrystallography. Acta Crystallogr, Sect F: Struct Biol Commun. 2015; 71:622–642. [PubMed: 26057787]
74. 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, Sect D: Biol Crystallogr. 2015; 71:1657–1667. [PubMed: 26249347]

75. Halle B. Biomolecular cryocrystallography: Structural changes during flash-cooling. Proc Natl Acad Sci U S A. 2004; 101:4793–4798. [PubMed: 15051877]
76. Fraser JS, Bedem Hvd, 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. [PubMed: 21918110]
77. Juers DH, Matthews BW. Cryo-cooling in macromolecular crystallography: advantages, disadvantages and optimization. Q Rev Biophys. 2004; 37:105–119. [PubMed: 15999418]
78. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T. Hidden alternative structures of proline isomerase essential for catalysis. Nature. 2009; 462:669–673. [PubMed: 19956261]
79. Chapman HN, Fromme P, Barty A, White TA, Kirian RA, Aquila A, Hunter MS, Schulz J, DePonte DP, Weierstall U, et al. Femtosecond X-ray protein nanocrystallography. Nature. 2011; 470:73–77. [PubMed: 21293373]
80. Estermann MA, Steurer W. Diffuse scattering data acquisition techniques. Phase Transitions. 1998; 67:165–195.
81. Meisburger SP, Ando N. Correlated Motions from Crystallography beyond Diffraction. Acc Chem Res. 2017; 50:580–583.
82. Wall, ME. Micro and Nano Technologies in Bioanalysis: Methods and Protocols. In: Foote, RS., Lee, JW., editors. Methods in Molecular Biology. Vol. 544. Humana Press; Totowa, NJ: 2009. p. 269-279.
83. Wall, ME. PhD thesis. Princeton University; New Jersey, USA: 1996. Diffuse features in X-ray diffraction from protein crystals.
84. Benschoten AHV, 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. [PubMed: 27035972]
85. Yefanov O, Gati C, Bourenkov G, Kirian RA, White TA, Spence JCH, Chapman HN, Barty A. Mapping the continuous reciprocal space intensity distribution of X-ray serial crystallography. Phil Trans R Soc B. 2014; 369:20130333. [PubMed: 24914160]
86. Welberry TR, Heerdegen AP, Goldstone DC, Taylor IA. Diffuse scattering resulting from macromolecular frustration. Acta Crystallogr, Sect B: Struct Sci. 2011; 67:516–524.
87. Wall ME, Benschoten AHV, 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. [PubMed: 25453071]
88. Ando N, Brignole EJ, Zimanyi CM, Funk MA, Yokoyama K, Asturias FJ, Stubbe J, Drennan CL. Structural interconversions modulate activity of Escherichia coli ribonucleotide reductase. Proc Natl Acad Sci U S A. 2011; 108:21046–21051. [PubMed: 22160671]
89. Maeda H, Dudareva N. The shikimate pathway and aromatic amino acid biosynthesis in plants. Annu Rev Plant Biol. 2012; 63:73–105. [PubMed: 22554242]
90. Fitzpatrick PF. Mechanism of aromatic amino acid hydroxylation. Biochemistry. 2003; 42:14083–14091. [PubMed: 14640675]
91. Schuller DJ, Grant GA, Banaszak LJ. The allosteric ligand site in the Vmax-type cooperative enzyme phosphoglycerate dehydrogenase. Nat Struct Biol. 1995; 2:69–76. [PubMed: 7719856]
92. Aravind L, Koonin EV. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J Mol Biol. 1999; 287:1023–1040. [PubMed: 10222208]
93. Chipman DM, Shaanan B. The ACT domain family. Curr Opin Struct Biol. 2001; 11:694–700. [PubMed: 11751050]
94. Lang EJ, Cross PJ, Mittelstädt G, Jameson GB, Parker EJ. Allosteric ACTion: the varied ACT domains regulating enzymes of amino-acid metabolism. Curr Opin Struct Biol. 2014; 29:102–111. [PubMed: 25543886]
95. Herrmann KM, Weaver LM. The shikimate pathway. Annu Rev Plant Physiol Plant Mol Biol. 1999; 50:473–503. [PubMed: 15012217]
96. Jiang M, Zhang H. Engineering the shikimate pathway for biosynthesis of molecules with pharmaceutical activities in E. coli. Curr Opin Biotechnol. 2016; 42:1–6. [PubMed: 26921705]

97. Ogino T, Garner C, Markley JL, Herrmann KM. Biosynthesis of aromatic compounds: 13C NMR spectroscopy of whole Escherichia coli cells. Proc Natl Acad Sci U S A. 1982; 79:5828–5832. [PubMed: 6136965]

98. Shumilin IA, Kretsinger RH, Bauerle RH. Crystal structure of phenylalanine-regulated 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from Escherichia coli. Structure. 1999; 7:865–875. [PubMed: 10425687]

99. Cross PJ, Pietersma AL, Allison TM, Wilson-Coutts SM, Cochrane FC, Parker EJ. Neisseria meningitidis expresses a single 3-deoxy-d-arabino-heptulosonate 7-phosphate synthase that is inhibited primarily by phenylalanine. Protein Sci. 2013; 22:1087–1099. [PubMed: 23754471]

100. Shumilin IA, Bauerle R, Wu J, Woodard RW, Kretsinger RH. Crystal structure of the reaction complex of 3-deoxy-d-arabino-heptulosonate-7-phosphate synthase from Thermotoga maritima refines the catalytic mechanism and indicates a new mechanism of allosteric regulation. J Mol Biol. 2004; 341:455–466. [PubMed: 15276836]

101. Cross PJ, Dobson RCJ, Patchett ML, Parker EJ. Tyrosine latching of a regulatory gate affords allosteric control of aromatic amino acid biosynthesis. J Biol Chem. 2011; 286:10216–10224. [PubMed: 21282100]

102. Konarev PV, Volkov VV, Sokolova AV, Koch MH, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr. 2003; 36:1277–1282.

103. Cross PJ, Parker EJ. Allosteric inhibitor specificity of Thermotoga maritima 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase. FEBS Lett. 2013; 587:3063–3068. [PubMed: 23916814]

104. Cross PJ, Allison TM, Dobson RCJ, Jameson GB, Parker EJ. Engineering allosteric control to an unregulated enzyme by transfer of a regulatory domain. Proc Natl Acad Sci U S A. 2013; 110:2111–2116. [PubMed: 23345433]

105. Bashiri G, Johnston JM, Evans GL, Bulloch EMM, Goldstone DC, Jirgis ENM, Kleinboelting S, Castell A, Ramsay RJ, Manos-Turvey A, et al. Structure and inhibition of subunit I of the anthranilate synthase complex of Mycobacterium tuberculosis and expression of the active complex. Acta Crystallogr, Sect D: Biol Crystallogr. 2015; 71:2297–2308. [PubMed: 26527146]

106. Vivan AL, Caceres RA, Abrego JRB, Borges JC, Ruggiero Neto J, Ramos CHI, de Azevedo WF, Basso LA, Santos DS. Structural studies of prephenate dehydratase from Mycobacterium tuberculosis H37Rv by SAXS, ultracentrifugation, and computational analysis. Proteins. 2008; 72:1352–1362. [PubMed: 18384085]

107. Nazmi AR, Lang EJM, Bai Y, Allison TM, Othman MH, Panjikar S, Arcus VL, Parker EJ. Interdomain conformational changes provide allosteric regulation en route to chorismate. J Biol Chem. 2016; 291:21836–21847.

108. Jiao W, Hutton RD, Cross PJ, Jameson GB, Parker EJ. Dynamic cross-talk among remote binding sites: the molecular basis for unusual synergistic allostery. J Mol Biol. 2012; 415:716–726. [PubMed: 22154807]

109. Munack S, Roderer K, Ökvist M, Kamarauskaite J, Sasso S, van Eerde A, Kast P, Krengel U. Remote control by inter-enzyme allostery: a novel paradigm for regulation of the shikimate pathway. J Mol Biol. 2016; 428:1237–1255. [PubMed: 26776476]

110. Sasso S, Ökvist M, Roderer K, Gamper M, Codoni G, Krengel U, Kast P. Structure and function of a complex between chorismate mutase and DAHP synthase: efficiency boost for the junior partner. EMBO J. 2009; 28:2128–2142. [PubMed: 19556970]

111. Fitzpatrick PF. Tetrahydropterin-dependent amino acid hydroxylases. Annu Rev Biochem. 1999; 68:355–381. [PubMed: 10872454]

112. Fitzpatrick PF. Allosteric regulation of phenylalanine hydroxylase. Arch Biochem Biophys. 2012; 519:194–201. [PubMed: 22005392]

113. Flydal MI, Martinez A. Phenylalanine hydroxylase: function, structure, and regulation. IUBMB Life. 2013; 65:341–349. [PubMed: 23457044]

114. Arturo EC, Gupta K, Héroux A, Stith L, Cross PJ, Parker EJ, Loll PJ, Jaffe EK. First structure of full-length mammalian phenylalanine hydroxylase reveals the architecture of an autoinhibited tetramer. Proc Natl Acad Sci U S A. 2016; 113:2394–2399. [PubMed: 26884182]

115. Meisburger SP, Taylor AB, Khan CA, Zhang S, Fitzpatrick PF, Ando N. Domain movements upon activation of phenylalanine hydroxylase characterized by crystallography and chromatography-

coupled small-angle X-ray scattering. J Am Chem Soc. 2016; 138:6506–6516. [PubMed: 27145334]

116. Jaffe EK, Stith L, Lawrence SH, Andrake M, Dunbrack RL. A new model for allosteric regulation of phenylalanine hydroxylase: implications for disease and therapeutics. Arch Biochem Biophys. 2013; 530:73–82. [PubMed: 23296088]
117. Gjetting T, Petersen M, Guldberg P, Güttler F. Missense mutations in the N-terminal domain of human phenylalanine hydroxylase interfere with binding of regulatory phenylalanine. Am J Hum Genet. 2001; 68:1353–1360. [PubMed: 11326337]
118. Tan K, Li H, Zhang R, Gu M, Clancy ST, Joachimiak A. Structures of open (R) and close (T) states of prephenate dehydratase (PDT)–implication of allosteric regulation by L-phenylalanine. J Struct Biol. 2008; 162:94–107. [PubMed: 18171624]
119. Zhang S, Fitzpatrick PF. Identification of the allosteric site for phenylalanine in rat phenylalanine hydroxylase. J Biol Chem. 2016; 291:7418–7425. [PubMed: 26823465]
120. Patel D, Kopec J, Fitzpatrick F, McCorvie TJ, Yue WW. Structural basis for ligand-dependent dimerization of phenylalanine hydroxylase regulatory domain. Sci Rep. 2016:1–10. [PubMed: 28442746]
121. Zhang S, Roberts KM, Fitzpatrick PF. Phenylalanine binding is linked to dimerization of the regulatory domain of phenylalanine hydroxylase. Biochemistry. 2014; 53:6625–6627. [PubMed: 25299136]
122. Zhang S, Hinck AP, Fitzpatrick PF. The amino acid specificity for activation of phenylalanine hydroxylase matches the specificity for stabilization of regulatory domain dimers. Biochemistry. 2015; 54:5167–5174. [PubMed: 26252467]
123. Maeder M. Evolving factor analysis for the resolution of overlapping chromatographic peaks. Anal Chem. 1987; 59:527–530.
124. Kobe B, Jennings IG, House CM, Michell BJ, Goodwill KE, Santarsiero BD, Stevens RC, Cotton RG, Kemp BE. Structural basis of autoregulation of phenylalanine hydroxylase. Nat Struct Mol Biol. 1999; 6:442–448.
125. Li J, Dangott LJ, Fitzpatrick PF. Regulation of phenylalanine hydroxylase: conformational changes upon phenylalanine binding detected by hydrogen/deuterium exchange and mass spectrometry. Biochemistry. 2010; 49:3327–3335. [PubMed: 20307070]
126. Kantrowitz ER. Allostery and cooperativity in Escherichia coli aspartate transcarbamoylase. Arch Biochem Biophys. 2012; 519:81–90. [PubMed: 22198283]
127. Yates RA, Pardee AB. Control of pyrimidine biosynthesis in Escherichia coli by a feed-back mechanism. J Biol Chem. 1956; 221:757–770. [PubMed: 13357469]
128. Wang J, Stieglitz KA, Cardia JP, Kantrowitz ER. Structural basis for ordered substrate binding and cooperativity in aspartate transcarbamoylase. Proc Natl Acad Sci U S A. 2005; 102:8881–8886. [PubMed: 15951418]
129. Monod J, Wyman J, Changeux JP. On the nature of allosteric transitions: a plausible model. J Mol Biol. 1965; 12:88–118. [PubMed: 14343300]
130. Gerhart JC, Schachman HK. Distinct subunits for the regulation and catalytic activity of aspartate transcarbamylase. Biochemistry. 1965; 4:1054–1062. [PubMed: 5320387]
131. Weber K. New structural model of E. coli aspartate transcarbamylase and the amino-acid sequence of the regulatory polypeptide chain. Nature. 1968; 218:1116–1119. [PubMed: 4872216]
132. Wiley DC, Lipscomb WN. Crystallographic determination of symmetry of aspartate transcarbamylase. Nature. :1119–1121.
133. Moody MF, Vachette P, Foote AM. Changes in the X-ray solution scattering of aspartate transcarbamylase following the allosteric transition. J Mol Biol. 1979; 133:517–532. [PubMed: 395314]
134. Wiley DC, Evans DR, Warren SG, McMurray CH, Edwards BFP, Franks WA, Lipscomb WN. The 5.5 Å resolution structure of the regulatory enzyme, aspartate transcarbamylase. Cold Spring Harbor Symp Quant Bioly. 1972; 36:285–290.
135. Ke HM, Lipscomb WN, Cho YJ, Honzatko RB. Complex of N-phosphonacetyl-L-aspartate with aspartate carbamoyltransferase. X-ray refinement, analysis of conformational changes and catalytic and allosteric mechanisms. J Mol Biol. 1988; 204:725–747. [PubMed: 3066911]

136. Stevens RC, Gouaux JE, Lipscomb WN. Structural consequences of effector binding to the T state of aspartate carbamoyltransferase: crystal structures of the unligated and ATP-and CTP-complexed enzymes at 2. 6-Å resolution. Biochemistry. 1990; 29:7691–7701. [PubMed: 2271528]

137. Jin L, Stec B, Lipscomb WN, Kantrowitz ER. Insights into the mechanisms of catalysis and heterotropic regulation of Escherichia coli aspartate transcarbamoylase based upon a structure of the enzyme complexed with the bisubstrate analogue N-phosphonacetyl-L-aspartate at 2. 1 Å. Proteins. 1999; 37:729–742. [PubMed: 10651286]

138. Svergun DI, Barberato C, Koch MHJ, Fetler L, Vachette P. Large differences are observed between the crystal and solution quaternary structures of allosteric aspartate transcarbamylase in the R state. Proteins. 1997; 27:110–117. [PubMed: 9037716]

139. Chen, P-c, Hub, JS. Interpretation of solution X-ray scattering by explicit-solvent molecular dynamics. Biophys J. 2015; 108:2573–2584. [PubMed: 25992735]

140. Fetler L, Tauc P, Hervé G, Moody MF, Vachette P. X-ray scattering titration of the quaternary structure transition of aspartate transcarbamylase with a bisubstrate analogue: influence of nucleotide effectors. J Mol Biol. 1995; 251:243–255. [PubMed: 7643401]

141. Fetler L, Tauc P, Vachette P. Carbamyl phosphate modifies the T quaternary structure of aspartate transcarbamylase, thereby facilitating the structural transition associated with cooperativity. J Appl Crystallogr. 1997; 30:781–786.

142. Newton CJ, Kantrowitz ER. The regulatory subunit of Escherichia coli aspartate carbamoyltransferase may influence homotropic cooperativity and heterotropic interactions by a direct interaction with the loop containing residues 230–245 of the catalytic chain. Proc Natl Acad Sci U S A. 1990; 87:2309–2313. [PubMed: 2179954]

143. Fetler L, Kantrowitz ER, Vachette P. Direct observation in solution of a preexisting structural equilibrium for a mutant of the allosteric aspartate transcarbamoylase. Proc Natl Acad Sci U S A. 2007; 104:495–500. [PubMed: 17202260]

144. Macol CP, Tsuruta H, Stec B, Kantrowitz ER. Direct structural evidence for a concerted allosteric transition in Escherichia coli aspartate transcarbamoylase. Nat Struct Biol. 2001; 8:423–426. [PubMed: 11323717]

145. Fetler L, Vachette P. Revisiting the allosteric mechanism of aspartate transcarbamoylase. Nat Struct Biol. 2002; 9:87–89. [PubMed: 11813011]

146. Macol CP, Kantrowitz ER, Tsuruta H, Stec B. Response to Fetler and Vachette. Nat Struct Mol Biol. 2002; 9:88–89.

147. Cockrell GM, Zheng Y, Guo W, Peterson AW, Truong JK, Kantrowitz ER. New Paradigm for allosteric regulation of Escherichia coli aspartate transcarbamoylase. Biochemistry. 2013; 52:8036–8047. [PubMed: 24138583]

148. Fetler L, Vachette P. The allosteric activator Mg-ATP modifies the quaternary structure of the R-state of Escherichia coli aspartate transcarbamylase without altering the T<–>R equilibrium. J Mol Biol. 2001; 309:817–832. [PubMed: 11397099]

149. Schneidman-Duhovny D, Hammel M, Sali A. FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Res. 2010; 38:W540–W544. [PubMed: 20507903]

150. Tsuruta H, Sano T, Vachette P, Taue P, Moody MF, Wakabayashi K, Amemiya Y, Kimura K, Kihara H. Structural kinetics of the allosteric transition of aspartate transcarbamylase produced by physiological substrates. FEBS Lett. 1990; 263:66–68. [PubMed: 2185037]

151. Tsuruta H, Vachette P, Sano T, Moody MF, Amemiya Y, Wakabayashi K, Kihara H. Kinetics of the quaternary structure change of aspartate transcarbamylase triggered by succinate, a competitive inhibitor. Biochemistry. 1994; 33:10007–10012. [PubMed: 8060968]

152. Tsuruta H, Nagamura T, Kimura K, Igarashi Y, Kajita A, Wang ZX, Wakabayashi K, Amemiya Y, Kihara H. Stopped-flow apparatus for X-ray scattering at subzero temperature. Rev Sci Instrum. 1989; 60:2356–2358.

