Published in final edited form as:
Chem Rev. 2016 June 08; 116(11): 6305–6322. doi:10.1021/acs.chemrev.5b00592.

Hybrid Approaches to Structural Characterization of Conformational Ensembles of Complex Macromolecular Systems Combining NMR Residual Dipolar Couplings and Solution X-ray Scattering

Vincenzo Venditti[*,†,‡], **Timothy K. Egner**[†], and **G. Marius Clore**[*,§]

[†]Department of Chemistry, Iowa State University, Ames, Iowa 50011, United States

[‡]Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa 50011, United States

[§]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, United States

Abstract

Solving structures or structural ensembles of large macromolecular systems in solution poses a challenging problem. While NMR provides structural information at atomic resolution, increased spectral complexity, chemical shift overlap, and short transverse relaxation times (associated with slow tumbling) render application of the usual techniques that have been so successful for medium sized systems (<50 kDa) difficult. Solution X-ray scattering, on the other hand, is not limited by molecular weight but only provides low resolution structural information related to the overall shape and size of the system under investigation. Here we review how combining atomic resolution structures of smaller domains with sparse experimental data afforded by NMR residual dipolar couplings (which yield both orientational and shape information) and solution X-ray scattering data in rigid-body simulated annealing calculations provides a powerful approach for investigating the structural aspects of conformational dynamics in large multidomain proteins. The application of this hybrid methodology is illustrated for the 128 kDa dimer of bacterial Enzyme I which exists in a variety of open and closed states that are sampled at various points in the catalytic cycles, and for the capsid protein of the human immunodeficiency virus.

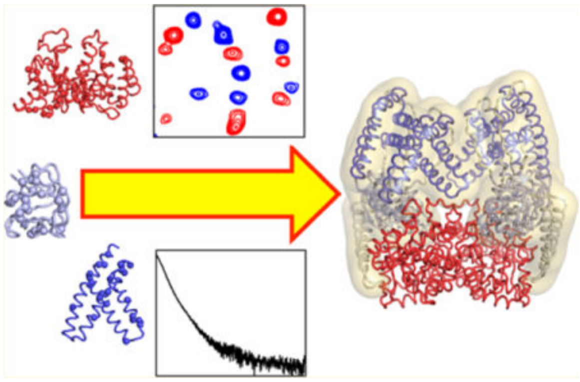
Graphical abstract

---

[*]**Corresponding Authors:** venditti@iastate.edu (V.V.); mariusc@mail.nih.gov (G.M.C.).

Notes

The authors declare no competing financial interest.



---

1. INTRODUCTION

Biological macromolecules sample a range of conformational states (or microstates).[1,2] The distribution of this ensemble of microstates is dynamic and highly sensitive to changes in external conditions such as binding state,[3] ligand concentration,[4] pH,[5–7] ionic strength,[8] and post-translational modifications.[9] Important biological processes, including enzymatic catalysis,[10–15] ligand binding,[16,17] allostery,[18] and signaling,[9,19] depend on the exact composition of the ensemble and on interconversion rates between microstates. Understanding the delicate balance between structure and dynamics that governs biological function represents a new frontier in modern structural biology and molecular biophysics and has driven several technical and conceptual advances in the field over the past few years.

Traditionally, high-resolution structures of conformational microstates have been obtained through the application of NMR and X-ray crystallographic studies in which the experimental conditions are adjusted to select one member, or a small subset of closely related members, of the conformational ensemble.[5,6] Although this approach allows determination of very precise structural models, trapping all different microstates accessed by a biopolymer can be a time-consuming and often impractical endeavor. More recently, computational strategies have been developed for interpreting structural data from highly heterogeneous systems, such as multidomain and intrinsically disordered proteins. Common aspects of these methods are the use of an ensemble-based representation for the system of interest, and the interpretation of experimental observables as a property of the overall ensemble instead of the individual microstates. Currently, structural ensembles are generated using one of two major approaches. The first calculates the ensemble by simulated annealing driven by the experimental data. Experimental restraints can be applied in either an ensemble-or time-average manner. The second involves first generating a large pool of possible structures and then selecting among these the most appropriate ensemble that fulfills the desired experimental observables. The computational strategies for generating dynamic structure ensembles have been extensively reviewed[20,21] and will not be discussed here.

There are a number of experimental methods that can provide structural and dynamical information to describe conformational ensembles. X-ray diffraction is affected by multiple sources of disorder (e.g., protein dynamics and crystal-lattice distortions) and has been used to model conformational heterogeneity in the crystal state.[10,22] Other methods, such as Förster resonance energy transfer (FRET)[23] and double electron–electron resonance (DEER) EPR spectroscopy,[24,25] have been used to investigate the structure and dynamics of biopolymers in solution. However, both FRET and DEER require covalent labeling of the molecule under investigation with fluorophores or paramagnetic centers, respectively, and, most importantly, only provide a single distance restraint per sample. These limitations have hampered the widespread use of FRET and DEER-derived distance restraints in calculations of conformational ensembles.

Undoubtedly, the most powerful experimental technique to investigate conformational ensembles is NMR spectroscopy. Indeed a large variety of NMR parameters can be measured that report on the structure and dynamics of biopolymers over a wide range of time scales and amplitudes of motion, and there are a number of methods to treat several types of NMR-derived restraints in ensemble calculations, including chemical shifts,[26] nuclear Overhauser enhancement (NOE) distance restraints,[27–31] generalized order parameters (S^2),[30,32,33] scalar three-bond J couplings,[34,35] residual dipolar couplings (RDC),[36–40] chemical shift anisotropy (CSA),[37] paramagnetic relaxation enhancement (PRE),[41–43] and hydrogen/deuterium exchange protection factors.[44] The biggest advantage of NMR over other solution techniques is that it provides structural restraints for hundreds to thousands of individual atoms, or small groups of atoms, per sample per experiment.[45] However, when investigating high molecular weight systems, such as large complexes and multidomain proteins, line broadening (due to increases in rotational correlation time) and resonance overlap (due to the increased number of NMR-active nuclei) can drastically reduce the number of analyzable cross-peaks in multidimensional NMR spectra, resulting in sparse data sets. In such cases, hybrid methods that couple sparse NMR data with structural information from other lower-resolution techniques, such as cryo-electron microscopy[46] or small/wide-angle X-ray scattering (SAXS/WAXS),[47,48] have been shown to be very powerful in solving the structure of complex molecular systems.

Here we review the use simulated annealing driven by experimental NMR and SAXS/WAXS data for the quantitative investigation of the structure and dynamics of complex macromolecules at atomic resolution.

2. OVERVIEW OF THE USE OF SAXS/WAXS IN STRUCTURAL BIOLOGY

In structural biology, the term low resolution often has negative implications. Most instrumental techniques strive for the highest resolution possible with the gold standard being atomic resolution. Techniques such as crystallography and NMR easily achieve this goal but often come up short when looking at complex systems such as protein assemblies, nucleic acids, and transmembrane proteins. Additionally, analyses of highly flexible systems, such as intrinsically disordered proteins, suffer greatly or may even be impossible using only crystallography or NMR. Lower-resolution techniques such as atomic force microscopy, cryoelectron microscopy, and SAXS/WAXS are emerging as extremely useful tools for structural investigation of these difficult systems.[49–51]

SAXS/WAXS is a solution state technique where a highly collimated X-ray beam is scattered by an incident sample containing the analyte of interest. This is a contrast technique that is very sensitive to changes in electron density. Because the analyte is tumbling in solution, its three-dimensional (3D) structure is reduced to a one-dimensional (1D) set of spherical shells and is represented as a one-dimensional plot of scattering intensity I against the momentum transfer vector q, given by

$$q=4\pi\sin(\theta)/\lambda \quad (1)$$

where 2θ is the scattering angle and λ the wavelength of the incident X-ray beam.

3D structure envelopes can be calculated through ab initio modeling, but the resulting solution is not necessarily the correct one as the information content contained within a 1D data set is necessarily limited (see “Computational Modeling” section). Nevertheless, there are several powerful verification techniques that can be used to evaluate structural hypotheses.[52–54]

In this section of the review we focus on the basics of sample preparation and requirements, instrumental design and operation, the theoretical basis of electromagnetic radiation scattering, data analysis, and finally 3D modeling of molecular structure. While this paper covers only SAXS/WAXS, small angle neutron scattering (SANS) is very closely related and useful but will not be discussed here. There are many informative reviews on SANS in the literature.[55,56]

2.1. Sample Preparation for SAXS/WAXS

For a SAXS experiment to yield high quality data, a correspondingly high quality sample is required. The analyte of interest must be monodisperse and of the highest possible purity (>95%). Including size exclusion chromatography as a final step in purification is required to ensure purity and check for aggregation.[57] In addition, the use of techniques such as multiangle light scattering combined with size exclusion chromatography can be useful for checking sample quality. Even a small degree of aggregation present in the sample can skew the results of analysis. This necessitates collecting data at multiple concentrations to determine the ideal conditions for each individual analyte. For proteins, the ideal concentration is generally between 1 and 10 mg/mL. A more in-depth discussion of this will be given in the Data Analysis section.

As with many other popular spectroscopic methods, SAXS/WAXS is a contrast technique. The data resulting from these experiments (I_{total}) are defined as the sum of the intensities of scattering due to the buffer (I_{b}) and the analyte (I_{a}):

$$I_{\text{total}}=I_{\text{b}}+I_{\text{a}} \quad (2)$$

In most cases this is not an issue as it is fairly straightforward to produce a second sample with an identical buffer system, just excluding the protein to be studied. These two samples are run under identical conditions and their corresponding data subtracted from one another:

$$I_{\mathrm{a}}=I_{\mathrm{total}}-I_{\mathrm{b}} \quad (3)$$

There are, however, some other important considerations that need to be taken into account. The contrast between the analyte and bulk solvent is dependent upon differences in electron density.[58] Water, with low concentrations of salts, has an electron density of ~0.33 e/Å whereas proteins have a slightly higher electron density of ~0.44 e/Å.[59] Thus, salt concentration should be kept below 1 M and ideally less than 100 mM. In addition to signal from bulk solvent and analyte, scattering by ordered layers of solvent around the analyte must be taken into account. This implies that the presence of detergents would be detrimental to data quality, in terms of adherence to proteins and in formation of bicelles/micelles, and must be used at very low to null concentrations.[60] Overall the signal collected in these experiments (I_{total}) is the sum of scattering intensities due to the buffer (I_{b}), protein (I_{a}), and ordered water layer around the protein (I_{w}).

$$I_{\mathrm{total}}=I_{\mathrm{b}}+I_{\mathrm{a}}+I_{\mathrm{w}} \quad (4)$$

Finally, because the sample is subjected to high energy X-rays, radiation damage can occur. Radical species scatter light very strongly and so it is particularly important to minimize radiation damage.[61] This is usually achieved in one of two ways. The X-ray source can be pulsed so as to let the sample recover from each exposure or a unidirectional constant flow cell can be utilized.[62] Flowing the solution through the exposure area will limit the amount of radiation any given portion of the sample receives but require much more material. A static cell may require as little as 1 μL whereas a flow cell may require up to 100 μL.

2.2. Data Collection: Instrument Setup

The main components of a SAXS instrument are the X-ray source, the collimator, the sample holder, and the detector (Figure 1).[63] X-rays from the source, whose energy is around the 1 Å range, are focused by the collimator. The spatially coherent beam interacts with the sample and is scattered by an angle 2θ, and then travels toward the detector.[64] The gathered data is sent to a computer for storage and analysis.

X-rays can be generated from a variety of sources. Historically, high voltage vacuum tubes were used but are not the best choice because of energy inefficiencies and the relatively low energy photons they produce.[65] The most common sources today are synchrotron beamlines that produce high energy photons by accelerating electrons in a magnetic field.[66] These are particularly useful because many different instruments can be set up to use the same radiation source simultaneously.

Focusing X-rays using lenses is near impossible and reflective focusing is only a little better.[67] Instead, collimation is used as a subtractive means of colinearizing photons in the beam.[68] This is ideal because even with a focused photon source there will be changes in the diameter of the beam, as a function of distance from the lens, which would interfere with the resulting data. A perfectly collimated photon source is one with no divergence; that is all photons travel a parallel path with respect to the others. In this way, the only signal on the detector outside of the beam's radius is due to scattering from the sample and not due to imperfections in the photon source.

The beam interacts with the sample (whose general criteria has been covered above) and is scattered (as described below) by an angle 2θ. The sample holder must either interact minimally with the incident photons both in terms of scattering and absorbance,[69] or must do so in a very predictable and reproducible way so that its effect can be subtracted out of the data.

2.3. Data Analysis

The data from a SAXS/WAXS experiment consist of a plot of scattering intensity over reciprocal space. As discussed above, the scattering intensity of the analyte, $I(q)$, is the difference between the scattering of the analyte in the sample matrix and the matrix itself. The independent variable q is defined by eq 1. Occasionally in the literature, the variables s and h are used for q or $q/2\pi$, respectively, so one must be careful to note the definition of the momentum transfer variable. q has units of Å^{-1} which roughly translates to 2π times the inverse resolution, being that $I(q)$ is the Fourier transform of the electron density function $\rho(\mathbf{r})$.

The curve $I(q)$ can generally thought to be broken into three regions; low q, medium q, and high q (Figure 2a). Low and medium q data are collected at a small angle (hence SAXS), while high q data are collected at wide angle (hence WAXS)

Low q refers to the range $q < 0.1$ Å^{-1} and reflects the size of the particle of interest. This metric is called the radius of gyration, R_{gyr} (defined as the root-mean-square distance of all atoms from their relative location to the particle's center of mass). R_{gyr} is a good indicator of overall size and can be calculated by regression to I(0) using either the Guinier[70] or Debye[71,72] approximations:

$$\text{Guinier:} \ln I(q) = \ln I(0) - \frac{q^2 {R_{\mathrm{gyr}}}^2}{3} \qquad (5)$$

$$\text{Debye:} I(q) = \frac{2I(0)}{q^4 {R_{\mathrm{gyr}}}^4} \left(q^2 {R_{\mathrm{gyr}}}^2 - 1 + e^{-q^2 {R_{\mathrm{gyr}}}^2}\right) \qquad (6)$$

R_{gyr} can also be calculated by other methods that will be covered below.

The medium q range ($0.1 < q < 0.5$ Å^{-1}) is important for defining the shape of the particle. In Figure 2b we show a few examples of particle shapes and their corresponding $I(q)$ curve.

It is relatively difficult to determine the shape of a particle just based on visual inspection, but luckily there are computational procedures to aid in this respect (see below).

The high q range of the curve ($q > 0.5$ Å^{-1}) is dominated by the atomic resolution structure of the particle of interest, and is usually characterized by a lower signal-to-noise ratio.

