Structural Characterization of Highly Flexible Proteins by Small-Angle Scattering

7

Tiago N. Cordeiro, Fátima Herranz-Trillo, Annika Urbanek, Alejandro Estaña, Juan Cortés, Nathalie Sibille, and Pau Bernadó

Abstract

Intrinsically Disordered Proteins (IDPs) are fundamental actors of biological processes. Their inherent plasticity facilitates very specialized tasks in cell regulation and signalling, and their malfunction is linked to severe pathologies. Understanding the functional role of disorder requires the structural characterization of IDPs and the complexes they form. Small-angle Scattering of X-rays (SAXS) and Neutrons (SANS) have notably contributed to this structural understanding. In this review we summarize the most relevant developments in the field of SAS studies of disordered proteins. Emphasis is given to ensemble methods and how SAS data can be combined with computational approaches or other biophysical information such as NMR. The unique capabilities of SAS enable its application to extremely challenging disordered systems such as low-complexity regions, amyloidogenic proteins and transient biomolecular complexes. This reinforces the fundamental role of SAS in the structural and dynamic characterization of this elusive family of proteins.

T.N. Cordeiro • A. Urbanek • N. Sibille • P. Bernadó (✉)
Centre de Biochimie Structurale. INSERM, CNRS, Université de Montpellier.29, rue de Navacelles, 34090 Montpellier, France
e-mail: pau.bernado@cbs.cnrs.fr

F. Herranz-Trillo
Centre de Biochimie Structurale. INSERM, CNRS, Université de Montpellier.29, rue de Navacelles, 34090 Montpellier, France

Department of Pharmacy and Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen, Denmark

A. Estaña
Centre de Biochimie Structurale. INSERM, CNRS, Université de Montpellier.29, rue de Navacelles, 34090 Montpellier, France

LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

J. Cortés
LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

B. Chaudhuri et al. (eds.), *Biological Small Angle Scattering: Techniques, Strategies and Tips*, Advances in Experimental Medicine and Biology 1009, DOI 10.1007/978-981-10-6038-0_7

Keywords
Intrinsically disordered proteins • Small-angle x-ray scattering • Small-angle neutron scattering • Ensemble methods • Protein-protein interactions • Nuclear magnetic resonance • Low-complexity regions • Random coil models • Amyloids

7.1 Introduction

Intrinsically disordered Proteins or Regions (IDPs/IDRs) have emerged as key actors for a large variety of biological functions such as cell signalling and regulation (Kriwacki et al. 1996; Wright and Dyson 1999, 2015; Dunker et al. 2002). The main feature of IDPs and IDRs is their lack of permanent secondary or tertiary structure that provides them with an inherent malleability enabling highly specialized biological functions (Dunker et al. 2002). Eukaryotic genomes are highly enriched in genes coding for disordered proteins, and this observation has been linked to the major complexity of these organisms. The capacity of IDPs to adapt their conformation to specifically recognize one or several partners, and the low to moderate affinity for partners make IDPs ideal for protein-protein interactions (Tompa et al. 2015). In fact, it has been shown that interactome hubs are enriched in this family of proteins (Dunker et al. 2005; Kim et al. 2008). Partner recognition is normally performed through conserved and partially structured motifs of the protein, and their individual properties can be modulated by post-translational modifications (PTM) or alternative splicing. IDRs are highly flexible regions connecting well-folded globular proteins forming the so-called multi-domain proteins. Multi-domain protein topology, which is highly prevalent in eukaryotes, enables the presence of multiple biological activities performed by the globular domains in close proximity (Hawkins and Lamb 1995; Levitt 2009). In many of these cases, IDRs behave as entropic linkers with an inherent plasticity that can be tuned depending on the length and the specific amino acid sequence of the region.

The biological relevance of IDPs has fostered their structural characterization (Eliezer 2009). Identification of the conformational preferences of binding motifs, the detection of transient long-range contacts within the chain, the structural perturbations exerted by PTM, the shape of biomolecular complexes with disordered partners, and the spatial distribution of globular domains in multi-domain protein are structural features that must be characterized to understand the molecular bases of biological function. This characterization is far from being trivial as the inherent disorder of IDPs/IDRs precludes their crystallization. Nuclear Magnetic Resonance (NMR) has become the only technique that can provide atomic-resolution information on IDPs (Dyson and Wright 2004). Novel NMR experiments and modelling strategies have been developed to interpret experimental parameters in terms of structure (Jensen et al. 2009, 2013, 2014; Wright and Dyson 2015). However, some structural aspects remain elusive by NMR such as the overall shape and size of disordered proteins or the distribution of interdomain position in multi-domain proteins.

Small-Angle Scattering of X-rays or Neutrons have emerged as powerful techniques to probe the structure and dynamics of biomolecules in solution at low resolution (Feigin and Svergun 1987; Svergun and Koch 2003; Koch et al. 2003; Putnam et al. 2007; Jacques and Trewhella 2010). SAS provides unique information about the overall size and shape of individual macromolecules or their complexes in a rapid manner. Additionally, structural changes upon environmental perturbations, such as interactions with other molecules, can be addressed straightforwardly. Major advances in instrumentation and computational methods in the last decade

have led to a tremendous increase in the applications of SAS in structural biology (Petoukhov and Svergun 2007; Mertens and Svergun 2010; Pérez and Nishino 2012; Rambo and Tainer 2013; Graewert and Svergun 2013). One of the major advances of SAXS in the last decade has been its extension to address biomolecular dynamics (Doniach 2001; Bernadó and Svergun 2012a, b; Receveur-Brechot and Durand 2012; Kikhney and Svergun 2015; Kachala et al. 2015). Although used in the past to study protein flexibility (Aslam et al. 2003), the availability of robust protocols to interpret SAS data in terms of ensembles of conformations have generalized these studies and, therefore, have enriched the spectrum of applications of the technique (Bernadó and Blackledge 2010).

The aim of this chapter is to provide an overview of the recent developments of SAS to study highly disordered proteins. A special emphasis is put on biological scenarios involving highly flexible proteins for which SAS is a crucial tool to tackle the structural/dynamic bases of biological function.

7.2 Scattering Properties of IDPs

The fact that IDPs sample an astronomical number of conformations has a strong impact on the scattering profiles measured and their comprehensive analysis in terms of structure. The experimental SAXS profile of an IDP corresponds to the average of all the conformations that the protein adopts in solution, inducing special features to the curves. Figure 7.1a displays the synthetic SAXS curves for seven conformations of $p15^{PAF}$, a 111 residue-long IDP, selected from a large pool of 5,000 (De Biasio et al. 2014). The individual conformations display several features along the complete momentum transfer range simulated. The initial part of the simulated curves, containing the lowest resolution structural information, presents distinct slopes indicating a large variety of possible sizes and shapes that an unstructured chain can adopt. The SAXS profile, obtained after averaging curves for the 5,000 conformations, presents a smoother behavior with essentially no features (Fig. 7.1b).

Traditionally, Kratky plots ($I(s)\cdot s^2$ as a function of s, where $I(s)$ is the scattering intensity and the momentum transfer s is defined as $s = 4\pi\sin(\theta)\,\lambda^{-1}$, 2θ is the scattering angle; and λ is the X-ray wavelength) have been used to qualitatively identify disordered states and distinguish them from globular particles. The scattering intensity of a globular protein behaves approximately as $1/s^4$ for s significantly greater than $1/R_g$, conferring a bell-shaped Kratky plot with a well-defined maximum. Conversely, 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 negligible thickness, $t << 1/s$, the Kratky plot 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. The Kratky representation has the capacity to enhance particular features of scattering profiles that allows an easier identification of different degrees of compactness (Doniach 2001). This is shown in Fig. 7.1c where different degrees of compactness for the conformations are observed. Multi-domain proteins, which are composed by two or more globular domains connected by intrinsically disordered linkers, present a dual (folded/disordered) behavior, and consequently SAXS profiles and Kratky plots display contributions from both structurally distinct regions. Pair-wise distance distributions, $p(r)$, derived from disordered proteins also present specific properties (Fig. 7.1d). The most characteristic feature is the smooth decrease towards large intramolecular distances. Maximum intramolecular distances, D_{max}, which represent the maximum distance within one of the accessible conformations of the protein, are very large in disordered proteins. It is worth noting that due to the low population of highly extended conformations in the ensembles, experimental D_{max} values are systematically underestimated (Bernadó 2010).

Unstructured proteins, due to the presence of extended conformations, are characterized by large average sizes compared to globular

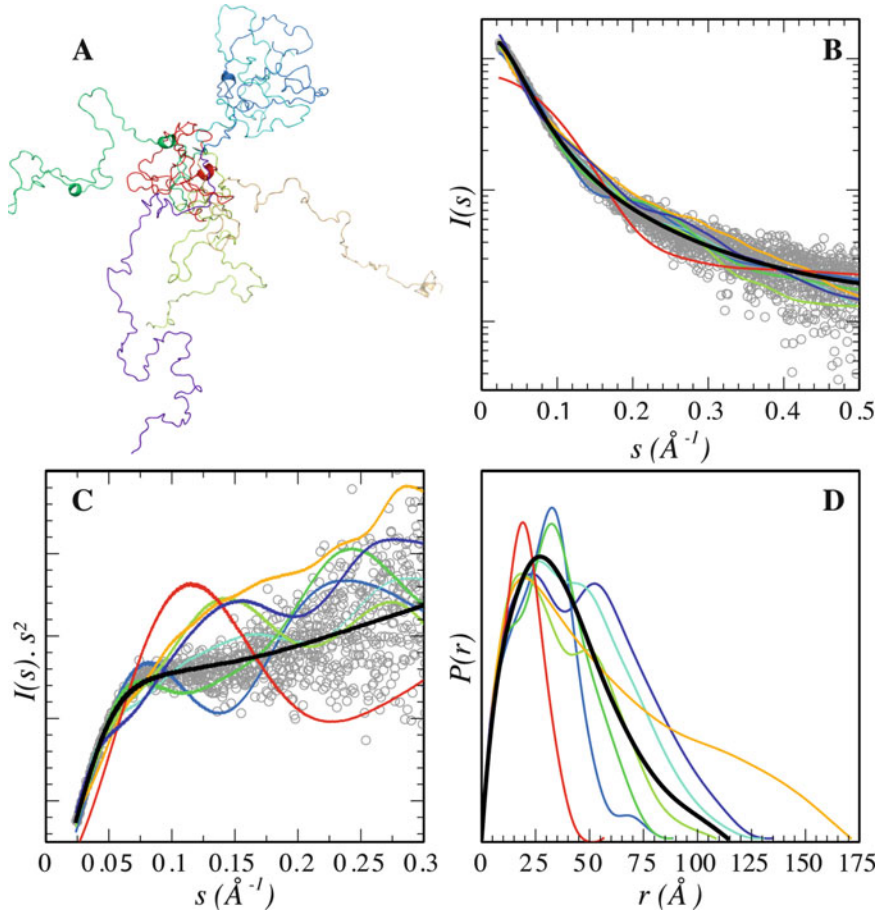

Fig. 7.1 (**a**) Seven representative conformers randomly selected from an ensemble of 5000 explicit all-atoms models generate for $p15^{PAF}$ (De Biasio et al. 2014). *Solid lines* correspond to their computed SAXS *curves* (**b**) and Kratky plots (**c**) and are *colored* as in planel *A*. *T*. The average over the ensemble of 5,000 conformations yields a featureless curve that is in very good agreement with the experimental data (*gray circles*). (**d**) p (r) functions computed for the seven conformers and the complete ensembles in the same color code as in panels (**a–c**)