153. Dreyfus M, Fries J, Tauc P, Herve G. Solvent effects on allosteric equilibria: stabilization of T and R conformations of Escherichia coli aspartate transcarbamylase by organic solvents. Biochemistry. 1984; 23:4852–4859. [PubMed: 6388636]

154. West JM, Xia J, Tsuruta H, Guo W, O'Day EM, Kantrowitz ER. Time evolution of the quaternary structure of Escherichia coli aspartate transcarbamoylase upon reaction with the natural substrates and a slow, tight-binding inhibitor. J Mol Biol. 2008; 384:206–218. [PubMed: 18823998]
155. Nordlund P, Reichard P. Ribonucleotide reductases. Annu Rev Biochem. 2006; 75:681–706. [PubMed: 16756507]
156. Ingram G, Kinnaird J. Ribonucleotide reductase: a new target for antiparasite therapies. Parasitol Today. 1999; 15:338–342. [PubMed: 10407382]
157. Hertel LW, Boder GB, Kroin JS, Rinzel SM, Poore GA, Todd GC, Grindey GB. Evaluation of the antitumor activity of gemcitabine (2′, 2′-difluoro-2′-deoxycytidine). Cancer Res. 1990; 50:4417–4422. [PubMed: 2364394]
158. Bonate PL, Arthaud L, Cantrell WR, Stephenson K, Secrist JA, Weitman S. Discovery and development of clofarabine: a nucleoside analogue for treating cancer. Nat Rev Drug Discovery. 2006; 5:855–863. [PubMed: 17016426]
159. Mathews CK. DNA precursor metabolism and genomic stability. FASEB J. 2006; 20:1300–1314. [PubMed: 16816105]
160. Cotruvo JA Jr, Stubbe J. Class I ribonucleotide reductases: metallocofactor assembly and repair in vitro and in vivo. Annu Rev Biochem. 2011; 80:733–767. [PubMed: 21456967]
161. Stubbe J. Ribonucleotide reductases: the link between an RNA and a DNA world? Curr Opin Struct Biol. 2000; 10:731–736. [PubMed: 11114511]
162. Licht S, Gerfen GJ, Stubbe J. Thiyl radicals in ribonucleotide reductases. Science. 1996; 271:477–481. [PubMed: 8560260]
163. Cooperman BS, Kashlan OB. A comprehensive model for the allosteric regulation of class la ribonucleotide reductases. Adv Enzyme Regul. 2003; 43:167–182. [PubMed: 12791390]
164. Fairman JW, Wijerathna SR, Ahmad MF, Xu H, Nakano R, Jha S, Prendergast J, Welin RM, Flodin S, Roos A, et al. Structural basis for allosteric regulation of human ribonucleotide reductase by nucleotide-induced oligomerization. Nat Struct Mol Biol. 2011; 18:316–322. [PubMed: 21336276]
165. Rofougaran R, Vodnala M, Hofer A. Enzymatically active mammalian ribonucleotide reductase exists primarily as an $\alpha 6\beta 2$ octamer. J Biol Chem. 2006; 281:27705–27711. [PubMed: 16861739]
166. Rofougaran R, Crona M, Vodnala M, Sjöberg BM, Hofer A. Oligomerization status directs overall activity regulation of the Escherichia coli class Ia ribonucleotide reductase. J Biol Chem. 2008; 283:35310–35318. [PubMed: 18835811]
167. Zimanyi CM, Ando N, Brignole EJ, Asturias FJ, Stubbe J, Drennan CL. Tangled up in knots: structures of inactivated forms of E. coli class Ia ribonucleotide reductase. Structure. 2012; 20:1374–1383. [PubMed: 22727814]
168. Ando N, Li H, Brignole EJ, Thompson S, McLaughlin MI, Page JE, Asturias FJ, Stubbe J, Drennan CL. Allosteric inhibition of human ribonucleotide reductase by dATP entails the stabilization of a hexamer. Biochemistry. 2016; 55:373–381. [PubMed: 26727048]
169. Eriksson M, Uhlin U, Ramaswamy S, Ekberg M, Regnström K, Sjöberg BM, Eklund H. Binding of allosteric effectors to ribonucleotide reductase protein R1: reduction of active-site cysteines promotes substrate binding. Structure. 1997; 5:1077–1087. [PubMed: 9309223]
170. Nordlund P, Eklund H. Structure and function of the Escherichia coli ribonucleotide reductase protein R2. J Mol Biol. 1993; 232:123–164. [PubMed: 8331655]
171. Uhlin U, Eklund H. Structure of ribonucleotide reductase protein R1. Nature. 1994; 370:533–539. [PubMed: 8052308]
172. Minnihan EC, Ando N, Brignole EJ, Olshansky L, Chittuluru J, Asturias FJ, Drennan CL, Nocera DG, Stubbe J. Generation of a stable, aminotyrosyl radical-induced $\alpha 2\beta 2$ complex of Escherichia coli class Ia ribonucleotide reductase. Proc Natl Acad Sci U S A. 2013; 110:3835–3840. [PubMed: 23431160]
173. Johansson R, Jonna VR, Kumar R, Nayeri N, Lundin D, Sjöberg BM, Hofer A, Logan DT. Structural mechanism of allosteric activity regulation in a ribonucleotide reductase with double ATP cones. Structure. 2016; 24:906–917. [PubMed: 27133024]

174. Jonna VR, Crona M, Rofougaran R, Lundin D, Johansson S, Brännström K, Sjöberg BM, Hofer A. Diversity in overall activity regulation of ribonucleotide reductase. J Biol Chem. 2015; 290:17339–17348. [PubMed: 25971975]
175. Reichard P, Baldesten A, Rutberg L. Formation of deoxycytidine phosphates from cytidine phosphates in extracts from Escherichia coli. J Biol Chem. 1961; 236:1150–1157. [PubMed: 13740426]
176. Thelander L. Physicochemical characterization of ribonucleoside diphosphate reductase from Escherichia coli. J Biol Chem. 1973; 248:4591–4601. [PubMed: 4578086]
177. Brown NC, Reichard P. Ribonucleoside diphosphate reductase: formation of active and inactive complexes of proteins B1 and B2. J Mol Biol. 1969; 46:25–38. [PubMed: 4902211]
178. Hassan AQ, Stubbe J. Mapping the subunit interface of ribonucleotide reductase (RNR) using photo cross-linking. Bioorg Med Chem Lett. 2008; 18:5923–5925. [PubMed: 18762419]
179. Seyedsayamdost MR, Chan CT, Mugnaini V, Stubbe J, Bennati M. PELDOR spectroscopy with DOPA-β2 and NH_2Y-α2s: distance measurements between residues involved in the radical propagation pathway of E. coli ribonucleotide reductase. J Am Chem Soc. 2007; 129:15748–15749. [PubMed: 18047343]
180. Aye Y, Stubbe J. Clofarabine 5′-di and-triphosphates inhibit human ribonucleotide reductase by altering the quaternary structure of its large subunit. Proc Natl Acad Sci U S A. 2011; 108:9815–9820. [PubMed: 21628579]
181. Aravind L, Wolf YI, Koonin EV. The ATP-cone: an evolutionarily mobile, ATP-binding regulatory domain. J Mol Microbiol Biotechnol. 2000; 2:191–194. [PubMed: 10939243]
182. Parker MJ, Zhu X, Stubbe J. Bacillus subtilis class Ib ribonucleotide reductase: high activity and dynamic subunit interactions. Biochemistry. 2014; 53:766–776. [PubMed: 24401092]
183. Fischbach MA, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem Rev. 2006; 106:3468–3496. [PubMed: 16895337]
184. Robbins T, Liu YC, Cane DE, Khosla C. Structure and mechanism of assembly line polyketide synthases. Curr Opin Struct Biol. 2016; 41:10–18. [PubMed: 27266330]
185. Weissman KJ. The structural biology of biosynthetic megaenzymes. Nat Chem Biol. 2015; 11:660–670. [PubMed: 26284673]
186. Whicher JR, Dutta S, Hansen DA, Hale WA, Chemler JA, Dosey AM, Narayan ARH, Håkansson K, Sherman DH, Smith JL, et al. Structural rearrangements of a polyketide synthase module during its catalytic cycle. Nature. 2014; 510:560–564. [PubMed: 24965656]
187. Reimer JM, Aloise MN, Harrison PM, Schmeing TM. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. Nature. 2016; 529:239–242. [PubMed: 26762462]
188. Dutta S, Whicher JR, Hansen DA, Hale WA, Chemler JA, Congdon GR, Narayan ARH, Håkansson K, Sherman DH, Smith JL, et al. Structure of a modular polyketide synthase. Nature. 2014; 510:512–517. [PubMed: 24965652]
189. Shen B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr Opin Chem Biol. 2003; 7:285–295. [PubMed: 12714063]
190. Khosla C, Herschlag D, Cane DE, Walsh CT. Assembly line polyketide synthases: mechanistic insights and unsolved problems. Biochemistry. 2014; 53:2875–2883. [PubMed: 24779441]
191. Weissman KJ. Uncovering the structures of modular polyketide synthases. Nat Prod Rep. 2015; 32:436–453. [PubMed: 25310997]
192. Edwards AL, Matsui T, Weiss TM, Khosla C. Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase. J Mol Biol. 2014; 426:2229–2245. [PubMed: 24704088]
193. Staunton J, Caffrey P, Aparicio JF, Roberts GA, Bethell SS, Leadlay PF. Evidence for a double-helical structure for modular polyketide synthases. Nat Struct Biol. 1996; 3:188–192. [PubMed: 8564546]
194. Aparicio JF, Caffrey P, Marsden A, Staunton J, Leadlay PF. Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase. J Biol Chem. 1994; 269:8524–8528. [PubMed: 8132579]

195. Kim CY, Alekseyev VY, Chen AY, Tang Y, Cane DE, Khosla C. Reconstituting modular activity from separated domains of 6-deoxyerythronolide B synthase. Biochemistry. 2004; 43:13892–13898. [PubMed: 15518537]
196. Tang Y, Kim CY, Mathews II, Cane DE, Khosla C. The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. Proc Natl Acad Sci U S A. 2006; 103:11124–11129. [PubMed: 16844787]
197. Tang Y, Chen AY, Kim CY, Cane DE, Khosla C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. Chem Biol. 2007; 14:931–943. [PubMed: 17719492]
198. Alekseyev VY, Liu CW, Cane DE, Puglisi JD, Khosla C. Solution structure and proposed domain–domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. Protein Sci. 2007; 16:2093–2107. [PubMed: 17893358]
199. Keatinge-Clay AT, Stroud RM. The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases. Structure. 2006; 14:737–748. [PubMed: 16564177]
200. Keatinge-Clay A. Crystal structure of the erythromycin polyketide synthase dehydratase. J Mol Biol. 2008; 384:941–953. [PubMed: 18952099]
201. Rambo RP, Tainer JA. Super-resolution in solution X-ray scattering and its applications to structural systems biology. Annu Rev Biophys. 2013; 42:415–441. [PubMed: 23495971]
202. Gokhale RS, Tsuji SY, Cane DE, Khosla C. Dissecting and exploiting inter-modular communication in polyketide synthases. Science. 1999; 284:482–485. [PubMed: 10205055]
203. Wu N, Cane DE, Khosla C. Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. Biochemistry. 2002; 41:5056–5066. [PubMed: 11939803]
204. Davison J, Dorival J, Rabeharindranto H, Mazon H, Chagot B, Gruez A, Weissman KJ. Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module. Chem Sci. 2014; 5:3081–3095.
205. Herbst DA, Jakob RP, Zähringer F, Maier T. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. Nature. 2016; 531:533–537. [PubMed: 26976449]
206. Marahiel MA, Stachelhaus T, Mootz HD. Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem Rev. 1997; 97:2651–2674. [PubMed: 11851476]
207. Tanovic A, Samel SA, Essen LO, Marahiel MA. Crystal structure of the termination module of a nonribosomal peptide synthetase. Science. 2008; 321:659–663. [PubMed: 18583577]
208. Drake EJ, Miller BR, Shi C, Tarrasch JT, Sundlov JA, Allen CL, Skiniotis G, Aldrich CC, Gulick AM. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. Nature. 2016; 529:235–238. [PubMed: 26762461]
209. Haque AS, Patel KD, Deshmukh MV, Chhabra A, Gokhale RS, Sankaranarayanan R. Delineating the reaction mechanism of reductase domains of nonribosomal peptide synthetases from mycobacteria. J Struct Biol. 2014; 187:207–214. [PubMed: 25108240]
210. Chhabra A, Haque AS, Pal RK, Goyal A, Rai R, Joshi S, Panjikar S, Pasha S, Sankaranarayanan R, Gokhale RS. Nonprocessive [2+ 2] e-off-loading reductase domains from mycobacterial nonribosomal peptide synthetases. Proc Natl Acad Sci U S A. 2012; 109:5681–5686. [PubMed: 22451903]
211. Jeffries CM, Lu Y, Hynson RM, Taylor JE, Ballesteros M, Kwan AH, Trewhella J. Human cardiac myosin binding protein C: structural flexibility within an extended modular architecture. J Mol Biol. 2011; 414:735–748. [PubMed: 22041450]
212. Møller M, Nielsen SS, Ramachandran S, Li Y, Tria G, Streicher W, Petoukhov MV, Cerione RA, Gillilan RE, Vestergaard B. Small angle X-ray scattering studies of mitochondrial glutaminase C reveal extended flexible regions, and link oligomeric state with enzyme activity. PLoS One. 2013; 8:e74783. [PubMed: 24098668]
213. Mathiasen L, Valentini E, Boivin S, Cattaneo A, Blasi F, Svergun DI, Bruckmann C. The flexibility of a homeodomain transcription factor heterodimer and its allosteric regulation by DNA binding. FEBS J. 2016; 283:3134–3154. [PubMed: 27390177]

214. Konarev PV, Kachalova GS, Ryazanova AY, Kubareva EA, Karyagina AS, Bartunik HD, Svergun DI. Flexibility of the linker between the domains of DNA methyltransferase SsoII revealed by small-angle X-ray scattering: Implications for transcription regulation in SsoII restriction–modification system. PLoS One. 2014; 9:e93453. [PubMed: 24710319]
215. Kikhney AG, Svergun DI. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. FEBS Lett. 2015; 589:2570–2577. [PubMed: 26320411]
216. Vestergaard B. Analysis of biostructural changes, dynamics, and interactions–small-angle X-ray scattering to the rescue. Arch Biochem Biophys. 2016; 602:69–79. [PubMed: 26945933]
217. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North ACT. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5. 5-Å resolution, obtained by X-ray analysis. Nature. 1960; 185:416–422. [PubMed: 18990801]
218. Perutz MF. Stereochemistry of cooperative effects in haemoglobin: haem–haem interaction and the problem of allostery. Nature. 1970; 228:726–734. [PubMed: 5528785]
219. Cooper A, Dryden DTF. Allostery without conformational change. Eur Biophys J. 1984; 11:103–109. [PubMed: 6544679]
220. Popovych N, Sun S, Ebright RH, Kalodimos CG. Dynamically driven protein allostery. Nat Struct Mol Biol. 2006; 13:831–838. [PubMed: 16906160]
221. Keen DA, Goodwin AL. The crystallography of correlated disorder. Nature. 2015; 521:303–309. [PubMed: 25993960]
222. Clarage JB, Clarage MS, Phillips WC, Sweet RM, Caspar DLD. Correlations of atomic movements in lysozyme crystals. Proteins: Struct, Funct Bioinf. 1992; 12:145–157.
223. Doucet J, Benoit JP, Cruse WBT, Prange T, Kennard O. Coexistence of A-and B-form DNA in a single crystal lattice. Nature. 1989; 337:190–192. [PubMed: 2911354]
224. Doucet J, Benoit JP. Molecular dynamics studied by analysis of the X-ray diffuse scattering from lysozyme crystals. Nature. 1987; 325:643–646.
225. Faure P, Micu A, Pérahia D, Doucet J, Smith JC, Benoit JP. Correlated intramolecular motions and diffuse X-ray scattering in lysozyme. Nat Struct Mol Biol. 1994; 1:124–128.
226. Phillips GN, Fillers JP, Cohen C. Motions of tropomyosin. Crystal as metaphor. Biophys J. 1980; 32:485–502. [PubMed: 7248457]
227. Chacko S, Phillips GN Jr. Diffuse X-ray scattering from tropomyosin crystals. Biophys J. 1992; 61:1256–1266. [PubMed: 1600083]
228. Kolatkar AR, Clarage JB, Phillips GN. Analysis of diffuse scattering from yeast initiator tRNA crystals. Acta Crystallogr, Sect D: Biol Crystallogr. 1994; 50:210–218. [PubMed: 15299461]
229. Thüne T, Badger J. Thermal diffuse X-ray scattering and its contribution to understanding protein dynamics. Prog Biophys Mol Biol. 1995; 63:251–276. [PubMed: 8599030]
230. Benoit JP, Doucet J. Diffuse scattering in protein crystallography. Q Rev Biophys. 1995; 28:131–169. [PubMed: 7568674]
231. Faure P, Pérez J, Doucet J, Benoit JP. X-ray diffuse scattering and molecular dynamics in proteins. J Phys IV-C9. 1994; 4:293–298.
232. Mozzarelli A, Rossi aGL. Protein function in the crystal. Annu Rev Biophys Biomol Struct. 1996; 25:343–365. [PubMed: 8800474]
233. Artymiuk PJ, Blake CCF, Grace DEP, Oatley SJ, Phillips DC, Sternberg MJE. Crystallographic studies of the dynamic properties of lysozyme. Nature. 1979; 280:563–568. [PubMed: 460438]
234. Kundu S, Melton JS, Sorensen DC, Phillips GN. Dynamics of proteins in crystals: comparison of experiment with simple models. Biophys J. 2002; 83:723–732. [PubMed: 12124259]
235. Sternberg MJE, Grace DEP, Phillips DC. Dynamic information from protein crystallography: An analysis of temperature factors from refinement of the hen egg-white lysozyme structure. J Mol Biol. 1979; 130:231–253. [PubMed: 469942]
236. Wilson MA, Brunger AT. The 1.0 Å crystal structure of Ca^{2+}-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. J Mol Biol. 2000; 301:1237–1256. [PubMed: 10966818]