While it is difficult to access the quality of SAXS data by a cursory glance at $I(q)$, there is one important thing to consider. $I(q)$ is the product of a form factor,[73] $F(q)$, that describes the shape of the scattering particle, and a structure factor, $S(q)$, that describes the interaction between different particles in the sample.[74]

$$I(q)=F(q)\cdot S(q) \quad (7)$$

Ideally $S(q)$ equals unity for every value of q, and $I(q)$ only reflects the particle shape. At low q, $S(q)$ is very sensitive to oligomerization state and the presence of aggregation, and therefore it is usually good practice to check the early points of the SAXS profile to assess the aggregation state of the sample. In particular an increase in $I(q)$ is diagnostic of an aggregated sample,[75] while, in contrast, a dip in the $I(q)$ curve indicates repulsion between molecules.[76] A more quantitative way of evaluating the aggregation state is by using the Guinier plot. For $S(q) = 1$, the Guinier transformation (eq 5) yields a straight line for $0.65/R_{\mathrm{gyr}} < q < 1.3/R_{\mathrm{gyr}}$.[77] Curvature in this plot can be attributed to sample imperfections and a residual plot can help identify such errors (Figure 3).

SAXS/WAXS experiments are limited in their information richness relative to similar instrumental techniques such as X-ray crystallography because of solution state tumbling which averages out the three-dimensional electron density map to a one-dimensional set of spherical shells. In crystallography, the Patterson function is an autocorrelation function that shows all atom pairs (including self-pairs) and provides the best solution to crystal structures without solving the phase problem.[78] The SAXS analog is the pair-distribution function,[79] $P(r)$, the interpretation of which is a histogram of all atom–atom correlations plotted against the distance between the atoms (Figure 4). $P(r)$ is calculated by an indirect Fourier transform of $I(q)$ over $0 < r < D_{\mathrm{max}}$:[80]

$$P(r)=\frac{r}{2\pi^2}\int_0^\infty qI(q)\sin(qr)\,\mathrm{d}q \quad (8)$$

where r is the distance between scattering elements (atoms in the case of a protein) and D_{max} is the maximum diameter of the particle. Because the value of D_{max} may not be known prior to calculation of $P(r)$, iterative optimization can be used to define the ideal bounds for the integral. The pair-distribution function is useful for determining the shape of the analyte as well as its size. Spherical proteins will give a symmetrical, bell-shaped $P(r)$ distribution while increasing deviations will result in tailing of the peak shape to higher values of r. When $P(r) = 0$, $r = D_{\mathrm{max}}$ as this is the point where there are no larger atom–atom pairwise distances. Once D_{max} has been determined, another method of approximating R_{gyr} can be calculated using the equation

$$R_{\mathrm{gyr}}{}^{2}=\frac{\int_{0}^{D_{\max}} r^{2} P(r)\,\mathrm{d}r}{\int_{0}^{D_{\max}} P(r)\,\mathrm{d}r} \quad (9)$$

2.4. Computational Modeling

Noncomputational SAXS analysis is useful for basic characterization of macromolecules, but the real value lies in computational modeling. It is possible to define molecular envelopes that describe the low-resolution shape of a protein or complex as well as carry out hypothesis testing to differentiate between discrepant models.[81,82] However, one needs to take care when considering possible envelopes calculated from SAXS data. 1D SAXS data cannot be used to directly calculate a 3D envelope. Extrapolating into higher dimensional space requires some level of "guess and check" and it is possible (and likely) to come up with many different envelopes that fit the SAXS curve quite well. No matter the quality of the SAXS data, an envelope can still be calculated! This is not to say that computational modeling of SAXS data is unreliable; on the contrary, it has been successfully used to resolve differences between solid and solution state structures, and to model missing residues in crystal structures.[83,84] In addition, the chance of accidentally misinterpreting data is almost eliminated when SAXS is used in combination with other techniques such as NMR or crystallography (see below). But it is important to keep in mind the limitations and dangers of overanalyzing results derived from SAXS.

The SAXS curve is back-calculated from a structural model using the equation

$$I(q)=\langle |A(q)|^{2}\rangle_{\Omega} \quad (10)$$

where $\langle\ \rangle_{\Omega}$ denotes the average over solid angle Ω (the average due to molecular tumbling in solution) and $A(q)$ is given by

$$A(q)=\sum_{j} f_{j}^{\mathrm{eff}}(q)\mathrm{e}^{iqx_{j}}+\sum_{k} f_{k}^{\mathrm{sph}}(q)\mathrm{e}^{iqy_{k}} \quad (11)$$

where $\boldsymbol{q}$ is the reciprocal space scattering vector with amplitude q; j sums over all atoms; $f_{j}^{\mathrm{eff}}(q)$ is the effective atomic scattering form factor; x_j is the position of atom j; k sums over points representing boundary-associated solvent; and $f_{k}^{\mathrm{sph}}(q)$ and x_k are, respectively, the positions and scattering form factors of these points. In eq 11 the first sum describes the scattering from each solute atom, while the second sum describes scattering from ordered solvent molecules bound to the surface of the macromolecule.[85] Details about back-calculation of SAXS profiles, as well as of the mathematical tricks implemented for speeding up computation, are reviewed elsewhere.[86–88]

In the context of analyzing conformational ensembles, the experimental SAXS curve is a population-weighted average of the SAXS profile of each conformer.[47] Therefore, given known structures for distinct conformational states, relative populations can be determined by simple linear combination of back-calculated scattering curves to maximize agreement with the experimental $I(q)$.

3. OVERVIEW OF THE USE OF RESIDUAL DIPOLAR COUPLINGS IN STRUCTURAL BIOLOGY

Of the various types of data that can be obtained by solution state NMR, residual dipolar couplings (RDCs) are very useful for structure refinement because they yield long-range orientational information.[89,90] RDCs describe the orientation of bond vectors relative to the static magnetic field.[91] To observe RDCs, the molecule under investigation must tumble anisotropically in solution so that the RDCs are not averaged out to zero. Weak alignment (of the order of 10^{-3}) can be readily achieved using various dilute alignment media.[92] This section of the paper will focus on the theoretical basis for RDCs, alignment media, alignment tensor prediction, and use of RDC restraints in structure calculation.

3.1. Theoretical Background

Dipolar couplings are through space interactions between magnetic nuclei. The dipolar interaction between two spins (D_{AB}) of gyromagnetic ratios γ_A and γ_B at a given distance, r, is given by

$$D_{\mathrm{AB}}(\vartheta)=D_{\max}\left\langle\frac{3\cos^2\vartheta-1}{2}\right\rangle \quad (12)$$

where ϑ is the angle between the internuclear bond vector and the external magnetic field (Figure 5) and $D_{\max}$ is the maximum value of the dipolar coupling, given by

$$D_{\max}=-\frac{(\mu_0\gamma_{\mathrm{A}}\gamma_{\mathrm{B}}h)}{(8\pi^3r^3)} \quad (13)$$

where μ_0 is the magnetic permeability in a vacuum and h is Planck's constant.[93] $\langle\rangle$ in eq 12 denotes the average over all possible orientations of the internuclear bond vector relative to the external magnetic field. In isotropic solution all orientations are possible, and D_{AB} averages to zero. Addition of an alignment medium breaks the orientational symmetry and results in nonzero values of D_{AB}.

Eq 12 can be recast in terms of a molecular coordinate frame, called the alignment tensor, that describes the relative orientation of the molecule with respect to the alignment medium:

$$D_{\mathrm{AB}}(\theta,\phi)=D_a^{\mathrm{AB}}\left[(3\cos^2\theta-1)+\frac{3}{2}(\sin^2\theta\cos 2\phi)\right] \quad (14)$$

where θ is the angle between the internuclear bond vector and the z axis of the alignment tensor, ϕ the angle between the xy plane projection of the internuclear bond vector and the x axis (Figure 5), D_a^{AB} the magnitude of the axial component of the alignment tensor, and η the rhombicity.[94]

3.2. Alignment Media

The maximum strength of a dipolar coupling for completely aligned samples is of the order of 20 kHz for ^{1}H–^{15}N and 40 kHz for ^{1}H–^{13}C dipolar interactions. This would cause the NMR spectrum to be so complex and broad that it would be uninterpretable. To circumvent this problem, dilute alignment media are used resulting in an ordering of ~10^{-3} with maximum residual dipolar couplings (RDC) of about 20 Hz.[95,96] There are many different alignment media and these have been extensively reviewed.[97–103] Here we give a brief outline of the most popular media employed in structural biology applications.

Bicelles were the first media to be used for RDC measurements.[95] They are relatively easy to prepare and alignment is somewhat tunable based on the concentration of lipids. Most commonly, dimyristoylphosphatidylcholine (DMPC) and dihexanoylphosphatidylcholine (DHPC) are used to make bicelles. At low temperatures, bicelles are isotropically oriented but at temperatures greater than 30 °C they transition to a nematic liquid crystalline phase. The normal vector of the bicelles is oriented perpendicular to the applied magnetic field. Using DMPC and DHPC results in purely steric alignment but trace amounts of charged lipids can be incorporated to produce varying degrees of electrostatic alignment.[97] This makes bicelles quite useful because multiple RDC data sets can be acquired under similar solution conditions.

The second alignment medium to be described comprised rod-shaped, filamentous bacteriophages and viruses.[96,104] A commonly used phage is pf1 that is approximately 20 000 Å long and 60 Å in diameter.[98] These virus particles, which are commercially available, spontaneously align in a magnetic field making them especially convenient for NMR. Their surface is covered in negative charges so proteins are aligned through both steric and electrostatic interactions. This can cause problems with positively charged proteins because they may align too much. The extent of alignment can be tuned by varying the concentration of phage and salt. Typically, the concentration of phage employed is 10–20 mg/mL with 100 mM or less NaCl at pH 6.5 to 8.

Polyacrylamide gel can be used for alignment of a protein when the gel is mechanically stressed.[105,106] The pores in the gel become elongated when compressed which allows for steric alignment. First the acrylamide is polymerized using 0.1% w/v ammonium persulfate and 0.5% w/v tetramethylethylenediamine in a tube, generally, with an inner diameter between 3.5 mm to 8 mm. The gel is then washed and dehydrated. The gel is rehydrated in an NMR tube using the protein solution and then compressed using the plunger of a Shigemi NMR tube. Alternatively, radial compression can be achieved by pushing a gel with a diameter larger than the NMR tube through a funnel into the NMR tube. The alignment tensors of radially and axially compressed samples are of opposite sign and do not provide new data. Electrostatic alignment using negative or positive charges is also possible if 2-

acrylamido-2-methyl-1-propanesulfonic acid or diallyldimethylammonium chloride, respectively is substituted for acrylamide during polymerization.[107] Either dialysis or dissolving the gel and centrifuging readily accomplish sample recovery.

3.3. Measuring RDCs

Measurement of RDCs is based on various common *J* scalar coupling experiments as well as some more sophisticated pulse schemes. RDCs appear in NMR spectra much in the same way that *J*-couplings do (Figure 6). The RDC between two nuclei is additive with that of the corresponding *J* coupling (Figure 6). To determine the magnitude of an RDC, one must know the sum of the dipolar and *J* coupling as well as the *J* coupling by itself, as shown in the equation below:

$$T=J+D \quad (15)$$

where J and D are the size, in Hz, of the J coupling and RDC, respectively, and T is the observed splitting.[108] This constitutes the basis of measuring RDCs in all of the available types of experiments. To determine the value of J, the experiment is carried out on a sample of the macromolecule under isotropic conditions. Using the same conditions, save for the inclusion of an alignment medium to produce anisotropy, the experiment is repeated to observe the new splitting that is equal to T. Subtraction of the splitting T from J yields the value of D.

The most commonly collected RDC for proteins is the ^{1}H–^{15}N backbone amide set because of its sensitivity and spectral resolution even for large, slow tumbling proteins. However, many other nuclei pairs provide useful structural information including ^1H_N-$^1H_\alpha$, ^1H_N-$^{13}C_\alpha$, $^1H_\alpha$-$^{13}C_\alpha$, ^{15}N-$^{13}C_\alpha$, ^{15}N–^{3}C′, $^{13}C_\alpha$-^{13}C′, $^{13}C_{\text{methyl}}$-^{13}C, $^1H_{\text{methyl}}$-$^{13}C_{\text{methyl}}$, and ^{1}H–^{1}H.[104,109–114] It is generally more straightforward to collect and analyze data from nuclei pairs that have a relatively large *J*-coupling, such as ^{1}H–^{15}N and ^{1}H–^{13}C, because there is almost no possibility of an unresolved doublet. However, there is a greater chance of spectral overlap with other signals due to the larger *J*-coupling. With smaller *J*-coupled systems, such as ^{13}C–^{13}C and ^{15}N–^{13}C, there is less of a problem with signals interfering with one another and more of a problem with unresolved doublets.[102]

There are two main techniques for collecting RDC data: IPAP and quantitative *J*-modulation.[115,116] Of these, there are important extensions that apply to various challenges. IPAP stands for in-phase/antiphase, and as the name implies, this experiment allows for the collection of the in-phase doublet spectrum, and then subsequently in an interleaved manner, the antiphase doublet spectrum.[109] The most commonly used example of this strategy is in the ^{15}N-IPAP heteronuclear single quantum coherence (HSQC) experiment, where the spectrum yields data just as a regular ^{1}H–^{15}N HSQC but includes the $^1J_{HN}$ splitting in the nitrogen dimension. This data collection scheme is unique in that it allows for the differentiation between the upper and lower components of a doublet. When the two spectra are summed together, the low field component of the doublet is added together while the high field component is canceled out. Likewise, the difference of the spectra results in canceling the lower field component, while adding the higher component. This technique is particularly useful when there is concern of spectral overlap due to doubling the number of peaks in the coupled spectrum.

The quantitative *J*-correlation experiment measures the *J* coupling between two nuclei by exploiting differences in signal intensities due to evolution of the scalar coupling.[117] Two experiments are acquired in an interleaved fashion. The first is a decoupled reference spectrum. The second spectrum allows the coupling to evolve during a spin-echo period. The ratio of the peak intensities is proportional to the *J* coupling between the nuclei. Alternatively, the time delay during the spin–echo period can be varied through a series of experiments so as to collect multiple peaks of varying intensities.[118] These intensities are fit to a trigonometric function to obtain the value of the coupling constant. The second method is generally more accurate, mostly by virtue of using more data points. The results of either method are resilient to signal loss effects, such as relaxation, that might skew peak integrations because each spectrum is normalized to a reference experiment that is identical to the attenuated experiment. More sophisticated versions of the basic quantitative *J* experiment have been developed and are primarily focused on expanding the applicability to larger systems. ARTSY (amide RDC by TROSY spectroscopy) is one such experiment based on a two-dimensional ^{1}H–^{15}N transverse relaxation optimized (TROSY) HSQC.[119] Here, the normal interleaved reference/attenuated data collection scheme is utilized. The main difference is that in the reference experiment the ^{1}H signal is allowed to dephase for half of the spin-echo period and in the attenuated experiment it dephases for the entire spin-echo period. As was the case before, taking the ratio of the attenuated to reference peak intensities will allow for the calculation of J or $J + D$ in isotropic and aligned media, respectively.

