



HAL
open science

An arsenal of methods for the experimental characterization of intrinsically disordered proteins – How to choose and combine them?

Antoine Schramm, Christophe Bignon, Stefania Brocca, Rita Grandori, Carlo
Santambrogio, Sonia Longhi

► To cite this version:

Antoine Schramm, Christophe Bignon, Stefania Brocca, Rita Grandori, Carlo Santambrogio, et al..
An arsenal of methods for the experimental characterization of intrinsically disordered proteins –
How to choose and combine them?. Archives of Biochemistry and Biophysics, 2019, pp.108055.
10.1016/j.abb.2019.07.020 . hal-02341641

HAL Id: hal-02341641

<https://hal.science/hal-02341641>

Submitted on 21 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

An arsenal of methods for the experimental characterization of intrinsically disordered proteins – how to choose and combine them?

Antoine Schramm¹, Christophe Bignon¹, Stefania Brocca², Rita Grandori², Carlo Santambrogio² and Sonia Longhi^{1*}

¹CNRS and Aix-Marseille Univ, Laboratoire Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, Marseille, France

²Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy

*To whom correspondence should be addressed.

Sonia LONGHI

CNRS and Aix-Marseille Université, AFMB, UMR 7257, Case 932

163, Avenue de Luminy, Case 932, 13288, Marseille, France

E-mail: Sonia.Longhi@afmb.univ-mrs.fr

Tel: (33) 4 91 82 55 80

Fax: (33) 4 91 26 67 20

Submitted to the ABB special issue on IDPs

ABSTRACT

In this review, we detail the most common experimental approaches to assess and characterize protein intrinsic structural disorder, with the notable exception of NMR and EPR spectroscopy, two ideally suited approaches that will be described in depth in two other reviews within this special issue. We discuss the advantages, the limitations, as well as the caveats of the various methods. We also describe less common and more demanding approaches that enable achieving further insights into the conformational properties of IDPs. Finally, we present recent developments that have enabled assessment of structural disorder in living cells, and discuss the currently available methods to model IDPs as conformational ensembles.

Key words. Experimental assessment of protein disorder; degree of compactness; secondary and tertiary structure content; ensemble models; *in-vivo* assessment of disorder.

Abbreviations. IDP, intrinsically disordered protein; IDR, intrinsically disordered region; 3D, three-dimensional; NMR, nuclear magnetic resonance; EPR, electron paramagnetic resonance; SDS-PAGE, sodium dodecyl sulfate polyacrylamide electrophoresis; TFE, trifluoro ethanol; MoRE, Molecular Recognition Element; SEC, size-exclusion chromatography; R_s , Stokes radius; Da, Dalton; MM, molecular mass; SEC-MALLS, SEC coupled to multiple angle laser light scattering; FPLC, fast protein liquid chromatography; HPLC, high-pressure liquid chromatography; DLS, dynamic light scattering; AUC, analytical ultracentrifugation; SV, sedimentation velocity; SE, sedimentation equilibrium; SAXS, small-angle X-ray scattering; R_g , radius of gyration; CI, compaction index; RC, random coil; PMG, premolten globule; CD, circular dichroism; UV, ultra violet; MRE; mean residue ellipticity; IR, infrared; SDSL, site-directed spin-labeling; FTIR, Fourier Transform Infrared; ATR, attenuated total reflection; DSC, differential scanning microcalorimetry; ANS, 1-anilino-8-naphthalenesulfonic acid; DSF, differential scanning fluorimetry; TSA, thermal shift assay; ESI, electrospray ionization; MS, mass spectrometry; CSD, charge state distribution; SASA, solvent-accessible surface area; IM, ion mobility; CCS, collision cross-section; HDX, hydrogen-deuterium exchange; FRET, Fluorescence resonance energy transfer; smFRET, single-molecule FRET; nsFCS, nanosecond fluorescence correlation spectroscopy; PEG, polyethylene glycol; TCQ, tryptophan triplet cysteine quenching; THF, tetrahydrofuran; ITC, isothermal titration calorimetry; GFP, green fluorescent protein; MeV, measles virus; DEER, double electron electron resonance.