proteins. The radius of gyration, R_g, which can be directly obtained from a SAXS curve using a classical Guinier approximation, is the most common descriptor to quantify the overall size of molecules in solution (Guinier 1939). In the Guinier approximation, it is assumed that at very small angles ($s < 1.3/R_g$) the scattering intensity can be represented as $I(s) = I(0) \exp(-(sR_g)^{2/3})$, and the R_g is obtained by a simple linear fit in logarithmic scale. Debye's equation (Eq. 1), which describes the scattering from an ensemble of monodisperse random coils, can be more precise than Guinier's approximation to derive R_g values as its validity extends to larger momentum transfer ranges (Calmettes et al. 1994).

$$\frac{I(s)}{I(0)} = \frac{2}{x^2}(x - 1 + e^{-x}); x = s^2 \mathrm{Rg}^2 \qquad (7.1)$$

Alternatively, the *p(r)* function calculated from the complete scattering profile using a Fourier transformation also yields precise R_g values for disordered proteins.

The experimental R_g is a single value representation of the size of the molecule, which for

disordered states represents a *z*-average over all accessible conformations in solution (Feigin and Svergun 1987). The most common quantitative interpretation of R_g for unfolded proteins, which is based on Flory's studies in polymer science, relates this parameter to the length of the protein chain through a power law (Flory 1953),

$$R_g = R_0 \cdot N^{\nu} \tag{7.2}$$

where N is the number of residues in the polymer chain, R_0 is a constant that depends on several factors, in particular, on the persistence length, and ν is an scaling exponent. For an excluded-volume polymer, Flory estimated ν to be ≈ 0.6, and more accurate theoretical estimates established a value of 0.588 (LeGuillou and Zinn-Justin 1977). 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, and a R_0 value of 1.927 ± 0.27 (Kohn et al. 2004). The agreement between the ν value obtained experimentally and the theoretical models suggest the random coil nature of the chemically denatured proteins. However, the question whether the conformational sampling in the chemically denatured state is equivalent to that found for IDPs in native conditions must be clarified (Stumpe and Grubmüller 2007 and references therein). Using atomistic ensemble models of several disordered proteins, Flory's equation has been parametrized as a function of the number of residues of the IDP, *N* (Bernadó and Blackledge 2009):

$$R_g = (2.54 \pm 0.01) \cdot N^{(0.522 \pm 0.01)} \tag{7.3}$$

The exponential value obtained from the parametrization, $\nu = 0.522 \pm 0.01$, is notably smaller than that derived from the dataset of denatured proteins, $\nu = 0.598 \pm 0.028$, indicating that IDPs are more compact than chemically denatured proteins. This observation is in line with NMR studies that indicated that urea denatured proteins have an enhanced sampling (around 15%) of extended conformations compared with IDPs (Meier et al. 2007).

Table 7.1 compiles R_g data from 74 IDPs from the literature, which are plotted as a function of the chain length in Fig. 7.2. As expected, the R_gs collected display a correlation with the number of residues of the chain. This linear relationship is closer to the above-mentioned parametrization of IDPs than to that established for chemically denatured proteins. As some IDPs are expected to have certain populations of secondary or tertiary structure, this relationship can be used as an interpretative tool. Thus, deviations from the average IDP random coil model indicate enhanced degrees of compactness or extendedness within the protein.

7.3 Molecular Modelling of Intrinsically Disordered Proteins

The interpretation of the SAXS parameters such as R_g, *p(r)* and D_{max} from disordered proteins in terms of structure is limited to overall molecular information. In order to fully exploit the structural and dynamic information encoded in SAXS data, the use of realistic three-dimensional models is necessary. However, the generation of conformational ensembles of disordered proteins is extremely challenging (Zhou 2004). IDPs present a relatively flat (non-funnelled) energy landscape, with an extremely large number of local minima separated by low-energy barriers. This, combined with their large size, makes the analysis of their energy landscape a challenging problem for computational methods. Most of the available computational methods aim at collecting an ensemble representation of IDPs (Bernadó et al. 2007; Jensen et al. 2014; Wright and Dyson 2015). This requires an extensive and statistically correct exploration of the conformational space to obtain a representative set of states. Three main families of approaches have been proposed to generate conformational ensembles: molecular dynamics (MD) simulations, Monte Carlo (MC) methods, and experimentally parametrized statistical approaches. These three families of methods are succinctly explained next.

Table 7.1 SAXS studies performed on IDPs

Protein	#Residues[a]	R_g^{exp} (Å)	R_g^{RC} (Å)[b]	References
CyaA toxin (1006–1707)	702	84	77.7	O'Brien et al. (2015)
MeCP2	486	62.5 ± 4.5	64.2	Yang et al. (2011)
Map2C	467	83 ± 1[d]	62.8	Borysik et al. (2015)
Rec1-resilin	310	43.4 ± 0.8	50.7	Balu et al. (2015)
Msh6 N-term	304	56 ± 2	50.2	Shell et al. (2007)
Ki-1/57	292	47.5 ± 1	49.2	Bressan et al. (2008)
Juxtanodin (1–282)	282	55.9	48.3	Ruskamo et al. (2012)
Dehydrin ERD10 (2–260)	259	60 ± 1.0[d]	46.2	Borysik et al. (2015)
MeCP2 (78–305)	228	37.0 ± 0.9	43.3	Yang et al. (2011)
PGC-1α 220 (2–220)	219	61.3	42.3	Devarakonda et al. (2011)
Synthetic Resilin	185	50 ± 5	38.8	Nairn et al. (2008)
CFTR (654–838)	185	32.5 ± 1.8	38.8	Marasini et al. (2013)
pCFTR (654–838)	185	29.1 ± 1.8	38.8	Marasini et al. (2013)
Dehydrin ERD14 (2–185)	184	60 ± 1.0[d]	38.6	Borysik et al. (2015)
Juxtanodin (1–170)	170	42.9	37.1	Ruskamo et al. (2012)
Myelin basic protein	170	33	37.1	Stadler et al. (2014)
AavLEA1	163	41 ± 1[d]	36.3	Borysik et al. (2015)
SRC-1 (617–769)	153	33.96	35.1	Pavlin et al. (2014)
Osteopontin	150	37.90 ± 0.08	34.7	Lenton et al. (2015)
Pig Calpastatin domain I	148	35.4	34.5	Konno et al. (1997)
HrpO	147	35.0	34.4	Gazi et al. (2008)
Ii-1	141[c]	41.0 ± 1	33.6	Boze et al. (2010)
α-Synuclein, pH 7.5	140	40 ± 1	33.5	Li et al. (2002)
α-Synuclein, pH 3.0	140	30 ± 1	33.5	Li et al. (2002)
N-tail nucleoprotein MV	139	27.2 ± 0.5	33.4	Longhi et al. (2003)
β-synuclein	137	49 ± 1	33.1	Uversky et al. (2002)
Human NHE1 cdt (5 °C)	131	37.1[d]	32.4	Nørholm et al. (2011)
Human NHE1 cdt (45 °C)	131	35.3[d]	32.4	Kjaergaard et al. (2010)
ERM transactivation domain	130	39.6 ± 0.7	32.2	Lens et al. (2010)
Human NL3 (731–848)	118	31.5 ± 1.0	30.6	Paz et al. (2008)
elF4E binding protein (4E–BP)	117	48.8 ± 0.2	30.5	Gosselin et al. (2011)
Juxtanodin (172–282)	111	31.5	29.7	Ruskamo et al. (2012)
$p15^{PAF}$(2–111)	110	28.1 ± 0.3	29.5	De Biasio et al. (2014)
Prothymosin α, pH 7.5	109	37.8 ± 0.9	29.4	Uversky et al. (1999)
Prothymosin α, pH 2.5	109	27.6 ± 0.9	29.4	Uversky et al. (1999)
*pa*NHE1 cdt (5 °C)	107	32.8[d]	29.1	Nørholm et al. (2011)
*pa*NHE1 cdt (45 °C)	107	32.9[d]	29.1	Kjaergaard et al. (2010)
N-protein of bacteriophage λ	107	33 ± 2[e]	29.1	Johansen et al. (2011a)
N-protein of bacteriophage λ	107	38 ± 3.5	29.1	Johansen et al. (2011b)
Human NCBD domain	105	33 ± 1	28.8	Borysik et al. (2015)
FEZ1 monomer	103	36 ± 1	28.5	Alborghetti et al. (2010)
HIV-1 Tat_{133}	101	33.0 ± 1.5	28.3	Foucault et al. (2010)
Human Calpastatin (137–237)	100	39.0 ± 1.5	28.1	Borysik et al. (2015)
p53 (1–93)	93	28.7 ± 0.3	27.1	Wells et al. (2008)
Sic1	92	34.7	26.9	Mittag et al. (2010)
pSic1 (hexaphosphorylated)	92	34.0	26.9	Mittag et al. (2010)
Juxtanodium (103–282)	79	37.4	38.1	Ruskamo et al. (2012)
PIR domain	75	26.5 ± 0.5	24.2	Moncoq et al. (2004)
N-term NRG1 type III	75	26.8[d]	24.2	Chukhlieb et al. (2015)

(continued)

Table 7.1 (continued)

