16

STRUCTURAL INSIGHTS INTO INTRINSICALLY DISORDERED PROTEINS BY SMALL-ANGLE X-RAY SCATTERING

Pau Bernadó[1] and Dmitri I. Svergun[2]

[1]*Institute for Research in Biomedicine, Parc Científic de Barcelona, Barcelona, Spain*
[2]*European Molecular Biology Laboratory, Hamburg Outstation, Hamburg, Germany*

ABSTRACT

Small-angle X-ray scattering (SAXS) is widely employed to study the overall solution structure and structural transitions of folded biologic macromolecules. At the same time, SAXS is one of the very few techniques yielding structural information about flexible macromolecules, unfolded and intrinsically unfolded proteins. In the past, SAXS was more often employed to qualitatively monitor the folding/unfolding processes by a few overall parameters; modern modeling methods allow one to extract much more detailed quantitative information for a better understanding of the biologic role of these proteins. In this chapter, after a brief introduction to the technical and experimental details of SAXS, its applications to the IDPs are described. Classical approaches based on the analysis of overall parameters are presented, and a recent development, the ensemble optimization method, is also explained in detail. The latter approach takes the conformational plasticity of IDPs into account by allowing for coexistence of multiple protein conformations in solution, and the analysis of the ensembles provides a new source of structural information.

Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation, Edited by Vladimir Uversky and Sonia Longhi

This approach and the combination of SAXS with other biophysical techniques, such as NMR, Föster Resonance Energy Transfer (FRET), and molecular simulations, promise structural insights into the detection of structural disorder and into monitoring of structural perturbations of IDPs upon environmental changes or binding to biologic partners.

16.1 INTRODUCTION

Diffraction methods are among the most powerful techniques to study macromolecular structure up to the atomic resolution. Large-scale structure determination initiatives using X-ray crystallography (41) yielded unprecedented numbers of high-resolution models for isolated proteins and/or their domains, and these numbers are expected to grow rapidly in the coming years. Much more difficulties have been encountered in the study of the proteins like intrinsically disordered proteins (IDPs). The inherent structural flexibility makes IDPs difficult (or impossible) to crystallize and also limits the amount of structural information, which can be obtained from solution nuclear magnetic resonance (NMR). The interest for this family of proteins has been enhanced by the discovery that they are fully active and play important biologic roles in several metabolic and signal transduction pathways even though they lack a permanent tertiary structure (14, 18, 19, 76, 87). In addition, some IDPs and nonnative states of proteins have been linked to the formation of toxic aggregates that cause important degenerative diseases (16).

The present chapter describes one of the most efficient methods for structural characterization of macromolecules in solution, a diffraction technique called small-angle X-ray scattering (SAXS). This method allows one to study the low-resolution structure of native particles in solutions and to analyze structural changes in response to variation of external conditions. Computational approaches are well established to retrieve low-resolution, three-dimensional structural models of proteins and complexes, and they are widely used in structural biology (59, 61, 72). Importantly, unlike most other structural methods, SAXS is applicable to equilibrium and nonequilibrium mixtures and monitor kinetic processes such as (dis)assembly (1) and (un)folding (17). In particular, the technique is effectively used to quantitatively characterize the overall structure of IDPs, and recent development of novel data analysis methods made it possible to also describe the flexibility of IDP ensembles in solution based on SAXS data. In this chapter, after a short reminder of the main theoretical and experimental aspects of X-ray scattering from macromolecular solutions, applications of SAXS to the analysis of IDPs will be presented.

16.2 BASICS OF X-RAY SCATTERING

This section will briefly describe the basic theoretical and experimental aspects of SAXS as applied to solutions of biologic macromolecules. The reader is

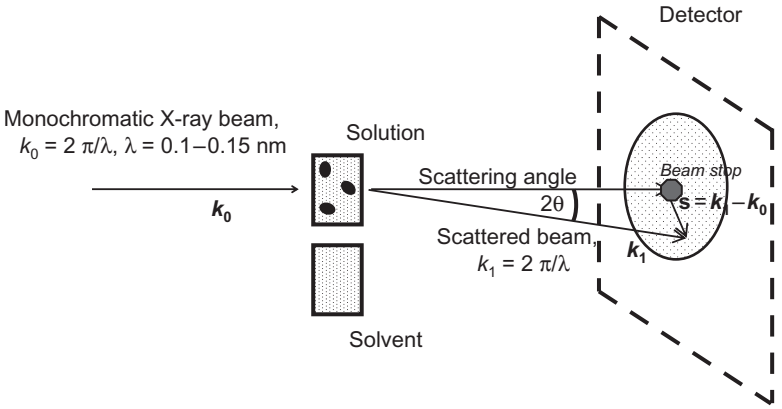

Figure 16.1 General scheme of a SAXS experiment.

referred to textbooks (20) or to recent reviews (36, 44, 73) for more detailed information.

In a SAXS experiment, a solution of macromolecules placed in a sample container (e.g., capillary or cuvette) is illuminated by a collimated monochromatic X-ray beam, and the isotropic intensity of the scattered beam is measured as a function of the scattering angle (Fig. 16.1). The solvent scattering is also measured and subtracted from that of the macromolecular solution. The difference is due to the dissolved particles and contains information about their structure. The SAXS experiments are usually performed using synchrotron radiation, and all major synchrotrons, like ESRF (Grenoble), ANL (Argonne), DESY (Hamburg), and Spring-8 (Himeji), have biologic SAXS beamlines, but home X-ray sources, although yielding much lower flux, can in some cases also be used. For structure analysis (shape, quaternary structure, flexibility studies), solutions containing single molecular species without aggregates are required. Sample monodispersity should usually be better than 90% and must be verified by other methods like native gel filtration, dynamic light scattering (DLS), or analytical ultracentrifugation (AUC) before going to a large-scale facility. Typical protein concentrations required for SAXS are in the range of about 0.5–10 mg/mL, and a concentration series is usually measured to get rid of interparticle interference. Knowledge about the absolute value of the solute concentration, with accuracy better than 10%, is required to estimate the molecular mass. The sample volume per measurement is about 20–100 μL depending on the setup so that about 2–5 mg of purified material is typically required for a complete structural study. Radiation damage is an issue for synchrotron studies, and the measured samples cannot be recovered and used for other experiments.

In biologic SAXS, relatively hard X-ray photons with wavelength λ, about 0.10–0.15 nm are typically employed. When the sample is illuminated by a monochromatic X-ray plane wave with wave vector $k_0 = |\mathbf{k_0}| = 2\pi/\lambda$, the elec-

trons within the object interacting with the incident radiation become sources of spherical waves. For structural studies, elastic scattering effects are relevant, where the modulus of the scattered wave $k_1 = |\mathbf{k_1}|$ is equal to k_0. The amplitude of the wave scattered by each atom is described by its scattering length f, which is proportional to the number of electrons.

The scattering process involves a transformation from the "real" space (coordinates $\mathbf{r}$ of the object) to the "reciprocal" space, that is, coordinates of scattering vectors $\mathbf{s} = (s,\Omega) = \mathbf{k_1} - \mathbf{k_0}$. Here, the momentum transfer is $s = 4\pi\ \lambda^{-1} \sin(\theta)$, where 2θ is the scattering angle and Ω is the direction of the scattering vector. The transformation is mathematically described by the Fourier operator

$$A(\mathbf{s}) = \Im[\rho(\mathbf{r})] = \int \rho(\mathbf{r}) \exp(i\mathbf{sr})\, d\mathbf{r}, \tag{16.1}$$

where the function $\rho(\mathbf{r})$ is the electron density distribution in the sample and $A(\mathbf{s})$ is the scattering amplitude. Experimentally, one measures the scattering intensity, $I(\mathbf{s}) = [A(\mathbf{s})]^2$. The real space resolution of the scattering data can be estimated as $d = 2\pi/s$, which is a counterpart of the well-known Bragg equation in crystallography. The scattering at small angles, that is, at small s, therefore provides information about large distances (much larger than the wavelength), that is, about overall particle structure.