When dealing with large systems, transverse (R_2) relaxation and spectral crowding becomes a considerable problem when measuring RDCs. One method to surmount both of these issues simultaneously is selective isotope labeling of methyl groups. During the production of the protein of interest, precursors are used to selectively label methyl groups of the side chains of isoleucine (I), leucine (L), valine (V), and alanine residues.[120–123] Spectra of these selectively labeled proteins are much less crowded than their uniformly labeled counterparts and, additionally, the relaxation characteristics of ^{13}CH$_3$ methyl groups are particularly ideal.[124,125] This enables high resolution measurement of RDC data in large systems that might otherwise suffer from very poor spectral quality. For ILV-selectively labeled proteins it is possible to collect ^{1}H$_{\mathrm{methyl}}$-^{13}C$_{\mathrm{methyl}}$ and ^{13}C$_{\mathrm{methyl}}$-^{13}C RDCs in spectra that are extended to a third dimension (^{13}C$^{\beta/\gamma}$) to further increase resolution of crowded spectra.[126] This experiment is again based on the quantitative *J* methodology and yields best results using multiple delay values during *J* evolution and curve fitting to obtain the coupling constant.

3.4. Data Analysis

To utilize RDCs as restraints in structure refinement, the alignment tensor must be determined. The alignment tensor comprises 5 terms: the magnitude of the axial component of the tensor (D_a), the rhombicity η of the tensor, and three parameters describing the orientation of the tensor. In simulated annealing refinement, the three orientation parameters can be represented by an orthogonal axis system that is treated as a rigid body and allowed to rotate during the course of the calculation. D_a and η can be determined directly from a histogram of the measured RDCs by noting that the maximum of the distribution (corresponding to D_{zz}) for a fixed distance interaction is given by $2D_a^A$, the minimum (corresponding to D_{yy}) by $-D_a(1 + 1.5\eta)$, and the mode (corresponding to D_{xx}) by $-D_a(1-1.5\eta)$, with the additional constraint that the sum of $D_{xx} + D_{yy} + D_{zz} = 0$ (Figure 7).[127] The robustness of this simple approach can be enhanced by application of maximum likelihood methods.[30,128] Alternatively, the values of D_a and η can also be optimized during simulated annealing.[30]

If a structure has already been determined, its quality can be assessed using singular value decomposition (SVD) to obtain the best-fit alignment tensor that minimizes the difference between the observed RDCs and those back-calculated from the structure.[93]

The quality of the RDC fit to the coordinates is assessed using an R-factor (spanning from 0 indicating no agreement to 1 for perfect agreement) that can be expressed as

$$R=\left\{\left\langle (D_{\mathrm{obs}}-D_{\mathrm{calc}})^2\right\rangle /(2\left\langle {D_{\mathrm{obs}}}^2\right\rangle)\right\}^{1/2} \quad (16)$$

where D_{obs} and D_{calc} are the observed and calculated RDCs, respectively.[129] $\langle {D_{\mathrm{obs}}}^2\rangle$ can either be the experimental value or can be calculated exactly by $2(D_a)^2(4 + 3\eta^2)/5$.[129] The latter, however, is not suitable for an ensemble of structures since each ensemble member will have different values of D_a and η.

The manner in which the protein transiently interacts with the alignment medium is determined by steric and electrostatic effects (see above). The degree to which these two factors affect the orientation of the macromolecule under investigation varies from one alignment medium to the next, and also depends on the shape and charge distribution of the protein. For primarily steric alignment, the orientation can be approximated very well by obstruction theory.[130] The probability to find the molecule in a certain orientation $\Omega = (\alpha, \beta, \gamma)$ is given by

$$P(\Omega)=\frac{L-l(\Omega)}{4\pi L-\int l(\Omega)\,\mathrm{d}\Omega}\cong\frac{L-l(\Omega)}{4\pi L} \quad (17)$$

where L is the distance between two parallel planes, $l(\Omega)$ is the length of the molecule in the direction orthogonal to the planes, and α, β, and γ are Euler angles that describe the orientation of the alignment tensor.[131] From this equation, it is immediately obvious that molecules will preferentially align with their long axis parallel to the planes. There are common programs to carry out this calculation such as PALES/SSIA[132] and Xplor-NIH.[30] Xplor-NIH can also readily carry out this calculation during simulated annealing which is extremely useful when refining a structure ensemble.

It is much more difficult to predict the orientation of a protein in an alignment media that is dominated by electrostatic interactions because the surface charge distribution of the molecule plays a big role in alignment. The alignment tensor of macromolecules that have a very uniform charge distribution, such as DNA, can be accurately described in electrostatic aligning media while most proteins suffer due to their nonuniform surface charge.[133]

Recently, it has been shown that it is also possible to perform structure calculations incorporating RDCs without using an alignment tensor.[134,135] The technique, dubbed the "ϑ method", gets around the necessity of the alignment tensor by using eq 12 directly in the refinement protocol (i.e., without recasting into eq 14). When using the ϑ method, the orientation of the molecule relative to the external magnetic field is optimized to maximize the linear correlation between experimental RDCs and the dipolar couplings back-calculated from the structure using eq 12. Therefore, even though the ϑ method eliminates the need for calculating the alignment tensor, it still requires optimization of four variable parameters (three rotational degrees of freedom and an RDC scaling factor) to describe the alignment of the molecule in the external magnetic field.

4. COMBINED USE OF SAXS AND RDC DATA FOR STRUCTURE DETERMINATION OF COMPLEX MOLECULAR SYSTEMS

Traditionally, NMR structure determination of macromolecules requires close to complete resonance assignments and acquisition of extensive data sets of NMR-derived structural restraints. In particular, short (up to ~6 Å) distance restraints, which are obtained by the analysis of NOE data, and restraints on the backbone torsion angles from three-bond scalar couplings and backbone chemical shifts, have been extensively employed to solve the 3D structures of globular proteins.[136] More recently, introduction of orientational restraints from RDC data measured in a dilute liquid crystalline media has enormously expanded the complexity of protein folds and assemblies amenable to structure determination by NMR.[95,137,138] Once local geometries are defined by NOE and dihedral angle restraints, RDC data can be used to orient rigid groups of atoms (such as secondary structure elements and domains) relative to one another. The beauty of this approach is that, if the structural domains can be treated as rigid bodies, only a small number of RDCs is required to orient the domains relative to one another.[139]

The main limitation of using RDCs to derive orientational information is the 4-fold degeneracy for orienting the alignment tensor that results in multiple solutions.[140] Ideally, NOE distance restraints[139] or a second noncolinear alignment tensor[141] can be used to resolve this ambiguity. However, these methods require measurements of interdomain NOEs or acquisition of an orthogonal set of RDC data, and are usually not applicable to larger systems for which only a very restricted set of NOEs may be measurable, and that may only be compatible with a single alignment medium. In such cases, an alternative strategy using hybrid methodology is required.

The most generally applicable hybrid method combines RDC data with SAXS/WAXS and treats structural domains as rigid bodies. RDCs provide orientational restraints and, in purely steric alignment media, shape information as well (see above), while SAXS provides complementary information on size and shape.[47] Conjoined rigid body/torsion angle simulated annealing driven by the RDC and SAXS data is then used to obtain structural solutions that are consistent with the experimental data. To use such a hybrid approach, the structures of the individual domains must be known. This can be achieved either by experimental techniques (i.e., X-ray crystallography or NMR) or, under limited circumstances, by homology modeling when the degree of sequence similarity is very high (>60%) and there are no gaps or insertions between the experimental structure and the domain being modeled. In any case, before their use in the simulated annealing protocol, the quality of the structures must be carefully assessed by comparison of the RDCs measured on the multidomain molecule with those back-calculated from the high-resolution structures of the building blocks. It is worth mentioning that inconsistencies between experimental and back-calculated RDCs do not necessarily mean that the structures of the isolated domains are of poor quality. Indeed they can simply be the result of a conformational change induced by interdomain interactions that are absent under the experimental conditions used to determine the 3D structures of the building blocks. Rigid-body simulated annealing driven by SAXS and sparse RDC data has been used to solve the structures of large multidomain proteins,[47,48,142] complex nucleic acids,[143–145] and protein oligomers.[146] Recently, the technology has been expanded to the refinement of conformational ensembles.[147]

4.1. Combined Use of SAXS and RDCs for Generation of Conformational Ensembles

A crucial step in setting up a structure calculation protocol is to decide whether to represent the molecule under investigation using a single structure or a conformational ensemble. As a general rule, an ensemble representation can only be invoked if the experimental data cannot be explained by refinement of a single structure. In addition, in the case of multidomain proteins, the presence of large amplitude interdomain motion can be confirmed by NMR relaxation measurements (sensitive to ps-ns time scale motions)[148,149] or by the analysis of experimental RDC data (sensitive to motions up to the ms time scale).[150,151]

When using a conformational ensemble the experimental data are considered as a global average representation of the system. Thus, at each step of the refinement protocol the RDC and SAXS data are back-calculated from the conformational ensemble as population-weighted averages over the ensemble members. This means that the relative populations of the ensemble members (also referred to as ensemble weights) must be optimized during the refinement. Broadly speaking, there are two ways to optimize ensemble weights. In several applications all ensemble members are given equal populations throughout the simulations.[20,21] These protocols rely on the fact that if a particular conformation is prevalent in the experimental sample, it will also be prevalent in the computational ensemble. The main limitation of such approach is that to properly describe relative populations between clusters of conformations, large ensemble sizes must be used. An alternative approach is to use a limited number of ensemble members and optimize their relative populations directly in the simulated annealing calculation.[142,147]

If interconversion between conformational states is in the fast-to-intermediate exchange regime (nano- to millisecond range), each ensemble member has its own alignment tensor.[142,147] Letting the alignment tensors float during refinement would add five variable parameters to the fit per ensemble member (see above), and would result in ill-defined and unstable fits. This fundamental problem can be circumvented by using a purely steric alignment medium so that the alignment tensor for each ensemble member can be directly calculated from its molecular coordinates at each step of the simulated annealing protocol.[131,142,147] We should also note that the tensor-free ϑ method[134] only reduces the number of parameters required to describe alignment from five to four per ensemble member (see above) and therefore does not provide a suitable alternative.

The optimal ensemble size (N_e) is usually determined empirically by searching for the smallest ensemble size that satisfies the experimental data.[20] To avoid data overfitting (for example by using an unreasonably large ensemble size), parameters for evaluating agreement between experimental and back-calculated data must be carefully chosen, and, where possible, additional data sets should be collected and reserved for cross-validation of the calculated conformational ensemble. In the context of ensemble refinement against RDC and SAXS data, the agreement between the structural ensemble and experimental RDCs is conveniently evaluated in terms of an R-factor (see above).[129] The target global R-factor value is given by the weighted average of the R-factors of the individual structural domains:

$$\text{global } R_{\text{target}} = \sum_i \frac{R_i N_i}{N} \qquad (18)$$

where R_i is the R-factor determined from the known structure of the ith structural domain using SVD, N_i is the number of RDCs measured for the ith structural domain, and N is the total number of experimental RDCs measured for the multidomain protein. Agreement between experimental and back-calculated SAXS data, is evaluated in terms of χ^2 with a target value of ~1.[152] Final R-factor and χ^2 values that are smaller than the corresponding target value are indicative of overfitting the data.

In the following sections the successful application of rigid-body ensemble refinement driven by RDC and SAXS data to the challenging cases of the 128 kDa dimer of bacterial Enzyme I and the HIV-1 capsid protein will be discussed.

4.2. Solution Structure and Dynamics of Bacterial Enzyme I

Enzyme I (EI) is the first protein in the phosphoenolpyruvate (PEP):sugar phosphotransferase system (PTS),[153] a key signal transduction pathway involved in the regulation of central carbon metabolism in bacteria. EI is responsible for both activation and regulation of the overall PTS.[19,154,155] EI is a large, dynamic protein that presented a real challenge to structural biologists. In this section we will analyze how only the integrated analysis of data from multiple techniques allowed for characterization of the structure and dynamics of EI.

From the structural point of view, EI is a 128 kDa dimer of identical subunits (Figure 8). Each subunit comprises two structural domains.[156] The N-terminal domain (EIN) is further divided in two subdomains, named EIN$^{\alpha}$ and EIN$^{\alpha/\beta}$, respectively. EIN$^{\alpha/\beta}$ contains the active site residue (His189) that is autophosphorylated by phosphoenolpyruvate (PEP). EIN$^{\alpha}$

provides the binding site for HPr (the second protein in the PTS pathway).[157,158] The C-terminal domain (EIC) is responsible for protein dimerization and contains the binding site for PEP.[17,159]

Several X-ray structures of EI have been solved showing that the enzyme can adopt different conformations (Figure 8). Indeed, the free EI from *Staphylococcus aureus*[160] and *Staphylococcus carnosus*[161] display two distinct open states in which the His189 is positioned more than 20 Å away from the PEP binding site on EIC, and the EIN domain adopts a conformation, named the A-state, in which the EIN^{α} and $\mathrm{EIN}^{\alpha/\beta}$ subdomains form extensive contacts with one another. On the other hand, the X-ray structure of a trapped phosphorylated intermediate of *Escherichia coli* EI,[162] solved by crystallizing EI from a solution containing PEP and Mg^{2+} and quenching the autophosphorylation reaction using oxalate, shows the enzyme in a closed conformation. In the closed state His189 is inserted into the PEP binding pocket on EIC and is positioned for in-line phosphoryl transfer from PEP to EIN. The open-to-closed transition is coupled to a conformational change in the EIN domain that involves a ~90° reorientation of EIN^{α} relative to $\mathrm{EIN}^{\alpha/\beta}$ (Figure 8). The EIN conformation observed in closed EI is referred to as B-state. Interestingly these rearrangements do not affect the local fold of the structural domains (EIN^{α}, $\mathrm{EIN}^{\alpha/\beta}$ and EIC) that display the same structure in all the crystal structures. However, an analysis of the EI structures on the basis of experimental SAXS data acquired for free EI indicates that none of these crystal structures corresponds to the solution structure of the enzyme (χ^2 for the fits to the SAXS data is >30 for all three crystal structures).[47,163]

To solve the solution structure of free EI, a conjoined rigid body/torsion angle/Cartesian coordinate simulated annealing refinement protocol driven by the experimental RDC and SAXS was employed in which the structural domains were treated as rigid bodies, the backbone of the linkers were given Cartesian degrees of freedom, and side chains were allowed torsion degrees of freedom.[47] Due to the large size of the enzyme, assignment of the NMR spectra was achieved by transferring the assignments of the isolated EIN domain onto the spectra of the full-length protein. At the time this work was performed, no assignments were available for EIC. Therefore, the structure calculation used only 58 backbone ^{1}H–^{15}N RDCs from the EIN domain (29 for EIN^{α} and 29 for $\mathrm{EIN}^{\alpha/\beta}$). These RDCs are fully consistent with the NMR and X-ray structures of isolated EIN in the A conformation but incompatible with the B conformation found in the structure of the closed state of EI; hence the EIN domain was held fixed to the NMR structure of free EIN. As mentioned earlier, the structure of the EIC dimer is the same in all the X-ray structures of EI; therefore, the EIC portion of the enzyme was kept fixed to the X-ray coordinates throughout the calculation. Given that one of the principal axes of the alignment tensor must coincide with the C_2 symmetry axis of the dimer,[164,165] the orientation of the EIN domains relative to the EIC dimer can be determined from sparse RDCs located only in the EIN domain. The results of the conjoined RDC/SAXS refinement revealed a new open structure for EI that is consistent with the experimental data (Figure 9). Interestingly, the data were fully satisfied by a single conformation, suggesting that the closed state, if at all present, is populated to a very minor extent (<5%).