237. Wilson MA, Brunger AT. Domain flexibility in the 1.75 Å resolution structure of Pb^{2+}-calmodulin. Acta Crystallogr, Sect D: Biol Crystallogr. 2003; 59:1782–1792. [PubMed: 14501118]
238. Chaudhry C, Horwich AL, Brunger AT, Adams PD. Exploring the structural dynamics of the E. coli chaperonin GroEL using translation-libration-screw crystallographic refinement of intermediate states. J Mol Biol. 2004; 342:229–245. [PubMed: 15313620]
239. Korostelev A, Noller HF. Analysis of structural dynamics in the ribosome by TLS crystallographic refinement. J Mol Biol. 2007; 373:1058–1070. [PubMed: 17897673]
240. Diamond R. On the use of normal modes in thermal parameter refinement: theory and application to the bovine pancreatic trypsin inhibitor. Acta Crystallogr, Sect A: Found Crystallogr. 1990; 46:425–435.
241. Chen X, Poon BK, Dousis A, Wang Q, Ma J. Normal-mode refinement of anisotropic thermal parameters for potassium channel KcsA at 3. 2 Å crystallographic resolution. Structure. 2007; 15:955–962. [PubMed: 17698000]
242. Poon BK, Chen X, Lu M, Vyas NK, Quiocho FA, Wang Q, Ma J. Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42-Å crystallographic resolution. Proc Natl Acad Sci U S A. 2007; 104:7869–7874. [PubMed: 17470791]
243. Edwards C, Palmer SB, Emsley P, Helliwell JR, Glover ID, Harris GW, Moss DS. Thermal motion in protein crystals estimated using laser-generated ultrasound and Young's modulus measurements. Acta Crystallogr, Sect A: Found Crystallogr. 1990; 46:315–320.
244. Nienhaus GU, Heinzl J, Huenges E, Parak F. Protein crystal dynamics studied by time-resolved analysis of X-ray diffuse scattering. Nature. 1989; 338:665–666.
245. Parak F, Knapp EW, Kucheida D. Protein dynamics: Mössbauer spectroscopy on deoxymyoglobin crystals. J Mol Biol. 1982; 161:177–194. [PubMed: 7154076]
246. Usha MG, Speyer J, Wittebort RJ. Dynamics of the hydrate and amide groups of crystalline ribonuclease and lysozyme. Chem Phys. 1991; 158:487–500.
247. 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. [PubMed: 26436197]
248. Stocker U, Spiegel K, Gunsteren WFv. On the similarity of properties in solution or in the crystalline state: A molecular dynamics study of hen lysozyme. J Biomol NMR. 2000; 18:1–12. [PubMed: 11061223]
249. Cerutti DS, Trong IL, Stenkamp RE, Lybrand TP. Dynamics of the streptavidin-biotin complex in solution and in its crystal lattice: distinct behavior revealed by molecular simulations. J Phys Chem B. 2009; 113:6971–6985. [PubMed: 19374419]
250. Vitkup D, Ringe D, Karplus M, Petsko GA. Why protein R-factors are so large: A self-consistent analysis. Proteins: Struct, Funct Bioinf. 2002; 46:345–354.
251. York DM, Wlodawer A, Pedersen LG, Darden TA. Atomic-level accuracy in simulations of large protein crystals. Proc Natl Acad Sci U S A. 1994; 91:8715–8718. [PubMed: 7521533]
252. Cerutti DS, Le Trong I, Stenkamp RE, Lybrand TP. Simulations of a protein crystal: explicit treatment of crystallization conditions links theory and experiment in the streptavidin-biotin complex. Biochemistry. 2008; 47:12065–12077. [PubMed: 18950193]
253. Cerutti DS, Freddolino PL, Duke RE, Case DA. Simulations of a protein crystal with a high resolution X-ray structure: evaluation of force fields and water models. J Phys Chem B. 2010; 114:12811–12824. [PubMed: 20860388]
254. Hafner J, Zheng W. All-atom modeling of anisotropic atomic fluctuations in protein crystal structures. J Chem Phys. 2011; 135:144114. [PubMed: 22010705]
255. Liu C, Janowski PA, Case DA. All-atom crystal simulations of DNA and RNA duplexes. Biochim Biophys Acta, Gen Subj. 2015; 1850:1059–1071.
256. Janowski PA, Liu C, Deckman J, Case DA. Molecular dynamics simulation of triclinic lysozyme in a crystal lattice. Protein Sci. 2016; 25:87–102. [PubMed: 26013419]
257. Blake CCF, Koenig DF, Mair GA, North ACT, Phillips DC, Sarma VR. Structure of hen egg-white lysozyme: a three-dimensional Fourier synthesis at 2 Å resolution. Nature. 1965; 206:757–761. [PubMed: 5891407]

258. Phillips DC. The hen egg-white lysozyme molecule. Proc Natl Acad Sci U S A. 1967; 57:483–495.

259. Mccammon JA, Gelin BR, Karplus M, Wolynes PG. The hinge-bending mode in lysozyme. Nature. 1976; 262:325–326. [PubMed: 958384]

260. Héry S, Genest D, Smith JC. X-ray diffuse scattering and rigid-body motion in crystalline lysozyme probed by molecular dynamics simulation. J Mol Biol. 1998; 279:303–319. [PubMed: 9636718]

261. Kuriyan J, Ösapay K, Burley SK, Brünger AT, Hendrickson WA, Karplus M. Exploration of disorder in protein structures by X-ray restrained molecular dynamics. Proteins: Struct, Funct Bioinf. 1991; 10:340–358.

262. Kraft P, Bergamaschi A, Broennimann C, Dinapoli R, Eikenberry EF, Henrich B, Johnson I, Mozzanica A, Schlepütz CM, Willmott PR, et al. Performance of single-photon-counting PILATUS detector modules. J Synchrotron Radiat. 2009; 16:368–375. [PubMed: 19395800]

263. Meinhold L, Smith JC. Fluctuations and correlations in crystalline protein dynamics: a simulation analysis of Staphylococcal nuclease. Biophys J. 2005; 88:2554–2563. [PubMed: 15681654]

264. Anfinsen CB. Principles that govern the folding of protein chains. Science. 1973; 181:223–230. [PubMed: 4124164]

265. Panick G, Malessa R, Winter R, Rapp G, Frye KJ, Royer CA. Structural characterization of the pressure-denatured state and unfolding/refolding kinetics of staphylococcal nuclease by synchrotron small-angle X-ray scattering and Fourier-transform infrared spectroscopy. J Mol Biol. 1998; 275:389–402. [PubMed: 9466917]

266. Roche J, Dellarole M, Caro JA, Guca E, Norberto DR, Yang Y, Garcia AE, Roumestand C, García-Moreno B, Royer CA. Remodeling of the folding free energy landscape of Staphylococcal nuclease by cavity-creating mutations. Biochemistry. 2012; 51:9535–9546. [PubMed: 23116341]

267. Arnone A, Bier CJ, Cotton FA, Day VW, Hazen EE, Richardson DC, Richardson J, Yonath A. A high resolution structure of an inhibitor complex of the extracellular nuclease of Staphylococcus aureus I. Experimental procedures and chain tracing. J Biol Chem. 1971; 246:2302–2316. [PubMed: 5555571]

268. Cotton FA, Hazen EE, Legg MJ. Staphylococcal nuclease: Proposed mechanism of action based on structure of enzyme—thymidine 3′,5′-bisphosphate—calcium ion complex at 1.5-Å resolution. Proc Natl Acad Sci U S A. 1979; 76:2551–2555. [PubMed: 288045]

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

270. Yuan T, Ouyang H, Vogel HJ. Surface exposure of the methionine side chains of calmodulin in solution. A nitride spin label and two-dimensional NMR study. J Biol Chem. :8411–8420.

271. Vetter SW, Leclerc E. Novel aspects of calmodulin target recognition and activation. Eur J Biochem. 2003; 270:404–414. [PubMed: 12542690]

272. Stratton MM, Chao LH, Schulman H, Kuriyan J. Structural studies on the regulation of Ca^{2+}/calmodulin dependent protein kinase II. Curr Opin Struct Biol. 2013; 23:292–301. [PubMed: 23632248]

273. Meador WE, Means AR, Quiocho FA. Modulation of calmodulin plasticity in molecular recognition on the basis of X-ray structures. Science. 1993; 262:1718–1721. [PubMed: 8259515]

274. McLoughlin SY, Jackson C, Liu JW, Ollis DL. Growth of Escherichia coli coexpressing phosphotriesterase and glycerophosphodiester phosphodiesterase, using paraoxon as the sole phosphorus source. Appl Environ Microbiol. 2004; 70:404–412. [PubMed: 14711669]

275. Ghanem E, Li Y, Xu C, Raushel FM. Characterization of a phosphodiesterase capable of hydrolyzing EA 2192; the most toxic degradation product of the nerve agent VX. Biochemistry. 2007; 46:9032–9040. [PubMed: 17630782]

276. Daumann LJ, Larrabee JA, Ollis D, Schenk G, Gahan LR. Immobilization of the enzyme GpdQ on magnetite nanoparticles for organophosphate pesticide bioremediation. J Inorg Biochem. 2014; 131:1–7. [PubMed: 24239906]

277. Jackson CJ, Carr PD, Liu JW, Watt SJ, Beck JL, Ollis DL. The structure and function of a novel glycerophosphodiesterase from Enterobacter aerogenes. J Mol Biol. 2007; 367:1047–1062. [PubMed: 17306828]

278. Jackson CJ, Carr PD, Kim HK, Liu JW, Ollis DL. The purification, crystallization and preliminary diffraction of a glycerophosphodiesterase from *Enterobacter aerogenes*. Acta Crystallogr, Sect F: Struct Biol Cryst Commun. 2006; 62:659–661.

279. Yip SHC, Foo JL, Schenk G, Gahan LR, Carr PD, Ollis DL. Directed evolution combined with rational design increases activity of GpdQ toward a non-physiological substrate and alters the oligomeric structure of the enzyme. Protein Eng, Des Sel. 2011; 24:861–872. [PubMed: 21979136]

280. Urzhumtsev A, Afonine PV, Van Benschoten AH, Fraser JS, Adams PD. From deep TLS validation to ensembles of atomic models built from elemental motions. Acta Crystallogr, Sect D: Biol Crystallogr. 2015; 71:1668–1683. [PubMed: 26249348]

281. Wedemeyer WJ, Welker E, Scheraga HA. Proline cis-trans isomerization and protein folding. Biochemistry. 2002; 41:14637–14644. [PubMed: 12475212]

282. Liu J, Farmer JD, Lane WS, Friedman J, Weissman I, Schreiber SL. Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. Cell. 1991; 66:807–815. [PubMed: 1715244]

283. Wang P, Heitman J. The cyclophilins. Genome Biol. 2005; 6:226. [PubMed: 15998457]

284. Gamble TR, Vajdos FF, Yoo S, Worthylake DK, Houseweart M, Sundquist WI, Hill CP. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. Cell. 1996; 87:1285–1294. [PubMed: 8980234]

285. Liu C, Perilla JR, Ning J, Lu M, Hou G, Ramalho R, Himes BA, Zhao G, Bedwell GJ, Byeon IJ, et al. Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. Nat Commun. 2016; 7:10714. [PubMed: 26940118]

286. 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. [PubMed: 23913260]

287. Wilson MA. Visualizing networks of mobility in proteins. Nat Methods. 2013; 10:835–837. [PubMed: 23985728]

288. Riccardi D, Cui Q, Phillips GN Jr. Evaluating elastic network models of crystalline biological molecules with temperature factors, correlated Motions, and diffuse X-ray scattering. Biophys J. 2010; 99:2616–2625. [PubMed: 20959103]

289. López-Blanco JR, Chacón P. New generation of elastic network models. Curr Opin Struct Biol. 2016; 37:46–53. [PubMed: 26716577]

290. Welberry, TR., Heerdegen, AP., Carr, PD. Aperiodic Crystals. Schmid, S.Withers, RL., Lifshitz, R., editors. Springer Netherlands; Dordrecht: 2013. p. 243-251.

291. Frank J, Agrawal RK. A ratchet-like inter-subunit reorganization of the ribosome during translocation. Nature. 2000; 406:318–322. [PubMed: 10917535]

292. Neutze R, Moffat K. Time-resolved structural studies at synchrotrons and X-ray free electron lasers: opportunities and challenges. Curr Opin Struct Biol. 2012; 22:651–659. [PubMed: 23021004]

293. Yano J, Kern J, Irrgang KD, Latimer MJ, Bergmann U, Glatzel P, Pushkar Y, Biesiadka J, Loll B, Sauer K, et al. X-ray damage to the Mn4Ca complex in single crystals of photosystem II: a case study for metalloprotein crystallography. Proc Natl Acad Sci U S A. 2005; 102:12047–12052. [PubMed: 16103362]

294. Grabolle M, Haumann M, Müller C, Liebisch P, Dau H. Rapid loss of structural motifs in the manganese complex of oxygenic photosynthesis by X-ray irradiation at 10–300 K. J Biol Chem. 2006; 281:4580–4588. [PubMed: 16352605]

295. Shen JR. The structure of photosystem II and the mechanism of water oxidation in photosynthesis. Annu Rev Plant Biol. 2015; 66:23–48. [PubMed: 25746448]

296. Kok B, Forbush B, McGloin M. Cooperation of charges in photosynthetic O2 Evolution–I. a linear four step mechanism. Photochem Photobiol. 1970; 11:457–475. [PubMed: 5456273]

297. Suga M, Akita F, Hirata K, Ueno G, Murakami H, Nakajima Y, Shimizu T, Yamashita K, Yamamoto M, Ago H, et al. Native structure of photosystem II at 1. 95 Å resolution viewed by femtosecond X-ray pulses. Nature. 2015; 517:99–103. [PubMed: 25470056]

298. Kern J, Alonso-Mori R, Tran R, Hattne J, Gildea RJ, Echols N, Glöckner C, Hellmich J, Laksmono H, Sierra RG, et al. Simultaneous femtosecond X-ray spectroscopy and diffraction of photosystem II at room temperature. Science. 2013; 340:491–495. [PubMed: 23413188]

299. Kupitz C, Basu S, Grotjohann I, Fromme R, Zatsepin NA, Rendek KN, Hunter MS, Shoeman RL, White TA, Wang D, et al. Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. Nature. 2014; 513:261–265. [PubMed: 25043005]

300. Sauter NK, Echols N, Adams PD, Zwart PH, Kern J, Brewster AS, Koroidov S, Alonso-Mori R, Zouni A, Messinger J, et al. No observable conformational changes in PSII. Nature. 2016; 533:E1–E2. [PubMed: 27193689]

301. Kuhl H, Kruip J, Seidler A, Krieger-Liszkay A, Bünker M, Bald D, Scheidig AJ, Rögner M. Towards structural determination of the water-splitting enzyme. J Biol Chem. 2000; 275:20652–20659. [PubMed: 10748017]

302. Ayyer K, Yefanov OM, Oberthür 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. [PubMed: 26863980]

303. Elser V, Millane RP. Reconstruction of an object from its symmetry-averaged diffraction pattern. Acta Crystallogr, Sect A: Found Crystallogr. 2008; 64:273–279.

304. Elser V. Phase retrieval by iterated projections. J Opt Soc Am A. 2003; 20:40–55.

305. Levantino M, Yorke BA, Monteiro DC, Cammarata M, Pearson AR. Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules. Curr Opin Struct Biol. 2015; 35:41–48. [PubMed: 26342489]

306. Knight J, Vishwanath A, Brody J, Austin RH. Hydrodynamic focusing on a silicon chip: mixing nanoliters in microseconds. Phys Rev Lett. 1998; 80:3863–3866.

307. Pollack L, Tate MW, Darnton NC, Knight J, Gruner SM, Eaton WA, Austin RH. Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle X-ray scattering. Proc Natl Acad Sci U S A. 1999; 96:10115–10117. [PubMed: 10468571]

308. Pollack L, Tate MW, Finnefrock AC, Kalidas C, Trotter S, Darnton NC, Lurio L, Austin RH, Batt CA, Gruner SM, et al. Time resolved collapse of a folding protein observed with small angle X-ray scattering. Phys Rev Lett. 2001; 86:4962–4965. [PubMed: 11384392]

309. Cho HS, Dashdorj N, Schotte F, Graber T, Henning R, Anfinrud P. Protein structural dynamics in solution unveiled via 100-ps time-resolved X-ray scattering. Proc Natl Acad Sci U S A. 2010; 107:7281–7286. [PubMed: 20406909]

310. Levantino M, Schirò G, Lemke HT, Cottone G, Glownia JM, Zhu D, Chollet M, Ihee H, Cupane A, Cammarata M. Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. Nat Commun. 2015; 6:6772. [PubMed: 25832715]

311. Doniach S. Changes in biomolecular conformation seen by small angle X-ray scattering. Chem Rev. 2001; 101:1763–1778. [PubMed: 11709998]

312. Pérez J, Nishino Y. Advances in X-ray scattering: from solution SAXS to achievements with coherent beams. Curr Opin Struct Biol. 2012; 22:670–678. [PubMed: 22954648]

313. Mertens HDT, Svergun DI. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. J Struct Biol. 2010; 172:128–141. [PubMed: 20558299]

314. Kabsch W. Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. J Appl Crystallogr. 1988; 21:916–924.

315. Kell GS. Isothermal compressibility of liquid water at 1 atm. J Chem Eng Data. 1970; 15:119–122.

Biographies

Nozomi Ando was born in Denver, grew up in Boston, and received her undergraduate degree in physics from MIT. She received her Ph.D. in physics from Cornell University,

where she worked with Sol Gruner and developed a love of X-ray scattering and synchrotrons. As an NIH postdoctoral fellow in Catherine Drennan's group at MIT, she became fascinated by metalloenzymes. Nozomi is currently an Assistant Professor in Chemistry at Princeton University, where her research program focuses on using non-conventional X-ray methods to study how enzymes work.

Steve P. Meisburger grew up in sunny San José, California. Following his undergraduate studies in physics at Carleton College, Steve received his Ph.D. in applied physics at Cornell University. As a graduate student in Lois Pollack's group, he studied ion interactions with RNA/DNA with time-resolved, cryo-, and anomalous X-ray scattering. Steve is currently an NIH postdoctoral fellow in the Ando group, where he is studying protein dynamics with diffuse scattering.

William C. Thomas was born in New Orleans and raised in Cleveland, Ohio. He received his undergraduate degrees in Biochemistry, Chemistry, and Biology from Oberlin College, where he conducted research in atmospheric chemistry in the group of Matthew Elrod. As a graduate student in the Ando group, he is studying allosteric enzymes with solution scattering methods.

Maxwell B. Watkins was born in Allentown, Pennsylvania and grew up in Emmaus, Pennsylvania. He received his undergraduate degree in Biochemistry from Lehigh University, where he studied membrane proteins in the group of Daniel Thévenin. As a graduate student in the Ando group, he is studying flexible enzymes with solution scattering methods.

Appendix A: Supplement to Physics of X-ray Scattering Section

A.1 Thomson Scattering

In the low-energy limit, the Klein-Nishina formula for the scattering of a photon by a free electron reduces to what is referred to as Thomson scattering in classical electrodynamics. In the classical theory, when an electromagnetic plane wave traveling in a direction $\hat{\mathbf{s}}$ is incident on an electron, the electron oscillates in the direction $\hat{\mathbf{e}}$ of the electric field ($\hat{\mathbf{e}} \perp \hat{\mathbf{s}}$), and emits X-rays in a dipole pattern. X-rays scattered in a direction $\hat{\mathbf{s}}'$ are recorded by a detector with solid angle $\Delta\Omega$ at an average rate[18]

$$J(\hat{\mathbf{s}}') = J_0 r_e^2 \Delta\Omega \left(1 - \left(\hat{\mathbf{e}} \cdot \hat{\mathbf{s}}'\right)^2\right), \quad (155)$$

where J_0 is the incident flux (photons per unit area per second), and r_e is the classical electron radius, $r_e = e^2/(4\pi\varepsilon_0 m_e c^2) = 2.818 \times 10^{-13}$ cm. The geometric term appearing in parentheses accounts for the dipole radiation pattern. The scattering from a free electron can be expressed equivalently as a differential scattering cross-section ($d\sigma/d\Omega$), equal to the scattering rate divided by the incident flux and the solid angle

$$\left(\frac{d\sigma}{d\Omega}\right)_{\mathrm{e}}=\frac{J\left(\hat{\mathbf{s}}'\right)}{J_0\Delta\Omega}=r_e^2\left(1-\left(\hat{\mathbf{e}}\cdot\hat{\mathbf{s}}'\right)^2\right). \quad (156)$$

The classical scattering cross-section of a single electron does not depend on the wavelength, λ, and when integrated over solid angle, has a fixed value of $\sigma=8\pi r_0^2/3$.