In the present chapter, we consider solution scattering studies where the intensity is isotropic and depends only on the momentum transfer s. To obtain the net signal from the macromolecules, the solvent scattering is subtracted from that of the solution, and the measured intensity corresponds to the excess density distribution $\Delta\rho(\mathbf{r}) = <\rho(\mathbf{r})> - \rho_s$, where ρ_s is the electron density of the solvent. Two major cases can be distinguished for biologic structural studies in solution:

1. The most important idealized case represents a dilute solution of randomly distributed noninteracting particles, all having the same structure. The intensity from the entire ensemble is proportional to the scattering from a single particle averaged over all orientations: $I(s) = \langle I(\mathbf{s})\rangle_\Omega = [F[\Delta\rho(\mathbf{r})]^2]_\Omega$. Typical objects of this kind are monodisperse dilute solutions of folded purified proteins, nucleic acids, or macromolecular complexes. Novel methods of SAXS data analysis allow for structure determination of such systems at low resolution (1–2 nm) (58, 70, 74). Note that for nonideal solutions of identical particles, which can interact with each other, the measured scattering intensity can be written as $I_S(s) = I(s) \times S(s)$. Here, $S(s)$ is the "structure factor" related to the interparticle interactions, which can, in most cases, be eliminated by measurements at different concentrations and extrapolation to infinite dilution.
2. Mixtures of particles of different types, where the scattering intensity will be a linear combination of their contributions

$$I(s) = \sum_{k=1}^{K} \nu_k I_k(s), \tag{16.2}$$

where K is the number of distinct particle types (components), and ν_k and $I_k(s)$ are the volume fraction and the scattering intensity from the kth component, respectively. Typical objects of this kind are equilibrium systems like oligomeric mixtures of proteins or nonequilibrium systems in the process of (dis) assembly or (un)folding. SAXS is employed to determine the volume fractions of the components in the system, if their scattering patterns are available, or to find the number of components based on the measurements at different conditions. IDPs, as we shall see below, represent a specific type of mixture with many components ($K \gg 1$) due to different configurations of individual molecules in solution.

It should also be noted that the basic equations and concepts valid for SAXS are also applicable for small-angle neutron scattering (SANS). The latter technique is also used in the structural studies of biomolecular solutions, but mostly for the analysis of complexes, where neutrons are able to provide additional information thanks to their sensitivity to hydrogen/deuterium composition of the solvent and of the solute. For IDPs, SAXS is more often used, which is faster, requires less material, and provides more precise data at higher scattering angles.

16.3 SCATTERING PROPERTIES OF IDPS

16.3.1 SAXS Profiles and Kratky Plots

The scattering profile measured for an IDP is the average of all these arising from the astronomical number of conformations that the protein adopts in solution. This particularity provides to the experimental SAXS data with very special trends. Figure 16.2A displays the synthetic SAXS curves for 10 randomly selected 100-residue-long polyalanine conformations from a large pool of 10,000 of them. The individual conformations present a large number of features along the simulated momentum transfer range. The initial part of the curve, describing the overall size of the conformation, also presents distinct slopes indicating a large variety of possible sizes that an unstructured chain can adopt. The average SAXS profile, obtained from the individual 10,000 conformations of the pool, presents a smoother behavior, with essentially no features.

Traditionally, Kratky plots ($I(s)\cdot s^2$ as a function of s) have been used to identify disordered states and distinguish them from globular particles. The Kratky representation enhances the particular features observed in scattering profiles allowing an easier identification of different kinds of particles (17). The scattering intensity of a globular protein behaves approximately as $1/s^4$ conferring a bell-shaped Kratky plot with a well-defined maximum. Conversely,

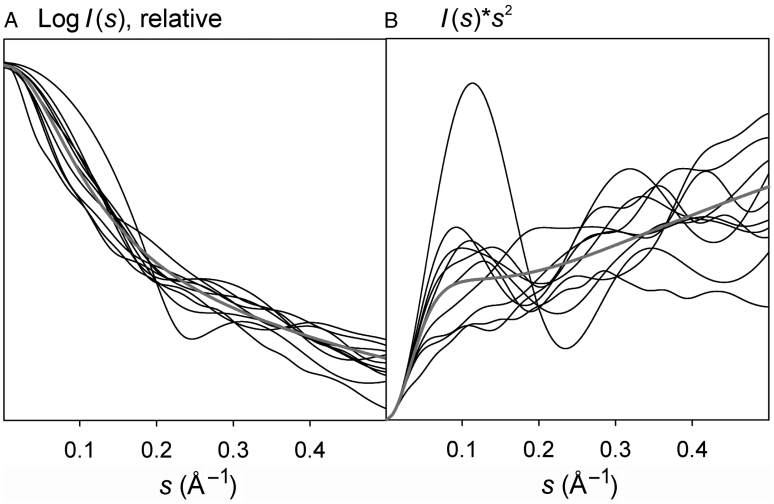

Figure 16.2 Effect of the conformational polydispersity of IDPs on SAXS profiles and their Kartky representations. (A) SAXS profiles (black) of 10 randomly selected 100-residue polyalanine chain conformations from a 10,000 pool of them. The average of the SAXS intensities from the 10,000 conformations is displayed with a thick gray line. (B) Kratky representation of the same data.

an ideal Gaussian chain has a $1/s^2$ dependence of $I(s)$ and, therefore presents a plateau at large s values. In the case of a chain with no thickness, the Kratky plot also presents a plateau over a specific range of s, which is followed by a monotonic increase. This last behavior is normally observed experimentally in unfolded proteins. To exemplify these results, we present the Kratky plots of the above-mentioned 10 individual SAXS profiles compared with the averaged one (Fig. 16.2B). A huge diversity of the Kratky plots for the different conformations is observed. Whereas some of them present a clear maximum, indicating a compact conformation, others present less evident shapes. The average SAXS profile presents a monotonic increase as expected for an unstructured system.

16.3.2 Radius of Gyration: A Single Parameter to Identify IDPs

Unstructured proteins, due to the possibility of adopting highly extended conformations, are characterized by large average sizes compared with globular proteins. One of the best known size characteristic of a protein, its radius of gyration R_g can be directly obtained from SAXS data using a classical Guinier approximation (28). The quantitative comparison of the experimentally measured R_g with these predicted by theoretical models, mainly derived from polymer physics, is used as a diagnostic of the unstructured nature of a

protein. In this section, we briefly describe the model values of R_g used for IDPs.

In the Guinier approximation, assuming that at very small angles ($s < 1.3/R_g$) the intensity is represented as $I(s) = I(0)\exp(-(sR_g)^2/3)$, and the R_g is obtained by a simple linear fit in the logarithmic scale (28). For unstructured proteins, alternative Debye approximation (13) and the transformation mentioned in Equation 16.2 (69) are also used. These two methods may provide more accurate R_g estimates for unstructured systems (57). The experimental R_g is a single value representation of the size of the molecule, which for the scattering from an ensemble of IDPs yields an average (so-called z-average, [20]) over all accessible conformations in solution. The most common quantitative R_g interpretation for unfolded proteins is based on Flory's equation (22), which relates it to the length of the protein chain through a power law

$$R_g = R_0 \cdot N^{\nu}, \tag{16.3}$$

where N is the number of residues in the polymer chain, a constant R_0 depends on several factors, in particular, on the persistence length, and ν is an exponential scaling factor. For an excluded-volume polymer, Flory estimated ν to be approximately 3/5, and more accurate estimates established a value of 0.588 (40). A recent compilation of R_g values measured for 26 chemically denatured proteins sampling broad range of chain lengths found a ν value of 0.598 ± 0.028 (37), in good agreement with hydrodynamic data (86), and an R_0 value of 1.927 ± 0.27. These results suggest the random coil nature of the chemically denatured proteins, at least in terms of the R_g parameter.

The comparison of R_gs measured for proteins with threshold values derived from different Flory's equation parameterizations is a very common strategy to identify IDPs. However, the question whether the conformational sampling in the denatured state is equivalent to that found in intrinsically disordered states is still unclear. Potential bias from the effect of denaturants such as urea or guanidinium chloride on the Ramachandran description of the backbone conformations has been postulated (67 and references therein). As a consequence of the potential differences at residue level in both environments, a modification of the overall properties could be envisioned, leading to differences in the SAXS properties. A recent NMR study based on the measurement of several residual dipolar couplings (RDCs) along the ubiquitin backbone indicated that the conformational sampling in chemically denatured proteins could be different from that found in IDPs (47). These results suggest that even in the absence of transiently structured regions, present Flory's law parameterizations may not be appropriate to estimate R_g for IDPs from the length of the protein, although more work on this field is needed to derive clearer conclusions.

16.4 APPLICATIONS OF SAXS TO IDPS

16.4.1 Biophysical Characterization of IDPs Using SAXS

SAXS has since decades been widely used to characterize unstructured protein states and IDPs. Normally, SAXS curves are analyzed in combination with other experimental techniques and bioinformatics tools to identify unstructured regions in proteins. Circular dichroism (CD), NMR, fluorescence spectroscopy, and hydrodynamic techniques such as size exclusion chromatography, AUC, or DLS have been used in combination with SAXS to identify proteins as IDPs. Since the pioneering study of prothymosin α (25), several other IDPs have been biophysically classified: C-terminal measles virus nucleoprotein (45), PIR domain (49), neuroligin 3 (56), synthetic resilin (54), Ki-1/57 (12), FEZ1 (39), and HrpO (26). In these cases, the monotonic increase of Kratky plots, as well as the large R_g values, and expanded $p(r)$ functions with too large D_{max} values for folded particles were the key features useful to identify them as unstructured proteins.