More recently, a study combining NMR relaxation dispersion measurements and SAXS confirmed that the closed structure is sampled at detectable populations only in the presence of PEP.[163] Indeed, PEP-binding suppresses conformational exchange in two loops of EIC that are part of the EIN/EIC interface in closed EI. This structural stabilization of the EIC domain by PEP binding activates the open-to-closed transition that allows EI to access the catalytically competent closed state. In addition, systematic analysis of SAXS profiles acquired for wild type EI (EI^{WT}) and two active site mutants on the basis of the solution structure of open EI and of the crystallographic closed state, revealed that the closed state of the enzyme is largely prevalent (best fit population ~60%) in the complex between PEP and the H189A (EI^{A}) mutant of the enzyme,[163] opening the way to structural characterization of closed EI in solution.[142] 23, 20, and 25 backbone 1H–^{15}N RDCs were measured for the EIN^{α}, $EIN^{\alpha/\beta}$, and EIC, respectively, by aligning the EI^{A}-PEP complex in neutral bicelles, and were used together with SAXS data to restrain a simulated annealing refinement calculation in which EIN^{α}, $EIN^{\alpha/\beta}$ and EIC were treated as rigid bodies with the linkers given Cartesian degrees of freedom. Refinement against the RDC or SAXS data individually converges to a single structure representation of EI^{A}-PEP that reproduces the experimental data set used in the refinement but not the omitted data set (Figure 10). A two-member ensemble, however, is required to simultaneously satisfy the RDC and SAXS data (Figures 10 and 11). The two conformations are approximately equally populated and correspond to the crystallographic closed state, which is competent for phosphoryl transfer from PEP to the EIN domain, and a partially closed state that likely represents an intermediate between the fully closed state, which is only transiently sampled in the EI^{WT}-PEP complex, and the fully open apo state.[142] The partially closed state revealed by the hybrid RDC/SAXS approach is likely involved in substrate-binding and product-release steps, and eluded characterization by other crystallographic and solution techniques.

4.3. Solution Structure and Dynamics of the HIV-1 Capsid Protein

The human immunodeficiency virus (HIV-1) capsid protein assembles into a cone that encloses viral RNA.[166–168] The full-length capsid protein (CA_{FL}) consists of an arrow-shaped N-terminal domain (NTD) and a globular C-terminal domain (CTD) that undergoes a monomer/dimer equilibrium (dimer dissociation constant, K_D, ~40 μM at 25 °C) in solution (Figure 12). The N- and C-terminal domains are separated by a short, flexible linker (Figure 12). In the context of the mature HIV capsid, the N-terminal domains assemble into ~250 hexameric[169] and 12 pentameric[166] rings that form the exterior of the capsid. The N-terminal domain rings are connected to one another by symmetric C-terminal domain dimers (Figure 12). Although assembly of the HIV-1 capsid plays a crucial role in the virus lifecycle and infectivity, obtaining a comprehensive structural characterization of full length capsid protein prior to assembly has been hampered by the fact that the system is highly dynamic and heterogeneous. Indeed, large amplitude motions between the N- and C-terminal domains, as well as the presence of a dynamic monomer/dimer equilibrium, result in severe line-broadening of the NMR resonances of the linker and dimer-interface regions, making the study of full length capsid protein by conventional NMR techniques impossible.

By using the hybrid approach described here, Deshmukh et al. quantitatively determined the conformational space spanned by the N-terminal domain relative to the C-terminal domain in both monomeric and dimeric capsid protein.[147] The RDC and SAXS data were acquired for the full-length capsid protein at different concentrations (ranging from 50 μM to 260 μM) and used simultaneously to refine a conformational ensemble that include both monomeric and dimeric species. In the calculation the relative populations of monomer and dimer were fixed based on the protein concentration and dimer K_D's (determined independently by analytical ultracentrifugation at the same temperatures as those used in the RDC and SAXS measurements), and an equal number of ensemble members was used to describe the monomeric and dimeric species (i.e., $N_{\mathrm{e}}^{\mathrm{monomer}}=N_{\mathrm{e}}^{\mathrm{dimer}}$). The optimal ensemble size $(N_{\mathrm{e}}=N_{\mathrm{e}}^{\mathrm{monomer}}=2N_{\mathrm{e}}^{\mathrm{dimer}})$ was determined empirically to be 10. Indeed, further increases in ensemble size did not result in any significant improvements in agreement with the experimental RDC and SAXS data (Figure 13). The results show that the conformational space sampled by the N-terminal domain relative to the C-terminal domain is different in the monomer and dimer with a distinct pattern of transient interactions between the two domains (Figure 14). This is due to the fact that much of the conformational space sampled by the N-teminal domain in the monomer is no longer accessible in the dimer due to steric clash with the C-terminal domain dimer. The conformational ensembles derived for the dimer and monomer are characterized by six (Figure 14A) and three (Figure 14B) main structural clusters, respectively. Interestingly, one of the clusters obtained for the protein dimer (cluster 6, which account for ~5% of the conformational ensemble of dimeric full length capsid protein) closely resembles the configuration sampled in both pentameric and hexameric capsid assemblies (Figure 14C).[147] These results suggest that the HIV-1 capsid assembles via conformational selection of a sparsely populated species and that stabilization of other clusters relative to cluster 6 may partially inhibit capsid assembly. In this regard the transient interactions between the N and C-terminal domains observed in cluster 2 of the dimer would be predicted to be enhanced by mutation of Pro38, Arg132 and Lys203 to a hydrophobic residue (alanine) consistent with the experimentally observed reduced capsid assembly rates for these three mutants.

5. CONCLUDING REMARKS

An atomic level understanding of protein conformational dynamics is crucial to modem structural biology and holds the promise to address unanswered questions on the functioning of important biological molecules such as enzymes, molecular machines, and allosteric systems.[170] This goal, however, is challenging as the great majority of biologically relevant, dynamic systems are large and multimeric and, therefore, elude structural characterization by conventional techniques. In this context the use of integrative approaches that combine structural data from multiple techniques have been shown to be very useful in determining 3D structural models of large and complex molecular systems.[171] Here, we have reviewed the application of a hybrid method that uses conjoined rigid body/torsion angle/Cartesian simulated annealing refinement driven by SAXS and NMR-derived RDC data to model conformational states in large multidomain proteins. This hybrid approach is streamlined because, once the high resolution structures of the rigid domains used as building-blocks are known from experimental or computational studies, only sparse RDC data complemented by SAXS data are required to reliably calculate conformational ensembles. This hybrid strategy has been successfully employed to obtain a quantitative description of the magnitude and distribution of interdomain motions in the HIV-1 capsid protein[147] and the complex between bacterial Enzyme I and PEP.[142] The methodology described here is readily transferred to the study of many other challenging systems, especially those involving large multidomain proteins and their complexes. Moreover, not only can this approach characterize the conformational space sampled by one domain relative to another as in the case of the HIV-1 capsid protein,[147] but it can also detect the simultaneous existence of distinct conformations and characterize their structures as in the case of the 128 kDa complex of Enzyme I (H189A) with PEP.[142] Further, in the case of the HIV-1 capsid protein the simultaneous existence of monomer and dimer can readily be handled.[147] In this regard it is worth noting that it would be impossible to characterize a mixture of coexisting states by crystallography as the crystallization procedure would only allow one state to crystallize out, and the probability of obtaining crystals of the various states is likely to be very small. Similarly, the coexistence of multiple conformational states renders interpretation of cryoelectron microscopy images extremely difficult if not impossible. Of course, any structural approach involving solution NMR does place certain limits on molecular size but, with deuteration and appropriate methyl-specific labeling, systems up to 200–300 kDa can potentially be tackled.

Acknowledgments

We thank Charles Schwieters for helpful discussions. This work was supported by the Intramural Program of the National Institute of Diabetes and Digestive and Kidney Diseases, NIH (G.M.C.) and by startup funding from Iowa State University (V.V.).

Biographies

Vincenzo Venditti is currently an Assistant Professor in the Department of Chemistry and in the Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology at Iowa State University (2015 to present). He received his Ph.D. degree from the Università degli Studi di Siena and was a postdoctoral fellow in the Laboratory of Chemical Physics at the NIH (2009–2014). His current interests lie in the use of NMR spectroscopy to study structure, dynamics, and interactions in large, multidomain systems.

Timothy K. Egner is currently a graduate student in the Department of Chemistry at Iowa State University. He received his B.S. in Biochemistry and Mathematics from North Central College in 2014. His current research focuses on the use of NMR spectroscopy for structural characterization of biological macromolecules.

G. Marius Clore is currently an NIH Distinguished Investigator in the Laboratory of Chemical Physics at the National Institute of Diabetes and Digestive and Kidney Disease, NIH. He received a B.Sc. (1st class honors) in Biochemistry from University College London (1976), an M.D. from University College Hospital Medical School (1979), and a Ph.D. from the National Institute for Medical Research, London (1982). He was a member of the Scientific Staff of the MRC National Institute for Medical Research, London (1980–1984); Head of the Biological NMR Group at Max Planck Institute for Biochemistry, Martinsried, Germany (1984–1988); and Senior Investigator (1988–2011), Chief of the Protein NMR Section (1991 to present), and NIH Distinguished Investigator (2011–present) in the Laboratory of Chemical Physics, NIDDK, at the National Institutes of Health. He is an elected Member of the United States National Academy of Sciences, a Fellow of the American Academy of Arts and Sciences, and a Foreign Member of the Academia Europeae. Among the prizes he has been awarded are the Royal Society of Chemistry Centenary Prize (2011) and the Biochemical Society Centenary Award and Francis Gowland Hopkins Memorial Lecture (2013).

References

1. Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. Science. 1991; 254:1598–1603. [PubMed: 1749933]
2. Dill KA. Polymer principles and protein folding. Protein Sci. 1999; 8:1166–1180. [PubMed: 10386867]
3. Tang C, Schwieters CD, Clore GM. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. Nature. 2007; 449:1078–1082. [PubMed: 17960247]
4. Tsai CJ, Ma B, Nussinov R. Folding and binding cascades: shifts in energy landscapes. Proc Natl Acad Sci U S A. 1999; 96:9970–9972. [PubMed: 10468538]
5. Reiter NJ, Blad H, Abildgaard F, Butcher SE. Dynamics in the U6 RNA intramolecular stem-loop: a base flipping conformational change. Biochemistry. 2004; 43:13739–13747. [PubMed: 15504036]
6. Venditti V, Clos L 2nd, Niccolai N, Butcher SE. Minimum-energy path for a u6 RNA conformational change involving protonation, base-pair rearrangement and base flipping. J Mol Biol. 2009; 391:894–905. [PubMed: 19591840]
7. Venditti V, Niccolai N, Butcher SE. Measuring the dynamic surface accessibility of RNA with the small paramagnetic molecule TEMPOL. Nucleic Acids Res. 2007; 36:e20. [PubMed: 18056080]
8. Suh JY, Tang C, Clore GM. Role of electrostatic interactions in transient encounter complexes in protein-protein association investigated by paramagnetic relaxation enhancement. J Am Chem Soc. 2007; 129:12954–12955. [PubMed: 17918946]
9. Tsai CJ, Del Sol A, Nussinov R. Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. Mol BioSyst. 2009; 5:207–216. [PubMed: 19225609]
10. 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]
11. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature. 2007; 450:913–916. [PubMed: 18026087]
12. Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson MA, Petsko GA, Karplus M, Hubner CG, Kern D. Intrinsic motions along an enzymatic reaction trajectory. Nature. 2007; 450:838–844. [PubMed: 18026086]
13. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, Kern D. Intrinsic dynamics of an enzyme underlies catalysis. Nature. 2005; 438:117–121. [PubMed: 16267559]
14. Tuttle LM, Dyson HJ, Wright PE. Side-chain conformational heterogeneity of intermediates in the Escherichia coli dihydrofolate reductase catalytic cycle. Biochemistry. 2013; 52:3464–3477. [PubMed: 23614825]
15. Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, Benkovic SJ, Wright PE. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. Science. 2011; 332:234–238. [PubMed: 21474759]
16. Clore GM. Interplay between conformational selection and induced fit in multidomain protein-ligand binding probed by paramagnetic relaxation enhancement. Biophys Chem. 2014; 186:3–12. [PubMed: 24070540]
17. Venditti V, Clore GM. Conformational selection and substrate binding regulate the monomer/dimer equilibrium of the C-terminal domain of Escherichia coli Enzyme I. J Biol Chem. 2012; 287:26989–26998. [PubMed: 22722931]

18. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. Nature. 2014; 508:331–339. [PubMed: 24740064]
19. Clore GM, Venditti V. Structure, dynamics and biophysics of the cytoplasmic protein-protein complexes of the bacterial phosphoenolpyruvate: sugar phosphotransferase system. Trends Biochem Sci. 2013; 38:515–530. [PubMed: 24055245]
20. Anthis NJ, Clore GM. Visualizing transient dark states by NMR spectroscopy. Q Rev Biophys. 2015; 48:35–116. [PubMed: 25710841]
21. Angyan AF, Gaspari Z. Ensemble-based interpretations of NMR structural data to describe protein internal dynamics. Molecules. 2013; 18:10548–10567. [PubMed: 23999727]
22. Burnley BT, Afonine PV, Adams PD, Gros P. Modelling dynamics in protein crystal structures by ensemble refinement. eLife. 2012; 1:e00311. [PubMed: 23251785]
23. Hillisch A, Lorenz M, Diekmann S. Recent advances in FRET: distance determination in protein-DNA complexes. Curr Opin Struct Biol. 2001; 11:201–207. [PubMed: 11297928]
24. Altenbach C, Kusnetzow AK, Ernst OP, Hofmann KP, Hubbell WL. High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. Proc Natl Acad Sci U S A. 2008; 105:7439–7444. [PubMed: 18490656]
25. Klug CS, Feix JB. Methods and applications of site-directed spin labeling EPR spectroscopy. Methods Cell Biol. 2008; 84:617–658. [PubMed: 17964945]
26. Cavalli A, Montalvao RW, Vendruscolo M. Using chemical shifts to determine structural changes in proteins upon complex formation. J Phys Chem B. 2011; 115:9491–9494. [PubMed: 21639128]
27. Bonvin AM, Boelens R, Kaptein R. Time- and ensemble-averaged direct NOE restraints. J Biomol NMR. 1994; 4:143–149. [PubMed: 22911161]
28. Kemmink J, van Mierlo CP, Scheek RM, Creighton TE. Local structure due to an aromatic-amide interaction observed by 1H-nuclear magnetic resonance spectroscopy in peptides related to the N terminus of bovine pancreatic trypsin inhibitor. J Mol Biol. 1993; 230:312–322. [PubMed: 7680725]
29. Vogeli B, Kazemi S, Guntert P, Riek R. Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs. Nat Struct Mol Biol. 2012; 19:1053–1057. [PubMed: 22940676]
30. Schwieters CD, Kuszewski JJ, Clore GM. Using Xplor-NIH for NMR molecular structure determination. Prog Nucl Magn Reson Spectrosc. 2006; 48:47–62.
31. Clore GM, Schwieters CD. Concordance of residual dipolar couplings, backbone order parameters and crystallographic B-factors for a small alpha/beta protein: a unified picture of high probability, fast atomic motions in proteins. J Mol Biol. 2006; 355:879–886. [PubMed: 16343537]
32. Best RB, Vendruscolo M. Determination of protein structures consistent with NMR order parameters. J Am Chem Soc. 2004; 126:8090–8091. [PubMed: 15225030]
33. Batta G, Barna T, Gaspari Z, Sandor S, Kover KE, Binder U, Sarg B, Kaiserer L, Chhillar AK, Eigentler A, Leiter E, Hegedus N, Pocsi I, Lindner H, Marx F. Functional aspects of the solution structure and dynamics of PAF–a highly-stable antifungal protein from Penicillium chrysogenum. FEBS J. 2009; 276:2875–2890. [PubMed: 19459942]
34. Lee JH, Li F, Grishaev A, Bax A. Quantitative residue-specific protein backbone torsion angle dynamics from concerted measurement of 3J couplings. J Am Chem Soc. 2015; 137:1432–1435. [PubMed: 25590347]
35. Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M. Simultaneous determination of protein structure and dynamics. Nature. 2005; 433:128–132. [PubMed: 15650731]
36. Clore GM, Schwieters CD. How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? J Am Chem Soc. 2004; 126:2923–2938. [PubMed: 14995210]
37. Hess B, Scheek RM. Orientation restraints in molecular dynamics simulations using time and ensemble averaging. J Magn Reson. 2003; 164:19–27. [PubMed: 12932451]
38. Louhivuori M, Otten R, Lindorff-Larsen K, Annila A. Conformational fluctuations affect protein alignment in dilute liquid crystal media. J Am Chem Soc. 2006; 128:4371–4376. [PubMed: 16569014]