It is often the case the a radiation source will not be completely polarized in a given direction, and Equation 156 must be modified accordingly. The scattering can be written explicitly as a linear combination of orthogonal polarization directions. Let $\hat{\mathbf{n}}$ be the vector normal to the plane of polarization ($\hat{\mathbf{n}}\perp\hat{\mathbf{e}}\perp\hat{\mathbf{s}}$), and p the fraction of in-plane photons. In Equation 156, we can rewrite $P(\hat{\mathbf{s}}',\hat{\mathbf{s}})\equiv\langle 1-(\hat{\mathbf{e}}\cdot\hat{\mathbf{s}}')^2\rangle_{\hat{\mathbf{e}}}$ as[314]

$$\begin{aligned}P\left(\hat{\mathbf{s}}',\hat{\mathbf{s}}\right)&=p\left(1-\left(\hat{\mathbf{s}}'\cdot(\hat{\mathbf{s}}\times\hat{\mathbf{n}})\right)^2\right)+(1-p)\left(1-\left(\hat{\mathbf{s}}'\cdot\hat{\mathbf{n}}\right)^2\right)\\&=p\left(1+\left(\hat{\mathbf{s}}'\cdot\hat{\mathbf{s}}\right)^2\right)+(1-2p)\left(1-\left(\hat{\mathbf{s}}'\cdot\hat{\mathbf{n}}\right)^2\right).\end{aligned} \quad (157)$$

When $p = 0.5$ we have an unpolarized source (such as an X-ray tube), whose polarization factor depends only on the scattering angle 2θ; $P=\frac{1}{2}(1+\cos^2(2\theta))$. When $p = 1$ we have a completely polarized source, such as a synchrotron. Synchrotron radiation is polarized in the plane of the ring ($\hat{n}$ points in the vertical direction), with $P = 1$ in the vertical scattering plane and $P = \cos^2(2\theta)$ in the horizontal scattering plane.

A.2 Resonant Scattering

Whereas the Thomson scattering cross-section of a free electron is the same regardless of X-ray energy, the atomic scattering factors depend on the X-ray photon energy $\hbar\omega$,[18]

$$f_n(\mathbf{q},\hbar\omega)=f_n^0(\mathbf{q})+\{f_n'(\hbar\omega)+i\,f_n''(\hbar\omega)\}. \quad (158)$$

The energy-dependent term $\{f'(\hbar\omega) + i\,f''(\hbar\omega)\}$ corrects for resonant scattering, and is significant in the vicinity of an X-ray absorption edge of the atom. The energy-independent term, $f_n^0(\mathbf{q})$, is the Fourier transform of the electron density around the atom

$$f_n^0(\mathbf{q})=\int_V\rho_n(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}d^3\mathbf{r}, \quad (159)$$

and at $\mathbf{q} = \mathbf{0}$, is equal to the number of electrons bound; $f_n^0(0)=Z_n$.

Resonant scattering plays an important role in several X-ray scattering techniques, such as *anomalous phasing* in MX and *anomalous contrast variation* in SAXS. In addition, X-ray Absorption Fine Structure (XAFS) measurements can be used to probe the energy-dependent oscillations in $f''(\hbar\omega)$, which are sensitive to the local environment of the resonant atom. In an anomalous scattering or diffraction experiment, one or more X-ray energies are chosen near the absorption edge for an atom of interest. To design the experiment and analyze the data, it is usually necessary to determine f' and f'' experimentally. First, f'' is measured by performing an XAFS experiment. Then, f' is calculated from f'' using the Kramers-Kronig relation,[18]

$$f'(\hbar\omega) = \frac{2}{\pi}\mathscr{P}\int_0^\infty \frac{\omega' f''(\hbar\omega')}{(\omega'^2 - \omega^2)} d\omega' \quad (160)$$

where $\mathscr{P}$ indicates that the Cauchy principal value of the integral is taken.

Although resonant scattering is important in many techniques, it is not central to the applications discussed in this review. Therefore, for notational convenience, we have assumed that the resonant contributions are small, making the approximation $f_n(\mathbf{q}, \hbar\omega) \approx f_n^0(\mathbf{q})$. The consequence of this approximation is that the molecular form factor description (Equation 7) is equivalent to the electron density description involving the Fourier transform of $\rho(\mathbf{r})$. Resonant scattering can be added back in to the theory by considering $\rho(\mathbf{r})$ to be a complex-valued quantity. Note that the autocorrelation function (Equation 9) should be modified in this case,

$$(\rho \star \rho)(\mathbf{r}) = \int \rho^*(\mathbf{u})\rho(\mathbf{r}+\mathbf{u})\, d^3\mathbf{u}, \quad (161)$$

where * is the complex conjugate. In addition, the number of electrons N_e in volume V, introduced in Section 2.1.2, would also be complex-valued, according to the definition,

$$N_e \equiv \int_V \rho(\mathbf{r})\, d^3\mathbf{r}. \quad (162)$$

Otherwise, the derivations of Section 2.1 require only minor modifications to account for a complex valued $\rho(\mathbf{r})$.

A.3 Liquid Structure Factor

The correlation function of a molecular liquid can be derived using a similar argument to that presented in Section 2.1.2. First, consider a one-component system composed of N_p identical, spherically symmetric particles with scattering factor $f_p(q)$. The ensemble average of the squared form factor is

$$\left\langle |F(\mathbf{q})|^2 \right\rangle_{\Pi} = \left\langle \sum_{n=1}^{N_p} \sum_{m=1}^{N_p} |f_p(|\mathbf{q}|)|^2 \, e^{i\mathbf{q}\cdot(\mathbf{r}_n - \mathbf{r}_m)} \right\rangle_{\Pi}. \quad (163)$$

Following the derivation leading to Equation 14, the intensity (Equation 17) is

$$I(\mathbf{q}) = \langle N_p \rangle \, |f_p(|\mathbf{q}|)|^2 \left\{ 1 + \int_V (\eta_p(\mathbf{r}) - n_p) \, e^{i\mathbf{q}\cdot\mathbf{r}} d^3\mathbf{r} \right\}, \quad (164)$$

where n_p is the average number density of particles, and $\eta_p(\mathbf{r})$ refers to the probability that a pair of particles are separated by a distance $\mathbf{r}$, which for a liquid, depends only on the distance $r = |\mathbf{r}|$ between them. Therefore, a pair correlation function $g(r) = \eta_p(\mathbf{r})/n_p$ is introduced, and the integral is changed to spherical coordinates using Equation 26. Then, Equation 164 can be written,

$$I(q) = \langle N_p \rangle \, |f_p(q)|^2 S(q), \quad (165)$$

where $S(q)$ is the structure factor,

$$S(q) = 1 - 4\pi n_p \int_0^\infty r^2 (1 - g(r)) \frac{\sin(qr)}{qr} dr. \quad (166)$$

The structure factor and the pair correlation function $g(r)$ are related by a Fourier-Bessel transform, which can be inverted to give $g(r)$ in terms of $S(q)$,

$$g(r) = 1 + \frac{1}{2\pi^2 n_p} \int_0^\infty q^2 (S(q) - 1) \frac{\sin(qr)}{qr} dq. \quad (167)$$

Forward scattering

For single-component liquids, the number of scattering electrons is $N_e = N_p \, |f_p(\mathbf{0})|$, and therefore Equation 19 relates the structure factor at $\mathbf{q} = \mathbf{0}$ to the variance in the number of particles, N_p,

$$S(0) = \frac{\left\langle (N_p - \langle N_\mathrm{p} \rangle)^2 \right\rangle}{\langle N_\mathrm{p} \rangle}. \quad (168)$$

For a non-interacting system (an ideal gas), the variance in the number of particles should equal the mean (according to Poisson statistics), so by Equation 168, $S(0) = 1$. Strongly interacting systems, such as liquids, have values of $S(0)$ that differ from unity. According to kinetic theory, the variance of the number of molecules is related to the isothermal compressibility, χ_T

$$\chi_T = \frac{1}{n_p}\left(\frac{\partial n_p}{\partial P}\right)_T = \frac{1}{n_p k_B T}\frac{\left\langle (N_\mathrm{p} - \langle N_\mathrm{p}\rangle)^2 \right\rangle}{\langle N_\mathrm{p}\rangle} \quad (169)$$

where n_p is the average number density, P is the pressure, T is the temperature, and k_B is the Boltzmann constant. Combining Equations 168 and 169,

$$S(0) = n_p\, k_B T \chi_T. \quad (170)$$

Liquid Water

Liquid water is omnipresent in biomolecular solutions and crystals, and its scattering contribution is important to consider. The forward scattering of liquid water can be calculated using its isothermal compressibility, which has been measured accurately using the speed of sound.[315] At 22 °C and atmospheric pressure, $\chi_T = 4.56 \times 10^{-10}$ Pa^{-1}, so by Equation 170,

$$(S(0))_{\mathrm{water},22^\circ\mathrm{C}} = 0.062. \quad (171)$$

The forward scattering cross-section per unit volume of a one-component liquid is

$$\frac{1}{V}\left(\frac{d\sigma(0)}{d\Omega}\right) = r_e^2 |f_p(0)|^2 S(0) \quad (172)$$

With $f_p(\mathbf{0}) = 10$ electrons for water and $r_e = 2.818 \times 10^{-13}$ cm,

$$\frac{1}{V}\left(\frac{d\sigma(0)}{d\Omega}\right)_{\mathrm{water},22^\circ\mathrm{C}} = 0.0164\,\mathrm{cm}^{-1} \quad (173)$$

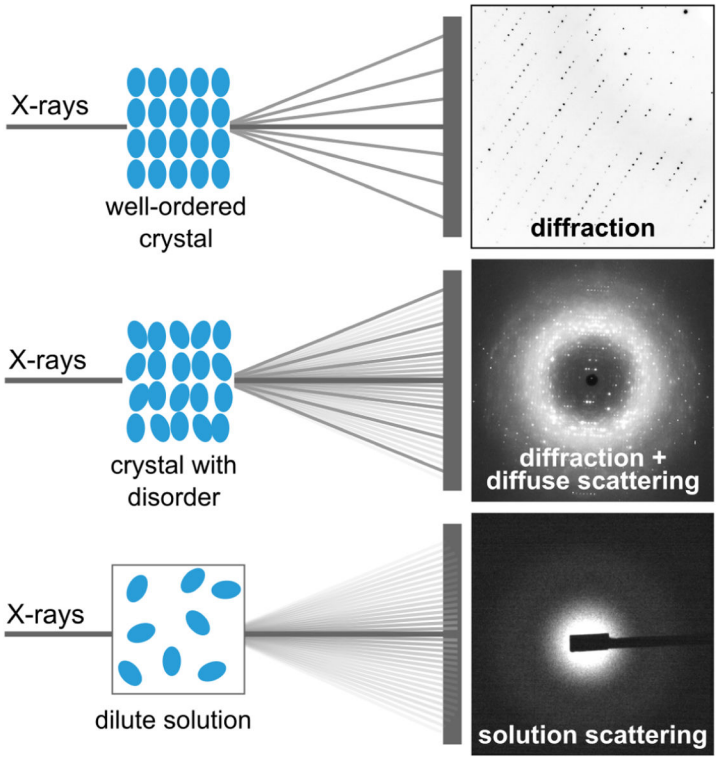


Figure 1.

Diffraction, diffuse scattering, and SAXS result from the elastic scattering of X-rays when they interact with electrons. The characteristic patterns they produce on a detector result from interference of the scattered waves, and they reflect the degree of disorder present in the sample. The diffuse scattering image was adapted with permission from Ref. 1. Copyright 1997 United States National Academy of Sciences.

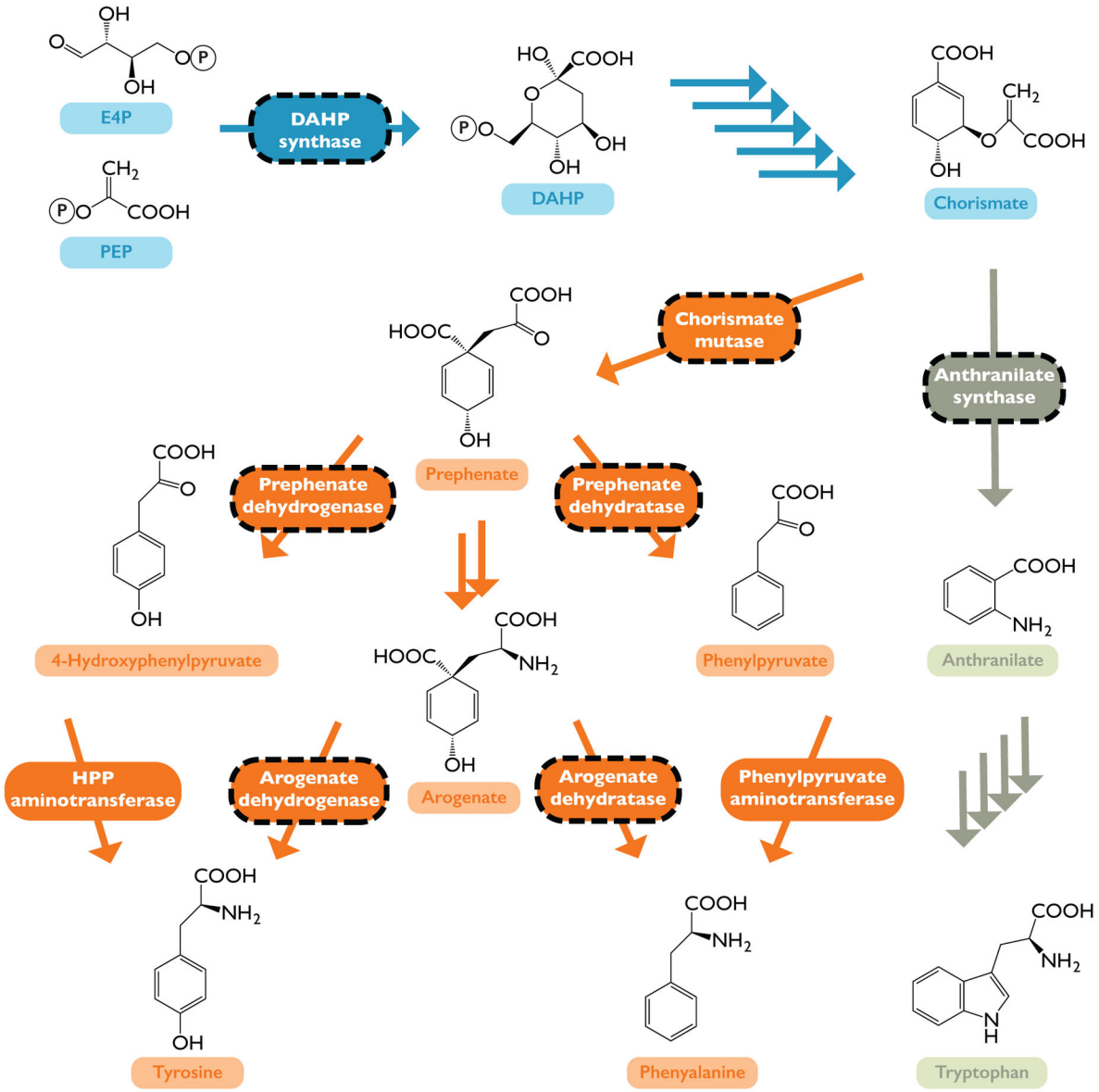


Figure 2.

The aromatic amino acid (AAA) biosynthetic pathway of plants and microorganisms is tightly regulated. The many regulated enzymes (black dotted lines) all share a common trait in that they are located at junction points within the pathway. All three AAAs originate from chorismate, synthesized via the shikimate pathway (blue). Phenylalanine and tyrosine are then formed from a branching pathway that begins with chorismate mutase (orange), while tryptophan is synthesized from a pathway beginning with anthranilate synthase (green).

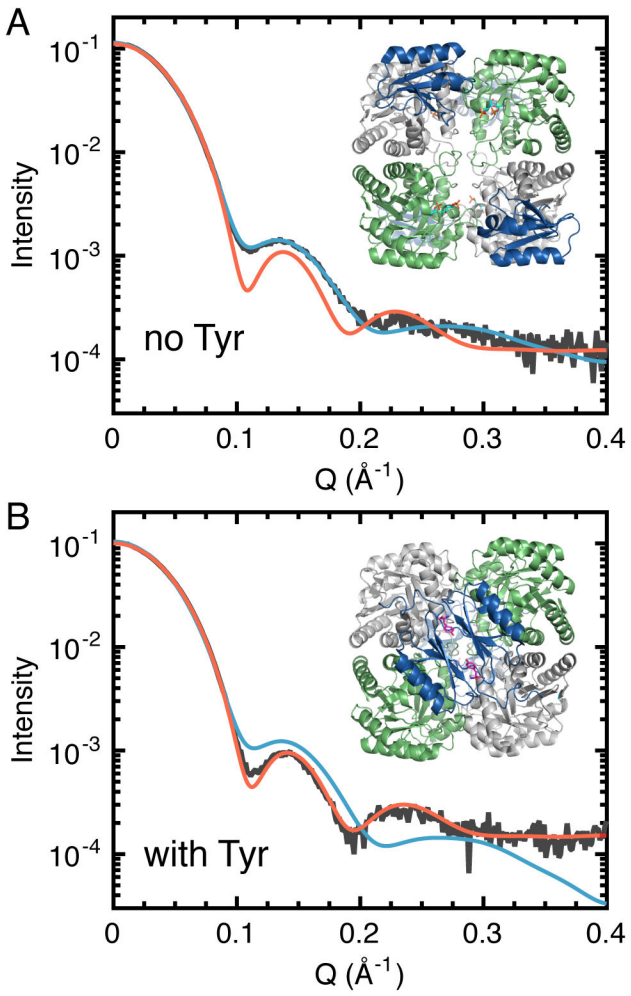


Figure 3.

SEC-SAXS studies reveal that *Tma*DAHP synthase is regulated by Tyr, an end-product of the AAA biosynthetic pathway, via the dimerization of ACT domains (blue domains in insets).[101] (A) The experimental scattering of *Tma*DAHP synthase in the absence of tyrosine was fitted via CRYSOL[2] to the crystal structure of the ligand-free enzyme in an open state (inset, PDB: 1RZM)[100] and found to be in close agreement. (B) Conversely, *Tma*DAHP synthase in the presence of Tyr was found to be better modeled by the Tyr-bound closed state (inset, PDB: 3PG9). In both plots, experimental data are in black, the theoretical profile of the open state is in blue, and the theoretical profile of the closed state is in orange. Adapted with permission from Ref. 101. Copyright 2011 American Society for Biochemistry and Molecular Biology.

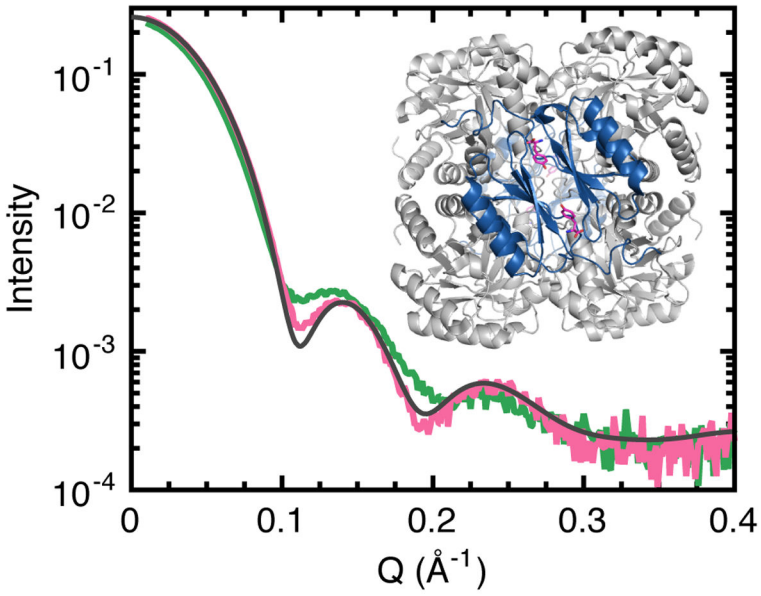


Figure 4.