An excellent example of the synergy between SAXS and other experimental and computational tools to structurally characterize IDPs is the C-terminal measles virus nucleoprotein case, N-tail (45), a 139-residue-long polypeptide. The random coil CD spectra and the low dispersion of the ^{1}H NMR spectra clearly suggested the unstructrured nature of the N-tail. These observations were in agreement with the notably large R_g obtained from Guinier analysis of the small-angle region, 27.5 ± 0.7 Å. However, N-tail Kratky plot presented a clear bump at $s = 0.08\,Å^{-1}$ followed by a flat region (Fig. 16.3). This dual behavior indicates a certain degree of compactness in a concrete region of the chain. A bioinformatics analysis predicts a 16-residue-long fragment to have a strong tendency to form an α-helix. In addition, an X-ray structure of this

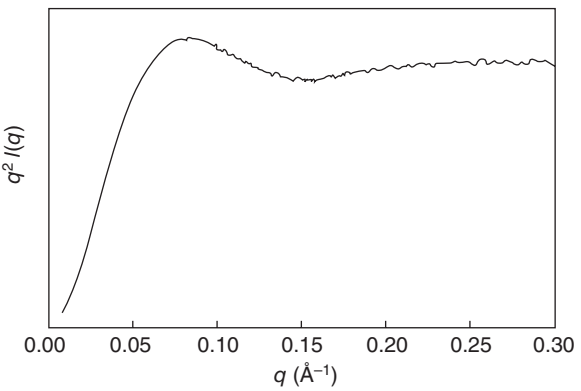

Figure 16.3 Kratky plot of the measles virus N-tail at 9 mg/mL. The bump at $q \approx 0.08\,Å^{-1}$ suggests the presence of a partially folded region within the disordered protein. Figure extracted with permission from Longhi et al. [45].

fragment with its natural partner, the C-terminal region of the phosphoprotein, was obtained in a helical conformation, confirming this stretch as a molecular recognition element (35). Interestingly, the same region of the protein belonging to the family-related Sendai virus was identified to have a large population, up to 75%, of helical conformations (31, 32), substantiating the SAXS results. Importantly, populations of nascent secondary structural elements such as the one found in the N-tail are often related with biomolecular recognition events involved in regulation and signaling (2, 23, 68).

There have been attempts to reconstruct low-resolution structures from SAXS data measured for disordered chains using standard programs. The resulting ab initio reconstructions display highly elongated shapes, dictated by the large D_{max} of IDPs. The validity of this single conformation approach to describe a highly plastic protein is uncertain, although it clearly helps in visualizing a largely diffused molecule.

Intrinsically disordered fragments are often attached to or tether folded domains. This architecture provides advantages in recognition events (18). There are several SAXS studies of partially folded proteins such as the transcriptional repressor CtBP (55), the C-terminus of Sendai virus phosphoprotein (5), α4 (66), N-terminus of Msh6 (64), p53 (75), Prion protein (43), or NEIL1 (78). The presence of highly dynamic fragments attached to globular particles induces a dual behavior in the Kratky representation that presents a maximum, corresponding to the folded part of the protein and a contribution with a continuous rise from the unfolded region. The relative number of amino acids from both distinct fragments dictates the actual features of the Kratky plot. When comparing the experimental scattering profile to the high-resolution structure of the globular domain, if available, normally a bad fit is obtained (55). When performing ab initio reconstructions from these mixed proteins, an additional density not corresponding to the folded part is obtained, which can be assigned to the flexible region (66, 78). Whether this density reflects the space sampling of the unstructured region is not clear.

16.4.2 Monitoring Environmental Changes in IDPs with SAXS

IDPs are often involved in signaling processes and must change their global properties upon environmental modifications within the cell in order to bind to or release from their natural partners. SAXS is a well-suited tool to rapidly monitor structural changes in proteins upon such environmental modifications. The changes, associated with varying pH (38), ionic strength (52), presence of specific ions (29), phosphorylation (29), or additives (30), must induce global size variation in IDPs in order to be monitored by SAXS. These global alterations are reflected again in the apparent R_g, D_{max}, and the appearance of Kratky plots.

Prothymosin α is an example of a protein studied by SAXS in different conditions (Fig. 16.4). Out of its 109 residues, roughly half are aspartic and glutamic acids, such that the protein is predicted to have a global charge of

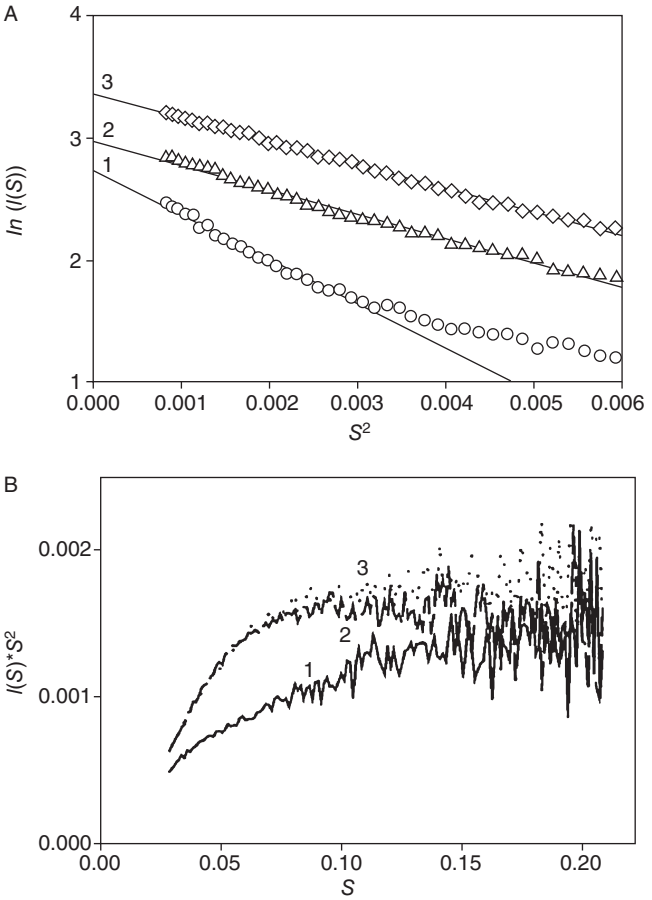

Figure 16.4 Guinier (A) and Kratky plot (B) representations of the SAXS data measured for prothymosin α under different experimental conditions: (1) pH 7.5, (2) pH 2.5, (3) pH 7.5, and 15 mM of $ZnCl_2$. Guinier plots have been scaled arbitrarily for a better visualization (note that the momentum transfer has been expressed units of $S = s/2\pi$). Reprinted from Uversky et al. [79], Copyright (2009), with permission from Elsevier.

–54 at neutral pH. SAXS curves of prothymosin α have been measured at nearly neutral, 7.5, and at acidic, 2.5, pHs (80). Despite Kratky plots indicating that the system was unstructured in both conditions, a dramatic R_g reduction was observed upon pH decrease, from 37.8 ± 0.9 to 27.6 ± 0.9 Å. Interestingly, a similar level of compaction, down to R_g of 28.1 ± 0.8 Å, was observed when, at neutral pH, 15 mM of Zn^{2+} was added to the sample (79). These dramatic changes in protein size can be explained by the decrease of the electrostatic repulsion within the chain at low pH, and with the presence of cations bound to the prothymosin α.

α-Synuclein (αS), a protein found in the aggregates involved in Parkinson's disease and other neurodegenerative pathologies, is a 140-amino acid IDP that has been widely studied and biophysically characterized in different conditions. Some biologic data indicate that the presence of certain metal ions accelerates fibrillation kinetics of αS. The direct interaction of this IDP with Cu^{2+} was confirmed by NMR (62), but the effect of this process on the structure of αS was unclear. SAXS experiments on αS in the presence or absence of Cu^{2+} revealed no global changes upon metal binding, indicating that the recognition process only induced local modifications (10). Through SAXS measurements, it was also demonstrated that the enhanced propensity to form aggregates of the two naturally occurring point mutants A30P and A53T was not provoked by distinct global properties in solution with respect to the wild type (42). Conversely, in acidic conditions, the R_g decreases from 40 ± 1 to 30 ± 1 Å (42), suggesting the presence of either partial folding or the formation of long-range contacts between remote regions of the chain as suggested in different NMR studies (3, 9).

16.4.3 A SAXS Perspective of Biomolecular Interactions of IDPs

The large conformational sampling observed by IDPs provides them with very specific binding properties and consequently pivotal roles in several signaling pathways (19). Interaction with a partner normally requires freezing the highly plastic IDP in a single conformation leading to a so-called folding upon binding (88). The entropic cost of this process makes IDPs very well suited for transient interactions required for signal transduction. Additionally, the large surface exposed by these proteins and the adaptability to different environments make IDPs highly selective or promiscuous partners, depending on the case. Studying the biomolecular recognition processes for IDPs is thus crucial for understanding of the biologic role of this family of proteins, and SAXS has been an important tool for this goal. Several studies were devoted to the interactions of different IDPs with other proteins (11, 24, 64) and with DNA (43, 75). We will give a more detailed explanation of two of them, Msh6 and p53, to exemplify the role that SAXS can have in this field.