Introduction	2
2. Obtaining hints of structural disorder	4
2.1. SDS-PAGE and limited proteolysis	4
2.2. Resistance to denaturing conditions	5
3. Assessing protein hydrodynamic properties	6
3.1. Size-exclusion chromatography and light scattering	6
3.2. Analytical ultracentrifugation (AUC)	8
3.3. Small-angle X-ray scattering (SAXS)	10
4. Assessing protein secondary structure content	13
4.1. Circular dichroism (CD) spectroscopy	13
4.2. Fourier transform infrared (FTIR) spectroscopy	14
5. Assessing protein tertiary structure	15
5.1. Near-UV CD spectroscopy	15
5.2. Differential scanning calorimetry (DSC)	16
5.3. Fluorescence spectroscopy	16
5.4. Electrospray ionization (ESI) mass spectrometry (MS)	17
5.5. Hydrogen-deuterium exchange (HDX) reaction	19
6. Approaches relying on protein labeling and/or site-directed mutagenesis	20
6.1. Fluorescence resonance energy transfer (FRET)	20
6.2. Vibrational spectroscopy of cyanylated cysteines	22
6.3. Tryptophan triplet cysteine quenching (TCQ)	23
7. <i>In-vivo</i> assessment of disorder	24
8. Modeling IDPs as conformational ensembles	25
9. Assessing binding events	27
10. Conclusions	28
Acknowledgments	29
Figure Legends	30
References	33

Introduction

Intrinsically disordered proteins (IDPs) or regions (IDRs) lack stable secondary and tertiary structures under physiological conditions and in the absence of binding partners. While the structure of folded proteins is dominated by enthalpy and can be represented by a countable number of well-defined models, the overall shape and transient structures in IDPs are imprinted to high entropy levels. IDPs should thus be regarded as “protein clouds”, *i.e.*

ensembles of conformers exhibiting a rapid conformational exchange and devoid of equilibrium geometry [1, 2]. Thereby, their description can't be based on a countable number of models, such as those provided by conventional high-resolution structural approaches for the determination of the three-dimensional (3D) structure of ordered macromolecules (*i.e.* X-ray crystallography, nuclear magnetic resonance NMR and cryo-electron microscopy).

By contrast, high-resolution structural information of IDPs in complex with binding partners or antibodies is available and can be easily retrieved from dedicated databases such as DisBind (<http://biophy.dzu.edu.cn/DisBind/>), DIBS (<http://dibs.enzim.ttk.mta.hu/>) [3] and FuzDB (<http://protodyn-database.org/>) [4]. However, it should be kept in mind that the structure of IDPs bound to their partners are neither representative of all the conformations sampled by the free protein in solution (*i.e.* they only represent one out of all the possible conformations), nor of the conformation that the IDP may adopt upon binding to a different target. Indeed, some IDPs were shown to exhibit an extreme polymorphism, *i.e.* they adopt a partner-dependent conformation [5, 6].

By contrast, methodological approaches that make use of protein samples in solution can give more representative information and are, thus, better suited for the study of IDPs. The extreme conformational heterogeneity of IDPs renders, however, their characterization inherently challenging and requires the combination of various, complementary approaches.

This review is meant to provide an outline of the experimental methods most commonly used to assess protein disorder, with the notable exception of NMR and electron paramagnetic resonance (EPR) spectroscopy that will be described in two other reviews of this special issue. The various experimental methods broadly vary in terms of the molecular detail and resolution provided, of protein amounts and instrumental requirements, and ease of implementation. **Figure 1** illustrates a general suggested procedure to assess protein structural disorder through experimental techniques of increasing level of complexity.

Before detailing the various experimental approaches that can be used to assess and characterize protein disorder, we would like to provide a list of “obvious” features that should alert biochemists and prompt them to suspect the presence of disorder. In particular, one can suspect that a protein or a protein region is intrinsically disordered if it presents at least one of the following properties: sequence hypervariability (as judged from multiple sequence alignment of homologous proteins), low sequence complexity and unusual amino acid composition with typical enrichment in proline, glycine and other disorder-promoting residues (glutamic acid, aspartic acid, serine, lysine, arginine, and glutamine), no or little predicted secondary structure, persistent solubility upon heat or acid treatment, ability to interact with multiple partners and very low extinction coefficient at 280 nm (because of depletion in aromatic residues).