Protein	#Residues[a]	R_g^{exp} (Å)	R_g^{RC} (Å)[b]	References
IB5	73[b]	27.9 ± 1.0	23.8	Boze et al. (2010)
ACTR (5 °C)	71	25.8[d]	23.5	Kjaergaard et al. (2010)
ACTR (45 °C)	71	23.8[d]	23.5	Kjaergaard et al. (2010)
PaaA2 (1–63)	70 [a]	22.15 ± 0.87[d]	23.3	Sterckx et al. (2014)
N-term VS virus phosphoprotein	68	26 ± 1[f]	23.0	Leyrat et al. (2011)
E3 ubiquitin ligase RNF4 (32–82)	57	25.8	21.0	Kung et al. (2014)
Histatin 5	24	13.3	13.3	Cragnell et al. (2016)
R/S peptide	24	12.6 ± 0.1	13.3	Rauscher et al. (2015)
Constructions of Tau protein				
Tau ht40	441	65 ± 3	61.0	Mylonas et al. (2008)
Tau K32	202	42 ± 3	40.6	Mylonas et al. (2008)
Tau K16	174	39 ± 3	37.5	Mylonas et al. (2008)
Tau K18	130	38 ± 3	32.2	Mylonas et al. (2008)
Tau ht23	352	53 ± 3	54.2	Mylonas et al. (2008)
Tau K27	171	37 ± 2	37.2	Mylonas et al. (2008)
Tau K17	143	36 ± 2	33.9	Mylonas et al. (2008)
Tau K19	99	35 ± 1	28.0	Mylonas et al. (2008)
Tau K44	283	52 ± 2	48.4	Mylonas et al. (2008)
Tau K10	167	40 ± 1	36.7	Mylonas et al. (2008)
Tau K25	185	41 ± 2	38.7	Mylonas et al. (2008)
Tau K23	254	49 ± 2	45.7	Mylonas et al. (2008)
Tau K32 AT8 AT100	202	41 ± 3	40.6	Mylonas et al. (2008)
Tau ht23 S214E	352	54 ± 3	54.2	Mylonas et al. (2008)
Tau ht23 AT8 AT100	352	52 ± 3	54.2	Mylonas et al. (2008)
Tau K18 P301L	130	35 ± 2	32.2	Mylonas et al. (2008)
Tau ht40 AT8 AT100 PHF1 (10 °C)	441	66 ± 3	61.0	Shkumatov et al. (2011)
Tau ht40 AT8 AT100 PHF1 (50 °C)	441	67 ± 3	61.0	Shkumatov et al. (2011)

[a]When present, purification tags or extra terminal residues resulting from cloning were considered as part of the protein
[b]Threshold R_g value obtained from the parametrization of Flory's relationship with the coil database
[c]Length of the most populated isoform of the samples was used
[d]R_g derived from averaging conformations selected with EOM
[e]Data measured by SANS in highly crowded conditions (130 mg/ml of BPTI)
[f]Data derived from the 10 mM Arg/Glu buffer

MD simulations analyze the evolution of the system under study by solving Newton's equations of motion (Karplus and McCammon 2002; Piana et al. 2014). Theoretically, MD is a suitable method to correctly sample the conformational space of IDPs. Nevertheless, in practice, the high-dimensionality and the wideness of the energy landscape hampers its exhaustive exploration. Several approaches have been proposed to enhance conformational exploration with MD methods. A particularly effective one is Replica Exchange MD (REMD) that runs multiple simulations in parallel with different settings (usually different temperatures) and exchanges states between these processes (Trakawa and Takada 2011; Zerze et al. 2015; Chebaro et al. 2015). Going further in this direction, a very recent method called Multiscale Enhanced Sampling (MSES) couples temperature replica exchange and Hamiltonian replica exchange, using a coarse-grained model to guide atomistic conformational sampling (Lee and Chen 2016). The performance of MD-based methods can also be improved by the integration of experimental data to restrain the exploration of the most relevant regions of the conformational space (Lindorff-Larsen et al. 2004; Dedmon et al. 2005; Wu et al. 2009).

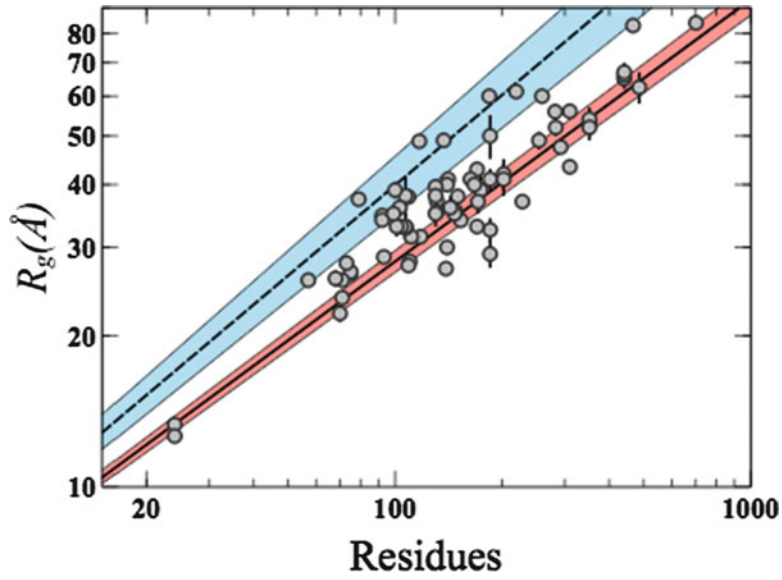

Fig. 7.2 R_g values from Table 7.1 (*gray dots*) as a function of the number of residues are plotted in Log-Log scale. *Straight lines* correspond to Flory's relationships parametrized for denatured proteins (*blue-dashed*) (Kohn et al. 2004) and IDPs (*red-solid*) (Bernadó and Blackledge 2009). *Colored bands* correspond to uncertainty of the parametrization for both models. Some IDPs are not fully disordered and are globally more extended or more compact than expected for a random coil. These structural features even if transient affect the experimental R_g

MC methods are a classical alternative to MD, the Markov chain Metropolis scheme (Metropolis et al. 1953) being the most widely used sampling technique (Vitalis and Pappu 2009a). The system is randomly perturbed and the new conformation is accepted with a probability that depends on the energy change between the new conformation and the previous one. Particular mention deserves a recently proposed variant called Hamiltonian Switch Metropolis Monte Carlo (HS-MMC), which has been specially conceived to study IDRs tethered to globular domains. Proteins including IDRs present energy minima due to the contact of the disordered and ordered regions. To avoid being trapped in such minima, the HS-MMC switches between an all-atom Hamiltonian to an excluded volume Hamiltonian to push the IDR away from the ordered domain.

Both, MD-based and MC-based approaches may suffer from inaccuracies of current energy models, which are better suited to globular proteins, and tend to provide structurally biased ensembles that do not properly reflect the conformational behaviour in solution of unstructured proteins (Best et al. 2014; Henriques et al. 2015). The development of more suitable forcefields and solvation models for IDPs are key issues for a correct performance of computational methods (Vitalis and Pappu 2009b; Emperador et al. 2015).

Knowledge-based statistical approaches are an alternative to physics-based energy functions. The most representative knowledge-based method for the generation of atomistic models of disordered proteins is Flexible-Meccano (FM) (Bernadó et al. 2005; Ozenne et al. 2012), although other similar methods have been described (Jha et al. 2005). The FM algorithm uses an amino acid-specific statistical coil derived from crystallographic structures. In FM, each conformation is built by assembling peptide plane units in a consecutive manner using a residue-specific coil library derived from crystallographic structures. To avoid the collapse of the chain, a coarse-grained description of side chains

is also used. Based on this set of conformations, experimentally measurable NMR parameters and SAXS curves can be estimated, which has permitted the validation of the resulting models (see below).

Despite the efforts that have been made to precisely describe the conformational states of disordered proteins, there are still many technical and conceptual issues that must be addressed to correctly describe their energy landscape and, as a consequence, their associated experimental observables.

7.4 Ensemble Approaches

IDPs sample a large number of conformations. Therefore, ensembles of conformations are the most appropriate framework to structurally represent this family of proteins. In recent years, structural biologists have addressed the challenge of describing dynamic systems in terms of ensembles of reliable conformations guided by experimental data that represents average values for the complete ensemble of conformations (Bernadó and Blackledge 2010). SAXS has not been exempt from this tendency and several approaches have been developed to characterize protein mobility: Ensemble Optimization Method (EOM) (Bernadó et al. 2007; Tria et al. 2015), Minimal Ensemble Search (MES) (Pelikan et al. 2009), Basis-Set Supported SAXS (BSS-SAXS) (Yang et al. 2010), Maximum Occurrence (MAX-Occ) (Bertini et al. 2010), Ensemble Refinement of SAXS (EROS) (Rozycki et al. 2011), Broad Ensemble Generator with Re-weighting (BEGR) (Daughdrill et al. 2012), and Bayesian Ensemble SAXS (BE-SAXS) (Antonov et al. 2016). These methods are based on a common strategy that consists of three consecutive steps: (i) computational generation of a large ensemble describing the conformational landscape available to the protein, (ii) computation of the theoretical SAXS curves from the individual conformations, and (iii) selection of a subensemble of conformations that collectively describes the experimental profile using multiparametric optimization methods. Despite the common philosophy, these programs present distinct features in the three steps. Readers are referred to the original articles for a detailed description of the approaches.

The availability of ensemble methods has revolutionized the study of flexible proteins by SAS. Ensemble methods provide a description in terms of statistical distributions of structural parameters or conformations that represents a crucial step forward with respect to traditional analysis based on averaged parameters extracted from raw data such as R_g or D_{max}. In that context, conformational perturbations exerted by temperature (Shkumatov et al. 2011; Kjaergaard et al. 2010), buffer composition (Leyrat et al. 2011), or mutations (Stott et al. 2010) can be monitored in terms of ensembles of accessible conformations.

The main approximation of ensemble methods is the discrete description of entities that probe an astronomical number of conformations. It is therefore reasonable to argue about the real meaning of the SAXS-derived ensembles. An additional problem is the statistically significant size of the derived ensembles based on experimental data with a very limited amount of information (Hammel 2012; Yang 2014). The described strategies use distinct philosophies to address these issues. In some cases such as EOM 2.0 and MES, programs search for the minimal number of conformations required to describe the data to limit or abolish over-fitting. BSS-SAXS and BE-SAXS use Bayesian statistics to address the confidence of the derived populations. In other cases, such as EROS and BEGR, populations of the conformers of an initially built ensemble are slightly modified (re-weighted) in order to describe the data.

SAXS-derived ensembles are representations of the conformational landscape sampled by proteins in solution, but not necessarily the exact states. Ensemble approaches are inherently ill-defined problems, and this is especially severe in SAXS that codes for a very limited amount of structural information. In that sense, it is more adequate to represent highly flexible proteins as distributions of accessible structural parameters

such as R_g or D_{max}. These representations are less prone to over-fitting artifacts (Bernadó et al. 2007).