SEC-SAXS studies show that allosteric regulation by Tyr can be conferred by fusing the ACT domain of *Tma*DAHP synthase to the normally unregulated *P. furiosa* enzyme.[104] Addition of Tyr to this chimeric construct leads to a change in scattering (green to pink), indicative of a structural change. The experimental scattering in the presence of Tyr is in close agreement to the theoretical profile calculated in CRYSOL[2] from the crystal structure of the Tyr-bound construct in the closed state (inset, PDB: 4GRS). Adapted with permission from Ref. 104. Copyright 2013 United States National Academy of Sciences.

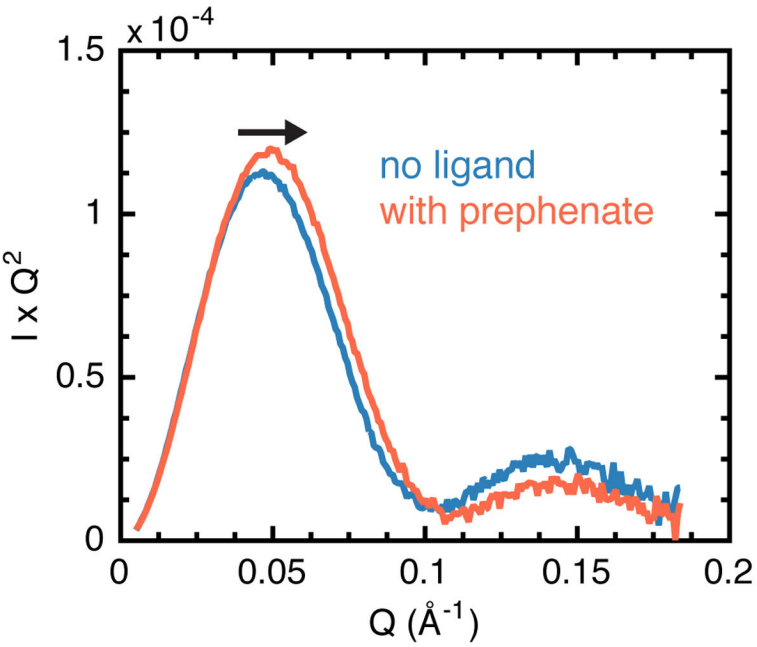


Figure 5.

SEC-SAXS studies reveal that the product of CM, prephenate, causes a tighter association between the DAHP synthase and CM domains of the DAHP synthase-CM fusion protein of *Geobacillus sp.*[107] This subtle compaction is best seen in the Kratky plot, where the main peak shifts to the right (blue to red) in the presence of prephenate (red line). Adapted with permission from Ref. 107. Copyright 2016 American Society for Biochemistry and Molecular Biology.

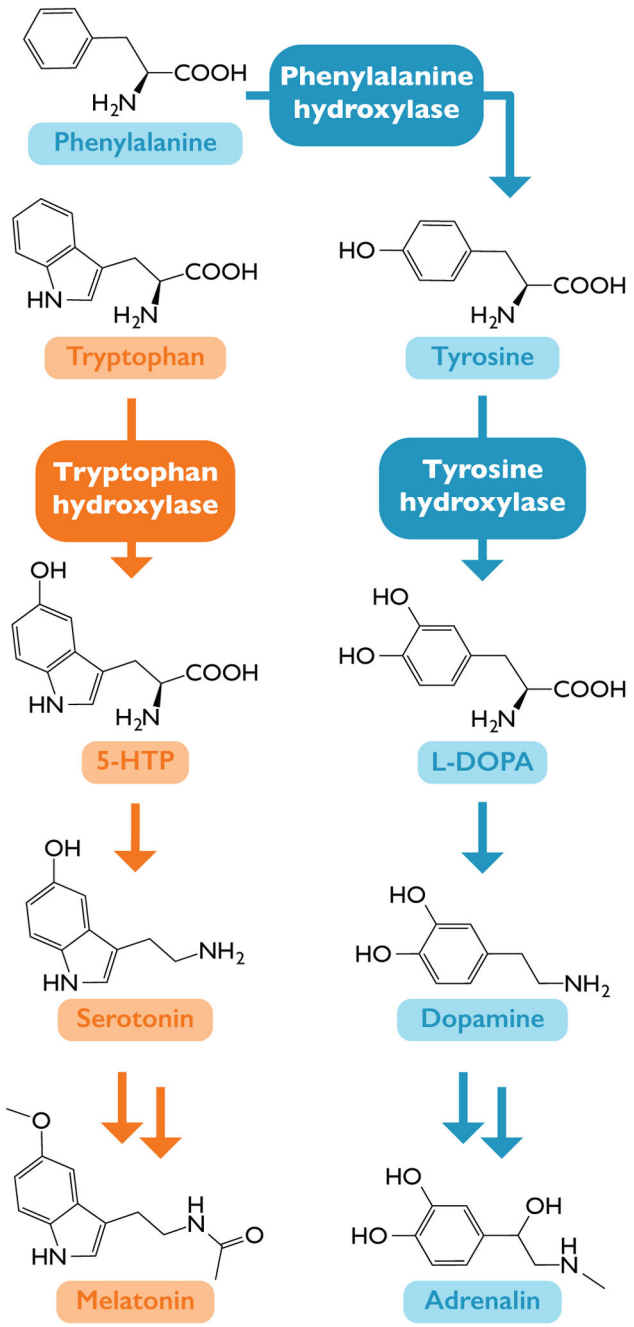


Figure 6.

The three AAAs are directed to the formation of important neurotransmitters in mammals, and the pterin-dependent hydroxylase enzymes are allosterically regulated via ACT domains.

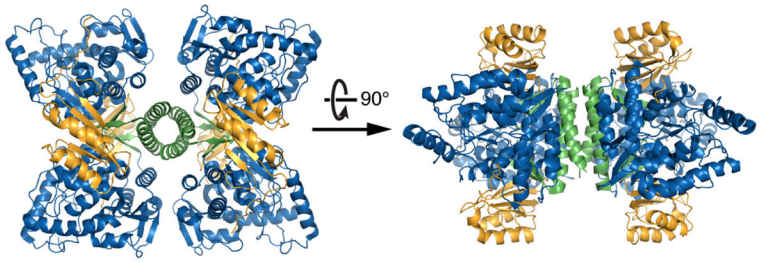


Figure 7.

The full-length crystal structure of rat PheH reveals a homotetramer in an inactive state. In this crystal structure, the ACT domains (orange) are too far apart to interact (PDB: 5DEN).[114]

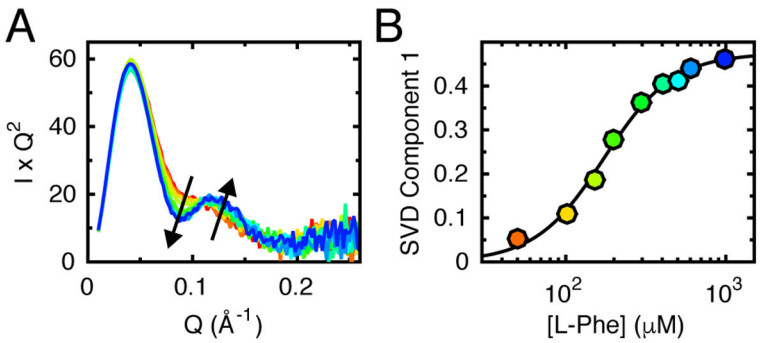


Figure 8.

PheH undergoes a cooperative structural change upon addition of the allosteric activator, Phe.[115] (A) Titration of 0–1 mM Phe to rat PheH leads to a change in the mid-*q* region in the Kratky plots (red to blue). (B) Singular value decomposition of this titration dataset reveals that this structural change is a cooperative two-state transition. Adapted with permission from Ref. 115. Copyright 2016 American Chemical Society.

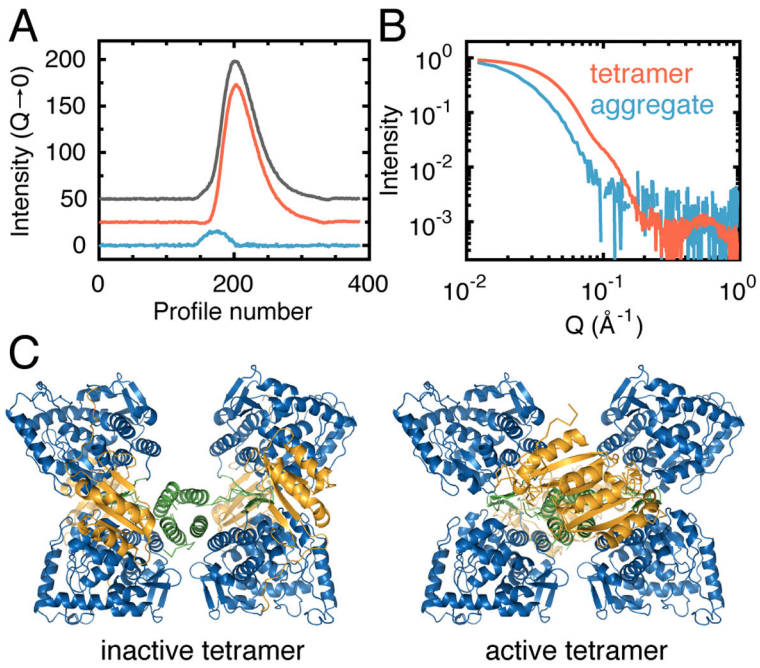


Figure 9.

SEC-SAXS coupled with EFA allows for structural modeling of the PheH tetramer.[115] (A) The elution profile of PheH (black) was separated via EFA into two sequential, overlapping peaks (blue and red). (B) These peaks were found to have scattering profiles indicative of aggregation (blue) and the PheH tetramer (red). (C) Rigid-body models generated from the EFA-separated scattering profiles indicate that the allosteric activation of PheH is consistent with the ACT domains (orange) rotating to dimerize with diagonal partners. Adapted with permission from Ref. 115. Copyright 2016 American Chemical Society.

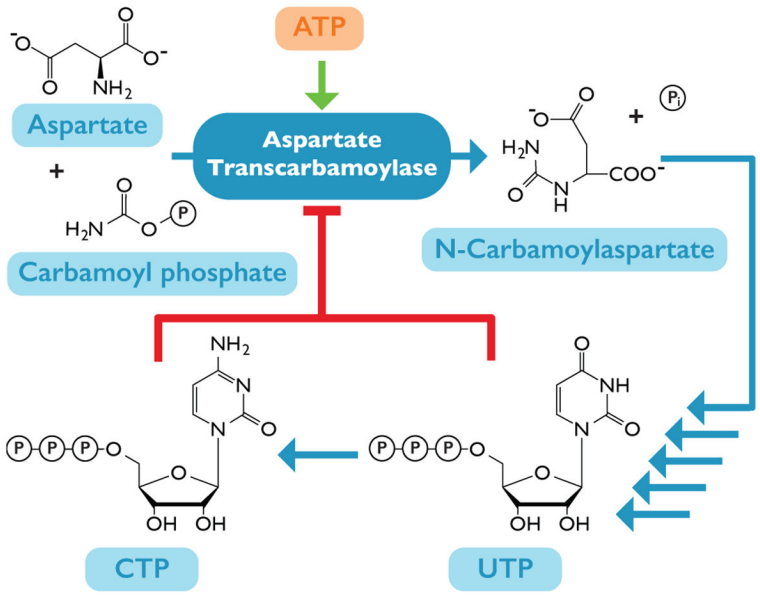


Figure 10.

ATCase catalyzes the first step in the biosynthetic pathway of the pyrimidines CTP and UTP. The enzyme is activated by the purine ATP and allosterically inhibited by CTP and UTP acting in tandem.

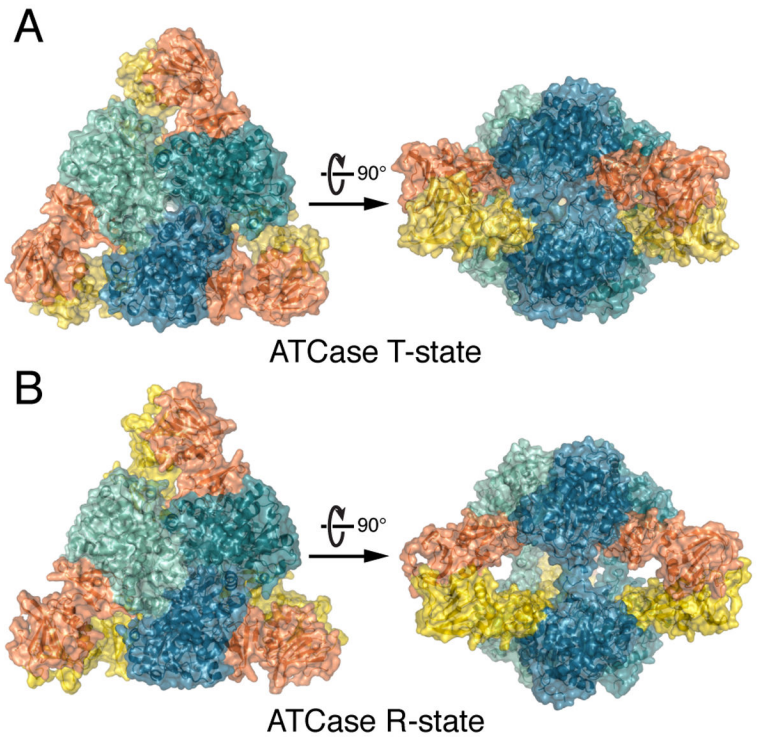


Figure 11.

Crystal structures of ATCase reveal two major conformations. (A) In the presence of the inhibitor CTP or in the absence of ligands, ATCase preferentially forms a closed heterododecamer of two catalytic trimers (blue) and three regulatory dimers (orange) known as the tense or T-state (PDB: 6AT1).[136] (B) Upon addition of the substrate analog PALA, ATCase instead forms a relaxed R-state (PDB: 1D09).[137] The PALA-bound structure involves rotation of both the regulatory dimers and the catalytic trimers, leading to an elongated and open structure relative to the T-state.

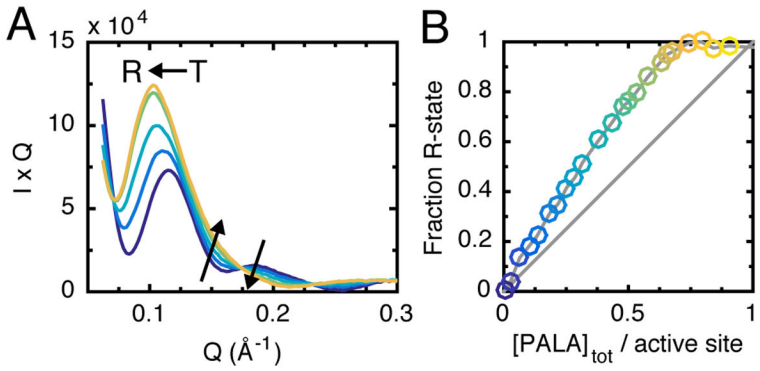


Figure 12.

SAXS reveals that ATCase undergoes a concerted transition between T and R-states that reaches saturation at substoichiometric concentrations of PALA. (A) Titration of PALA (purple to yellow) causes the subsidiary maximum to shift to lower *q*, indicative of a transition to a more open structure. Clear iso-scattering points suggest a two-state transition from the closed T-state to the open R-state. (B) SVD reveals that R-state saturation occurs before all available active sites are PALA bound, indicative of cooperative binding leading to an increased PALA binding affinity.[141] Adapted with permission from Ref. 141. Copyright 1997 International Union of Crystallography. http://journals.iucr.org

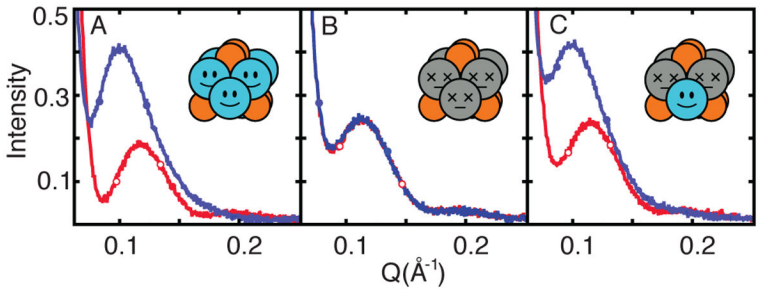


Figure 13.

SAXS studies of a hybrid ATCase provide evidence for concerted cooperativity. (A) ATCase with wild-type catalytic domains (blue spheres) undergoes a visible change in scattering profile between the ligand-free T-state (red curve) and PALA-bound R-state (blue curve). (B) A mutant ATCase consisting of mutant catalytic domains (brown spheres) incapable of binding substrate showed no difference between the ligand-free and PALA-present scattering profiles. (C) Addition of PALA to a hybrid ATCase with five mutant catalytic domains (brown spheres) and a single wild-type domain (blue sphere) leads to a change in scattering similar to that seen for the wild-type enzyme, indicating that only a single functional catalytic domain is necessary to cause the hybrid ATCase to shift from T- to R-state. Adapted with permission from Ref. 144. Copyright 2001 Macmillan Publishers Ltd: Nature.

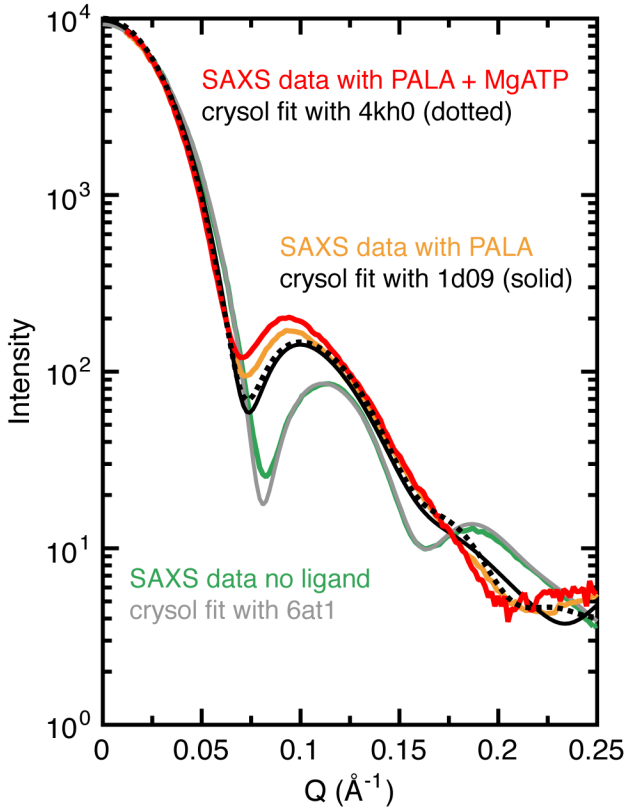


Figure 14.

Discrepancies between SAXS and crystallography have left open questions about the interpretation of the ATCase R-state structure. The scattering of ligand-free ATCase (green curve) agrees well with a theoretical profile of the ligand-free crystal structure (grey curve, PDB: 6AT1).[136] However, the scattering of ATCase with PALA (orange curve) or with PALA and Mg^{2+}-ATP (red curve) do not overlay with theoretical profiles of crystal structures with PALA-bound (black solid, PDB: 1D09)[137] or with PALA and Mg^{2+}-ATP bound (black dotted, PDB: 4KH0).[147] As noted by Svergun,[138] this discrepancy is likely due to an R-state more open in solution than in a crystal. Adapted with permission from Ref. 148. Copyright 2001 Elsevier.