The complex of Msh2 and Msh6 recognizes mismatched bases in DNA during mismatch repair. The N-terminal region of Msh6, a 304-residue-long IDP, recognizes PCNA, a homotrimeric protein that controls processivity of DNA polymerases. Shell and coworkers demonstrated this direct interaction with SAXS (64). A comparison of the R_g, Kratky plots, and $p(r)$ functions of the isolated partners and the complex showed that PCNA does not induce substantial structure to the N-terminal region of Msh6, which remains mainly disordered and proteolytically accessible upon binding. The interaction of the Msh2–Msh6 complex with PCNA was also addressed by SAXS. The interaction was shown to produce a heterocomplex that could be considered as a flexible dumbbell, where both globular regions are tethered by the N-terminal Msh6 fragment that acts as a molecular leash. These observations where

further confirmed in an additional experiment with a biologically active deletion mutant of Msh6 with a notably shorter N-terminal tail. In these conditions, the important size changes upon binding were enhanced and easily monitored by the $p(r)$ and D_{max} derived from the SAXS profiles.

The tumor suppressor p53 is a multifunctional protein that plays a crucial role in the processes like apoptosis control and DNA repair p53 is a homotetramer with two folded domains that are tethered and flanked by unstructured regions. In fact, 37% of the whole sequence can be considered an IDP, and, as a consequence, p53 is a highly flexible protein as shown by the SAXS data of the full-length and different constructs (75, 85). In the modeling of p53 against the SAXS data, high-resolution structures of the folded domains from crystallography and NMR were used as rigid bodies, and the unfolded domains were represented by chains of dummy residues following a well-established method (58). The free p53 is a rather open cross-like tetrameric assembly, which collapses in the presence of DNA to tightly embrace the latter (Fig. 16.5). This is how the flexibility of IDPs helps the protein to fulfil its function of constantly watching over the DNA in the cell. Interestingly, the SAXS model of the complex without the disordered N-terminal transactivation domains is in excellent agreement with an independent electron microscopy map of the same molecule (75). In a subsequent study (85), a yet more detailed

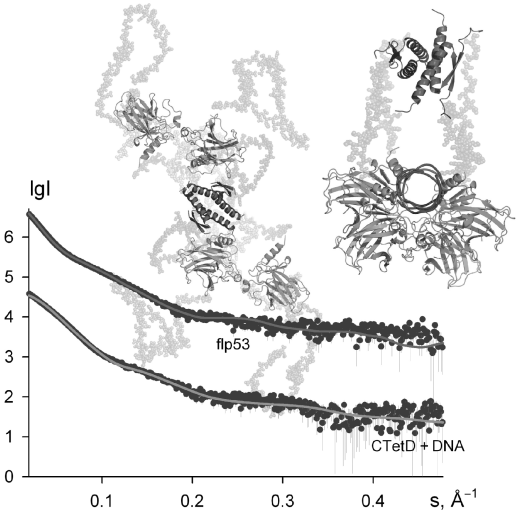

Figure 16.5 The models of tumor suppressor p53 generated from the SAXS curves (dots: experimental data; solid lines: computed patterns). In the free form (left), the four domains of p53 recognizing DNA (green and cyan) are arranged in two separated dimers, which wrap around DNA (pink) upon binding (right). The available high-resolution structures of the domains employed in the modeling are displayed as ribbons, the flexible portions of p53 as semitransparent beads. The N-terminal transactivation domains in the free full-length protein flp53 (tomato beads) are truncated in the complex with DNA (CTeD + DNA). See color insert.

information about the flexibility of the isolated transactivation domain and its structure within the full-length p53 complexed coactivator was obtained by a joint use of SAXS and NMR.

16.4.4 SAXS as a Tool to Validate Computational Models of Unstructured Proteins

The development of realistic ensemble models of unstructured states of proteins has been an important subject of research for many years (91). Initially, these models were focused on protein states under denaturing conditions for a better understanding of the folding process (65). In addition, molecular properties straightforwardly derived from these ensembles such as solvent accessibility serve as a basis for a quantitative interpretation of protein thermodynamics and stability (4, 63). More recently, structural models of IDPs have been developed in order to structurally characterize this family of proteins (5).

The quality of in silico structural models has been evaluated by their ability to properly reproduce experimental data measured for disordered proteins. SAXS has arguably been the most popular technique, although hydrodynamic measurements and NMR have also been used (48). Several types of models with different levels of structural resolution have been validated with previously mentioned compilations of R_g for denatured proteins (15, 21, 27, 33, 77, 83, 84, 90). In these studies, a remarkable agreement with the experimental data is obtained. However, for most of the models, an additional parameterization accounting for the exclusion or the solvation terms is required.

In more advanced studies, complete scattering profiles were considered. Extending the momentum transfer range simulated the development of more adequate ensemble models, which has to be tested at higher resolution and in a more stringent context. Zagrovic et al. addressed the conformational study of a synthetic peptide by combining state-of-the-art molecular dynamic simulations with the SAXS profile experimentally measured (89). In another example, scattering patterns of the reduced ribonuclease A in different denaturing conditions were well reproduced with a proper selection of the solvation term (83). A very simplistic structural model based on polymer theory was used to describe the SANS curve of denatured phosphoglycerate kinase (60).

Fewer studies of this kind were performed for IDPs. To explain the SAXS curve measured for the complex of $p27^{kip1}$/Cdk2/Cyclin A, mobility of the system was explicitly accounted for (24). Hundreds of snapshots of the molecular dynamics simulation of the complex were collected, and their theoretical SAXS profile computed and averaged yielded a curve with a better description of the experimental data than other single-conformation models.

The program flexible Meccano (FM) has been the most tested structural model for IDPs. FM assembles peptidic units, considered as rigid entities, in a consecutive way (5). The force field used for this algorithm includes a coil

description of the residue-specific Ramachandran space sampled by the amino acids, and a coarse-grained description of the side chains that avoids collapse within the chain. This program has been tested for a large number of IDPs, and it has successfully described SAXS and several NMR observables measured for these proteins. A detailed description of the algorithm and its applications to IDPs can be found in chapter 21. FM was used to describe the SAXS profile of Protein X, a three-helix bundle that has a 60-amino-acid-long tail attached at the N-terminus (5). The biophysical characterization of the transactivation domain of p53 was also addressed in the context of the full-length protein and when isolated (85). Importantly, in these cases, the same structural model was able to simultaneously describe NMR properties, which mainly report on the conformational (local) properties and SAXS curves that report on the size and shape (global) characteristics of the proteins. These results suggest the general applicability of the FM structural model to fully describe IDPs.

16.5 THE ENSEMBLE OPTIMIZATION METHOD (EOM)

The EOM has been developed as a general approach to address the structural characterization of IDPs by SAXS (6). Using this strategy, the experimental SAXS data are assumed to arise from a (undetermined a priori) number of coexisting conformational states generalizing the traditional "single-state" methods. The conformations belonging to the final ensemble are selected by a Monte-Carlo-type search based on the scattering patterns computed from a large representative random pool, becoming a data-driven optimized ensemble strategy. The price to pay in order to optimize the large number of degrees of freedom by EOM is that assumptions are required to generate the pool of unfolded chains. The randomization model is also used to establish a threshold for the later quantitative interpretation of the results in terms of individual selected structures, and in this sense the EOM is not entirely model independent. It is worth mentioning that EOM provides a new source of structural information for disordered and also for flexible multidomain proteins (7, 8, 81, 82). The major features of the EOM will be discussed in the following sections.

16.5.1 The EOM Algorithm

In the algorithm, schematically depicted in Figure 16.6, the solute is represented by an ensemble containing N different conformations of the same molecule (in practice, N is about 20–50). The appropriate ensemble is selected from a pool containing $M \gg N$ conformers, which should cover the conformational space of the molecule. A genetic algorithm (GA) (34) is then employed to select subsets of configurations that, after averaging their individual scattering profiles, fit the experimental data.