In disordered proteins, SAS reports on the overall size and shape of the protein in solution. The presence of extendedness or compactness can be probed by SAS, but regions causing these structural biases can not be identified unambiguously due to the low-resolution nature of the data. An interesting way to enrich the structural content of SAXS data to more precisely localize partially structured regions in IDPs has been proposed. In this strategy, SAXS curves are measured for multiple deletion mutants of the disordered chain, and simultaneously fitted in terms of a common ensemble. Note that this strategy is only valid if the structural elements of the full-length protein remain intact in the deletion mutants. This approach is described in detail in the original article of EOM where it was applied to a flexible multi-domain kinase (Bernadó et al. 2007). The most notable example of this approach was the study of two isoforms of Alzheimer-related tau, ht23 and ht40 (Mylonas et al. 2008). SAXS data for full-length ht23 and ht40 and for five and three deletion mutants for each isoform were measured, respectively. The simultaneous fit of all curves with EOM unambiguously identified the so-called repeat region as the source of residual secondary structure in tau. For the ht23 isoform, with three repeats, the maximum separation is found within the repeat domain itself. Conversely, the ht40 isoform, with four repeats, presented 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.

In addition to the previously described strategy, the information content of SAXS curves measured for IDPs can be enriched with data from other techniques. NMR is by far the most common technique used synergistically with SAXS (Sibille and Bernadó 2012), and a special section of this chapter is devoted to it. Additionally, experimental data coming from Electron Paramagnetic Resonance (EPR) and single molecule Fluorescence Resonance Energy Transfer (smFRET) have been synergistically used with SAXS to characterize disordered proteins (Boura et al. 2012). In that sense, the development of robust ways to integrate other biophysical measurements in ensemble approaches is an unavoidable future direction.

7.5 Application of SANS to Study IDPs

The physical bases and the structural information that can be derived form SAXS and SANS are equivalent and, in principle, both techniques can be used to characterize biomolecules. The more general use of SAXS is based on the higher beam intensity and more widespread availability of X-ray beam-lines. SANS however offers some advantages with respect to SAXS. The first one is the absence of radiation damage that makes it a non-destructive technique. The second one is based on the possibility to perform contrast-match experiments, where some components of the sample are made invisible by finely tuning the degree of deuteration of sample components and the ratio of H_2O/D_2O of the buffer. Contrast matching has been widely used in structural biology and the reader is refereed to the excellent reviews available on the subject (Jacrot 1976; Heller 2010; Gabel 2012).

Although SANS has been used to study IDPs or IDRs (Krueger et al. 2011) or radiation sensitive systems (Greving et al. 2010; Stanley et al. 2011; Perevozchikova et al. 2014), its main application is in contrast-match experiments. An example of this family of experiments is the study of the interaction of β-amyloid (1–40) with the detergent SDS (Jeng et al. 2006). In this study the use of deuterated SDS in SANS experiments enabled the characterization of the peptide in the presence of SDS and showed that β-amyloid adopts a short rod-like shape. Interestingly the authors observed that SDS suppress fibrils by forming complexes with a 30:1 molar ratio between detergent and protein.

IDPs may be particularly sensitive to the effects of molecular crowding, and contrast-

matching SANS experiments is a powerful tool to study these effects. By choosing appropriate levels of deuteration of the protein of interest and the crowding agent, the scattering contribution of the latter can be made negligible, enabling the study of the structural perturbations exerted by macromolecular crowding on disordered proteins. This strategy was used to study the disordered N-protein of bacteriophage λ in the presence of high concentrations of bovine pancreatic trypsin inhibitor (BPTI), a small globular protein (Johansen et al. 2011a; Goldenberg and Argyle 2014). In 46% D_2O, high concentrations of non-deuterated BPTI were contrast matched, and only the signal of the 85% deuterated N-protein was visible. The study demonstrated that molecular crowding exerted a compaction effect on N-protein. However, this effect was non-linear and the effects observed at 65 mg/mL of BPTI were equivalent to those at 130 mg/mL.

7.6 SAXS Studies of IDPs Within Macromolecular Assemblies

Given the structural plasticity of disordered proteins, they are regarded as interacting specialists and have a special place in cell signalling using short motifs or domain-sized disordered segments for partner recognition (Wright and Dyson 2015; Tompa et al. 2015). Providing a comprehensive description of the biomolecular recognition processes for IDPs is thus of great importance for the understanding of key biological functions that are orchestrated by this family of proteins. In this regard, SAXS is emerging as an extremely valuable tool for characterizing biomolecular interactions involving highly flexible proteins, which are highly challenging for crystallography. The application of solution NMR also encounters severe limitations when characterizing large biomolecular complexes. Conversely, SAXS, which is not limited by size, offers a source of structural and dynamic information that by itself or combined with NMR and/or X-ray crystallography is a promising alternative for the structural characterization of highly dynamic proteins and complexes in solution.

The binding of an IDP to its target produces specific SAXS signatures that enable the detection of the interaction. Typically, the mixing of two interacting partners will result in a rise of R_g values, otherwise, if the proteins do not form a complex the R_g value will follow an population-weighted average of the two. Insightful information can be obtained by inspecting the *p(r)* function in the absence and presence of the disordered interaction partner. The scattering of globular proteins generally give a symmetrical bell-shaped *p(r)* function, while the interaction with a sufficiently large IDP results in tailing of the peak shape to higher values of *r,* culminating at a large D_{max}. The Kratky and Porod-Debye analyses are also powerful indicators of flexibility within macromolecular assemblies (Rambo and Tainer 2011).

It is possible to define 3D molecular envelopes describing the low-resolution shapes of flexible macromolecular assemblies involving IDPs (Longhi et al. 2003; Marsh et al. 2010; Gosselin et al. 2011; Devarakonda et al. 2011). However, the resulting fixed low-resolution structure is not the most appropriate structural description of a disordered protein. Ensemble analysis of explicit or coarse-grained models should provide a more accurate characterization of flexible macromolecular assemblies, particularly if combined with available structure coordinates of folded-domains and data affordable by NMR. Moreover, weak or moderate affinities are a hallmark of many IDP molecular recognition events, which entails the presence of multiple species in solution at standard experimental concentrations. In these cases, any modeling strategy envisioned should account for this species polydispersity to reliably describe SAS data.

Several SAXS studies have been devoted to the interactions of different IDPs with other proteins or DNA, as the only source of structural information or in combination with high-resolution methods. Some examples will be presented.

The complex of Msh2 and Msh6 recognizes mismatched bases in DNA during mismatch repair. The N-terminal region of Msh6, a 304 residue long IDR, recognizes PCNA, a protein that controls processivity of DNA polymerases. Shell and co-workers have demonstrated this direct interaction with SAXS (Shell et al. 2007). 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 complex that could be considered as a highly flexible dumbbell where both globular domains are tethered by the N-terminal Msh6 fragment that acts as a molecular leash. These observations were further confirmed in an additional experiment with a biologically active deletion mutant of Msh6 containing a notably shorter N-terminal tail. Under these conditions the important size changes upon binding were easily monitored using the *p(r)* and D_{max}.

The tumor suppressor p53 is a multifunctional protein that plays a crucial role in processes like apoptosis control and DNA repair. P53 is a homotetramer with two folded domains that are tethered and flanked by unstructured regions. Rigid-body modeling of SAXS data measured for p53 suggests that the protein is a rather open cross-like tetrameric assembly, which collapses to tightly embrace DNA (Tidow et al. 2007). This is how the flexibility of IDPs helps the protein to fulfill its function.

Nuclear receptors (NR) are engaged in gene transcription regulation in response to binding of specific ligands. Signal transduction from ligand binding to gene expression requires the recruitment of co-regulator proteins. Most NR co-regulators function as flexible scaffolds for chromatin modifiers and transcriptional machinery (Millard et al. 2013). Structural analyses of their interaction have been restricted to small peptides of the regulators and the nuclear receptor ligand-binding domain. More recently, several SAXS studies have provided new insights into the stoichiometry and binding mode of the complexes formed by NR and disordered co-regulators (Jin et al. 2008; Rochel et al. 2011; Devarakonda et al. 2011; Pavlin et al. 2014).

β-thymosin/WH2 (β-t/WH2) domains are widespread short disordered regions (25–50 residues) able to recognize G-actin and regulate actin-assembly. When bound to G-actin the N-terminal half of β-t/WH2 adopts a well-ordered amphipathic helix. SAXS analysis at different ionic strengths revealed that the C-terminal regions of different β-t/WH2 domains display distinct dynamics, which correlate with functional differences. At low ionic strength β-t/WH2 sequesters G-actin in a polymerization incompetent state, where the dynamic interactions of the C-terminal part are restrained to a single conformational state. This SAXS-driven observation prompted the study of the β-t/WH2:G-actin complex by NMR at different ionic strengths revealing that a single intermolecular salt-bridge controls the assembly (Didry et al. 2012).

7.7 Low-Complexity Regions

Low-complexity regions (LCRs) are unusually simple protein sequences with a strong amino acid composition bias, and include homo-repeats of a single amino acid, short period repeats or aperiodic mosaics of a few residues (Wootton and Federhen 1993). Protein sequences from all three kingdoms of life contain LCRs, but they are more common in eukaryotes, with *Plasmodium falciparum* being an extreme case as ~90% of its proteins host LCRs (Marcotte et al. 1999; Pizzi and Frontali 2001). Despite their high abundance and doubtless biological relevance, not many studies are focused on the structural characterization of LCRs mainly due to the technical challenges they pose. LCRs are often inserted in IDRs precluding their crystallization (Haerty and

Golding 2010), and the NMR sequence assignment is complicated (or impossible) by the similarity of nuclear chemical environments. In that context, SAS is a powerful technique to investigate the structure and dynamics of this elusive family of proteins.

Prothymosin α was the first and probably the most well characterized LCR-hosting protein (Gast et al. 1995). Roughly half of the 109 residues of Prothymosin α are acidic (Asp and Glu), leading to a net charge of −44 at neutral pH. A SAXS analysis of prothymosin α at near neutral (7.5) and acidic (2.5) pH showed that while it was unstructured under both conditions, a dramatic reduction in R_g (from 37.8 ± 0.9 Å to 27.6 ± 0.8 Å) could be observed (Uversky et al. 1999). The pH-induced reduction in protein size can be explained by the neutralization of the negatively charged acidic residues due to the decrease in pH. A similar reduction of size was observed after the addition of 15 mM Zn^{2+} at neutral pH (R_g = 28.1 ± 0.8 Å) suggesting the electrostatic screening of the negative charges by cations (Uversky et al. 2000). A similar situation was observed for the basic proline-rich salivary proteins IB5 and Il-1. Proline-rich salivary proteins bind polyphenolic plant compounds (e.g. tannins) and form aggregates upon binding high concentrations of these compounds. At sequence level IB5 and Il-1 contain repetitions of Pro, Gly and Gln or Glu residues, and they are predicted to be disordered. Boze et al. used SAXS to study the conformations of IB5 and Il-1, and to see if there are any functional advantages linked to the respective conformations (Boze et al. 2010). Both IB5 and Il-1 showed an R_g larger than the theoretical one for IDPs of their lengths, R_g = 27.9 Å ± 1.0 Å and 41.0 Å ± 1.0 Å, respectively. In addition, the experimental D_{max} (110 ± 10 Å and 155 Å ± 10 Å, respectively) also indicated the presence of highly extended conformations that could contain poly-proline II helices.