It is worthy to emphasize that the compositional bias of IDPs may lead to errors in estimating their concentration using classical assays for protein quantification, such as the Bradford assay or ultraviolet absorbance at 280 nm [7]. This is well illustrated in the case of the intrinsically disordered, N-terminal region of glycine transporters that were found to exhibit a dynamic interaction (*e.g.* spectral evolution with time) with Coomassie G-250 resulting in interference with their Bradford assay [8]. Whether this behavior is unique to these IDPs or whether it can be regarded as a hallmark of protein disorder remains to be established and certainly deserves to be investigated.

Finally, we would like to take this opportunity to underline that whenever dealing with proteins containing more than two cysteines (and hence potentially able to form intramolecular disulfide bonds) it is advised to perform experiments in both oxidizing and reducing conditions. Indeed, so-called “conditionally disordered proteins” are proteins whose conformational state depends on environmental factors, such as pH, temperature or redox potential [9]. Examples of redox-dependent conditionally disordered proteins are Hsp33 [10], Cox17 [11] and CP12 [12]. A useful tool to obtain hints on possible redox-dependent conditional disorder is IUPred2A, a disorder predictor that takes into account the redox state and that is meant to identify protein regions featuring a redox-dependent conditional disorder [13].

2. Obtaining hints of structural disorder

2.1. SDS-PAGE and limited proteolysis

IDPs often display abnormal mobility in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) experiments and their apparent molecular mass (MM) determined by this technique is often 1.2–1.8 times higher than expected from sequence data or measured by mass spectrometry (MS). This aberrant electrophoretic migration is mainly due to the compositional bias typical of IDPs, e.g. high net charge and low hydrophobicity. As a result, IDPs bind less SDS than globular proteins and are thus delayed in their migration [14]. It has been suggested that the peculiar amino acid composition of IDPs leads to a larger Stokes radius (R_s) of the protein by the loss of the necklace-like structure of the protein-detergent complex, thus further slowing migration and yielding apparent higher MMs [15, 16]. Beyond the charge, other additional parameters can be responsible for the abnormal electrophoretic migration. One such a parameter is proline content: proline-rich proteins were shown to migrate differently than expected or even not as a single band ([17] and references therein). The high proline content presumably increases the rigidity of the protein, which makes the protein migrate more slowly in SDS-PAGE than globular proteins of the same MM. The appearance of two bands may result from conformational differences arising from cis-trans isomerization in proline-rich domain. An additional parameter is the degree of compaction of the protein in solution, as well exemplified in the case of the homologous pairs of Nipah virus and Hendra virus N_{TAIL} truncation variants, which were found to display a different migration in spite of their similar content in charged and hydrophobic residues [18]. Another parameter that affects the electrophoretic behavior is the linear distribution of opposite charge (referred to as κ) [19], as well illustrated by charge permutants of two IDPs, which were found to exhibit a different electrophoretic mobility - overall lower for high- κ variants - although they possess the same net charge [20] (**Figure 2A**). Differences in migration are probably due to the clustering of positively charged residues, which may hinder their interaction with the sulfate groups of the detergent [21].

Therefore, hints about intrinsic disorder may be retrieved by the analysis of protein as early as in the first purification steps. However, anomalous migration in SDS-PAGE can be regarded only as a hallmark of structural disorder, e.g. it does not prove that the protein under

study is an IDP. An additional step towards the goal of drawing reliable conclusions about the (dis)ordered state of a protein could be performing limited proteolysis.

Proteolytic sites are typically found at flexible loop regions (as indicated by high crystallographic temperature factors or B-values) that are also exposed to the solvent [22, 23], and are notably absent in regions of regular secondary structure, especially β -strands [24, 25]. Flexible loops protrude from the protein surface and are expected to bind and adapt to the protease active site as a result of a local unfolding [23]. Fontana and co-workers have shown that the higher the content in disorder, the faster the proteolysis [26, 27]. Furthermore, proteolysis hypersensitive sites often occur within protein regions characterized by missing electron density (as assessed by X-ray crystallography), thus allowing establishing a direct link between proteolysis sensitivity and disorder. Limited proteolysis was proved to be very useful and reliable to explore protein structure and dynamics, thereby complementing other physicochemical approaches. Striking advantages of this simple biochemical technique are the requirement of very small amounts of (unlabeled) protein and the ability to provide region-specific information. The main disadvantages of this technique reside in the dependence of cleavage sites in the protein sequence and in its inability to discriminate between disordered proteins and proteins with a high content in loops or linkers. To address the first limitation, it is advised to use multiple proteases with different sequence specificity and to favor those with low sequence-specificity, such as thermolysin, elastase and proteinase k, which better reports on the solvent accessibility and flexibility of cleavage sites (**Figure 2B**) [28]. Thermolysin also offers the advantage of retaining activity in the presence of trifluoroethanol (TFE) [29], thereby allowing limited proteolysis experiments to be carried out in the presence of this secondary-structure stabilizer. Because TFE mimics the hydrophobic environments experienced by proteins in protein-protein interactions, proteolysis-resistant fragments can be regarded as protein regions containing putative Molecular Recognition Elements (MoREs) (for examples see [30, 31]).