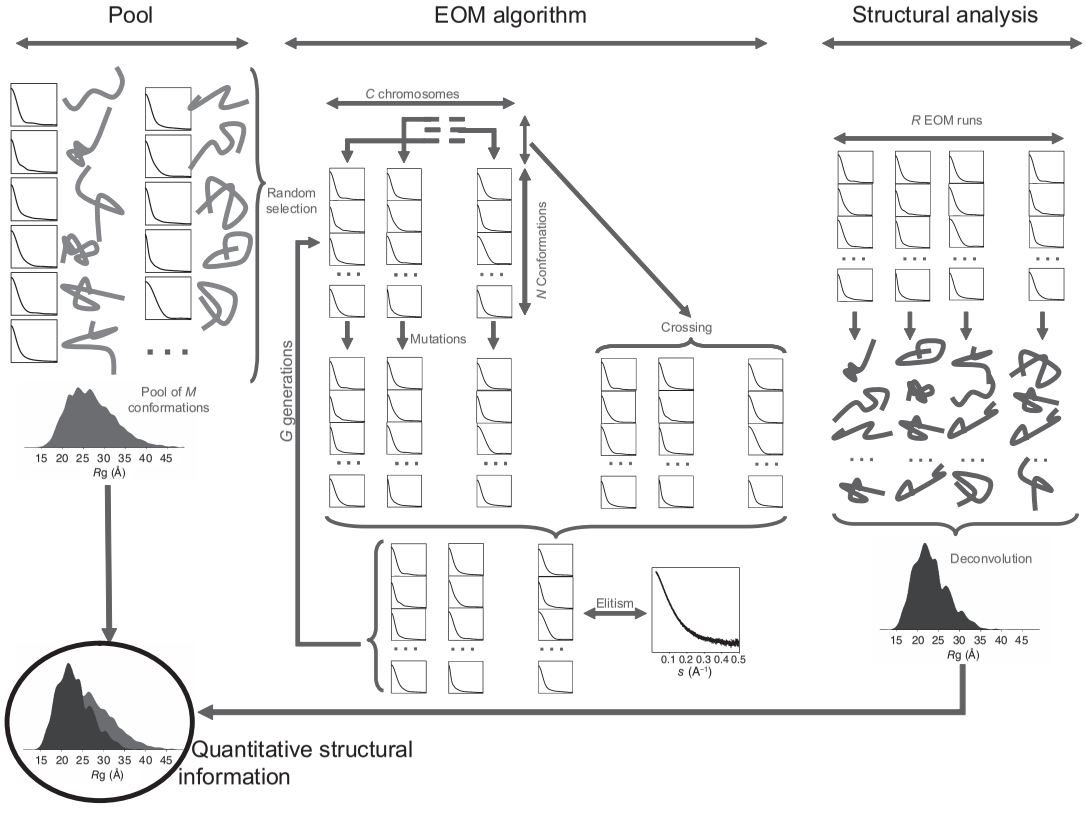

Figure 16.6 Schematic representation of the EOM scheme for the analysis of SAXS data in terms of R_g distributions. The M conformations/curves of the pool (random distribution), left part of the figure, are used to generate the initial C chromosomes and to feed the genetic operators (*mutations*, *crossing*, and *elitism*) along the GA process that runs for G generations. The complete process is repeated R independent times, and each run provides N selected structures/curves that fit the experimental profile. The structural analysis of the resulting conformations is displayed on the right part of the scheme; the R_g distribution of the selected ($N \times R$) conformations is compared with that derived from the pool that is considered as a complete conformational freedom scenario. From this, comparison is possible to derive a quantitative structural estimation of the IDP conformations coexisting in solution.

The scattering data from such an ensemble is computed by summing up the individual scattering patterns as in Equation 16.2. We can assume that the subsets are uniformly populated, so that the intensity of a subset $I(s)$ containing N conformers is

$$I(s) = \frac{1}{N}\sum_{n=1}^{N} I_n(s), \tag{16.4}$$

where $I_n(s)$ is the scattering from the nth conformer. The scattering curves from all the structures in the pool are first precomputed using the program CRYSOL (71), and the subsequent selection operators are preformed using these patterns, which significantly speeds up the calculations. The subset selection is performed using a GA, where each subset (called chromosome) contains N scattering profiles (genes) from different conformers. In the first generation, $C = 50$ chromosomes are generated by selecting the conformations randomly from the pool. In each generation G, these C chromosomes are submitted to *mutation* and *crossing*. In *mutation*, the genes of each chromosome are randomly exchanged for others either from the pool or from the chromosomes of the same generation. In the *crossing*, genes of two randomly selected chromosomes are exchanged. After the two genetic operations, the population is composed of $3C$ chromosomes. For each chromosome, the average (Eq. 16.4) of its individual SAXS profiles is compared with the experimental scattering yielding the fitness function (Eq. 16.5)

$$\chi^2 = \frac{1}{K-1}\sum_{j=1}^{K}\left[\frac{\mu I(s_j) - I_{\exp}(s_j)}{\sigma(s_j)}\right]^2, \tag{16.5}$$

where $I_{\exp}(s)$ is the experimental scattering, K is the number of experimental points, $\sigma(s)$ is the standard deviation, and μ is a scaling factor. The best C chromosomes yielding the lowest χ^2 are selected for further evolution in an *elitism* fashion, typically for up to G = 1000–5000 generations. The chromosome with the best agreement to the experimental data is collected for further structural analysis. The complete process is repeated R times in order to collect $R \times N$ final conformations. The distributions of the low-resolution parameters, R_g, maximum distance (D_{max}), and anisotropy in the selected conformations provide information about the structure of the solute, as discussed in the following sections.

16.5.2 Structural Information Coded in EOM Ensembles

It may seem that a subpopulation with $N \approx 50$ is far too small to explain the conformational behavior of a flexible protein sampling an astronomical amount of conformations in solution. However, these subpopulations are sufficient to depict the global properties of the protein in terms of size and shape (i.e., R_g, D_{max}, and anisotropy) rather than conformational details at residue level. The structures selected by the EOM are neither the most populated

conformations, not even are they claimed to really exist in solution. Instead, the EOM ensemble is a tool to describe the size and shape distributions sampled by the unfolded molecule. Selected ensembles derived from repeated EOM runs starting from different random seeds normally contain different conformations, but they all provide similar R_g, D_{max}, and anisotropy distributions. Therefore, the algorithm is able to find equivalent minima in terms of distributions but of course not in terms of individual molecular configurations: the latter is not identifiable given the low resolution of SAXS. Still, the distributions provided by EOM yield a major improvement over traditional approaches that condense all structural characteristics in the averaged R_g value. One may also argue that due to the restricted information content of the SAXS data, the EOM model containing a relatively small number of conformers adequately describe the entire system. Similar approaches using higher-resolution techniques (e.g., RDCs in NMR) require much larger ensembles to reach reasonable levels of convergence (5, 33).

It is very instructive to compare the EOM-derived distributions with those of the random ensembles (e.g., initial pools) to potentially detect the presence of nascent secondary structural elements or low populations of transient long-range contacts as alterations of the distributions with respect to the pool. Local transient structures display broader distributions of R_g and normally shifted toward larger values, whereas long-range contacts and residual tertiary structure provide more compact distributions than the random coils.

The EOM has been applied to the biophysical characterization of intrinsically unfolded neuligin 3 (56) and HrpO (26). For the neuroligin 3 case, the EOM suggests the presence of two different families of conformations coexisting in solution (Fig. 16.7). The first family includes relatively compact structures, with less conformational diversity than the starting pool. A second family of structures displayed extremely large values of R_g that were attributed to either a population of highly extended molecules or the presence of oligomers. Interestingly, this bimodal distribution disappeared upon the addition of urea, where a unimodal and highly extended ensemble of conformations was depicted by the EOM.

16.5.3 Multiple SAXS Curve Fitting with EOM

The EOM provides overall structural information, whereas the exact regions responsible for conformational restriction remains elusive. It appears, however, that using deletion mutants, more detail could be obtained even with such a low-resolution method as SAXS. Indeed, the scattering profiles from different deletion mutants of the full-length protein can be fitted simultaneously with EOM, and the sum of the individual figures of merit drives the optimization (6). In this way, contributions from different chain fragments to the SAXS data can be distinguished to identify the regions responsible for the increase or decrease in compactness of the protein. The SAXS resolution is therefore improved by the addition of multiple experimental curves (of course, this