Another example for a protein with low sequence complexity is the crosslinked elastomeric protein resilin. Resilin is rich in Gly, Ser and Pro residues and is present in most insects where it is critical for flight and jumping. Due to its low stiffness, high fatigue lifetime and high resilience, resilin has been of great interest for the production of biomaterials for biomedical applications. SAXS measurements on a recombinant resilin (rec1-resilin) comprising 18 copies of the N-terminal repeat motif (GGRPSDSYGAPGGGN), yielded an R_g of 43.4 ± 0.8 Å and a D_{max} of 200 Å (Balu et al. 2015). Since the expected R_g of a 310 residue-long protein with a compact structure would be ~19.6 Å and that of an IDP would be ~59.5 Å, these data suggest that rec1-resilin is largely unfolded but not completely disordered.

Homo-repeat proteins represent extreme cases for the structural characterization, and huntingtin protein (htt) is the prototypical example of this family of proteins. Htt exon 1 contains two homo-repeat regions consisting of poly-Gln and poly-Pro, respectively. SAXS has been used for the characterization of the overall properties of this protein. Kratky plots of htt exon1 constructs fused to thioredoxin showed broad peaks with less decrease at higher scattering angles, consistent with flexible or unfolded proteins. Interestingly, the observed radii of gyration for constructs with 16 Gln or 39 Gln were very similar, R_g = 49 Å or 52 Å, respectively (Owens et al. 2015). These data were consistent with previous experimental observations and MD simulations indicating that poly-Gln tracts form disordered, collapsed globular structures in solution (Vitalis et al. 2008; Dougan et al. 2009).

An example for a protein with low sequence complexity that forms higher order structures is silk fibroin. The heavy chain of silk fibroin from *Bombyx mori* is 5,263 amino acids long and dominated (94%) by the repetition of Gly-X repeats where X is mainly Ala (65%), Ser (23%), or Tyr (9%) (Zhou et al. 2001). In a pioneering study, Greving et al. characterized native (SF) and reconstituted silk fibroin (RSF) by SANS (Greving et al. 2010). Their study identified significant differences between the molecular weights and R_gs of native and reconstituted silk fibroin, as well as for RSF samples prepared under different conditions.

7.8 Application of SAS to Aggregating IDPs

Through a complex self-assembly process, some IDPs form amyloid fibrils that are the hallmark of disorders such as Alzheimer's, Parkinson's or diabetes (Chiti and Dobson 2006). Fibrillation is a very complex process that involves multiple oligomeric species that are transformed following intricate pathways. Interestingly, there are indications that soluble oligomers, often precursors of fibrils, are the main cause of cytotoxicity and neuronal damage. The structural analysis of the distinct species involved in fibrillation is a major challenge due to their instability, low relative concentration, difficulties of isolation, and the equilibrium between species of very different sizes, present at any time point during the fibrillation process.

The aggregation process of huntingtin exon 1 (htt, see above) has been followed by SANS (Stanley et al. 2011). For this study, the aggregation of a pathological truncated version of the protein encompassing 42 consecutive Gln residues ($NtQ_{42}P_{10}$) was followed in a time-dependent manner. The study indicated that the aggregation begins with dimeric and trimeric forms of $NtQ_{42}P_{10}$. In a very fast process, large oligomeric species, estimated to be 14-mers, are formed, and after that point the protein becomes predominately fibrillar. The same group addressed the effects of the Gln tract length on the fibrillation mechanism by investigating a non-pathological htt version ($NtQ_{22}P_{10}$) also using SANS (Perevozchikova et al. 2014). Their results indicate that the length of the homo-repeat dictates the size of the initial species and the aggregation pathway followed by htt. In these studies, however, the complex nature of the samples, which are composed by multiple oligomeric species, is not taken into account, and the average values of the derived parameters are analysed. The quantitative interpretation of SAS data of fibrillating proteins requires their decomposition into species-pure SAS profiles, reporting on the structure of each species present, and their relative concentration, reporting on the kinetics of the process. In a pioneering study this challenging SAXS data decomposition was successfully achieved (Giehm et al. 2011). In this study, the fibrillation of α-synuclein was monitored using SAXS by measuring profiles every 30 min during 24 h. Measured profiles contain contributions from all species present at each time-point. Singular Value Decomposition (SVD) analysis indicated that three species are enough to describe the complete SAXS dataset, which were assigned to monomer/dimer, mature fibril, and a third species of unknown nature. Using SAXS profiles measured separately for the two extreme components (monomer/dimer and mature fibril) as fixed contributions, the SAXS curve for the third component was derived (Vestergaard et al. 2007). This decomposed species turned out to be a building block for fibril formation. From the analysis of the time-dependent relative populations, the kinetics of the aggregation was also derived. Despite the novelty of the study, it seems clear that more robust and objective approaches to decompose complex SAXS data from aggregating proteins are necessary.

7.9 Combined Use of SAXS and Nuclear Magnetic Resonance

NMR is the only technique that can provide atomic-resolution information of IDPs (Dyson and Wright 2004). The first step to study an IDP by NMR is to assign a resonance frequency to all magnetically active nuclei (^{1}H, ^{15}N, and ^{13}C) of the protein. Due to very low amide proton dispersion, assignment of IDP spectra is challenging. However, the use of high magnetic field spectrometers and several methodological developments allow to routinely assign NMR frequencies (Narayanan et al. 2010). NMR is a highly versatile technique and multiple observables reporting on protein structure and dynamics can be measured. In the following paragraphs the most relevant ones will be succinctly described.

- **Chemical shifts (CSs):** CSs correspond to the resonance frequencies of the nuclei, and are the primary information that can be derived from a NMR experiment. CSs are very sensitive to the chemical environment of nuclei and reveal the presence of secondary structural elements. A chemical shift index (CSI) has been established to highlight regions that deviate from purely random coil to form secondary structural elements (Wishart and Sykes 1994; Wishart et al. 1995). With the discovery of IDPs, the interest in using CSs to detect partially structured elements has been renewed, and several databases have appeared based on small synthetic peptides (Schwarzinger et al. 2000; Kjaergaard et al. 2011) or IDPs (Tamiola et al. 2010) to identify these regions.
- **Scalar $^{3}J_{HNHA}$Couplings:** The scalar couplings between H^{N} and H^{α} are sensitive to backbone conformations and can be converted into the Φ torsion angle using Karplus-like relationships (Karplus 1959). Secondary structures can be unveiled using $^{3}J_{HNHA}$. Values below 5 Hz suggest the presence of α-helices, whereas values above 8 Hz suggest extended regions. Values lying in between correspond to random coil conformations. In combination with other NMR observables, $^{3}J_{HNHA}$ can be useful to discriminate between extended structures and poly-proline-II helices (Oh et al. 2012).
- **Residual Dipolar Couplings (RDCs):** RDCs, measured in magnetically aligned samples, are the most sensitive experimental measurement to probe conformational sampling in IDPs (Jensen et al. 2009). Negative NH RDC values are observed in random coils (Louhivuori et al. 2003). Interestingly, more positive and more negative RDCs than expected for a random coil are associated to α-helices and extended conformations, respectively (Mohana-Borges et al. 2004). This is an excellent indication to qualitatively assess the presence of distinct types of secondary structural elements. More quantitative interpretation of RDCs can be derived when applying atomistic models of disordered chains (Jha et al. 2005; Bernadó et al. 2005; Marsh et al. 2008). The measurement of multiple backbone RDCs enriches the description of residue-specific structural preferences (Jensen et al. 2008).
- **Paramagnetic Relaxation Enhancement (PRE) experiments:** PREs are measured as enhanced relaxation rates in residues that are close (within 15–30 Å) to a paramagnetic tag engineered in a specific position of the chain (Gillespie and Shortle 1997). Although nitroxyl spin labels are normally used, paramagnetic cations can also be attached providing, in addition to PREs, other structural observables such as pseudocontact shifts (PCS) (Otting 2008). Therefore, PREs are suitable observables to probe transient long-range interactions in IDPs.

The complementarity between NMR and SAS is based on the distinct resolution of the information provided. Whereas SAS probes the overall properties of molecules, NMR information reports on atomic or residue-specific information. Therefore, the simultaneous description of both observables strongly suggests the appropriateness of the derived model. In that context, SAXS can be used to validate structural models of IDPs refined with NMR data. In this approach, the residue-specific conformational preferences of an IDP are refined using RDCs and CSs using Flexible-Meccano (Bernadó et al. 2005; Ozenne et al. 2012). The final model contains percentages of secondary structural elements in localized regions that have been imposed to properly describe the NMR data. The resulting ensemble can be validated by simply comparing the average SAXS curve computed from the ensemble with that experimentally measured. This strategy has been applied to the partially folded Sendai virus PX (Bernadó et al. 2005), the transactivation domain of p53 (Wells et al. 2008), the K18 construct of Tau protein (Mylonas et al. 2008; Mukrasch et al. 2007), and the oncogene $p15^{PAF}$ (De Biasio et al. 2014). A similar approach has been performed to study PaaA2

antitoxin (Sterckx et al. 2014). In this last study the NMR derived ensemble was used as starting pool for a SAXS EOM refinement demonstrating that the protein exists in solution as two preformed helices, connected by a flexible linker.