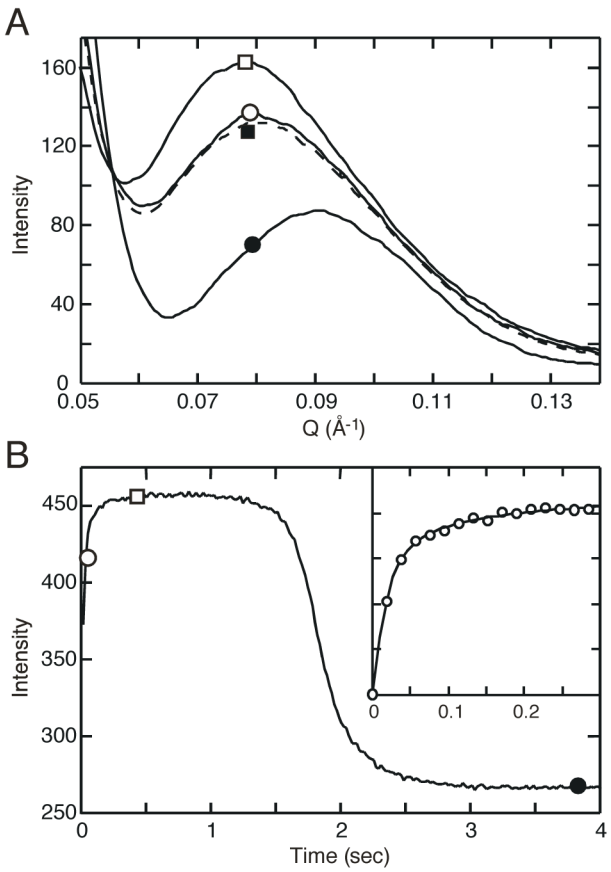


Figure 15.

Time-resolved SAXS studies provide insight into the kinetics of the ATCase transition.[154] (A) A time series of SAXS patterns of ATCase with substrate at 5 °C captures the change in conformation at 38 ms (open circles), 380 ms (open squares), and 3800 ms (filled circles) after mixing. The long dashed curve (filled squares) is a linear combination of 33% T-state and 67% R-state. (B) Integrated scattering intensities plotted as a function of time indicate a rapid, initial transition from T- to R-state after mixing of Asp with ATCase and CP, reaching a steady-state phase with majority of the population in the R-state from 0.5–1.5 s. The population then reverts to the T-state after 1.5 s as substrates are exhausted. Adapted with permission from Ref. 154. Copyright 2008 Elsevier.

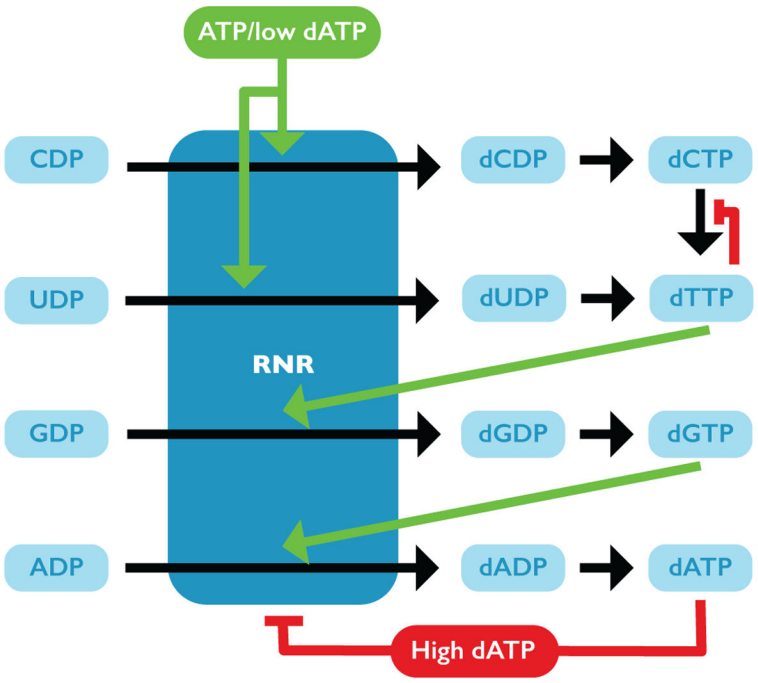


Figure 16.

Ribonucleotide reductase (RNR) is located at a critical junction in nucleotide metabolism and is tightly regulated by various nucleotide effectors. Class I RNRs act on ribonucleoside diphosphates. The pathway products dTTP, dGTP, and dATP at low concentration act as specificity effectors by increasing the affinity for specific substrates. ATP and dATP at high concentrations act as activity effectors, increasing and decreasing catalytic turnover, respectively.

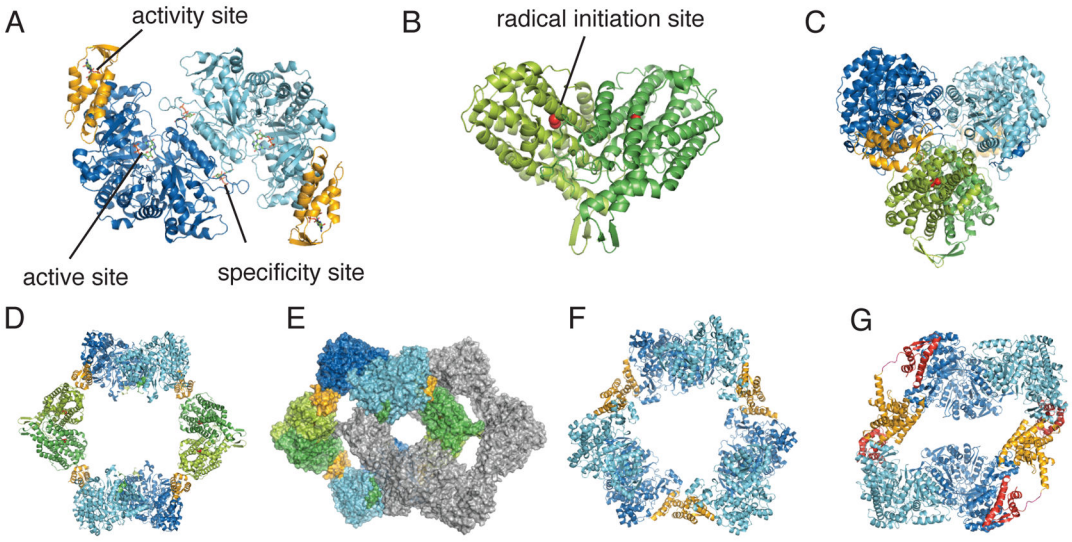


Figure 17.

Class Ia RNRs can form a variety of active and inhibited oligomers. (A) The α dimer of *E. coli* RNR contains the active site as well as allosteric sites. The activity site is located in the N-terminal ATP-cone domain (orange), which plays a role in higher order oligomerization (PDB: 3R1R).[169] (B) The β dimer of *E. coli* RNR initiates radical chemistry via a diferric metallocofactor (red spheres) (PDB: 1RIB).[170] (C) A low-resolution model of the active *E. coli* RNR depicts an $\alpha_2\beta_2$ complex in which the subunits are docked along their symmetry axes.[88,171,172] (D) *E. coli* RNR forms an inhibited $\alpha_4\beta_4$ complex in the presence of saturating concentrations of dATP.[88] (E) Under crowding conditions, *E. coli* RNR forms concatenated $\alpha_4\beta_4$ rings.[167] (F) Human RNR forms an α_6 ring in the presence of dATP, precluding interaction with the β_2 dimer (PDB: 5D1Y).[168] (G) *P. aeruginosa* RNR forms an inactive α_4 ring in the presence of dATP. The enzyme contains two sequential ATP-cone domains: one that is deficient in dATP-binding (red) and an N-terminal copy that can bind two dATP molecules (orange) (PDB: 5IM3).[173]

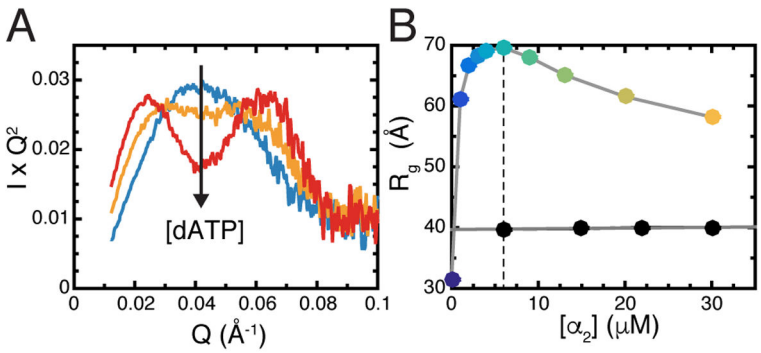


Figure 18.

SAXS titration studies of *E. coli* RNR show that a large, non-globular oligomer with 1:1 subunit stoichiometry forms at inhibiting levels of dATP.[88] (A) The Kratky representation of 0 μM (blue), 12 μM (orange), and 40 μM (red) dATP titrated into a 6 μM solution of α_2 and β_2 in the presence of CDP. As [dATP] increases, the Kratky plot shifts from a single dominant peak to a double-humped curve characteristic of a non-globular structure. (B) Titration of 0–30 μM α_2 (violet to yellow) in a solution of 6 μM β_2 and saturating dATP and CDP reveals a maximum R_g value at the equimolar point (dotted line), indicative of 1:1 subunit stoichiometry in the inhibited complex. Adapted with permission from Ref. 88. Copyright 2011 United States National Academy of Sciences.

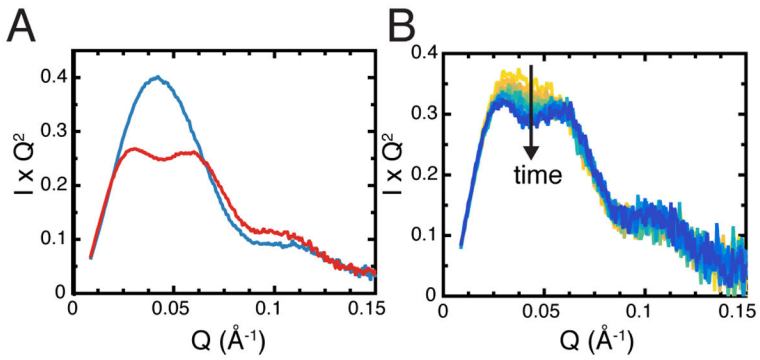


Figure 19.

SAXS studies of a trapped radical mutant of *E. coli* RNR reveal an active $\alpha_2\beta_2$ complex that is kinetically stabilized even under strongly inhibiting conditions. (A) An unnatural amino tyrosine (NH_2Y_{730}) leads to a radical trap at this position in the α subunit. Mixing this mutant α with β lacking a radical leads to a Kratky plot reminiscent of an inhibited $\alpha_4\beta_4$ species at high protein concentrations (red). By contrast, mixing the mutant α and β with an intact radical leads to a Kratky plot with a single dominant peak, indicative of an active $\alpha_2\beta_2$ species even at high protein concentrations (blue). (B) A time-course over 22 minutes (yellow to blue) of the mutant RNR in the presence of inhibiting levels of dATP displays a gradual shift from a monomodal to a bimodal Kratky curve, consistent with a remarkably slow conversion from $\alpha_2\beta_2$ to the inhibited $\alpha_4\beta_4$. Adapted with permission from Ref. 172.

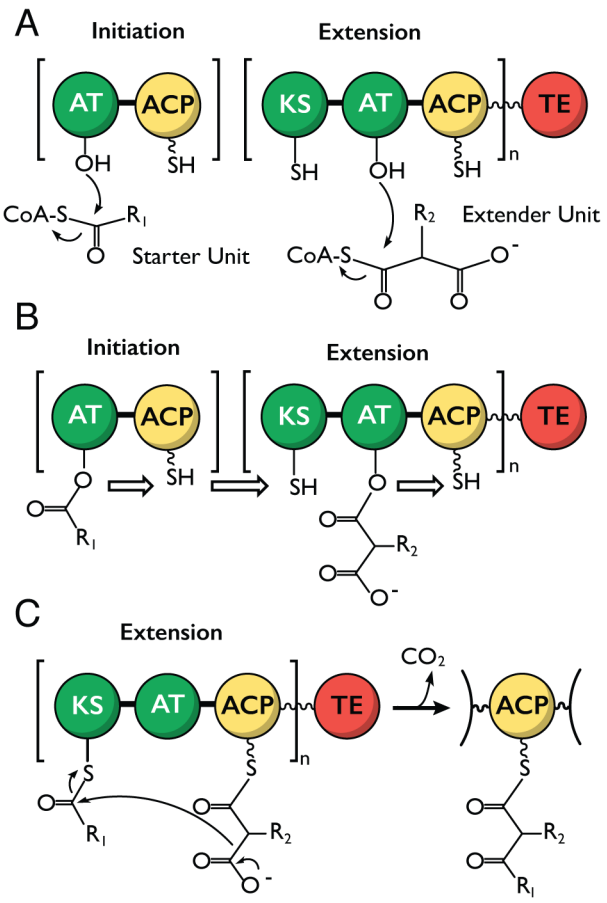


Figure 20.

Initiation and elongation processes in a minimal Type I PKS. (A) Mechanism of loading of starter and extender units to PKS modules. The CoA-linked starter or extender unit is first covalently attached to the AT domain, with the CoA acting as the leaving group. (B) The starter unit is transferred to the Ppant arm (only the terminal thiol is shown) of the initiation module's ACP domain and then transferred to the KS domain of the first extension module, where it is covalently attached via a thioester bond. Extender units are transferred to the Ppant arm of the ACP domain in the same module. The now loaded substrate is then transferred to the subsequent module's KS domain in a similar manner. (C) Mechanism of the carbon-carbon bond-forming reaction between the already attached individual units.[190]

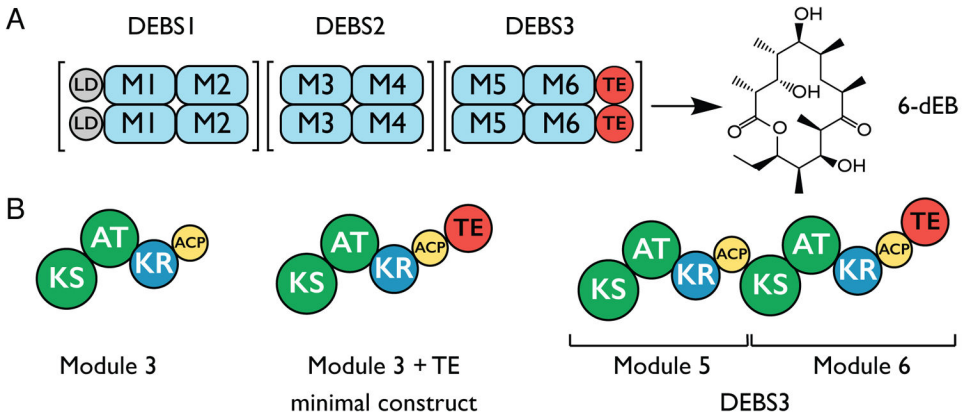


Figure 21.

Overall organization of the DEBS pathway, which produces 6-deoxyerythronolide B (dEB), and constructs studied by SEC-SAXS.[192] (A) The pathway contains a loading (LD) didomain, 6 extension modules (M1 – M6), and a cyclizing thioesterase (TE) domain. DEBS modules are organized into three distinct complexes, DEBS1, DEBS2 and DEBS3. Each complex is homodimeric, and each chain contains two different modules. (B) Domain composition of the major constructs described in this review, namely module 3 (M3), the minimal M3 + TE construct, and DEBS3 (a bimodule of M5 and M6).[191,192]

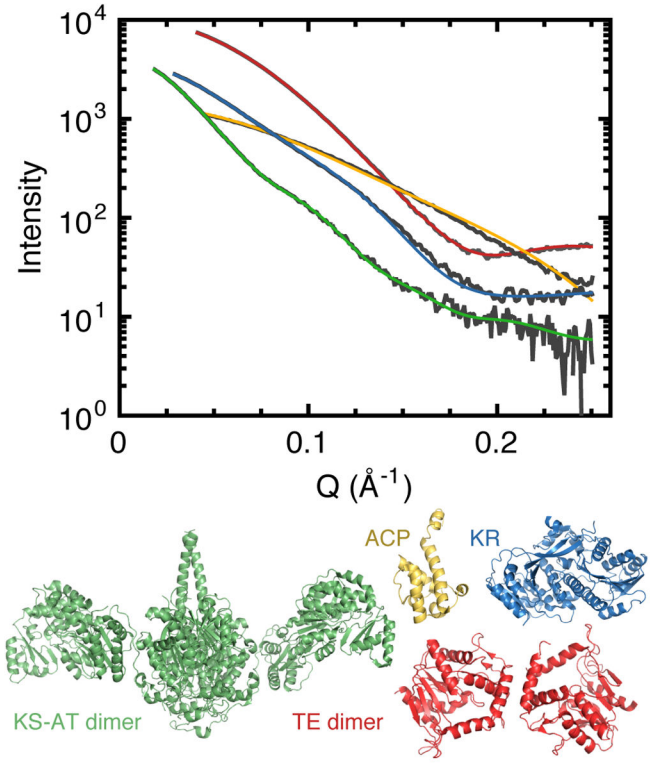


Figure 22.

SEC-SAXS studies of DEBS components show agreement with crystal structures. (Top) Solution scattering profiles (black) of individual DEBS domains compared to CRYSOL fits from the existing crystal structures.[192] The fits shown include the KS-AT didomain (green), ACP domain (yellow), KR domain (cyan), and TE domain (red). (Bottom) Ribbon diagrams of the crystal structures (PDB codes 2JU2, 2FR0, 1MO2 and 2QQ3) corresponding to the CRYSOL fits.[197–200] Adapted with permission from Ref. 192. Copyright 2014 Elsevier.

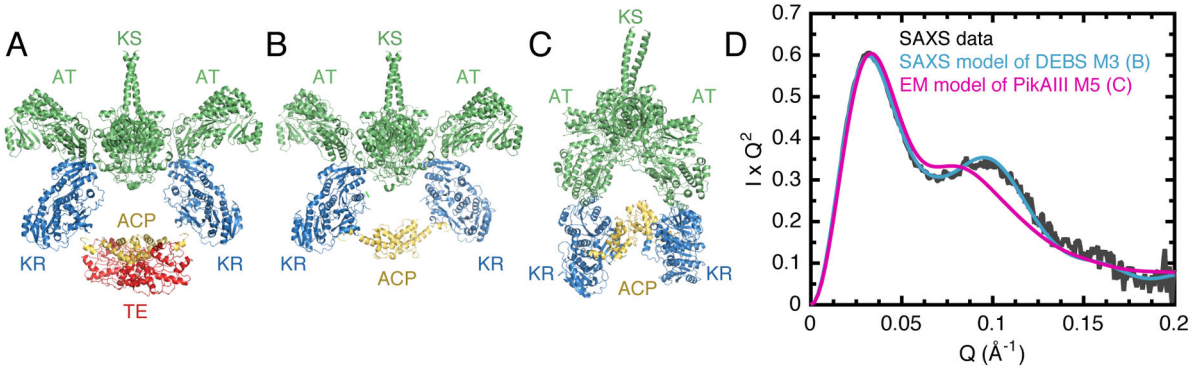


Figure 23.