39. Montalvao RW, De Simone A, Vendruscolo M. Determination of structural fluctuations of proteins from structure-based calculations of residual dipolar couplings. J Biomol NMR. 2012; 53:281–292. [PubMed: 22729708]

40. De Simone A, Montalvao RW, Vendruscolo M. Determination of conformational equilibria in proteins using residual dipolar couplings. J Chem Theory Comput. 2011; 7:4189–4195. [PubMed: 22180735]

41. Clore GM, Iwahara J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. Chem Rev. 2009; 109:4108–4139. [PubMed: 19522502]

42. Iwahara J, Clore GM. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. Nature. 2006; 440:1227–1230. [PubMed: 16642002]

43. Iwahara J, Schwieters CD, Clore GM. Ensemble approach for NMR structure refinement against 1H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. J Am Chem Soc. 2004; 126:5879–5896. [PubMed: 15125681]

44. Best RB, Vendruscolo M. Structural interpretation of hydrogen exchange protection factors in proteins: characterization of the native state fluctuations of CI2. Structure. 2006; 14:97–106. [PubMed: 16407069]

45. Clore GM. Visualizing lowly-populated regions of the free energy landscape of macromolecular complexes by paramagnetic relaxation enhancement. Mol BioSyst. 2008; 4:1058–1069. [PubMed: 18931781]

46. Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S, Lange A. Atomic model of the type III secretion system needle. Nature. 2012; 486:276–279. [PubMed: 22699623]

47. Schwieters CD, Suh JY, Grishaev A, Ghirlando R, Takayama Y, Clore GM. Solution structure of the 128 kDa enzyme I dimer from Escherichia coli and its 146 kDa complex with HPr using residual dipolar couplings and small- and wide-angle X-ray scattering. J Am Chem Soc. 2010; 132:13026–13045. [PubMed: 20731394]

48. Takayama Y, Schwieters CD, Grishaev A, Ghirlando R, Clore GM. Combined use of residual dipolar couplings and solution X-ray scattering to rapidly probe rigid-body conformational transitions in a non-phosphorylatable active-site mutant of the 128 kDa enzyme I dimer. J Am Chem Soc. 2011; 133:424–427. [PubMed: 21162528]

49. Shinsky SA, Cosgrove MS. Unique role of the WD-40 repeat protein 5 (WDR5) subunit within the mixed lineage leukemia 3 (MLL3) histone methyltransferase complex. J Biol Chem. 2015; 290:25819–25833. [PubMed: 26324722]

50. Heinrich F, Chakravarthy S, Nanda H, Papa A, Pandolfi PP, Ross AH, Harishchandra RK, Gericke A, Losche M. The PTEN tumor suppressor forms homodimers in solution. Structure. 2015; 23:1952–1957. [PubMed: 26299948]

51. Xia Z, Bell DR, Shi Y, Ren P. RNA 3D structure prediction by using a coarse-grained model and experimental data. J Phys Chem B. 2013; 117:3135–3144. [PubMed: 23438338]

52. Hura GL, Menon AL, Hammel M, Rambo RP, Poole FL 2nd, Tsutakawa SE, Jenney FE Jr, Classen S, Frankel KA, Hopkins RC, Yang SJ, Scott JW, Dillard BD, Adams MW, Tainer JA. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat Methods. 2009; 6:606–612. [PubMed: 19620974]

53. Tsutakawa SE, Hura GL, Frankel KA, Cooper PK, Tainer JA. Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. J Struct Biol. 2007; 158:214–223. [PubMed: 17182256]

54. Horejs C, Pum D, Sleytr UB, Tscheliessnig R. SAXS: Structure Verification of an S-Layer Protein using a Fractal Mean Force Potential. Biophys J. 2010; 98:760a.

55. Gutberlet T, Heinemann U, Steiner M. Protein crystallography with neutrons–status and perspectives. Acta Crystallogr, Sect D: Biol Crystallogr. 2001; 57:349–354. [PubMed: 11173501]

56. Breyton C, Gabel F, Lethier M, Flayhan A, Durand G, Jault JM, Juillan-Binard C, Imbert L, Moulin M, Ravaud S, Hartlein M, Ebel C. Small angle neutron scattering for the study of solubilised membrane proteins. Eur Phys J E: Soft Matter Biol Phys. 2013; 36:71.

57. Rambo RP, Tainer JA. Improving small-angle X-ray scattering data for structural analyses of the RNA world. RNA. 2010; 16:638–646. [PubMed: 20106957]

58. Grishaev A, Anthis NJ, Clore GM. Contrast-matched small-angle X-ray scattering from a heavy-atom-labeled protein in structure determination: application to a lead-substituted calmodulin-peptide complex. J Am Chem Soc. 2012; 134:14686–14689. [PubMed: 22908850]

59. Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys. 2007; 40:191–285. [PubMed: 18078545]

60. Grishaev A. Sample preparation, data collection, and preliminary data analysis in biomolecular solution X-ray scattering. Curr Protoc Protein Sci. 2012 Chapter 17, Unit 17.14.10.1002/0471140864.ps1714s70.

61. Garman EF, Weik M. Radiation damage to macromolecules: kill or cure? J Synchrotron Radiat. 2015; 22:195–200. [PubMed: 25723921]

62. Lipfert J, Millett IS, Seifert S, Doniach S. Sample holder for small-angle x-ray scattering static and flow cell measurement. Rev Sci Instrum. 2006; 77:046108.

63. Koch MHJ. Instruments and methods for small-angle scattering with synchrotron radiation. Makromol Chem, Macromol Symp. 1988; 15:19–90.

64. Pontoni D, Narayanan T, Rennie A. High-dynamic range SAXS data acquisition with an X-ray image intensifier. J Appl Crystallogr. 2002; 35:207–211.

65. Otendal M, Tuohimaa T, Vogt U, Hertz HM. A 9 keV electron-impact liquid-gallium-jet x-ray source. Rev Sci Instrum. 2008; 79:016102. [PubMed: 18248074]

66. Koch MHJ. SAXS Instrumentation for Synchrotron Radiation then and now. J Phys Conf Ser. 2010; 247:1–23.

67. Hamley IW, Pedersen JS. Analysis of Neutron and X-ray Reflectivity Data. I. Theory. J Appl Crystallogr. 1994; 27:29–35.

68. Chu B, Harney PJ, Li Y, Linliu K, Yeh F. A laser-aided prealigned pinhole collimator for synchrotron x rays. Rev Sci Instrum. 1994; 65:597–602.

69. Block ID, Scheffold F. Modulated 3D cross-correlation light scattering: improving turbid sample characterization. Rev Sci Instrum. 2010; 81:123107. [PubMed: 21198014]

70. 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]

71. Calmettes P, Durand D, Desmadril M, Minard P, Receveur V, Smith JC. How random is a highly denatured protein? Biophys Chem. 1994; 53:105–113. [PubMed: 17020841]

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

73. Groenewegen P, Feil D. Molecular Form Factors in X-ray Crystallography. Acta Crystallogr, Sect A: Cryst Phys, Diffr, Theor Gen Crystallogr. 1969; 25:444–450.

74. Hilfiker R, Eicke H, Sager W, Steeb C, Hofmeier U, Gehrke R. Form and structure factors of water/AOT/Oil micro-emulsions from synchrotron SAXS. Ber Bunsenges Phys Chem. 1990; 94:677–683.

75. Huang HK, Sheu HS, Chuang WT, Jeng US, Su AC, Wu WR, Liao KF, Chen CY, Chang SY, Lai HM. Correlated changes in structure and viscosity during gelatinization and gelation of tapioca starch granules. IUCrJ. 2014; 1:418–428.

76. Zhang F, Skoda MW, Jacobs RM, Martin RA, Martin CM, Schreiber F. Protein interactions studied by SAXS: effect of ionic strength and protein concentration for BSA in aqueous solutions. J Phys Chem B. 2007; 111:251–259. [PubMed: 17201449]

77. Grant TD, Luft JR, Carter LG, Matsui T, Weiss TM, Martel A, Snell EH. The accurate assessment of small-angle X-ray scattering data. Acta Crystallogr, Sect D: Biol Crystallogr. 2015; 71:45–56. [PubMed: 25615859]

78. Patterson AL. A Fourier series method for the determination of the components of interatomic distances in crystals. Phys Rev. 1934; 46:372–376.

79. Debye P, Bueche AM. Scattering by an Inhomogeneous Solid. J Appl Phys. 1949; 20:518–525.

80. Glatter O. A new method for the evaluation of small-angle scattering data. J Appl Crystallogr. 1977; 10:415–421.
81. Sousa MC, Trame CB, Tsuruta H, Wilbanks SM, Reddy VS, McKay DB. Crystal and solution structures of an HslUV protease-chaperone complex. Cell. 2000; 103:633–643. [PubMed: 11106733]
82. Mendillo ML, Putnam CD, Kolodner RD. Escherichia coli MutS tetramerization domain structure reveals that stable dimers but not tetramers are essential for DNA mismatch repair in vivo. J Biol Chem. 2007; 282:16345–16354. [PubMed: 17426027]
83. Vestergaard B, Sanyal S, Roessle M, Mora L, Buckingham RH, Kastrup JS, Gajhede M, Svergun DI, Ehrenberg M. The SAXS solution structure of RF1 differs from its crystal structure and is similar to its ribosome bound cryo-EM structure. Mol Cell. 2005; 20:929–938. [PubMed: 16364917]
84. Petoukhov MV, Eady NA, Brown KA, Svergun DI. Addition of missing loops and domains to protein models by x-ray solution scattering. Biophys J. 2002; 83:3113–3125. [PubMed: 12496082]
85. Schwieters CD, Clore GM. A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. Biochemistry. 2007; 46:1152–1166. [PubMed: 17260945]
86. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J. 1999; 76:2879–2886. [PubMed: 10354416]
87. Heller WT. ELLSTAT: shape modeling forsolution small-angle scattering of proteins and protein complexes with automated statistical characterization. J Appl Crystallogr. 2006; 39:671–675.
88. Walther D, Cohen FE, Doniach S. Reconstruction of low-resolution three-dimensionaldensity maps from one-dimensional small-angle X-ray solution scattering data for biomolecules. J Appl Crystallogr. 2000; 33:350–363.
89. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. Proc Natl Acad Sci U S A. 1995; 92:9279–9283. [PubMed: 7568117]
90. Eletsky A, Pulavarti SV, Beaumont V, Gollnick P, Szyperski T. Solution NMR experiment for measurement of 15N-1H residual dipolar couplings in large proteins and supramolecular complexes. J Am Chem Soc. 2015; 137:11242–11245. [PubMed: 26293598]
91. Smith SA, Palke WE, Gerig JT. The Hamiltonians of NMR Part II. Concepts Magn Reson. 1992; 4:181–204.
92. Bax A, Grishaev A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. Curr Opin Struct Biol. 2005; 15:563–570. [PubMed: 16140525]
93. Prestegard JH, Bougault CM, Kishore AI. Residual dipolar couplings in structure determination of biomolecules. Chem Rev. 2004; 104:3519–3540. [PubMed: 15303825]
94. Lipsitz RS, Tjandra N. Residual dipolar couplings in NMR structure analysis. Annu Rev Biophys Biomol Struct. 2004; 33:387–413. [PubMed: 15139819]
95. Tjandra N, Bax A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science. 1997; 278:1111–1114. [PubMed: 9353189]
96. Clore GM, Starich MR, Gronenborn AM. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. J Am Chem Soc. 1998; 120:10571–10572.
97. Losonczi JA, Prestegard JH. Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules. J Biomol NMR. 1998; 12:447–451. [PubMed: 9835051]
98. Hansen MR, Hanson P, Pardi A. Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. Methods Enzymol. 2000; 317:220–240. [PubMed: 10829283]
99. Bertini I, Felli IC, Luchinat C. Lanthanide induced residual dipolar couplings for the conformational investigation of peripheral $^{15}NH_2$ moieties. J Biomol NMR. 2000; 18:347–355. [PubMed: 11200529]
100. Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S. Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or

embedding of oriented purple membranes. J Biomol NMR. 2000; 18:303–309. [PubMed: 11200524]

101. Prestegard JH, Kishore AI. Partial alignment of biomolecules: an aid to NMR characterization. Curr Opin Chem Biol. 2001; 5:584–590. [PubMed: 11578934]

102. Bax A, Kontaxis G, Tjandra N. Dipolar couplings in macromolecular structure determination. Methods Enzymol. 2001; 339:127–174. [PubMed: 11462810]

103. Maltsev AS, Grishaev A, Roche J, Zasloff M, Bax A. Improved cross validation of a static ubiquitin structure derived from high precision residual dipolar couplings measured in a drug-based liquid crystalline phase. J Am Chem Soc. 2014; 136:3752–3755. [PubMed: 24568736]

104. Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Biol. 1998; 5:1065–1074. [PubMed: 9846877]

105. Chou JJ, Gaemers S, Howder B, Louis JM, Bax A. A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. J Biomol NMR. 2001; 21:377–382. [PubMed: 11824758]

106. Tycko R, Blanco FJ, Y I. Alignment of biopolymers in strained gels: a new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. J Am Chem Soc. 2000; 122:9340–9341.