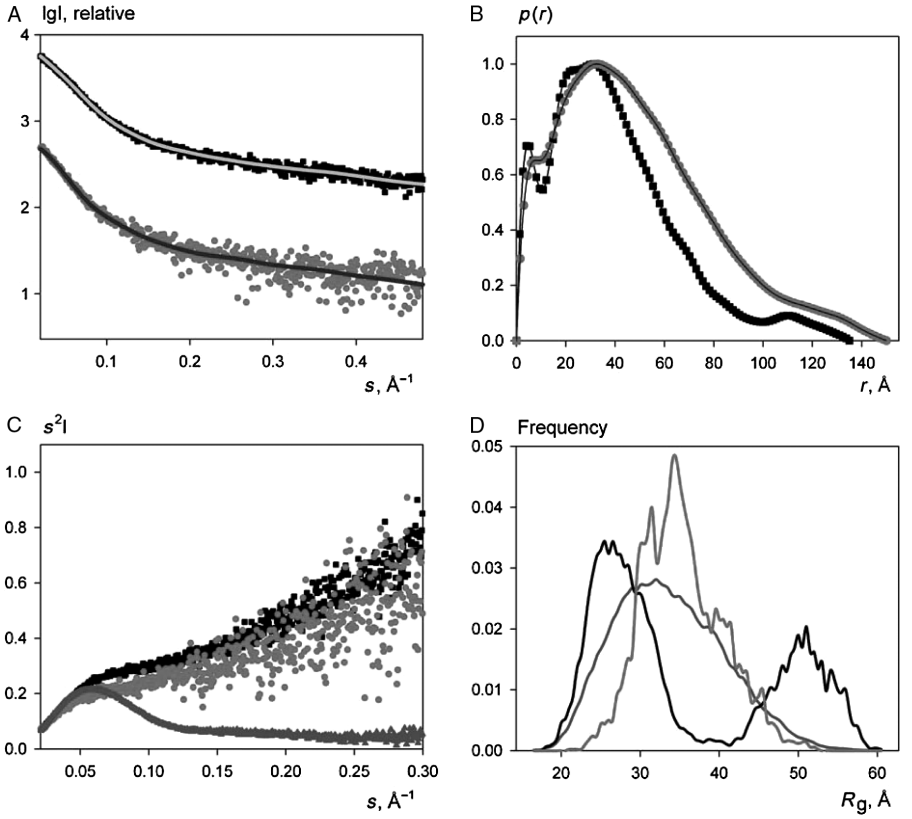

Figure 16.7 SAXS measurements and analysis on neuroligin 3. (A) X-ray scattering intensities in native (black squares) and denaturing (gray circles) conditions. EOM fits to the profiles are displayed in solid lines. (B) $p(r)$ distribution functions for both data sets. (C) Kratky plots of both data sets (same color code) compared with that obtained for the globular protein BSA (triangles). (D) R_g distributions derived from the EOM analysis of both data sets (same color code). The dual and the monotonic behavior of neuroligin 3 in native and denaturing conditions respectively are observed when compared from the distribution of the randomly generated conformations of the pool. Figure reprinted with permission from Paz et al. [56].

strategy is only valid if the structural elements of the full-length protein remain intact in the deletion mutants).

Mylonas et al. (53) applied the multiple SAXS curve fitting with EOM to the SAXS analysis of tau, an IDP involved in neuronal microtubule stabilization, and found abnormal deposits in the brain of Alzheimer's disease patients (46). Two tau isoforms were studied, and SAXS data for the full-length and several different deletion mutants for each isoform were used (Fig. 16.8). The EOM unambiguously identifies the so-called repeat region as the source of residual secondary structure in tau, in perfect agreement with previous NMR

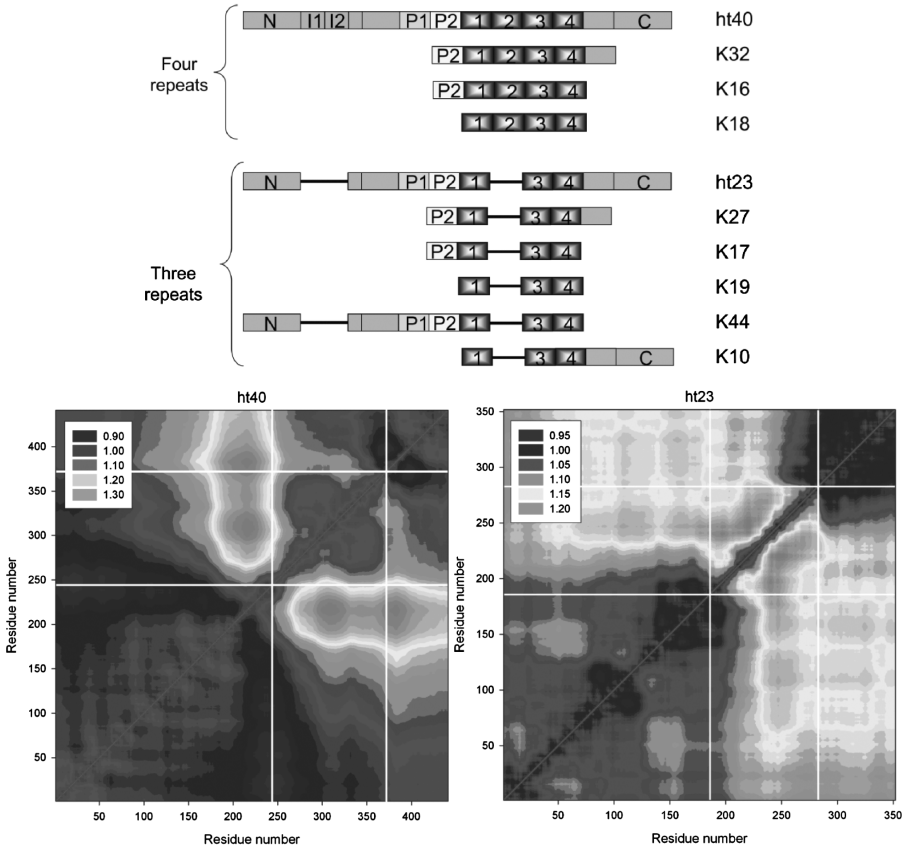

Figure 16.8 Bar diagrams of isoforms of tau protein used for the simultaneous SAXS curve analysis with EOM (top panel). Two families are depicted with four (ht40) and three (ht23) repeat regions. Cα–Cα distance plot of ht40 (middle panel) and ht23 (bottom panel) using multiple curve fitting. Each plot point shows the ratio of the average distance of the selected structures with that obtained from the pool. White lines indicate the residues belonging to the repeat domain. Figure extracted with permission from Mylonas et al. [53]. See color insert.

data indicating the presence of turns and extended fragments in this region (50, 51). The multiple curve fitting, through the averaged Cα–Cα inter-residue distance matrix identifies a distinct conformational behavior depending on the number of repeats present in the isoforms. For the isoform ht23, with three repeats, the maximum separation is found within the repeat domain itself. The full-length ht40 isoform with four domains reveals an enhanced separation between the repeat domain and the preceding region. These results suggest that the different number of turns (one per repeat) may lead to different global arrangements of the chain in that region enhancing or shortening the average interdomain distances expected from a random coil.

16.6 CONCLUSIONS

High-resolution methods are barely applicable to study flexible molecules like IDPs in solution, and among the low-resolution structural techniques, not many can be used. Due to its versatility, speed, and ability to deal with disordered matter and mixtures, SAXS is one of the major structural players in the IDP field. For decades, SAXS was employed to unfolded proteins at a rather simplistic level, by qualitative analysis using Kratky plots and the only parameter, R_g being quantitatively determined. Although even these approaches yielded valuable information about numerous IDPs in solution, a genuine breakthrough was brought up by recent developments in SAXS allowing for comprehensive quantitative analysis of the scattering patterns. Ensemble optimization coupled with Molecular Dynamics (MD) simulations and the joint use with local information obtained by NMR or FRET promises make SAXS an extremely effective tool in the characterization of structure and dynamics of IDPs in solution.

ACKNOWLEDGMENTS

Financial support form the Spanish Ministry of Education-FEDER (BIO2007-63458) is gratefully acknowledged. P. B. holds a Ramón y Cajal contract that is partially financed by the Spanish Ministry of Education and the by funds provided to the Institute for Research in Biomedicine by the Generalitat de Catalunya.

REFERENCES

1. Akiyama, S., A. Nohara, and Y. Maéda. 2008. Assembly and disassembly dynamics of the cyanobacterial periodosome. Mol Cell **29**:703–16.
2. Baker, J. M. R., R. P. Hudson, V. Kanelis, W.-Y. Choy, P. H. Thibodeau, P. J. Thomas, and J. D. Forman-Kay. 2007. CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. Nature Struct Mol Biol **14**:738–45.
3. Bernadó, P., C. W. Bertoncini, C. Griesinger, M. Zweckstetter, and M. Blackledge. 2005. Defining long-range order and local disorder in native α-synuclein using residual bipolar couplings. J Am Chem Soc **127**:17968–9.
4. Bernadó, P., M. Blackledge, and J. Sancho. 2006. Sequence-specific solvent accessibilities of protein residues in unfolded protein ensambles. Biophys J **91**:4536–43.
5. Bernadó, P., L. Blanchard, P. Timmins, D. Marion, R. W. H. Ruigrok, and M. Blackledge. 2005. A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. Proc Natl Acad Sci U S A **102**:17002–7.
6. Bernadó, P., E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun. 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc **129**:5656–64.