The best manner to exploit the complementarity of both techniques is to integrate the experimental data into the same refinement protocol. Some of these integrative approaches have been applied to IDPs. One of these is ENSEMBLE, a program that derives ensembles of disordered proteins by collectively describing SAXS curves in addition to several NMR observables: CS, J-couplings, RDCs, PREs, Nuclear Overhauser effects, hydrodynamic radius, solvent accessibility restraints, hydrogen-exchange protection factors, and ^{15}N R_2 relaxation rates (Marsh et al. 2007; Krzeminski et al. 2013). A large number of random structures are computed with FOLDTRAJ or TRADES (Feldman and Hogue 2000, 2002), and a Monte Carlo algorithm is used to select a subset of these structures that are collectively consistent with the experimental restraints. This subset is used as a basis for the generation of new structures, and the process is repeated until a final ensemble consistent with all of the experimental measurements is obtained. This approach addresses the intrinsic problem of under-restraining and consequent over-fitting by finding the smallest ensemble that is consistent with all experimental restraints imposed. ENSEMBLE has been recently applied to characterize the protein Sic1 and its hexaphosphorylated version pSic1 by combining SAXS data with several NMR parameters, including CS, PREs, RDCs, and ^{15}N R_2 (Mittag et al. 2010). Moreover, a structural model of the complex between pSic1, which contains several binding regions, and its partner Cdc4 was generated by combining restraints of the free form of pSic1 with sparse NMR data of the complex suggesting a fuzzy interaction. ASTEROIDS is another program that allows the synergistic interpretation of NMR and SAXS data (Jensen et al. 2008). The power of ASTEROIDS is illustrated in a recent study of tau and α-synuclein using NMR (CSs, RDCs, PREs) and SAXS data (Schwalbe et al. 2014). Using extensive cross-validation, the authors showed that five different types of independent experimental parameters are predicted more accurately by selected ensembles than by statistical coil descriptions. With this method, they could highlight that tau and α-synuclein sample poly-proline II region in the aggregation-nucleation sites.

7.10 Final Remarks

Disordered systems represent a challenge for structural biology. The presence of multiple conformational states hampers the application of traditional approaches and has fostered the development of strategies with the capacity to capture their inherent plasticity. Despite the limited amount of information coded, SAS has played (and will play) a crucial role in the structural biology of disordered proteins. The appearance of several ensemble approaches to describe SAS data in terms of multiple conformations has been a revolution in the field as they have opened the characterization of the size and shape properties of highly flexible proteins in solution. These methods, however, are model-dependent as their results can be biased depending on the approach used to compute atomic structures of disordered states. New strategies will have to be developed to accurately explore the energetically plausible conformational landscape of disordered proteins without dramatically increasing computational costs. A related aspect that remains poorly understood in IDPs is hydration. The accurate description of surrounding water molecules in disordered chains is necessary to describe their scattering properties.

Alternatively, the integration of experimental information derived from other biophysical methods, mainly NMR, enriches the resolution of the derived conformational ensembles. This is a research field that despite its continuous evolution will require additional efforts. Indeed, a database, pE-DB (http://pedb.vib.be), has been created to compile structural ensembles derived either from NMR, SAXS or their combination to foster integrative approaches and the

development of validation strategies in disordered proteins (Varadi et al. 2014).

Disordered biomolecular complexes exemplify the challenges of the field and the need to properly combine experimental and computational approaches. Their characterization requires the detailed description of both the interacting surface and the disordered regions. In that context, SAS can be the key technique to merge in a unified picture the information derived from high-resolution techniques such as NMR and crystallography. Moreover, many of the involving IDPs have a low or moderate affinity causing species polydispersity in typical SAS experiments. This incorporates an additional level of complexity that has to be overcome.

In the last decade there has been a general realization that protein dynamics is fundamental to understand biomolecular function. SAS has a brilliant future in that emerging field. The exploitation of SAS information to address this recent field in structural biology will require further experimental, computational and conceptual developments.

Acknowledgements This work was supported by the European Research Council under the European Union's H2020 Framework Programme (2014-2020) / ERC Grant agreement n° [648030], the French Infrastructure for Integrated Structural Biology (FRISBI – ANR-10-INSB-05-01), and Labex EpiGenMed, an « Investissements d'avenir » program, (ANR-10-LABX-12-01) to PB. FHT is supported by INSERM and the Sapere Aude programme SAFIR of the University of Copenhagen. AU is supported by a grant from the Fondation pour Recherche Médicale.

References

Alborghetti MR, Furlan AS, Silva JC, Paes Leme AF, Torriani ICL, Kobarg J (2010) Human FEZ1 protein forms a disulfide bond mediated dimer: implications for cargo transport. J Proteome Res 9:4595–4603

Antonov LD, Olsson S, Boomsma W, Hamelryck T (2016) Bayesian inference of protein ensembles from SAXS data. Phys Chem Chem Phys 18:5832–5838

Aslam M, Guthridge JM, Hack BK, Quigg RJ, Holers VM, Perkins SJ (2003) The extended multidomain solution structures of the complement protein Crry and its chimeric conjugate Crry-Ig by scattering, analytical ultracentrifugation and constrained modelling: implications for function and therapy. J Mol Biol 329:525–550

Balu R, Knott R, Cowieson NP, Elvin CM, Hill AJ, Choudhury NR, Dutta NK (2015) Structural ensembles reveal intrinsic disorder for the multi-stimuli responsive bio-mimetic protein Rec1-resilin. Sci Rep 5:10896

Bernadó P (2010) Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. Eur Biophys J 39:769–780

Bernadó P, Blackledge M (2009) A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. Biophys J 97:2839–2845

Bernadó P, Blackledge M (2010) Structural biology: proteins in dynamic equilibrium. Nature 468:1046–1048

Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci U S A 102:17002–17007

Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129:5656–5664

Bernadó P, Svergun DI (2012a) Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. Mol BioSyst 8:151–167

Bernadó P, Svergun DI (2012b) Analysis of intrinsically disordered proteins by small-angle X-ray scattering. Methods Mol Biol 896:107–122

Bertini I, Giachetti A, Luchinat C, Parigi G, Petoukhov MV, Pierattelli R, Ravera E, Svergun DI (2010) Conformational space of flexible biological macromolecules from average data. J Am Chem Soc 132:13553–13558

Best RB, Zheng W, Mittal J (2014) Balanced protein-water interactions improve properties of disordered proteins and non-specific protein associantion. J Chem Theory Comput 10:5113–5124

Borysik AJ, Kovacs D, Guharoy M, Tompa P (2015) Ensemble methods enable a new definition for the solution to gas-phase transfer of intrinsically disordered proteins. J Am Chem Soc 137:13807–13817

Boura E, Różycki B, Chung HS, Herrick DZ, Canagarajah B, Cafiso DS, Eaton WA, Hummer G, Hurley JH (2012) Solution structure of the ESCRT-I and -II supercomplex: implications for membrane budding and scission. Structure 20:874–886

Boze H, Marlin T, Durand D, Pérez J, Vernhet A, Canon F, Sarni-Manchado P, Cheynier V, Cabane B (2010) Proline-rich salivary proteins have extended conformations. Biophys J 99:656–665

Bressan GC, Silva JC, Borges JC, Dos Passos DO, Ramos CHI, Torriani IL, Kobarg J (2008) Human regulatory protein Ki-1/57 has characteristics of an intrinsically unstructured protein. J Proteome Res 7:4465–4474

Calmettes P, Durand D, Desmadril M, Minard P, Receveur V, Smith JC (1994) How random is a highly denatured protein? Biophys Chem 53:105–114

Chebaro Y, Ballard AJ, Chakraborty D, Wales DJ (2015) Intrinsically disordered energy landscapes. Sci Rep 5:10386

Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75:333–366

Chukhlieb M, Raasakka A, Ruskamo S, Kursula P (2015) The N-terminal cytoplasmic domain of neuregulin 1 type III is intrinsically disordered. Amino Acids 47:1567–1577

Cragnell C, Durand D, Cabane B, Skepö M (2016) Coarse-grained modelling of the intrinsically disordered protein Histatin 5 in solution. Monte Carlo simulations in combination with SAXS. Proteins 84 (6)777–791

Daughdrill GW, Kashtanov S, Stancik A, Hill SE, Helms G, Muschol M, Receveur-Bréchot V, Ytreberg FM (2012) Understanding the structural ensembles of a highly extended disordered protein. Mol BioSyst 8:308–319

De Biasio A, Ibáñez de Opakua A, Cordeiro TN, Villate M, Merino N, Sibille N, Lelli M, Diercks T, Bernadó P, Blanco FJ (2014) p15PAF is an intrinsically disordered protein with nonrandom structural preferences at sites of interaction with other proteins. Biophys J 106:865–874

Dedmon M, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM (2005) Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. J Am Chem Soc 127:4767

Devarakonda S, Gupta K, Chalmers MJ, Hunt JF, Griffin PR, Van Duyne GD, Spiegelman BM (2011) Disorder-to-order transition underlies the structural basis for the assembly of a transcriptionally active PGC-1α/ERRγ complex. Proc Natl Acad Sci U S A 108:18678–18683

Didry D, Cantrelle F-X, Husson C, Roblin P, Moorthy AME, Perez J, Le Clainche C, Hertzog M, Guittet E, Carlier M-F, van Heijenoort C, Renault L (2012) How a single residue in individual β-thymosin/WH2 domains controls their functions in actin assembly. EMBO J 31:1000–1013

Doniach S (2001) Changes in biomolecular conformation seen by small angle X-ray scattering. Chem Rev 101:1763–1778

Dougan L, Li J, Badilla CL, Berne BJ, Fernandez JM (2009) Single homopolypeptide chains collapse into mechanically rigid conformations. Proc Natl Acad Sci U S A 106:12605–12610

Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z (2002) Intrinsic disorder and protein function. Biochemistry 41:6573–6582

Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 272:5129–5148

Dyson HJ, Wright PE (2004) Unfolded proteins and protein folding studied by NMR. Chem Rev 104:3607–3622

Eliezer D (2009) Biophysical characterization of intrinsically disordered proteins. Curr Opin Struct Biol 19:23–30

Emperador A, Sfriso P, Villarreal MA, Gelpí JL, Orozco M (2015) PACSAB: coarse-grained force field for the study of protein-protein interactions and conformational sampling in multiprotein systems. J Chem Theory Comput 11:5929–5938

Feigin LA, Svergun DI (1987) Structure analysis by small-angle X-ray and neutron scattering. Plenum Press, New York

Feldman HJ, Hogue CW (2000) A fast method to sample real protein conformational space. Proteins 39:112–131

Feldman HJ, Hogue CW (2002) Probabilistic sampling of protein conformations: new hope for brute force? Proteins 46:8–23

Flory PJ (1953) Principles of polymer chemistry. Cornell University Press, Ithaca

Foucault M, Mayol K, Receveur-Bréchot V, Bussat M-C, Klinguer-Hamour C, Verrier B, Beck A, Haser R, Gouet P, Guillon C (2010) UV and X-ray structural studies of a 101-residue long tat protein from a HIV-1 primary isolate and of its mutated, detoxified, vaccine candidate. Proteins 78:1441–1456

Gabel F (2012) Small angle neutron scattering for the structural study of intrinsically disordered proteins in solution: a practical guide. Methods Mol Biol 896:123–135

Gast K, Damaschun H, Eckert K, Schulze-Forster K, Maurer HR, Müller-Frohne M, Zirwer D, Czarnecki J, Damaschun G (1995) Prothymosin alpha: a biologically active protein with random coil conformation. Biochemistry 34:13211–13218