107. Ulmer TS, Ramirez BE, Delaglio F, Bax A. Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. J Am Chem Soc. 2003; 125:9179–9191. [PubMed: 15369375]

108. Brunner E. Residual Dipolar Couplings in Protein NMR. Concepts Magn Reson. 2001; 13:238–259.

109. Ottiger M, Delaglio F, Bax A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. J Magn Reson. 1998; 131:373–378. [PubMed: 9571116]

110. Tjandra N, Marquardt J, Clore GM. Direct refinement against proton-proton dipolar couplings in NMR structure determination of macromolecules. J Magn Reson. 2000; 142:393–396. [PubMed: 10648162]

111. Kontaxis G, Bax A. Multiplet component separation for measurement of methyl 13C-1H dipolar couplings in weakly aligned proteins. J Biomol NMR. 2001; 20:77–82. [PubMed: 11430758]

112. Meier S, Haussinger D, Jensen P, Rogowski M, Grzesiek S. High-accuracy residual ^1H_N-^{13}C and ^1H_N-^1H_N dipolar couplings in perdeuterated proteins. J Am Chem Soc. 2003; 125:44–45. [PubMed: 12515503]

113. Permi P. Measurement of residual dipolar couplings from $^1H\alpha$ to $^{13}C\alpha$ and ^{15}N using a simple HNCA-based experiment. J Biomol NMR. 2003; 27:341–349. [PubMed: 14512731]

114. Vogeli B, Yao L, Bax A. Protein backbone motions viewed by intraresidue and sequential HN-Hα residual dipolar couplings. J Biomol NMR. 2008; 41:17–28. [PubMed: 18458825]

115. Tjandra N, Grzesiek S, Bax A. Magnetic Field Dependence of Nitrogen-Proton J Splittings in 15N-Enriched Human Ubiquitin Resulting from Relaxation Interference and Residual Dipolar Coupling. J Am Chem Soc. 1996; 118:6264–6272.

116. Tjandra N, Bax A. Large variations in $^{13}C\alpha$ chemical shift anisotropy in proteins correlate with secondary structure. J Am Chem Soc. 1997; 119:9576–9577.

117. Bax A, Vuister GW, Grzesiek S, Delaglio F, Wang AC, Tschudin R, Zhu G. Measurement of homo- and heteronuclear J couplings from quantitative J correlation. Methods Enzymol. 1994; 239:79–105. [PubMed: 7830604]

118. Liu Y, Prestegard JH. Measurement of one and two bond N-C couplings in large proteins by TROSY-based J-modulation experiments. J Magn Reson. 2009; 200:109–118. [PubMed: 19581113]

119. Fitzkee NC, Bax A. Facile measurement of ^1H-^{15}N residual dipolar couplings in larger perdeuterated proteins. J Biomol NMR. 2010; 48:65–70. [PubMed: 20694505]

120. Rosen MK, Gardner KH, Willis RC, Parris WE, Pawson T, Kay LE. Selective methyl group protonation of perdeuterated proteins. J Mol Biol. 1996; 263:627–636. [PubMed: 8947563]

121. Gardner KH, Kay LE. The use of ^2H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. Annu Rev Biophys Biomol Struct. 1998; 27:357–406. [PubMed: 9646872]

122. Godoy-Ruiz R, Guo C, Tugarinov V. Alanine methyl groups as NMR probes of molecular structure and dynamics in high-molecular-weight proteins. J Am Chem Soc. 2010; 132:18340–18350. [PubMed: 21138300]

123. Clark L, Zahm JA, Ali R, Kukula M, Bian L, Patrie SM, Gardner KH, Rosen MK, Rosenbaum DM. Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote Pichia pastoris. J Biomol NMR. 2015; 62:239–245. [PubMed: 26025061]

124. Ollerenshaw JE, Tugarinov V, Kay LE. Methyl TROSY: explaination and experimental verification. Magn Reson Chem. 2003; 41:843–852.

125. Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. Cross-correlated relaxation enhanced ^{1}H-^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. J Am Chem Soc. 2003; 125:10420–10428. [PubMed: 12926967]

126. Guo C, Godoy-Ruiz R, Tugarinov V. High resolution measurement of methyl $^{13}C_m$-^{13}C and ^1H_m-$^{13}C_m$ residual dipolar couplings in large proteins. J Am Chem Soc. 2010; 132:13984–13987. [PubMed: 20860373]

127. Clore GM, Gronenborn AM, Bax A. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. J Magn Reson. 1998; 133:216–221. [PubMed: 9654491]

128. Warren JJ, Moore PB. Application of dipolar coupling data to the refinement of the solution structure of the sarcin-ricin loop RNA. J Biomol NMR. 2001; 20:311–323. [PubMed: 11563555]

129. Clore GM, Garrett DS. R-factor, free R, and complete cross-validation for dipolar coupling refinement of NMR structures. J Am Chem Soc. 1999; 121:9008–9012.

130. Zweckstetter M, Bax A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. J Am Chem Soc. 2000; 122:3791–3792.

131. Huang JR, Grzesiek S. Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin. J Am Chem Soc. 2010; 132:694–705. [PubMed: 20000836]

132. Zweckstetter M. NMR: prediction of molecular alignment from structure using the PALES software. Nat Protoc. 2008; 3:679–690. [PubMed: 18388951]

133. Zweckstetter M, Hummer G, Bax A. Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. Biophys J. 2004; 86:3444–3460. [PubMed: 15189846]

134. Camilloni C, Vendruscolo M. A tensor-free method for the structural and dynamical refinement of proteins using residual dipolar couplings. J Phys Chem B. 2015; 119:653–661. [PubMed: 24824082]

135. Camilloni C, Vendruscolo M. Reply to “Comment on ’A tensor-free method for the structural and dynamic refinement of proteins using residual dipolar couplings’”. J Phys Chem B. 2015; 119:8225–8226.

136. Clore GM, Gronenborn AM. Determining the structures of large proteins and protein complexes by NMR. Trends Biotechnol. 1998; 16:22–34. [PubMed: 9470228]

137. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. Use of dipolar ^{1}H-^{15}N and ^{1}H-^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat Struct Biol. 1997; 4:732–738. [PubMed: 9303001]

138. Clore GM, Gronenborn AM. New methods of structure refinement for macromolecular structure determination by NMR. Proc Natl Acad Sci U S A. 1998; 95:5891–5898. [PubMed: 9600889]

139. Clore GM. Accurate and rapid docking of protein-protein complexes on the basis of intermolecular nuclear overhauser enhancement data and dipolar couplings by rigid body minimization. Proc Natl Acad Sci U S A. 2000; 97:9021–9025. [PubMed: 10922057]

140. Ramirez BE, Bax A. Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium. J Am Chem Soc. 1998; 120:9106–9107.

141. Al-Hashimi HM, Valafar H, Terrell M, Zartler ER, Eidsness MK, Prestegard JH. Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. J Magn Reson. 2000; 143:402–406. [PubMed: 10729267]

142. Venditti V, Schwieters CD, Grishaev A, Clore GM. Dynamic equilibrium between closed and partially closed states of the bacterial Enzyme I unveiled by solution NMR and X-ray scattering. Proc Natl Acad Sci U S A. 2015; 112:11565–11570. [PubMed: 26305976]

143. Wang YX, Zuo X, Wang J, Yu P, Butcher SE. Rapid global structure determination of large RNA and RNA complexes using NMR and small-angle X-ray scattering. Methods. 2010; 52:180–191. [PubMed: 20554045]

144. Burke JE, Butcher SE. Nucleic acid structure characterization by small angle X-ray scattering (SAXS). Curr Protoc Nucleic Acid Chem. 2012 Chapter 7, Unit 7.18.10.1002/0471142700.nc0718s51.

145. Burke JE, Sashital DG, Zuo X, Wang YX, Butcher SE. Structure of the yeast U2/U6 snRNA complex. RNA. 2012; 18:673–683. [PubMed: 22328579]

146. Rossi P, Shi L, Liu G, Barbieri CM, Lee HW, Grant TD, Luft JR, Xiao R, Acton TB, Snell EH, Montelione GT, Baker D, Lange OF, Sgourakis NG. A hybrid NMR/SAXS-based approach for discriminating oligomeric protein interfaces using Rosetta. Proteins: Struct, Funct Genet. 2015; 83:309–317. [PubMed: 25388768]

147. Deshmukh L, Schwieters CD, Grishaev A, Ghirlando R, Baber JL, Clore GM. Structure and dynamics of full-length HIV-1 capsid protein in solution. J Am Chem Soc. 2013; 135:16133–16147. [PubMed: 24066695]

148. Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM. Deviations from the simple two parameter model free approach to the interpretation of 15N nuclear magnetic relaxation of proteins. J Am Chem Soc. 1990; 112:4989–4991.

149. Baber JL, Szabo A, Tjandra N. Analysis of slow interdomain motion of macromolecules using NMR relaxation data. J Am Chem Soc. 2001; 123:3953–3959. [PubMed: 11457145]

150. Braddock DT, Cai M, Baber JL, Huang Y, Clore GM. Rapid identification of medium- to large-scale interdomain motion in modular proteins using dipolar couplings. J Am Chem Soc. 2001; 123:8634–8635. [PubMed: 11525687]

151. Braddock DT, Louis JM, Baber JL, Levens D, Clore GM. Structure and dynamics of KH domains from FBP bound to single-stranded DNA. Nature. 2002; 415:1051–1056. [PubMed: 11875576]

152. Schwieters CD, Clore GM. Using small angle solution scattering data in Xplor-NIH structure calculations. Prog Nucl Magn Reson Spectrosc. 2014; 80:1–11. [PubMed: 24924264]

153. Chauvin F, Brand L, Roseman S. Enzyme I: the first protein and potential regulator of the bacterial phosphoenolpyruvate: glycose phosphotransferase system. Res Microbiol. 1996; 147:471–479. [PubMed: 9084757]

154. Meadow ND, Fox DK, Roseman S. The bacterial phosphoenolpyruvate: glycose phosphotransferase system. Annu Rev Biochem. 1990; 59:497–542. [PubMed: 2197982]

155. Venditti V, Ghirlando R, Clore GM. Structural basis for Enzyme I inhibition by alpha-ketoglutarate. ACS Chem Biol. 2013; 8:1232–1240. [PubMed: 23506042]

156. LiCalsi C, Crocenzi TS, Freire E, Roseman S. Sugar transport by the bacterial phosphotransferase system. Structural and thermodynamic domains of enzyme I of Salmonella typhimurium. J Biol Chem. 1991; 266:19519–19527. [PubMed: 1655788]

157. Garrett DS, Seok YJ, Liao DI, Peterkofsky A, Gronenborn AM, Clore GM. Solution structure of the 30 kDa N-terminal domain of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR. Biochemistry. 1997; 36:2517–2530. [PubMed: 9054557]

158. Garrett DS, Seok YJ, Peterkofsky A, Clore GM, Gronenborn AM. Identification by NMR of the binding surface for the histidine-containing phosphocarrier protein HPr on the N-terminal domain of enzyme I of the Escherichia coli phosphotransferase system. Biochemistry. 1997; 36:4393–4398. [PubMed: 9109646]

159. Seok YJ, Lee BR, Zhu PP, Peterkofsky A. Importance of the carboxyl-terminal domain of enzyme I of the Escherichia coli phosphoenolpyruvate: sugar phosphotransferase system for phosphoryl donor specificity. Proc Natl Acad Sci U S A. 1996; 93:347–351. [PubMed: 8552636]

160. Oberholzer AE, Schneider P, Siebold C, Baumann U, Erni B. Crystal structure of enzyme I of the phosphoenolpyruvate sugar phosphotransferase system in the dephosphorylated state. J Biol Chem. 2009; 284:33169–33176. [PubMed: 19801641]

161. Marquez J, Reinelt S, Koch B, Engelmann R, Hengstenberg W, Scheffzek K. Structure of the full-length enzyme I of the phosphoenolpyruvate-dependent sugar phosphotransferase system. J Biol Chem. 2006; 281:32508–32515. [PubMed: 16867985]
162. Teplyakov A, Lim K, Zhu PP, Kapadia G, Chen CC, Schwartz J, Howard A, Reddy PT, Peterkofsky A, Herzberg O. Structure of phosphorylated enzyme I, the phosphoenolpyruvate:sugar phosphotransferase system sugar translocation signal protein. Proc Natl Acad Sci U S A. 2006; 103:16218–16223. [PubMed: 17053069]
163. Venditti V, Tugarinov V, Schwieters CD, Grishaev A, Clore GM. Large interdomain rearrangement triggered by suppression of micro- to millisecond dynamics in bacterial Enzyme I. Nat Commun. 2015; 6:5960. [PubMed: 25581904]
164. Bewley CA, Clore GM. J Am Chem Soc. 2000; 122:6009–6016.
165. Clore GM, Bewley CA. Using conjoined rigid body/torsion angle simulated annealing to determine the relative orientation of covalently linked protein domains from dipolar couplings. J Magn Reson. 2002; 154:329–335. [PubMed: 11846592]
166. Pornillos O, Ganser-Pornillos BK, Yeager M. Atomic-level modelling of the HIV capsid. Nature. 2011; 469:424–427. [PubMed: 21248851]
167. Dahirel V, Shekhar K, Pereyra F, Miura T, Artyomov M, Talsania S, Allen TM, Altfeld M, Carrington M, Irvine DJ, Walker BD, Chakraborty AK. Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. Proc Natl Acad Sci U S A. 2011; 108:11530–11535. [PubMed: 21690407]
168. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, Ahn J, Gronenborn AM, Schulten K, Aiken C, Zhang P. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature. 2013; 497:643–646. [PubMed: 23719463]
169. Pornillos O, Ganser-Pornillos BK, Kelly BN, Hua Y, Whitby FG, Stout CD, Sundquist WI, Hill CP, Yeager M. X-ray structures of the hexameric building block of the HIV capsid. Cell. 2009; 137:1282–1292. [PubMed: 19523676]
170. van den Bedem H, Fraser JS. Integrative, dynamic structural biology at atomic resolution–it's about time. Nat Methods. 2015; 12:307–318. [PubMed: 25825836]
171. Ward AB, Sali A, Wilson AI. Integrative structural biology. Science. 2013; 339:913–915. [PubMed: 23430643]
172. Valentini E, Kikhney AG, Previtali G, Jeffries CM, Svergun DI. SASBDB, a repository for biological small-angle scattering data. Nucleic Acids Res. 2015; 43:D357–D363. [PubMed: 25352555]
173. Bewley CA, Gustafson KR, Boyd MR, Covell DG, Bax A, Clore GM, Gronenborn AM. Solution structure of cyanovirin-N, a potent HIV-inactivating protein. Nat Struct Biol. 1998; 5:571–578. [PubMed: 9665171]
174. Liao DI, Silverton E, Seok YJ, Lee BR, Peterkofsky A, Davies DR. The first step in sugar transport: crystal structure of the amino terminal domain of enzyme I of the E. coli PEP: sugar phosphotransferase system and a model of the phosphotransfer complex with HPr. Structure. 1996; 4:861–872. [PubMed: 8805571]
175. Schwieters CD, Clore GM. Reweighted atomic densities to represent ensembles of NMR structures. J Biomol NMR. 2002; 23:221–225. [PubMed: 12238594]
176. Du S, Betts L, Yang R, Shi H, Concel J, Ahn J, Aiken C, Zhang P, Yeh JI. Structure of the HIV-1 full-length capsid protein in a conformationally trapped unassembled state induced by small-molecule binding. J Mol Biol. 2011; 406:371–386. [PubMed: 21146540]

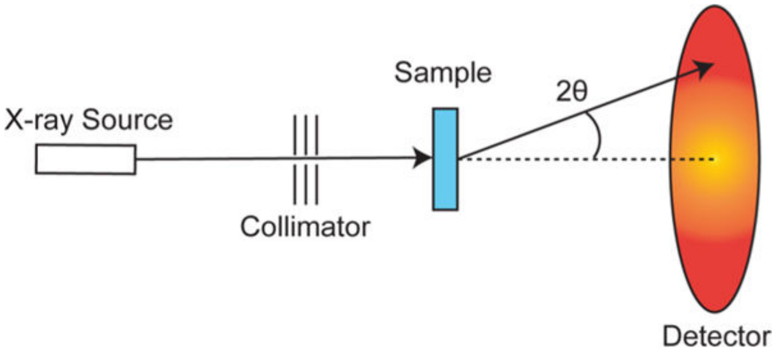


Figure 1.
Schematic representation of a SAXS instrument (see main text).