SEC-SAXS experiments enable rigid-body modeling of a DEBS module and comparison to a cryo-EM structure of a similar module. (A) SAXS-derived model of the minimal DEBS M3 + TE construct derived by treating the individual domain crystal structures shown in Figure 3 as rigid bodies. (B) SAXS model for DEBS M3, created in a similar manner.[192] (C) The EM model of PikAIII module 5 (M5) displays a notably different architecture.[188] (D) Experimental SAXS data of DEBS M3 shown as Kratky plots. Theoretical profile of the SAXS-derived DEBS M3 model (blue) provides a better fit than the PikAIII EM model (pink). Adapted with permission from Ref. 192. Copyright 2014 Elsevier.

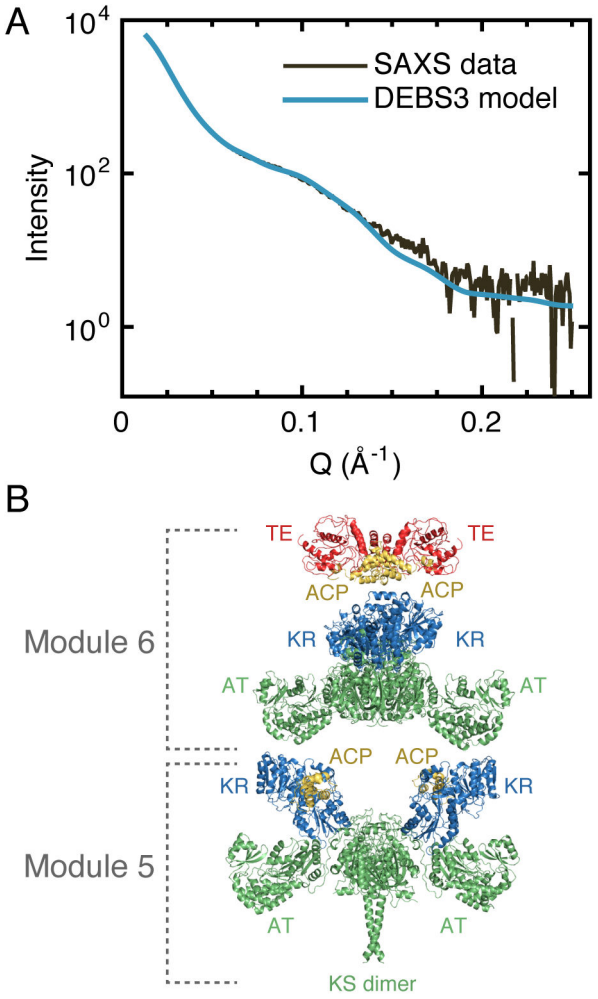


Figure 24.

SEC-SAXS enables rigid-body modeling of a full bimodular complex, DEBS3. (A) Experimental scattering of the DEBS3 bimodule model (black) was used to generate rigid-body models. The theoretical scattering from a representative structure obtained through rigid-body modeling is shown in blue. (B) Rigid-body model of DEBS3 generated from the crystal structures shown in Figure 3.[192] Adapted with permission from Ref. 192. Copyright 2014 Elsevier.

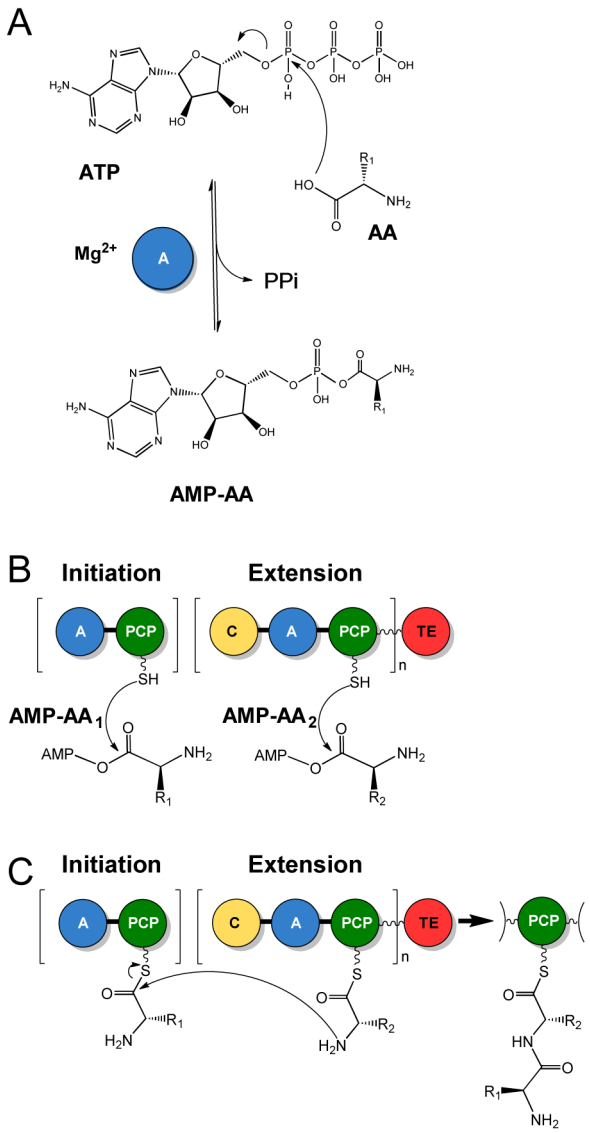


Figure 25.

Initiation and elongation processes in a minimal NRPS. (A) The amino acid to be incorporated is first activated by the A domain using one molecule of ATP to form an amino acid-AMP adduct (AMP-AA). (B) The activated starter unit is covalently linked through a thioester bond to the initiation module's PCP domain. Extender units are loaded onto their respective modules in the same way as the starter unit. (C) Mechanism of the amide bond-forming reaction between individual units.[185]

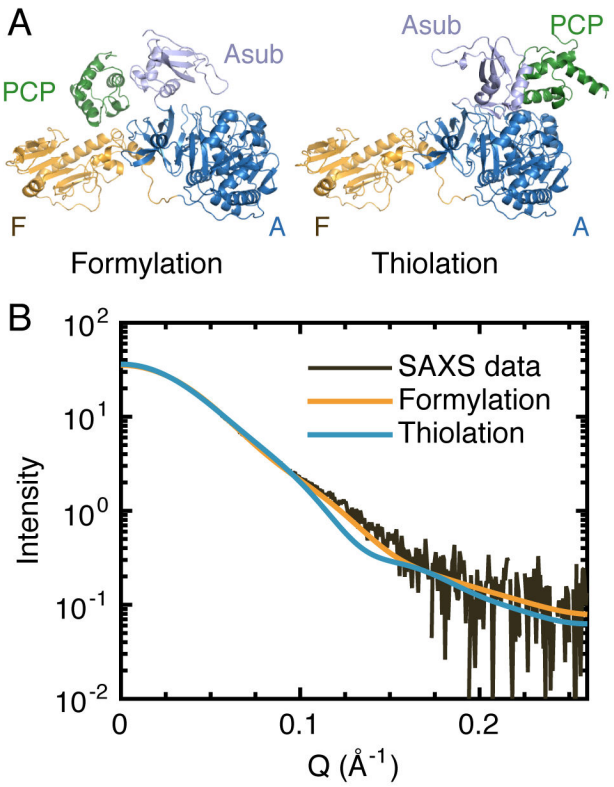


Figure 26.

SAXS experiments provide evidence for a linear arrangement of the F and A domains in the NRPS module, LrgA. (A) Crystal structures of the formylation and thiolation states of LgrA, with the individual domains colored as follows: PCP (green), F (yellow), A (blue), and the A subdomain (lavender). In both structures, the F and A domains adopt the extended conformation. (B) The theoretical scattering of the formylation and thiolation states both show good agreement to the experimental data at low *q*.[187] Adapted with permission from Ref. 187. Copyright 2016 Macmillan Publishers Ltd: Nature.

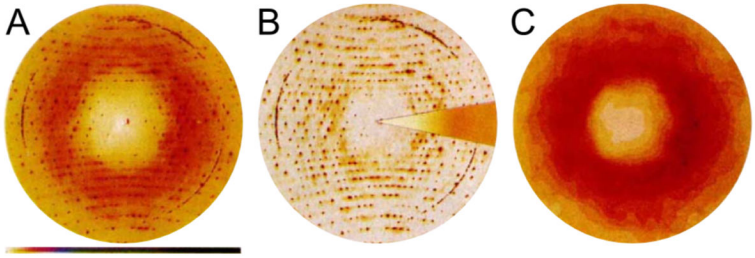


Figure 27.

Diffuse scattering from insulin.[3] (A) A 40-hour still diffraction image of insulin collected on X-ray film. Bragg reflections are overexposed. The sharp arcs are artifacts from the beamline setup. (B) Bragg reflections and haloes digitally separated by subtracting the smoothly varying diffuse scattering component of the data. Inset: Rotationally symmetric Compton and water scattering. (C) Variational scattering evaluated from difference between (A) and two components in (B). Adapted with permission from Ref. 3. Copyright 1988 Macmillan Publishers Ltd: Nature.

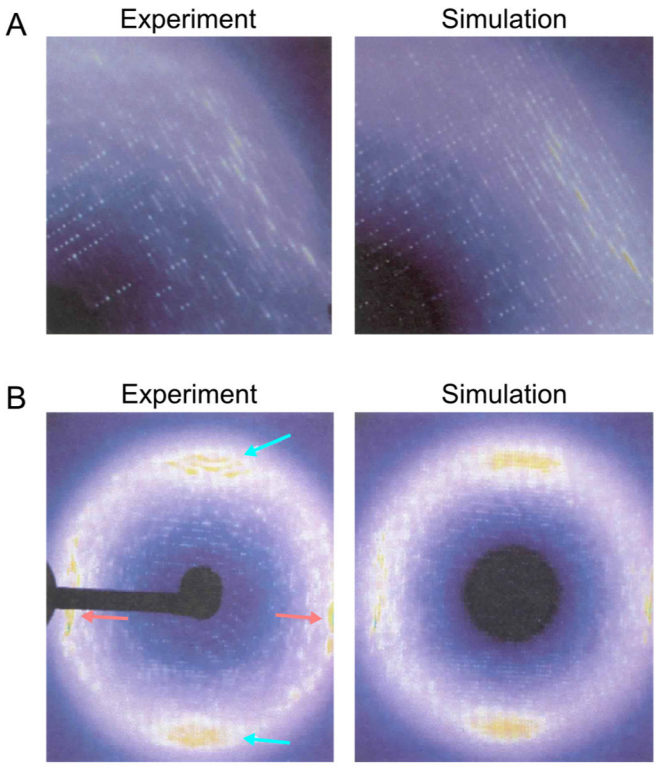


Figure 28.

Diffuse scattering from yeast tRNA crystals reveals lattice-coupled motions as well as intramolecular motions.[228] (A) Simulated scattering from anisotropic lattice-coupled motions agrees well with the experimentally observed diffuse streaks between Bragg peaks. (B) The cloudy diffuse scattering (see arrows) in the experimental data can be well described a model for short-range correlated motions of the anticodon arm. Adapted with permission from Ref. 228. Copyright 1994 International Union of Crystallography. http://journals.iucr.org

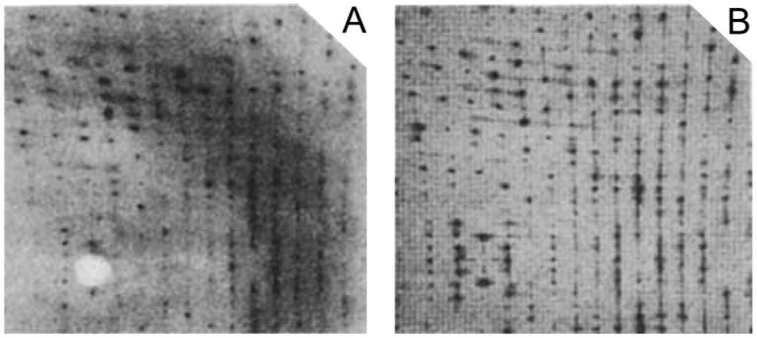


Figure 29.

Diffuse scattering streaks indicate lattice disorder in orthorhombic lysozyme crystals. [224] (A) Still diffraction image recorded on X-ray film shows diffuse streaks between Bragg reflections along the reciprocal lattice planes perpendicular to $\mathbf{a}^*$ and $\mathbf{c}^*$. (B) Simulation of the diffuse scattering produced by small correlated displacements of pairs of molecules along **a** and **c** axes in the crystal lattice reproduces the streaked features. Adapted with permission from Ref. 224. Copyright 1987 Publishers Ltd: Nature.

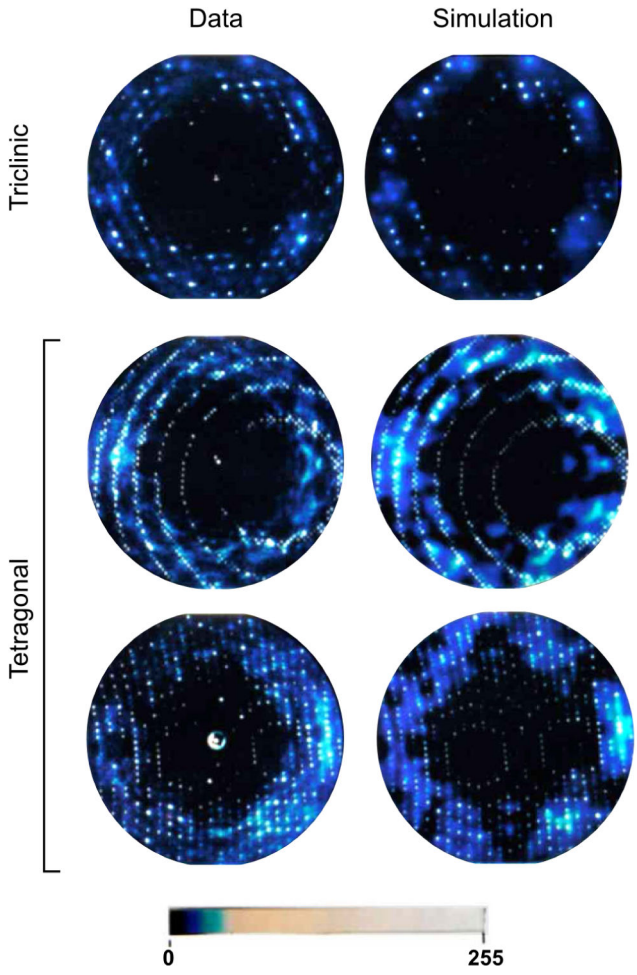


Figure 30.

Diffuse scattering from tetragonal and triclinic lysozyme crystals collected on X-ray film (left) compared with simulated scattering from LLM models (right).[222] Top: Still diffraction image from triclinic lysozyme compared with LLM model. Middle and bottom: Still diffraction images from tetragonal lysozyme separated by a 90 degree rotation compared with simulated scattering from the LLM model. Adapted with permission from Ref. 222. Copyright 1992 John Wiley & Sons, Inc.

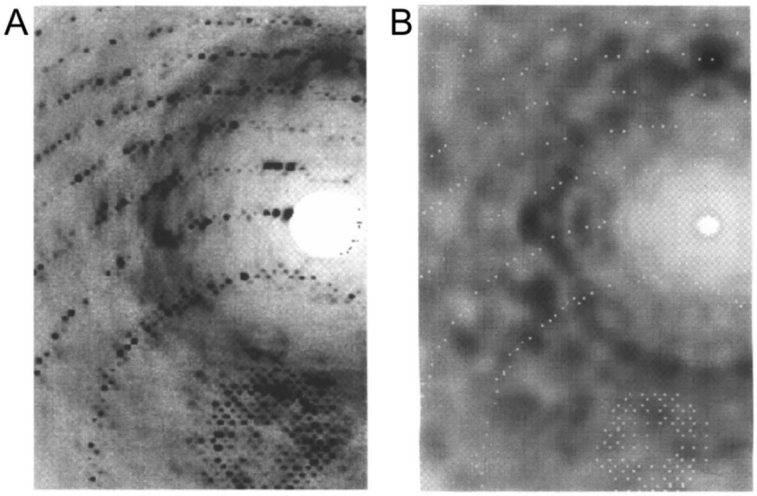


Figure 31.

Diffuse scattering from tetragonal lysozyme is compared with simulated scattering from TLS refinement.[7] (A) Still diffraction image collected on an image-plate detector. (B) Simulated scattering taking into account both rigid-body rotation and translation achieved satisfactory agreement. Adapted with permission from Ref 7. Copyright 1996 International Union of Crystallography. http://journals.iucr.org

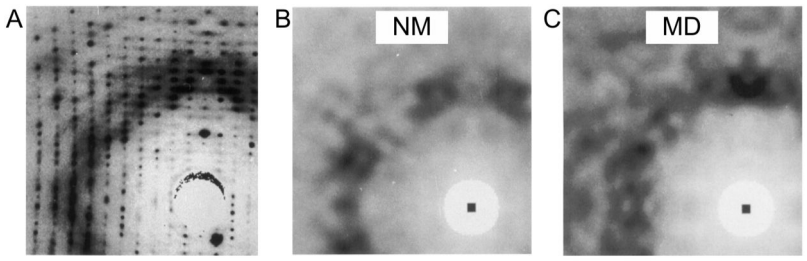


Figure 32.

(A) Experimental diffuse scattering from orthorhombic lysozyme compared with diffuse scattering patterns calculated from (B) NM analysis and (C) MD simulations.[225] Adapted with permission from Ref. 225. Copyright 1994 Macmillan Publishers Ltd: Nature.

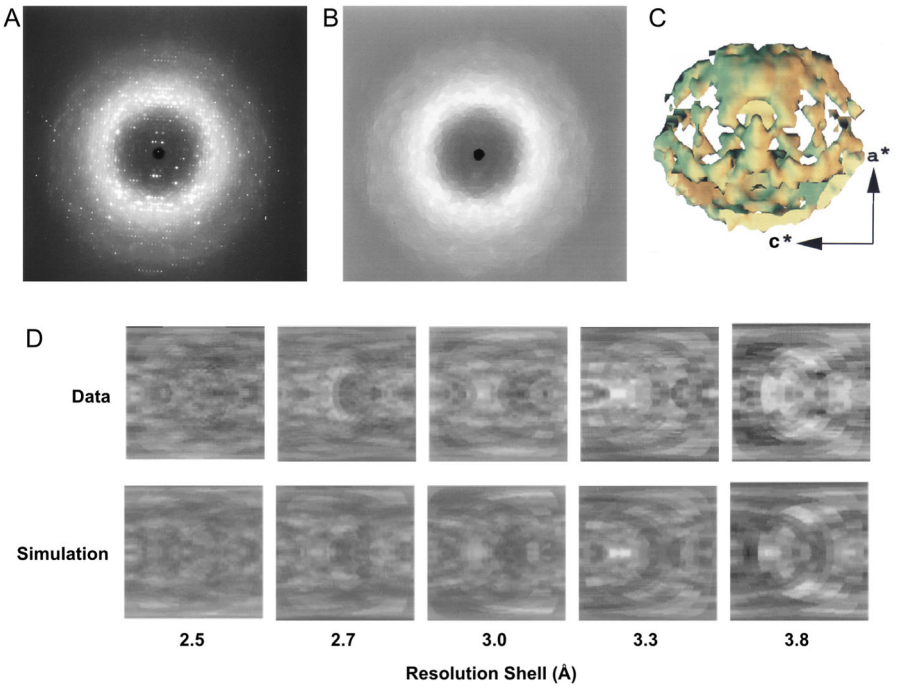


Figure 33.