Gazi AD, Bastaki M, Charova SN, Gkougkoulia EA, Kapellios EA, Panopoulos NJ, Kokkinidis M (2008) Evidence for a coiled-coil interaction mode of disordered proteins from bacterial type III secretion systems. J Biol Chem 283:34062–14068

Giehm L, Svergun DI, Otzen DE, Vestergaard B (2011) Low-resolution structure of a vesicle disrupting alpha-synuclein oligomer that accumulates during fibrillation. Proc Natl Acad Sci U S A 108:3246–3251

Gillespie JR, Shortle D (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. J Mol Biol 268:158–169

Goldenberg DP, Argyle B (2014) Minimal effects of macromolecular crowding on an intrinsically disordered protein: a small-angle neutron scattering study. Biophys J 106:905–914

Gosselin P, Oulhen N, Jam M, Ronzca J, Cormier P, Czjzek M, Cosson B (2011) The translational repressor 4E-BP called to order by eIF4E: new structural insights by SAXS. Nucleic Acids Res 39:3496–3503

Graewert MA, Svergun DI (2013) Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). Curr Opin Struct Biol 23:748–754

Greving I, Dicko C, Terry A, Callow P, Vollrath F (2010) Small angle neutron scattering of native and reconstituted silk fibroin. Soft Matter 6:4389

Guinier A (1939) La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. Ann Phys (Paris) 12:161–237

Haerty W, Golding GB (2010) Low-complexity sequences and single amino acid repeats: not just "junk" peptide sequences. Genome 53:753–762

Hammel M (2012) Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). Eur Biophys J 41:789–799

Hawkins AR, Lamb HK (1995) The molecular biology of multidomain proteins. Eur J Biochem 232:7–18

Heller WT (2010) Small-angle neutron scattering and contrast variation: a powerful combination for studying biological structures. Acta Crystallogr D Biol Crystallogr 66:1213–1217

Henriques J, Cragnell C, Skepö M (2015) Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. J Chem Theory Comput 11:3420–3431

Jacques DA, Trewhella J (2010) Small-angle scattering for structural for structural biology-expanding the frontier while avoiding the pitfalls. Protein Sci 19:642–657

Jacrot B (1976) The study of biological structures by neutron scattering from solution. Rep Prog Phys 39:911–953

Jeng U-S, Lin T-L, Lin JM, Ho DL (2006) Contrast variation SANS for the solution structure of the β-amyloid peptide 1–40 influenced by SDS surfactants. Phys B Condens Matter 385-386:865–867

Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RW, Blackledge M (2008) Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. J Am Chem Soc 130:8055–8061

Jensen MR, Markwick PR, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, Bernadó P, Blackledge M (2009) Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. Structure 17:952–960

Jensen MR, Ruigrok RWH, Blackledge M (2013) Describing intrinsically disordered proteins at atomic resolution by NMR. Curr Opin Struct Biol 23:426–435

Jensen MR, Zweckstetter M, Huang JR, Blackledge M (2014) Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. Chem Rev 114:6632–6660

Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: resolving the reconciliation problem. Proc Natl Acad Sci U S A 102:13099–13104

Jin KS, Park JK, Yoon J, Rho Y, Kim J-H, Kim EE, Ree M (2008) Small-angle X-ray scattering studies on structures of an estrogen-related receptor alpha ligand binding domain and its complexes with ligands and coactivators. J Phys Chem B 112:9603–9612

Johansen D, Jeffries CMJ, Hammouda B, Trewhella J, Goldenberg DP (2011a) Effects of macromolecular crowding on an intrinsically disordered protein characterized by small-angle neutron scattering with contrast matching. Biophys J 100:1120–1128

Johansen D, Trewhella J, Goldenberg DP (2011b) Fractal dimension of an intrinsically disordered protein: small-angle X-ray scattering and computational study of the bacteriophage λ N protein. Protein Sci 20:1955–1970

Kachala M, Valentini E, Svergun DI (2015) Application of SAXS for the structural characterization of IDPs. Adv Exp Med Biol 870:261–289

Karplus M (1959) Contact electron-spin coupling of nuclear magnetic moments. J Chem Phys 30

Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9:646–652

Kikhney AG, Svergun DI (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. FEBS Lett 589:2570–2577

Kim PM, Sboner A, Xia Y, Gerstein M (2008) The role of disorder in interaction networks: a structural analysis. Mol Syst Biol 4:179

Kjaergaard M, Brander S, Poulsen FM (2011) Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. J Biomol NMR 49:139–149

Kjaergaard M, Nørholm AB, Hendus-Altenburger R, Pedersen SF, Poulsen FM, Kragelund BB (2010) Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II? Protein Sci 19:1555–1564

Koch MHJ, Vachette P, Svergun DI (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q Rev Biophys 36:147–227

Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci U S A 101:12491–12496

Konno T, Tanaka N, Kataoka M, Takano E, Maki M (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

Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc Natl Acad Sci U S A 93:11504–11509

Krueger S, Shin J-H, Raghunandan S, Curtis JE, Kelman Z (2011) Atomistic ensemble modeling and small-angle neutron scattering of intrinsically disordered protein complexes: applied to minichromosome maintenance protein. Biophys J 101:2999–3007

Krzeminski M, Marsh JA, Neale C, Choy WY, Forman-Kay JD (2013) Characterization of disordered proteins with ENSEMBLE. Bioinformatics 29:398–399

Kung CC-H, Naik MT, Wang S-H, Shih H-M, Chang C-C, Lin L-Y, Chen C-L, Ma C, Chang C-F, Huang T-H (2014) Structural analysis of poly-SUMO chain recognition by the RNF4-SIMs domain. Biochem J 462:53–65

Lee KH, Chen J (2016) Multiscale enhanced sampling of intrinsically disordered protein conformations. J Comput Chem 37:550–557

LeGuillou JC, Zinn-Justin J (1977) Critical exponents for the n-vector model in three dimensions from field theory. Phys Rev Lett 39:95–98

Lens Z, Dewitte F, Monté D, Baert J-L, Bompard C, Sénéchal M, Van Lint C, de Launoit Y, Villeret V, Verger A (2010) Solution structure of the N-terminal transactivation domain of ERM modified by SUMO-1. Biochem Biophys Res Commun 399:104–110

Lenton S, Seydel T, Nylander T, Holt C, Härtlein M, Teixeira S, Zaccai G (2015) Dynamic footprint of sequestration in the molecular fluctuations of osteopontin. J R Soc Interface 12:0506

Levitt M (2009) Nature of the protein universe. Proc Natl Acad Sci U S A 106:11079–11084

Leyrat C, Jensen MR, Ribeiro EA Jr, Gérard FC, Ruigrok RW, Blackledge M, Jamin M (2011) The N(0)-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient α-helices. Protein Sci 20:542–556

Li J, Uversky VN, Fink AL (2002) Conformational behavior of human alpha-synuclein is modulated by familial Parkinson's disease point mutations A30P and A53T. Neurotoxicology 23:553–567

Lindorff-Larsen K, Kristjansdottir S, Teilum K, Fieber W, Dobson C, Poulsen F, Vendruscolo M (2004) Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme a binding protein. J Am Chem Soc 126:3291

Longhi S, Receveur-Bréchot V, Karlin D, Johansson K, Darbon H, Bhella D, Yeo R, Finet S, Canard B (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 278:18638–18648

Louhivuori M, Pääkkönen K, Fredriksson K, Permi P, Lounila J, Annila A (2003) On the origin of residual dipolar couplings from denatured proteins. J Am Chem Soc 125:15647–15650

Marasini C, Galeno L, Moran O (2013) A SAXS-based ensemble model of the native and phosphorylated regulatory domain of the CFTR. Cell Mol Life Sci 70:923–933

Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D (1999) A census of protein repeats. J Mol Biol 293:151–160

Marsh JA, Baker JM, Tollinger M, Forman-Kay JD (2008) Calculation of residual dipolar couplings from disordered state ensembles using local alignment. J Am Chem Soc 130:7804–7805

Marsh JA, Dancheck B, Ragusa MJ, Allaire M, Forman-Kay JD, Peti W (2010) Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. Structure 18:1094–1103

Marsh JA, Neale C, Jack FE, Choy WY, Lee AY, Crowhurst KA, Forman-Kay JD (2007) Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. J Mol Biol 367:1494–1510

Meier S, Grzesiek S, Blackledge M (2007) Mapping the conformational landscape of urea-denatured ubiquitin using residual dipolar couplings. J Am Chem Soc 129:9799–9807

Mertens HD, Svergun DI (2010) Structural characterization of proteins and complexes using small-angle X-ray scattering. J Struct Biol 172:128–141

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087–1092

Millard CJ, Watson PJ, Fairall L, Schwabe JWR (2013) An evolving understanding of nuclear receptor coregulator proteins. J Mol Endocrinol 51:T23–T36

Mittag T, Marsh J, Grishaev A, Orlicky S, Sicheri F, Tyers M, Forman-Kay JD (2010) Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. Structure 18:494–506

Mittal A, Lyle N, Harmon TS, Pappu RV (2014) Hamiltonian switch metropolis Monte Carlo simulations for improved conformational sampling of intrinsically disordered regions tethered to ordered domains of proteins. J Chem Theory Comput 10:3550–3562

Mohana-Borges R, Goto NK, Kroon GJ, Dyson HJ, Wright PE (2004) Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. J Mol Biol 340:1131–1142

Moncoq K, Broutin I, Craescu CT, Vachette P, Ducruix A, Durand D (2004) SAXS study of the PIR domain from the Grb14 molecular adaptor: a natively unfolded protein with a transient structure primer? Biophys J 87:4056–4064

Mukrasch MD, Markwick P, Biernat J, Bergen M, Bernadó P, Griesinger C, Mandelkow E, Zweckstetter M, Blackledge M (2007) Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. J Am Chem Soc 129:5235–5243

Mylonas E, Hascher A, Bernadó P, Blackledge M, Mandelkow E, Svergun DI (2008) Domain conformation of tau protein studied by solution small-angle X-ray scattering. Biochemistry 47:10345–10353

Nairn KM, Lyons RE, Mulder RJ, Mudie ST, Cookson DJ, Lesieur E, Kim M, Lau D, Scholes FH, Elvin CM

(2008) A synthetic resilin is largely unstructured. Biophys J 95:3358–3365

Narayanan RL, Durr UHN, Bibow S, Biernat J, Mandelkow E, Zweckstetter M (2010) Automatic assignment of the intrinsically disordered protein Tau with 441-residues. J Am Chem Soc 132:11906–11907