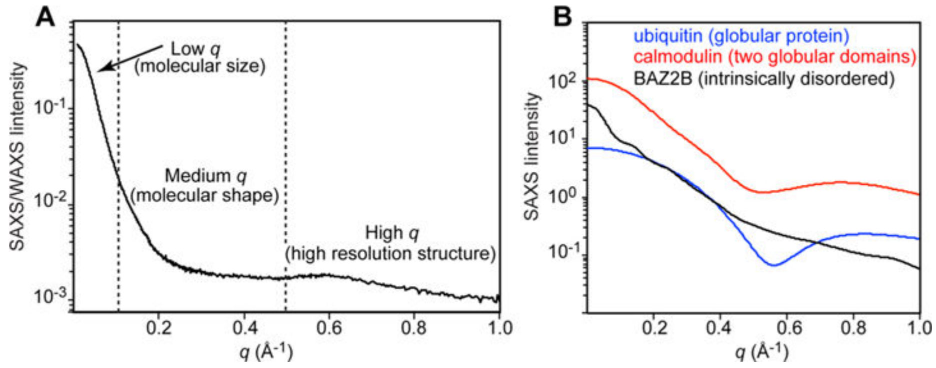


Figure 2.

(A) SAXS/WAXS curve acquired for Enzyme I of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. Vertical dashed lines separate the low, medium and high *q* regions. (B) SAXS curves acquired for ubiquitin (blue), calmodulin (red), and the bromodomain of protein 2B (BAZ2B – black). Data displayed in (B) were downloaded from the Small Angle Scattering Biological Data Bank (SASBDB).[172]

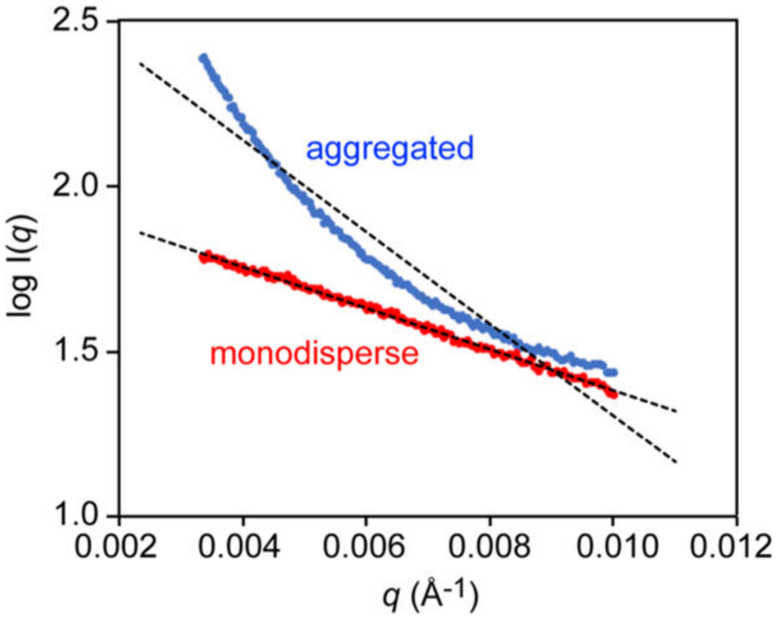


Figure 3.

Examples of Guinier plots for monodisperse (red) and aggregated (blue) samples.

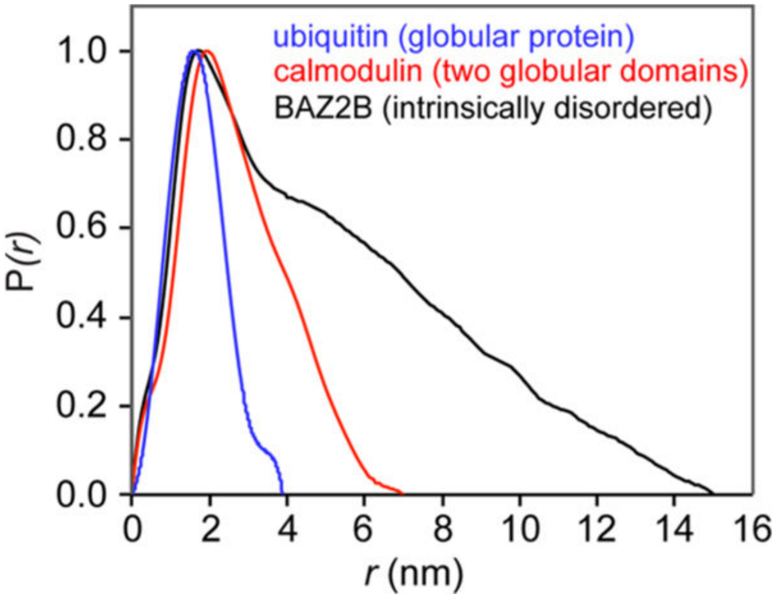


Figure 4.

Pair-distribution functions, $P(r)$, obtained from the SAXS curves acquired for ubiquitin (blue), calmodulin (red), and the bromodomain of protein 2B (BAZ2B – black). The experimental SAXS data were obtained from the SASBDB.[172]

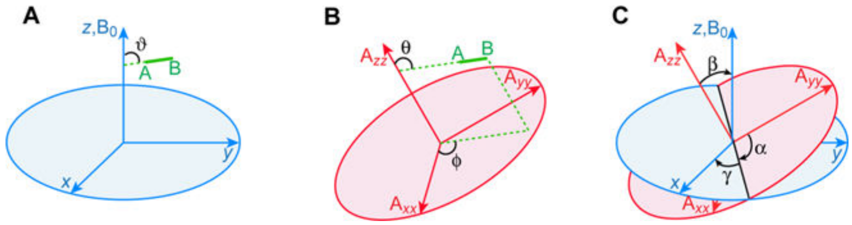


Figure 5.

Relationship between the bond vector A–B, the alignment tensor and the external magnetic field (B_0). (A) The angle ϑ describes the orientation of the bond vector relative to B_0; (B) the angles θ and ϕ define the orientation of the bond vector relative to the alignment tensor; and (C) the relationship between the alignment tensor and the external magnetic field is given by the Euler angles α, β and γ. The A-B vector is displayed green; the alignment tensor is colored red. The magnetic field B_0 is taken parallel to an external reference frame (blue).

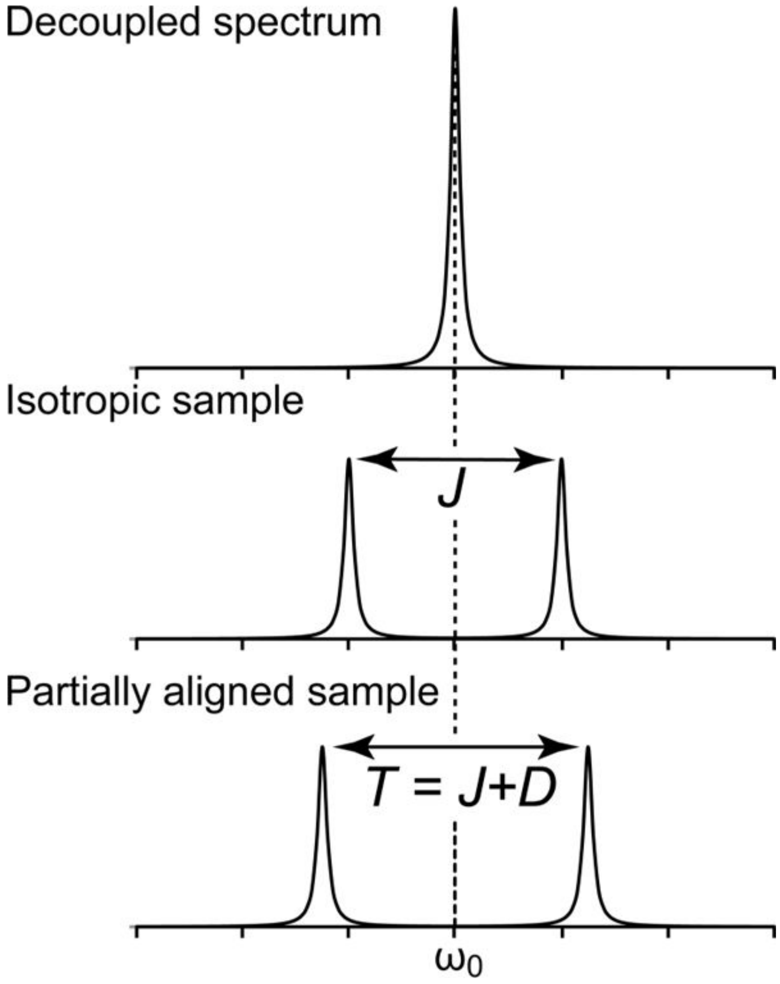


Figure 6.

Comparison between a NMR peak measured in the (A) absence of coupling (i.e., decoupled), (B) presence of scalar (*J*) coupling (isotropic sample), and (C) presence of scalar (*J*) and dipolar (*D*) coupling (partially aligned sample).

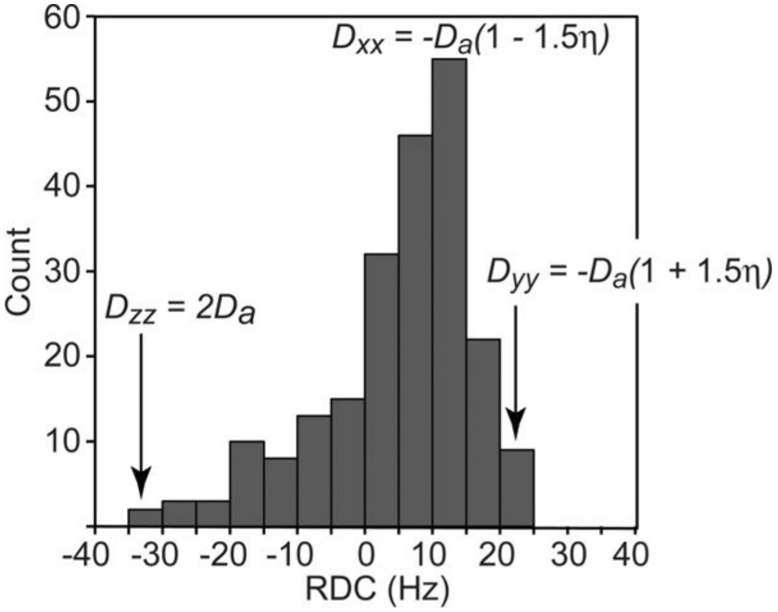


Figure 7.

Deriving the magnitude of the axial component and rhombicity of the alignment tensor from a histogram of normalized RDCs. The data are taken for the protein cyanovirin,[173] and the backbone Cα-Hα and Cα-C$'$ RDCs are normalized relative to the backbone N–H RDCs. The maximum, minimum and mode of the distribution correspond to $D_{zz} = 2D_a$, $D_{yy} = -$-$D_a(1 + 1.5\eta)$, and $D_{xx} = -D_a(1 - 1.5\eta)$, where D_a (scaled for N–H bond vectors) is the magnitude of the axial component of the alignment tensor and η the rhombicity. The sum of the three orthogonal components of the alignment tensor is equal to zero ($D_{zz} + D_{yy} + D_{xx} = 0$).

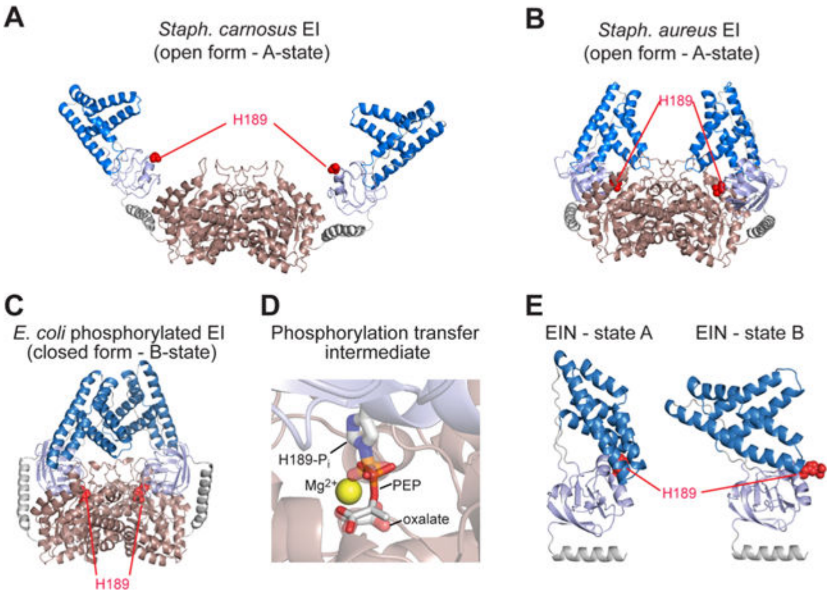


Figure 8.

X-ray structures of bacterial Enzyme I (EI). (A) EI from *Staph. carnosus* (2HRO);[161] (B) EI from *Staph. aureus* (2WQD);[160] (C) phosphorylated EI from *E. coli* (2HWG).[162] The C-terminal dimerization domain (EIC) is colored pink; the EIN^{α} and $\text{EIN}^{\alpha/\beta}$ subdomains of the N-terminal domain (EIN) are colored blue and light blue, respectively; the active site His189, located in the $\text{EIN}^{\alpha/\beta}$ subdomain, is shown as red spheres. (D) Structural model of phosphoenolpyruvate (PEP) bound to the active site of the *E. coli* EI structure. PEP is shown as solid sticks, and the Mg^{2+} ion is displayed as a yellow sphere. The phosphorylated His189 and the oxalate molecule in the X-ray structure of the phosphoryl transfer intermediate in the closed state (2HWG) are displayed as transparent sticks. (E) A[157,174] and B[162] conformations of the EIN domain seen in the open and closed states of EI, respectively.