7. Bernadó, P., Y. Pérez, D. I. Svergun, and M. Pons. 2008. Structural characterization of the active and inactive states of Src kinase in solution by small-angle X-ray scattering. J Mol Biol **376**:492–505.
8. Bertini, I., V. Calderone, M. Fragai, R. Jaiswal, C. Luchinat, M. Melikian, E. Mylonas, and D. I. Svergun. 2008. Evidence of reciprocal reorientation of the catalytic and hemopexin-like domains of full-length MMP-12. J Am Chem Soc **130**:7011–21.
9. Bertoncini, C. W., Y.-S. Jung, C. O. Fernández, W. Hoyer, C. Griesinger, T. M. Jovin, and M. Zweckstetter. 2005. Release of long-range tertiary interactions potentiates aggregation of natively unstructured α-synuclein. Proc Natl Acad Sci U S A **102**:1430–5.
10. Binolfi, A., R. M. Rasia, C. W. Bertoncini, M. Ceolin, M. Zweckstetter, C. Griesinger, T. M. Jovin, and C. O. Fernández. 2006. Interaction of α-synuclein with divalent metal ions reveals key differences: a link between structure, binding specificity and fibrillation enhancement. J Am Chem Soc **128**:9893–901.
11. Bourhis, J.-M., V. Receveur-Bréchot, M. Oglesbee, X. Zhang, M. Buccellato, H. Darbon, B. Cannard, S. Finet, and S. Longhi. 2005. The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. Protein Sci **14**:1975–92.
12. Bressan, G. C., J. C. Silva, J. C. Borges, D. O. dos Passos, C. H. I. Ramos, I. L. Torriani, and J. Kobarg. 2008. Human regulatory protein Ki-1/57 has characteristics of an intrinsically unstructured protein. J Proteome Res **7**:4465–74.
13. Calmettes, P., D. Durand, M. Desmadril, V. Receveur, and J. C. Smith. 1994. How random is a highly denatured protein? Biophys Chem **53**:105–14.
14. Dafforn, T. R., and C. J. I. Smith. 2004. Natively unfolded domains in endicytosis: hooks, lines and linkers. EMBO Rep **5**:1046–52.
15. Ding, F., R. K. Jha, and N. V. Dokholyan. 2005. Scaling behavior and structure of denatured proteins. Structure **13**:1047–54.
16. Dobson, C. M. 2004. Experimental investigation of protein folding and misfolding. Methods **34**:4–14.
17. Doniach, S. 2001. Changes in biomolecular conformation seen by small angle X-ray scattering. Chem Rev **101**:1763–78.
18. Dunker, A. K., C. J. Brown, J. D. Lawson, L. M. Iakoucheva, and Z. Obradovic. 2002. Intrinsic disorder and protein function. Biochemistry **41**:6573–82.
19. Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. Nature Rev **6**:197–208.
20. Feigin, L. A., and D. I. Svergun. 1987. Structure Analysis by Small-Angle X-ray and Neutron Scattering. Plenum Press, New York.
21. Fitzkee, N. C., and G. D. Rose. 2004. Reassessing random-coil statistics in unfolded proteins. Proc Natl Acad Sci U S A **101**:12497–502.
22. Flory, P. J. 1953. Principles of Polymer Chemistry. Cornell University Press, Ithaca, NY.
23. Fuxreiter, M., I. Simon, P. Friedrich, and P. Tompa. 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol **338**:1015–26.

24. Galea, C. A., A. Nourse, Y. Wang, S. G. Sivakolundu, W. T. Séller, and R. W. Kriwacki. 2008. Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27Kip1. J Mol Biol **376**:827–38.

25. Gast, K., H. Damaschun, H. K. Eckert, K. Schulze-Forster, H. R. Maurer, M. Müller-Frohne, D. Zirwer, J. Czarnecki, and G. Damaschun. 1995. Prothymosin α: a biologically active protein with random coil conformation. Biochemistry **34**:13211–8.

26. Gazi, A. D., M. Bastaki, S. N. Charola, E. A. Gkougkoulia, E. A. Kapellios, N. J. Panapoulos, and M. Kokkinidis. 2008. Evidence for a coiled-coil interaction mode of disordered proteins from bacterial type III secretion systems. J Biol Chem **283**:34062–8.

27. Goldenberg, D. P. 2003. Computational simulation of the statistical properties of unfolded proteins. J Mol Biol **326**:1615–33.

28. Guinier, A. 1939. La diffraction des rayons X aux trés petits angles: application a l'étude de phenomenes ultramicroscopiques. Ann Phys (Paris) **12**:161–237.

29. He, G., A. Ramachandran, T. Dahl, S. George, D. Schultz, D. Cookson, A. Veis, and A. George. 2005. Phosphorylation of phosphophoryn is crucial for its function as a mediator of biomineralization. J Biol Chem **280**:33109–14.

30. Hong, D.-P., A. L. Fink, and V. N. Uversky. 2008. Structural characteristics of α-synuclein oligomers stabilized by flavonoid baicalein. J Mol Biol **383**:214–23.

31. Houben, K., D. Marion, N. Tarbouriech, R. W. H. Ruigrok, and L. Blanchard. 2007. Interaction of the C-terminal domains of Sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. J Virol **81**:6807–16.

32. Jensen, M. R., K. Houben, E. Lescop, L. Blanchard, R. W. H. Ruigrok, and M. Blackledge. 2008. Quantitative conformational analysis of partially folded proteins from residual dipoar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. J Am Chem Soc **130**:8055–61.

33. Jha, A. K., A. Colubri, K. F. Freed, and T. R. Sosnick. 2005. Statistical coil model of the unfolded state: resolving the reconciliation problem. Proc Natl Acad Sci U S A **102**:13099–104.

34. Jones, G. 1998. Genetic and evolutionary algorithms. In Encyclopedia of Computational Chemistry. Wiley, Chichester, UK.

35. Kingston, R. L., D. J. Hamel, L. S. Gay, F. W. Dahlquist, and B. W. Matthews. 2004. Structural basis for the attachment of a paramyxoviral polymerase to its template. Proc Natl Acad Sci U S A **101**:8301–6.

36. Koch, M. H. J., P. Vachette, and D. I. Svergun. 2003. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q Rev Biophys **36**:147–227.

37. Kohn, J. E., I. S. Millet, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach, and K. W. Plaxco. 2004. Random coil behaviour and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci U S A **101**: 12491–6.

38. Konno, T., N. Tanaka, M. Kataoka, E. Takano, and M. Maki. 1997. A circular dichroism study of preferential hydration and alcohol effects on a denatured protein, pig calpastatin domain I. Biochim Biophys Acta **1342**:73–82.
39. Lanza, D. C. F., J. C. Silva, E. M. Assmann, A. J. C. Quaresma, G. C. Bressan, I. L. Torriani, and J. Kobarg. 2009. Human FEZ1 has characteristics of a natively unfolded protein and dimerizes in solution. Proteins **74**:104–21.
40. LeGuillou, J. C., and J. Zinn-Justin. 1977. Critical exponents for the n-vector model in three dimensions from field theory. Phys Rev Lett **39**:95–8.
41. Levitt, M. 2007. Growth of novel protein structural data. Proc Natl Acad Sci U S A **104**:3183–8.
42. Li, J., V. N. Uversky, and A. L. Fink. 2002. Conformational behavior of human α-synuclein is modulated by familial Parkinson's disease point mutations A30P and A53T. Neurotoxicology **23**:553–67.
43. Lima, L. M. T. R., Y. Cordeiro, L. W. Tinoco, A. F. Marques, C. L. P. Oliveira, S. Sampath, R. Kodali, G. Choi, D. Foguel, I. Torriani, B. Caughey, and J. L. Silva. 2006. Structural insights into the interaction between prion protein and nucleic acid. Biochemistry **45**:9180–7.
44. Lipfert, J., and S. Doniach. 2007. Small-angle X-ray scattering from RNA, proteins, and protein complexes. Annu Rev Biophys Biomol Struct **36**:307–27.
45. Longhi, S., V. Receveur-Brechot, D. Karlin, K. Johansson, H. Darbon, D. Bhella, R. Yeo, S. Finet, and B. Canard. 2003. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. J Biol Chem **278**:18638–48.
46. Mandelkow, E. M., and E. Mandelkow. 1998. Tau in Alzheimer's disease. Trends Cell Biol **8**:425–7.
47. Meier, S., S. Grzesiek, and M. Blackledge. 2007. Mapping the conformational landscape of urea-denatured ubiquitin using residual dipolar couplings. J Am Chem Soc **129**:9799–807.
48. Mittag, T., and J. D. Forman-Kay. 2007. Atomic-level characterization of disordered protein ensembles. Curr Opin Struct Biol **17**:3–14.
49. Moncoq, K., I. Broutin, C. T. Craescu, P. Vachette, A. Ducruix, and D. Durand. 2004. SAXS study of the PIR domain from the Grb14 molecular adaptor: a natively unfolded protein with transient structure primer? Biophys J **87**:4056–64.
50. Mukrasch, M. D., J. Biernat, M. von Bergen, C. Griesinger, E. Mandelkow, and M. Zweckstetter. 2005. Sites of tau important for aggregation populate β-structure and bind to microtubules and polyanions. J Biol Chem **280**:24978–86.
51. Mukrasch, M. D., P. Markwick, J. Biernat, M. von Bergen, P. Bernadó, C. Griesinger, E. Mandelkow, M. Zweckstetter, and M. Blackledge. 2007. Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. J Am Chem Soc **129**:5235–43.
52. Munishkina, L. A., A. L. Fink, and V. N. Uversky. 2004. Conformational prerequisites for formation of amyloid fibrils from histones. J Mol Biol **342**:1305–24.
53. Mylonas, E., A. Hascher, P. Bernadó, M. Blackledge, E. Mandelkow, and D. I. Svergun. 2008. Domain conformation of tau protein studied by solution small-angle X-ray scattering. Biochemistry **47**:10345–53.