Nørholm A-B, Hendus-Altenburger R, Bjerre G, Kjaergaard M, Pedersen SF, Kragelund BB (2011) The intracellular distal tail of the Na+/H+ exchanger NHE1 is intrinsically disordered: implications for NHE1 trafficking. Biochemistry 50:3469–3480

O'Brien DP, Hernandez B, Durand D, Hourdel V, Sotomayor-Pérez A-C, Vachette P, Ghomi M, Chamot-Rooke J, Ladant D, Brier S, Chenal A (2015) Structural models of intrinsically disordered and calcium-bound folded states of a protein adapted for secretion. Sci Rep 5:14223

Oh K-I, Jung Y-S, Hwang G-S, Cho M (2012) Conformational distributions of denatured and unstructured proteins are similar to those of 20 x 20 blocked dipeptides. J Biomol NMR 53:25–41

Otting G (2008) Prospects for lanthanides in structural biology by NMR. J Biomol NMR 42:1–9

Owens GE, New DM, West AP, Bjorkman PJ (2015) Anti-PolyQ antibodies recognize a short PolyQ stretch in both normal and mutant Huntingtin exon 1. J Mol Biol 427:2507–2519

Ozenne V, Bauer F, Salmon L, Huang J, Jensen MR, Segard S, Bernadó P, Charavay C, Blackledge M (2012) Flexible-meccano:a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28:1463–1470

Palazzesi F, Prakash MK, Bonomi M, Barducci A (2015) Accuracy of current all-atom force-fields in modeling protein disordered states. J Chem Theory Comput 11:2–7

Pavlin MR, Brunzelle JS, Fernandez EJ (2014) Agonist ligands mediate the transcriptional response of nuclear receptor heterodimers through distinct stoichiometric assemblies with coactivators. J Biol Chem 289:24771–24778

Paz A, Zeev-Ben-Mordehai T, Lundqvist M, Sherman E, Mylonas E, Weiner L, Haran G, Svergun DI, Mulder FAA, Sussman JL, Silman I (2008) Biophysical characterization of the unstructured cytoplasmic domain of the human neuronal adhesion protein neuroligin 3. Biophys J 95:1928–1944

Pelikan M, Hura GL, Hammel M (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering. Gen Physiol Biophys 28:174–189

Perevozchikova T, Stanley CB, McWilliams-Koeppen HP, Rowe EL, Berthelier V (2014) Investigating the structural impact of the glutamine repeat in huntingtin assembly. Biophys J 107:411–421

Pérez J, Nishino Y (2012) Advances in X-ray scattering: from solution SAXS to achievements with coherent beams. Curr Opin Struct Biol 22:670–678

Petoukhov MV, Svergun DI (2007) Analysis of X-ray and neutron scattering from biomacromolecular solutions. Curr Opin Struct Biol 17:562–571

Piana S, Klepeis JL, Shaw DE (2014) Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. Curr Opin Struct Biol 24:98–105

Pizzi E, Frontali C (2001) Low-complexity regions in plasmodium falciparum proteins. Genome Res 11:218–229

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

Rambo RP, Tainer JA (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. Biopolymers 95:559–571

Rambo RP, Tainer JA (2013) Super-resolution X-ray sacttering and its applications to structural systems biology. Annu Rev Biophys 42:415–441

Rauscher S, Gapsys V, Gajda MJ, Zweckstetter M, de Groot BL, Grubmüller H (2015) Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment. J Chem Theory Comput 11:5513–5524

Receveur-Brechot V, Durand D (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. Curr Protein Pept Sci 13:55–75

Rochel N, Ciesielski F, Godet J, Moman E, Roessle M, Peluso-Iltis C, Moulin M, Haertlein M, Callow P, Mély Y, Svergun DI, Moras D (2011) Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. Nat Struct Mol Biol 18:564–570

Rozycki B, Kim YC, Hummer G (2011) SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. Structure 19:109–116

Ruskamo S, Chukhlieb M, Vahokoski J, Bhargav SP, Liang F, Kursula I, Kursula P (2012) Juxtanodin is an intrinsically disordered F-actin-binding protein. Sci Rep 2:899

Schwalbe M, Ozenne V, Bibow S, Jaremko M, Jaremko L, Gajda M, Jensen MR, Biernat J, Becker S, Mandelkow E, Zweckstetter M, Blackledge M (2014) Predictive atomic resolution descriptions of intrinsically disordered hTau40 and α-Synuclein in solution from NMR and small angle scattering. Structure 22:238–249

Schwarzinger S, Kroon GJ, Foss TR, Wright PE, Dyson HJ (2000) Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. J Biomol NMR 18:43–48

Shell SS, Putnam CD, Kolodner RD (2007) The N terminus of Saccharomyces Cerevisiae Msh6 is an unstructured tether to PCNA. Mol Cell 26:565–578

Shkumatov AV, Chinnathambi S, Mandelkow E, Svergun DI (2011) Structural memory of natively unfolded tau protein detected by small-angle X-ray scattering. Proteins 79:2122–2131

Sibille N, Bernadó P (2012) Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. Biochem Soc Trans 40:955–962

Stadler AM, Stingaciu L, Radulescu A, Holderer O, Monkenbusch M, Biehl R, Richter D (2014) Internal nanosecond dynamics in the intrinsically disordered myelin basic protein. J Am Chem Soc 136:6987–6994

Stanley CB, Perevozchikova T, Berthelier V (2011) Structural formation of huntingtin exon 1 aggregates probed by small-angle neutron scattering. Biophys J 100:2504–2512

Sterckx YGJ, Volkov AN, Vranken WF, Kragelj J, Jensen MR, Buts L, Garcia-Pino A, Jové T, Van ML, Blackledge M, van Nuland NAJ, Loris R (2014) Small-angle X-ray scattering- and nuclear magnetic resonance-derived conformational ensemble of the highly flexible antitoxin PaaA2. Structure 22:854–865

Stott K, Watson M, Howe FS, Grossmann JG, Thomas JO (2010) Tail-mediated collapse of HMGB1 is dynamic and occurs *via* differential binding of the acidic tail to the A and B domains. J Mol Biol 403:706–722

Stumpe MC, Grubmüller H (2007) Interaction of urea with amino acids: implications for urea-induced protein denaturation. J Am Chem Soc 129:16126–16131

Svergun DI, Koch MHJ (2003) Small-angle scattering studies of biological macromolecules in solution. Rep Prog Phys 66:1735–1782

Tamiola K, Acar B, Mulder FAA (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. J Am Chem Soc 132:18000–18003

Tidow H, Melero R, Mylonas E, Freund SMV, Grossmann JG, Carazo JM, Svergun DI, Valle M, Fersht AR (2007) Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. Proc Natl Acad Sci U S A 104:12324–12329

Tompa P, Schad E, Tantos A, Kalmar L (2015) Intrinsically disordered proteins: emerging interaction specialists. Curr Opin Struct Biol 35:49–59

Trakawa T, Takada S (2011) Multiscale ensemble modeling of intrinsically disordered proteins: p53 N-terminal domain. Biophys J 101:1450–1458

Tria G, Mertens HD, Kachala M, Svergun DI (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCrJ 2:207–217

Uversky VN, Gillespie JR, Millett IS, Khodyakova AV, Vasilenko RN, Vasiliev AM, Rodionov IL, Kozlovskaya GD, Dolgikh DA, Fink AL et al (2000) Zn(2+)-mediated structure formation and compaction of the "natively unfolded" human prothymosin alpha. Biochem Biophys Res Commun 267:663–668

Uversky VN, Gillespie JR, Millett IS, Khodyakova AV, Vasiliev AM, Chernovskaya TV, Vasilenko RN, Kozlovskaya GD, Dolgikh DA, Fink AL et al (1999) Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. Biochemistry 38:15009–15016

Uversky VN, Li J, Souillac P, Millett IS, Doniach S, Jakes R, Goedert M, Fink AL (2002) Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of alpha-synuclein assembly by beta- and gamma-synucleins. J Biol Chem 277:11970–11978

Varadi M, Kosol S, Lebrun P, Valentini E, Blackledge M, Dunker AK, Felli IC, Forman-Kay JD, Kriwacki RW, Pierattelli R, Sussman J, Svergun DI, Uversky VN, Vendruscolo M, Wishart D, Wright PE, Tompa P (2014) pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. Nucleic Acids Res 42:D326–D335

Vestergaard B, Groenning M, Roessle M, Kastrup JS, van de Weert M, Flink JM, Frokjaer S, Gajhede M, Svergun DI (2007) A helical structural nucleus is the primary elongating unit of insulin amyloid fibrils. PLoS Biol 5:1089–1097

Vitalis A, Pappu RV (2009a) Methods for Monte Carlo simulations of biomacromolecules. Ann Rep Comput Chem 5:49–76

Vitalis A, Pappu RV (2009b) ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. J Comput Chem 30:673–699

Vitalis A, Wang X, Pappu RV (2008) Atomistic simulations of the effects of Polyglutamine chain length and solvent quality on conformational Equilibria and spontaneous Homodimerization. J Mol Biol 384:279–297

Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, E. Mylonas, Svergun DI, Blackledge M, Fersht AR (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci U S A 105:5762–5767

Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD (1995) (1)H, (13)C and (15)N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J Biomol NMR 5:332

Wishart DS, Sykes BD (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 4:171–180

Wootton JC, Federhen S (1993) Statistics of local complexity in amino acid sequences and sequence databases. Comput Chem 17:149–163

Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293:321–331

Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 16:18–29

Wu K-P, Weinstock DS, Narayanan C, Levy RM, Baum J (2009) Structural reorganization of α-synuclein at low

pH observed by NMR and REMD simulations. J Mol Biol 391:784

Yang S (2014) Methods for SAXS-based structure determination of biomolecular complexes. Adv Mater 26:7902–7910

Yang S, Blachowicz L, Makowski L, Roux B (2010) Multidomain assembled states of Hck tyrosine kinase in solution. Proc Natl Acad Sci U S A 107:15757–15762

Yang C, van der Woerd MJ, Muthurajan UM, Hansen JC, Luger K (2011) Biophysical analysis and small-angle X-ray scattering-derived structures of MeCP2-nucleosome complexes. Nucleic Acids Res 39:4122–4135

Zerze GH, Miller CM, Granata D, Mittal J (2015) Free energy surface of an intrinsically disordered protein: comparison between temperature replica exchange molecular dynamics and bias-exchange metadynamics. J Chem Theory Comput 11:2776–2782

Zhou H-X (2004) Polymer models of protein stability, folding, and interactions. Biochemistry 43:2141–2154

Zhou CZ, Confalonieri F, Jacquet M, Perasso R, Li ZG, Janin J (2001) Silk fibroin: structural implications of a remarkable amino acid sequence. Proteins 44:119–122