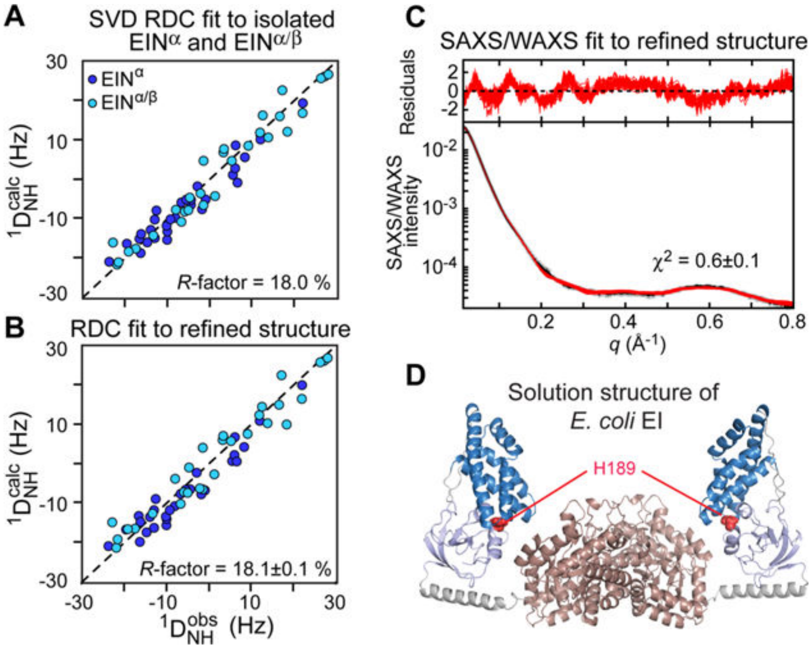


Figure 9.

Combined SAXS/RDC refinement of the *E. coli* EI structure. (A) Comparison of the observed and calculated RDCs obtained by SVD fits to the individual EIN$^{\alpha}$ (blue circles) and EIN$^{\alpha/\beta}$ (light blue circles) subdomains. (B) Comparison of the observed and calculated RDCs for the refined structure of the EI dimer. The resulting RDC *R*-factor is the same (within experimental error) as the RDC *R*-factor obtained from the SVD fits to the individual subdomains. (C) Agreement between the experimental and back calculated SAXS/WAXS curve. (D) Solution structure of unliganded *E. coli* EI obtained by combined SAXS/RDC refinement (2KX9)[47] Color coding as in Figure 8. Adapted from Schwieters et al. (2010).[47]

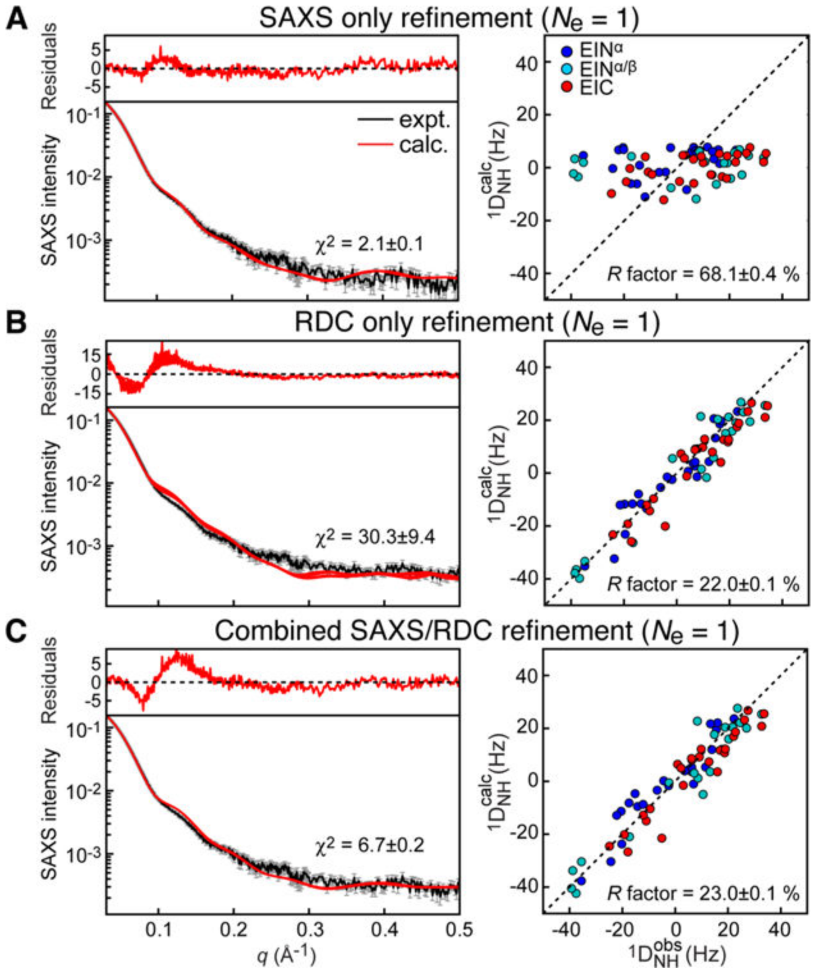


Figure 10.

Comparison between experimental and back-calculated SAXS (left) and RDC (right) data for a single-structure refinement of the EI^A-PEP complex. (**A)** Refinement using only SAXS data. (B) Refinement using only RDC data. (C) Combined SAXS/RDC refinement. Adapted from Venditti et al. (2015).[142]

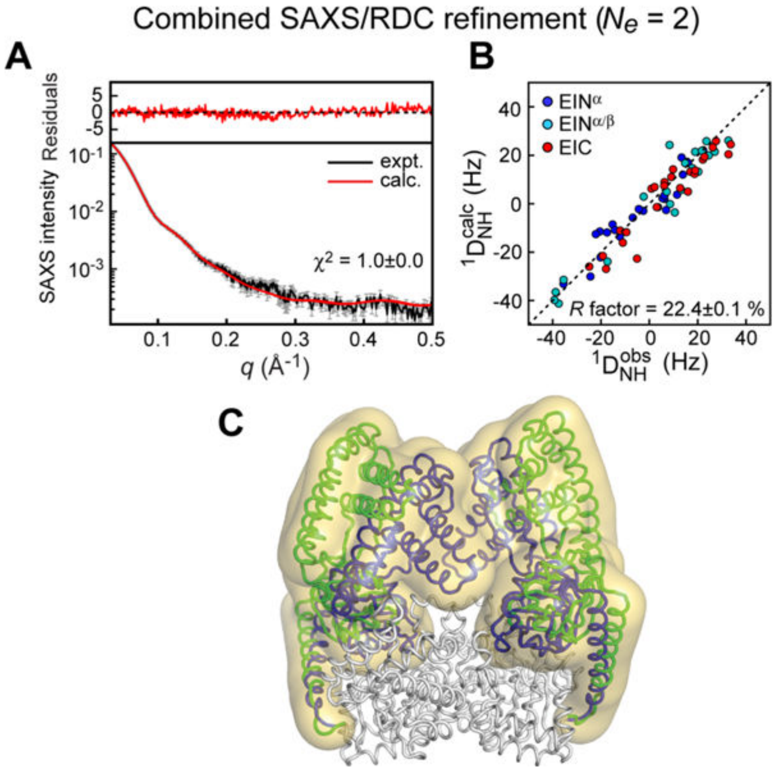


Figure 11.

Combined SAXS/RDC refinement of the EI^A-PEP complex using a two-member ensemble representation. (A) Agreement between experimental and back calculated SAXS curve. (B) Comparison of observed and calculated RDCs. (C) Structural ensemble obtained for EI^A-PEP complex. The overall distribution of EIN relative to EIC is shown as a reweighted atomic probability map[175] plotted at 2% of maximum (transparent yellow surface). Representative structures for the closed and partially closed EI conformations (PDB code 2N5T) are shown as blue and green ribbons, respectively. Adapted from Venditti et al. (2015).[142]

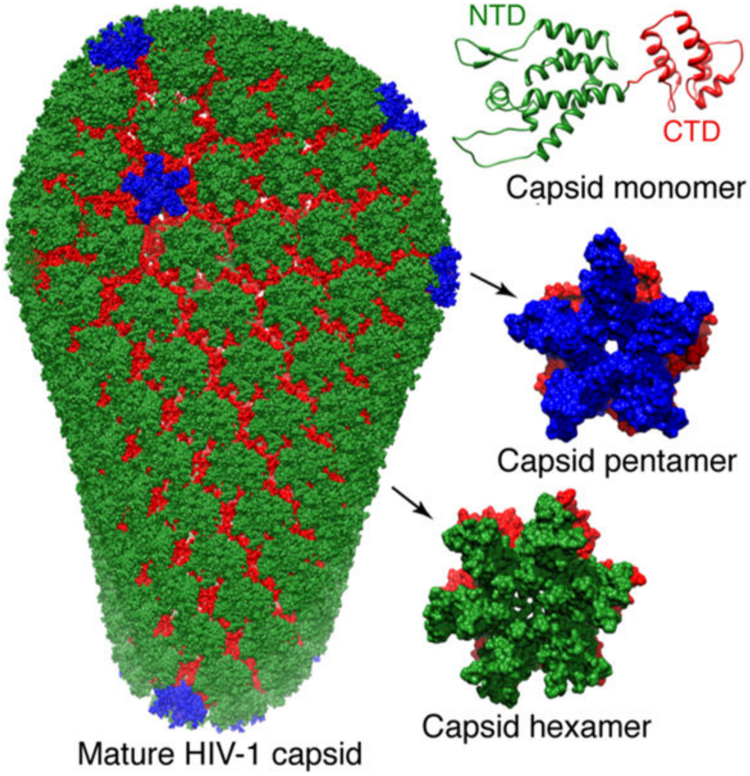


Figure 12.

HIV-1 capsid assembly. The capsid protein comprises N (green) and C (red) terminal domains (top right).[176] The N-terminal domains associate to form either pentamers[166] (middle right with N-terminal domains in blue) or hexamers[169] (bottom right with N-terminal domains in green) which assemble via the C-terminal domain dimers to form a cone comprising ~250 hexamers and exactly 12 pentamers (left).[166] Adapted from Deshmukh et al.[147]

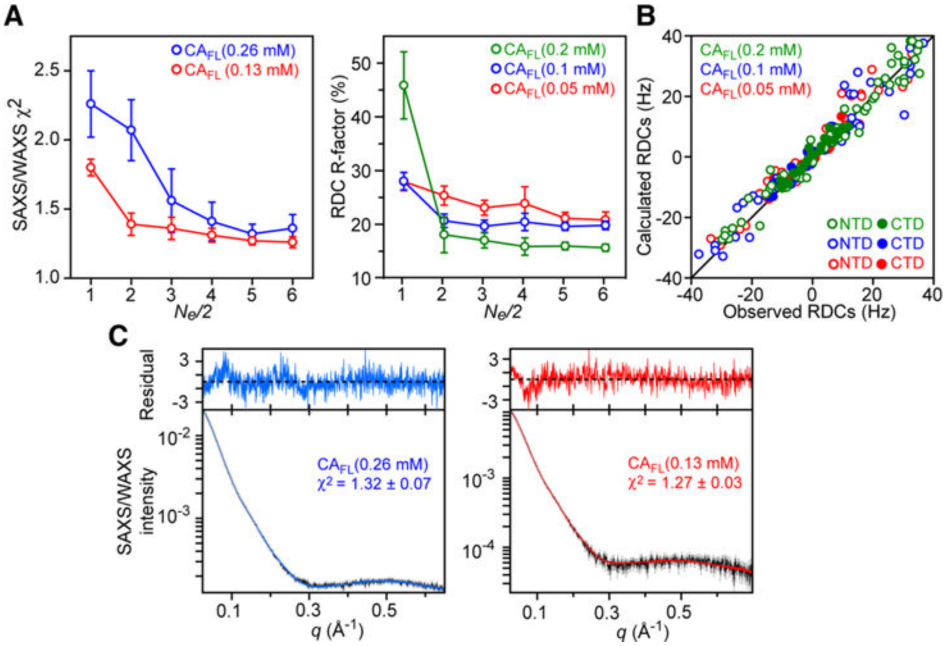


Figure 13.

RDC and SAXS/WAXS driven ensemble simulated annealing refinement of the HIV-1 capsid protein. Both monomer and dimer are included in the calculations and are represented by an equal number of ensemble members ($N_e/2$). RDC and SAXS/WAXS data at several different concentrations are treated simultaneously. (A) SAXS/WAXS χ^2 and RDC R-factor as a function of ensemble size. (B) Correlation between observed and back-calculated RDCs based on molecular shape at three concentrations of capsid protein. (C) Agreement between observed (black) and calculated SAXS/WAXS curves at two capsid protein concentrations. The residuals, given by $(I_i^{\text{calc}} - I_i^{\text{obs}})/I_i^{\text{err}}$, are plotted above the curves. Error bars: ± 1 standard deviation. Adapted from Deshmukh et al.[147]

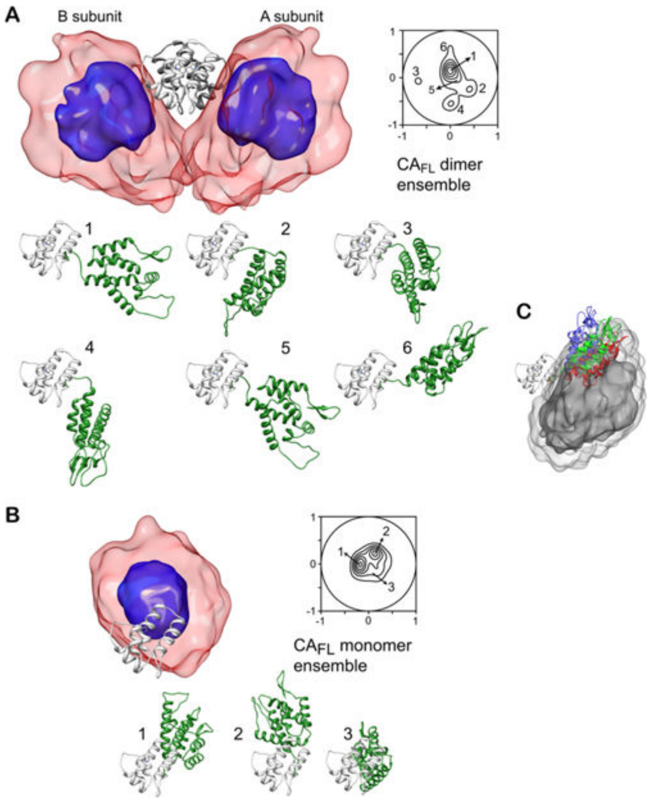


Figure 14.

Structural ensembles calculated for full-length wild type HIV-1 capsid protein. The dimer and monomer ensembles are shown in (A) and (B), respectively. The overall distribution of the N-terminal domain relative to the C-terminal domain (gray ribbon) is displayed as a reweighted atomic probability density map plotted at 50% (blue) and 10% (red transparent) of maximum. Projection contour maps showing the distribution of the position of the centroid of the N-terminal domain relative to the C-terminal domain are also shown. The dimer and monomer ensembles are characterized by six and three main clusters, respectively. (For clarity only a single subunit is shown for the six dimer clusters; the orientation of the C-terminal domain is the same throughout). (C) Position of the N-terminal domain in cluster 6 (red) of the capsid protein dimer ensemble compared to that in the pentamer (green) and hexamer (blue) with the C-terminal domain shown as a gray ribbon, and the atomic probability density map of the N-terminal domain in the capsid dimer plotted at 10% (dark gray) and 2% (light gray) of maximum. Adapted from Deshmukh et al.[147]