Diffuse scattering from SNase is attributed to LLM.[1] (A) Raw diffraction image from SNase crystal. (B) Diffraction image after polarization correction, solid angle normalization, and mode filtering to remove Bragg peaks. (C) 3D diffuse map. (D) Mercator projections of the experimental diffuse patterns are compared with simulated scattering from best-fit LLM model. Adapted from Ref. 1. Copyright 1997 United States National Academy of Sciences.

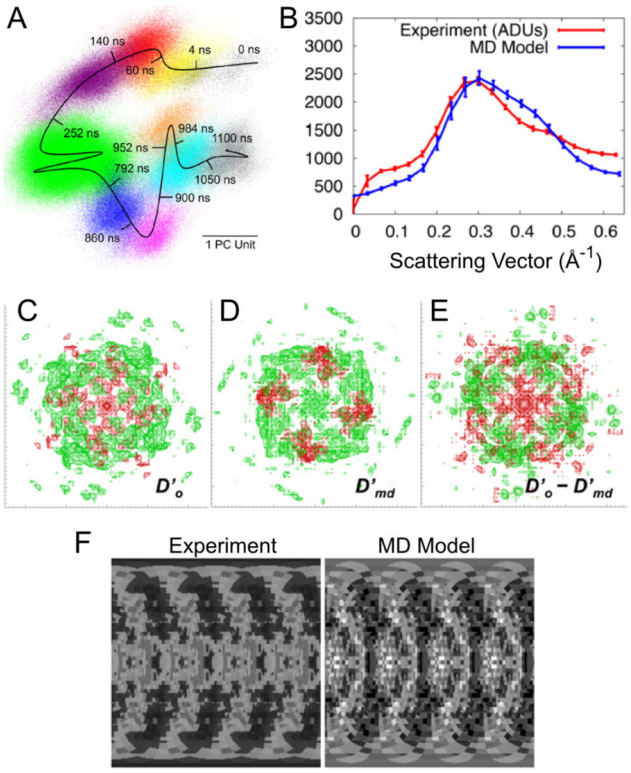


Figure 34.

Comparison of diffuse scattering from SNase with MD simulations.[87] (A) Scatter plot of structures extracted from the MD trajectory projected on the first two principal components of the C-alpha position covariance matrix. (B) Comparison of isotropic diffuse intensity for experimental data (red) and MD model (blue). (C) Isosurface in the experimental map ($D_0^{'}$) where positive intensity is shown in green and negative intensity is in red. (D) Isosurface of simulated map from the MD model ($D_{\mathrm{md}}^{'}$). (E) Difference intensity map ($D_0^{'}-D_{\mathrm{md}}^{'}$). The (F) experimental and (G) simulated anisotropic scattering in the 0.27Å^{-1} resolution shell, where there is the best agreement between the two. Adapted with permission from Ref. 87. Copyright 2014 United States National Academy of Sciences.

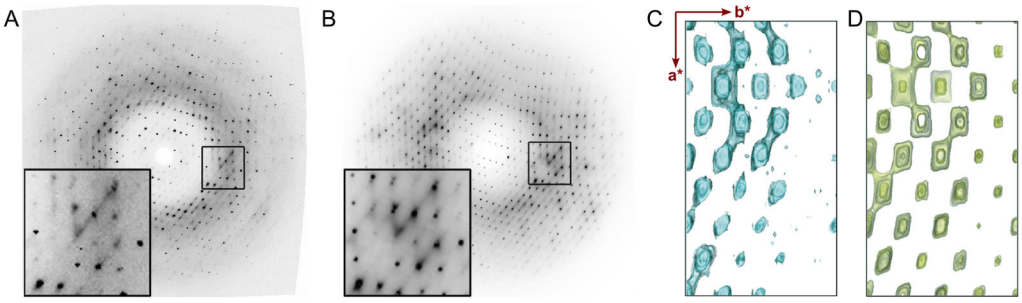


Figure 35.

Diffuse scattering from crystals of calmodulin were used to infer motions via LLM.[64] (A) Still diffraction image of calmodulin reveals streaked features in the solvent ring (inset) as well as broader large-scale diffuse scattering in the background. (B) Simulated scattering using an anisotropic LLM model reproduces the experimental scattering. The isosurface streaks in (C) the data and (D) simulation show good agreement. Adapted with permission from Ref. 64. Copyright 1997 Elsevier.

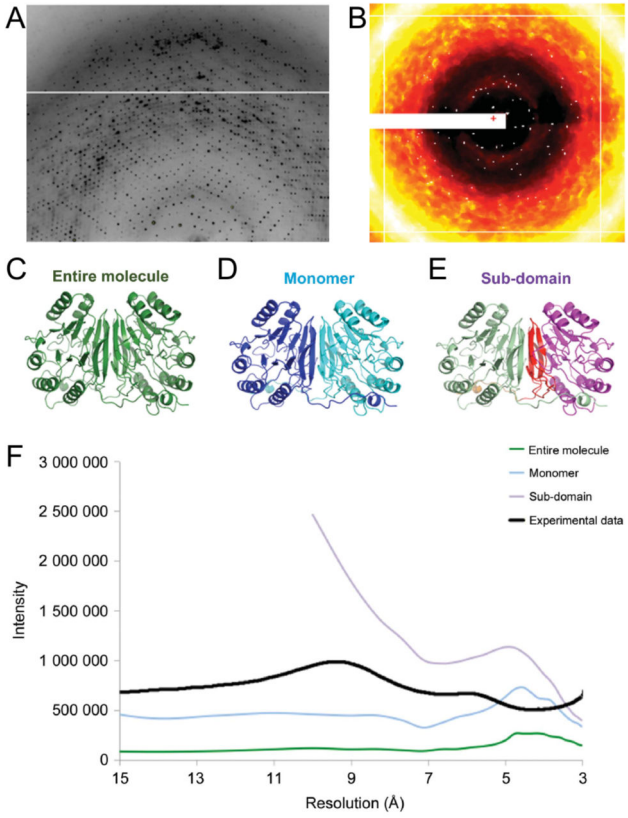


Figure 36.

Glycerophosphodiesterase (GpdQ) is an example where collective motions inferred from TLS refinement do not agree with the diffuse scattering.[74] (A) A still diffraction image from GpdQ shows significant diffuse scattering. (B) Mode-filtered GpdQ diffraction image. (C)–(E) TLS refinements were performed with treating either the dimer (shown in C), the monomer (shown in D), or each subdomain (shown in E) as separate rigid bodies. (F) Radial diffuse intensity profiles for experimental and simulated GpdQ data. None of the TLS-derived ensembles showed significant agreement with the experimental data. Adapted with permission from Ref. 74. Copyright 2015 International Union of Crystallography. http://journals.iucr.org

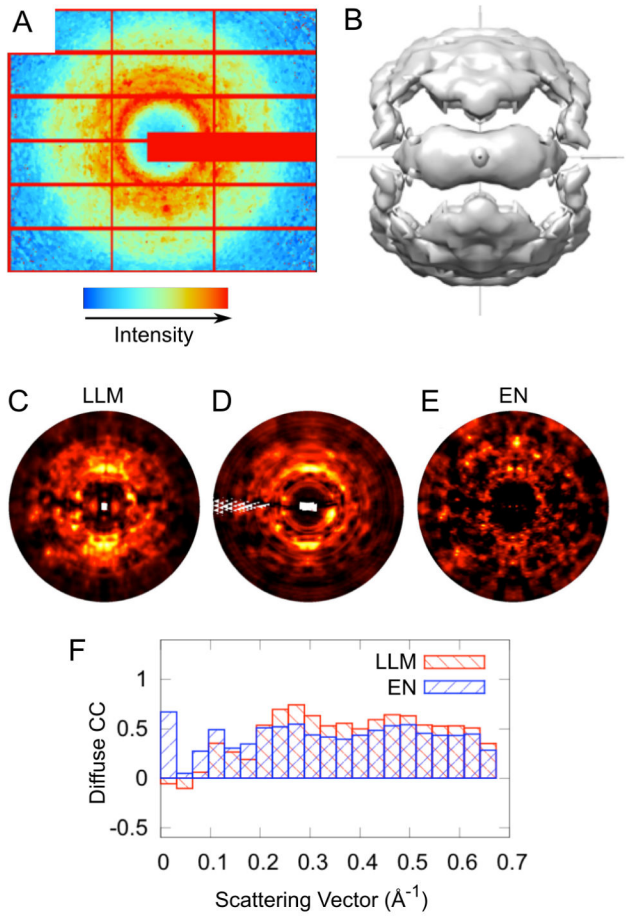


Figure 37.

The diffuse scattering of CypA was compared with three models of motion.[84] (A) Mode-filtered diffraction image. (B) The data from individual images are combined and symmetrized to yield a nearly complete 3D diffuse map. (C) Simulated diffraction images for CypA obtained with LLM model. (D) Integrated 3D diffuse data. (E) Simulated diffraction images for CypA obtained with an EN model. Lighter colors correspond to stronger intensity. (F) Linear correlation coefficients (CCs) between diffuse scattering data and LLM (red) or EN models (blue) computed by resolution shell for CypA. Adapted with permission from Ref. 84. Copyright 2016 United States National Academy of Sciences.

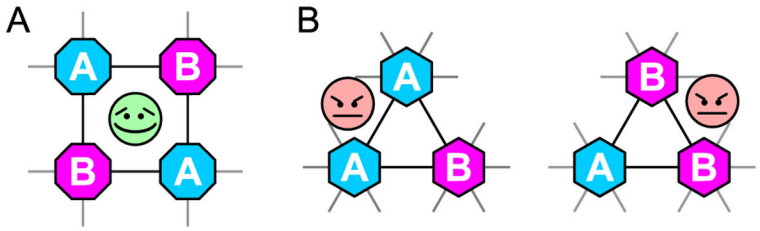


Figure 38.

Geometric frustration. Two conformations, A and B, prefer to be neighbors as AB rather than AA or BB. (a) The constraints can be satisfied on a square lattice. (b) The constraints cannot be satisfied on a triangular lattice. The minimum energy configuration shows correlated disorder.

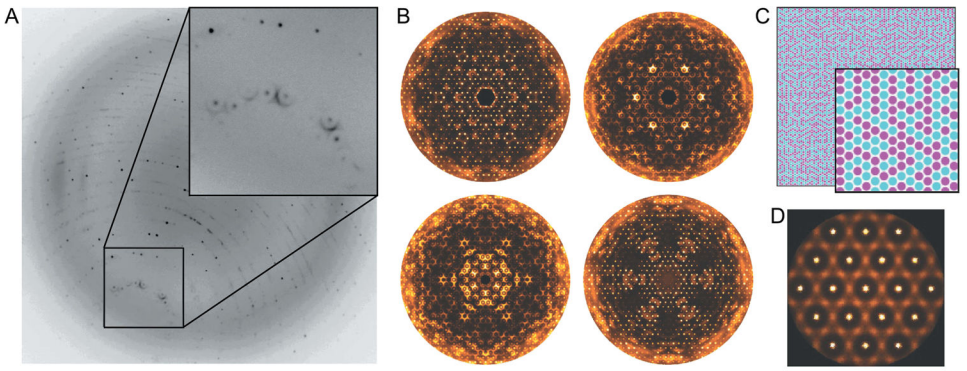


Figure 39.

Geometric frustration in a crystal of the N-terminal fragment of the FFV Gag protein leads to unusual diffuse scattering features.[86] (A) Zooming into a diffraction image shows the diffuse circular features that occur in the vicinity of the Bragg peaks. (B) Reciprocal lattice sections normal to **c**. (C)–(D) Simulating the scattering from the Ising model (shown in C) leads to a pattern (shown in D) that can explain the experimental data. Adapted with permission from Ref. 86. Copyright 2011 International Union of Crystallography. http://journals.iucr.org

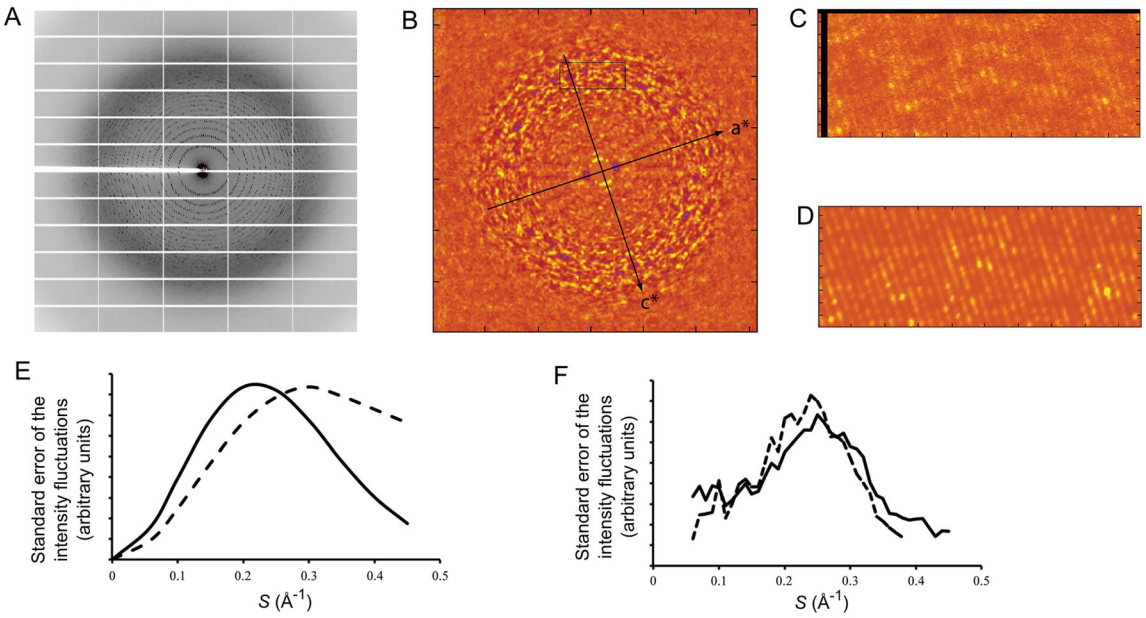


Figure 40.

Acoustic vibrations contribute to the diffuse scattering of 70S ribosome crystals. [66] (A) Still diffraction image of 70S ribosome from *T. thermophilus*. (B) The diffraction image after processing to subtract the circularly symmetric background and reduce noise by binning. The pattern shows mm symmetry (C) Unbinned but fully-processed data from the section of the image shown in (B). Ridges of intensity parallel to $\mathbf{c}^*$ are visible. (D) Simulated scattering using an acoustic vibration model. (E) Standard deviation of the intensity vs. scattering angle predicted for acoustic vibrations (solid line) and motions without long-ranged correlations (dashed line). (F) Standard deviation of the experimentally measured diffuse intensity vs. scattering angle (solid line) follows follows the same quantity simulated using the acoustic vibration model with the measured Bragg intensities (dashed line). Adapted with permission from Ref. 66. Copyright 2015 International Union of Crystallography. http://journals.iucr.org

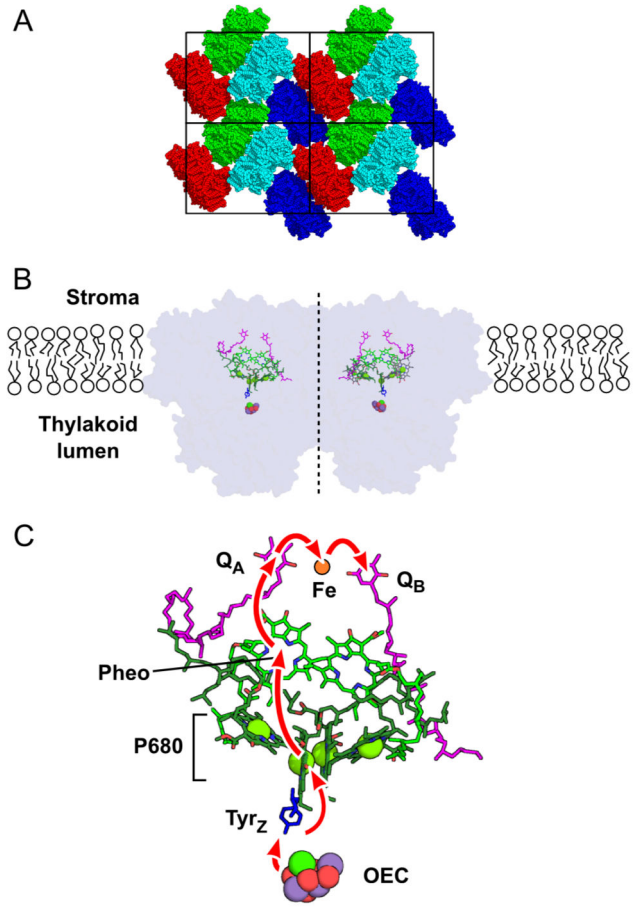


Figure 41.

Photosystem II (PSII). (A) Packing of PSII in the crystal lattice (**b**–**c** plane), PDB ID: 5E7C. (B) Orientation of PSII dimer in the thylakoid membrane (axis of symmetry indicated by a dashed line). (C) Key cofactors involved in the water-splitting reaction, drawn in the same orientation as the left side of the dimer in part (B). Red arrows show the path of electron transport.

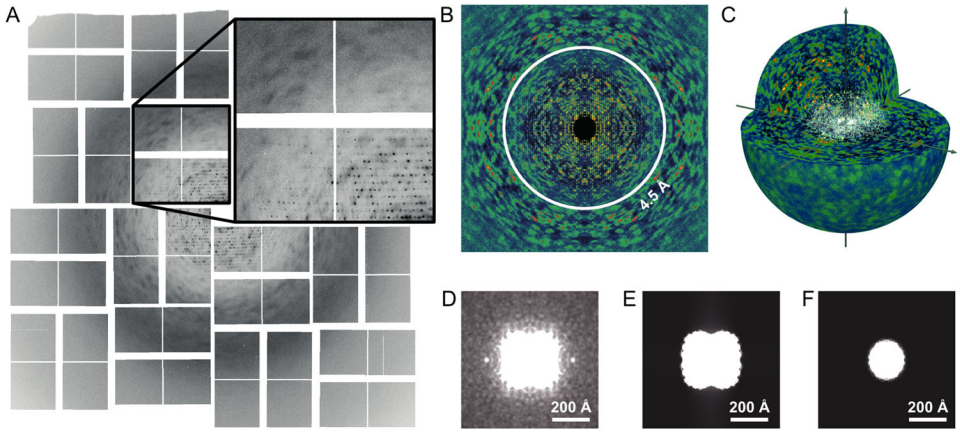


Figure 42.

Serial crystallography reveals diffuse scattering extending beyond the Bragg diffraction limit of PSII.[302] (A) A still diffraction pattern of a PSII microcrystal shows diffuse scattering beyond the resolution of the Bragg peaks. (B) Speckle patterns are clearly observed extending past the 4.5-Å resolution of Bragg diffraction (white circle) in a 2D slice through the diffraction volume. (C) The 3D diffraction volume assembled by merging 2,848 still images. (D) Projection of the 3D autocorrelation, the Fourier transform of the continuous diffraction intensities, along a crystal axis. (E)–(F) The equivalent projections of the autocorrelation functions calculated from the model of the PSII dimer (shown in E) agrees better with the experimentally derived autocorrelation than the monomer (shown in F). Adapted with permission from Ref. 302. Copyright 2016 Macmillan Publishers Ltd: Nature.