54. Nairn, K. M., R. E. Lyons, R. J. Mulder, S. T. Mudie, D. J. Cookson, E. Lesieur, M. Kim, D. Lau, F. H. Scholes, and C. M. Elvin. 2008. A synthetic resilin is largely unstructured. Biophys J **95**:3358–65.

55. Nardini, M., D. Svergun, P. V. Konarev, S. Spano, M. Fasano, C. Bracco, A. Pesce, A. Donadini, C. Cericola, F. Secundo, A. Luini, D. Corda, and M. Bolognesi. 2006. The C-terminal domain of the transcriptional corepressor CtBP is intrinsically unstructured. Protein Sci **15**:1042–50.

56. Paz, A., T. Zeev-Ben-Mordehai, M. Lundqvist, E. Sherman, E. Mylonas, L. Weiner, G. Haran, D. I. Svergun, F. A. A. Mulder, J. L. Sussman, and I. Silman. 2008. Biophysical characterization of the unstructured cytoplasmatic domain of the human neuronal adhesion protein neuroligin 3. Biophys J **95**:1928–44.

57. Pérez, J., P. Vachette, D. Russo, M. Desmandril, and D. Durand. 2001. Heat-induced unfolding of neocarzinostatin, a small all-β protein investigated by small-angle X-ray scattering. J Mol Biol **308**:721–43.

58. Petoukhov, M. V., and D. I. Svergun. 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J **89**:1237–50.

59. Petoukhov, M. V., and D. I. Svergun. 2007. Analysis of X-ray and neutron scattering from biomolecular solutions. Curr Opin Struct Biol **17**:562–71.

60. Petrescu, A.-J., V. Receveur, P. Calmettes, D. Durand, M. Desmandril, B. Roux, and J. C. Smith. 1997. Small-angle neutron scattering by strongly denatured protein: analysis using random polymer theory. Biophys J **72**:335–42.

61. Putnam, C. D., M. Hammel, G. L. Hura, and J. A. Tainer. 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining achúrate macromolecular structure, conformations and assemblies in solution. Q Rev Biophys **40**:191–285.

62. Rasia, R. M., C. W. Bertoncini, D. Marsh, W. Hoyer, D. Cherny, M. Zweckstetter, C. Griesinger, T. M. Jovin, and C. O. Fernández. 2005. Structural characterization of copper(II) binding to α-synuclein: insights into the bioinorganic chemistry of Parkinson's disease. Proc Natl Acad Sci U S A **102**:4294–9.

63. Robertson, A. D., and K. P. Murphy. 1997. Protein structure and the energetics of protein stability. Chem Rev **97**:1251–67.

64. Shell, S. S., C. D. Putnam, and R. D. Kolodner. 2007. The N terminus of *Saccharomyces cerevisiae* Msh6 is an ustructured tether to PCNA. Mol Cell **26**:565–78.

65. Shortle, D. 1996. The denatured state (the other half of the holding equation) and its role in protein stability. FASEB J **10**:27–34.

66. Smetana, J. H. C., C. L. P. Oliveira, W. Jablonka, T. A. Perthinez, F. R. G. Carneiro, M. Montero-Lomeli, I. Torriani, and N. I. T. Zanchin. 2006. Low resolution structure of the human α4 protein (IgBP1) and studies on the stability of α4 and of its yeast ortholog Tap42. Biochim Biophys Acta **1764**:724–34.

67. Stumpe, M. C., and H. Grubmüller. 2007. Interaction of urea with amino acids: implications for urea-induced protein denaturation. J Am Chem Soc **129**:16126–31.

68. Sugase, K., H. J. Dyson, and P. E. Wright. 2007. Mechanism of coupled folding of an intrinsically disordered protein. Nature **447**:1021–5.

69. Svergun, D. I. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr **25**:495–503.

70. Svergun, D. I. 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J **76**:2879–86.

71. Svergun, D. I., C. Barberato, and M. H. J. Koch. 1995. CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Cryst **28**:768–73.

72. Svergun, D. I., and M. H. J. Koch. 2002. Advances in structure analysis using small-angle scattering in solution. Curr Opin Struct Biol **12**:654–660.

73. Svergun, D. I., and M. H. J. Koch. 2003. Small-angle scattering studies of biological macromolecules in solution. Rep Prog Phys **66**:1735–82.

74. Svergun, D. I., Petoukhov, M. V., and M. H. J. Koch. 2001. Determination of domain structure of proteins from X-ray solution scattering. Biophys J **80**:2946–53.

75. Tidow, H., R. Melero, E. Mylonas, S. M. V. Freund, J. G. Grossmann, J. M. Carazo, D. I. Svergun, M. Valle, and A. R. Fersht. 2007. Quaternary structure of tumor suppressor p53 and a specific p53-DNA complex. Proc Natl Acad Sci U S A **104**:12324–9.

76. Tompa, P. 2002. Intrinsically unstructured proteins. Trends Biochem Sci **27**:527–33.

77. Tran, H. T., X. Wang, and R. V. Pappu. 2005. Reconciling observations of sequence-specific conformational propensities with the generic polymeric behaviour of denatured proteins. Biochemistry **44**:11369–80.

78. Tsutakawa, S. E., G. L. Hura, K. A. Frankel, P. K. Cooper, and J. A. Tainer. 2007. Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. J Struct Biol **158**:214–23.

79. Uversky, V. N., J. R. Gillespie, I. S. Millett, A. V. Khodyakova, R. N. Vasilenko, A. M. Vasiliev, I. L. Rodionov, G. D. Kozlovskaya, D. A. Dolgikh, A. L. Fink, S. Doniach, E. A. Permyanov, and V. M. Abramov. 2000. Zn^{2+}-mediated structure formation and compactation of the "natively unfolded" human prothymosin α. Biochem Biophys Res Commun **267**:663–8.

80. Uversky, V. N., J. R. Gillespie, I. S. Millett, A. V. Khodyakova, A. M. Vasiliev, T. V. Chernovskaya, R. N. Vasilenko, G. D. Kozlovskaya, D. A. Dolgikh, A. L. Fink, S. Doniach, and V. M. Abramov. 1999. Natively unfolded human prothymosin α adopts partially folded collapsed conformation at acidic pH. Biochemistry **38**:15009–16.

81. VanOudenhove, J., E. Anderson, S. Krueger, and J. L. Cole. 2009. Analysis of PKR structure by small-angle scattering. J Mol Biol **387**:910–20.

82. Wang, S., M. T. Overgaard, Y. X. Hu, and D. B. McKay. 2008. The *Bacillus subtilis* RNA helicase YxiN is distended in solution. Biophys J **94**:L01–3.

83. Wang, Y., J. Trewhella, J., and D. P. Goldenberg. 2008. Small-angle X-ray scattering of reduced ribonuclease A: effects of solution conditions and comparisons with a computational model of unfolded proteins. J Mol Biol **377**:1576–92.

84. Wang, Z., K. W. Plaxco, and D. E. Makarov. 2007. Influence of loca and residual structures on the scaling behavior and dimensions of unfolded proteins. Biopolymers **68**:321–8.

85. Wells, M., H. Tidow, T. J. Rutherford, P. Markwick, M. R. Jensen, E. Mylonas, D. I. Svergun, M. Blackledge, and A. R. Fersht. 2008. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci U S A **105**:5762–7.

86. Wilkins, D. K., S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, and L. J. Smith. 1999. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry **38**:16424–31.

87. Wright, P. E., and H. J. Dyson. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol **293**:321–31.

88. Wright, P. E., and H. J. Dyson. 2009. Linking holding and binding. Curr Opin Struct Biol **19**:31–8.

89. Zagrovic, B., J. Lipfert, E. J. Soria, I. S. Millett, W. F. van Gunsteren, S. Doniach, and V. S. Pande. 2005. Unusual compactness of a polyproline type II structure. Proc Natl Acad Sci U S A **102**:11698–703.

90. Zhou, H.-X. 2002. Dimensions of denatured protein chains from hydrodynamic data. J Phys Chem B **106**:5769–75.

91. Zhou, H.-X. 2004. Polymer models of protein stability, folding, and interactions. Biochemistry **43**:2141–54.