2.2. Resistance to denaturing conditions

The sequence compositional bias that typifies IDPs, combined with the lack of a tightly packed hydrophobic core, results in a different response to denaturing conditions compared to structured proteins. Thus, information about intrinsic disorder may be obtained from the analysis of protein conformational stability.

The lack of ordered structure renders IDPs rather insensitive to temperature increase. Consequently, IDPs can be easily separated from ordered proteins due to their resistance to heat-induced unfolding and aggregation. In structured proteins, upon thermal denaturation, hydrophobic patches, which are normally buried inside the folded structure, become exposed to the solvent and engage in uncontrolled inter- and intramolecular interactions that finally lead to the formation of insoluble aggregates. The high content of hydrophilic and charged amino acid maintains IDPs soluble even at high temperatures. This property can be exploited to purify disordered proteins [20, 32-34]. Noteworthy, a few IDPs were shown to become more ordered upon heating, due to the increased strength of the hydrophobic interaction at higher temperatures, leading to a stronger hydrophobic driving force for folding [35, 36].

Because of the lack of ordered structure, IDPs are also resistant to acidic treatment. In ordered proteins, the protonation of negatively charged side chains, induced by a decrease in

pH, leads to charge imbalances [37], disruption of salt bridges [38] and causes the dissociation of subunits [39] and the release of cofactors [40]. To accommodate the charge imbalance resulting from acidification, the chain often adopts a more open conformation that precludes aggregation. By contrast, many IDPs were shown to resist to precipitation induced by acidification [41]. Notably, in extremely acidic and/or alkaline conditions, IDPs can undergo partial folding as a result of the minimization of the large net charge present at neutral pH, thereby decreasing charge-charge intramolecular repulsion and permitting hydrophobic-driven collapse to a partially-folded intermediate [42].

More recent studies however have unveiled that, under acidic conditions, various IDPs may undergo precipitation as a function of their net charge per residue (NCPR). The higher the NCPR, the more intense the protein aggregation, which, for any given protein, occurs at the isoelectric point (pI) [43]. Noteworthy, a similar effect can be observed also at alkaline pH, giving rise to a symmetric behavior with respect to the pI. The phenomenon underlying this behavior is rather simple to explain: as the pI approaches, the intramolecular charge repulsions decreases, promoting the collapse of hydrophobic residues that leads to both folding [42, 44] and aggregation [43]. Rather than being insensitive to pH changes, IDPs appear capable of undergoing intense and reversible conformational changes coupled to changes in protonation state, due to the high number of charges present in their linear sequence and to the reversibility of their protonation state.

Similarly to heat, relative insensitivity towards denaturing agents, such as urea and guanidinium hydrochloride, is a hallmark of disorder. Because of the lack of ordered structure, addition of denaturants leads only to a limited expansion of the chain (for an example see [45]). The extent of chain expansion can be documented through hydrodynamic analyses as described below.

3. Assessing protein hydrodynamic properties

3.1. Size-exclusion chromatography and light scattering

One of the easiest approaches to document the extent of protein compaction is size-exclusion chromatography (SEC). SEC allows separating proteins according to their hydrodynamic volume and their shape. Following calibration of SEC columns with globular protein standards, IDPs are eluted with an apparent molecular mass higher than calculated from the amino acid sequence. Upon protein injection (typically in the 0.5 – 1 mg range for analytical purposes), the R_s can be determined from the elution volume using a calibration curve obtained using globular proteins of known MM. The R_s (in Å) of the MM standards (in Daltons) is calculated according to [46]:

$$\log (R_s^{Obs}) = 0.369 * (\log MM) - 0.254 \quad (\text{Eq. 1})$$

Note that this equation can be used irrespective of the chemical nature of the matrix of the column, provided that (i) the matrix is suitable to allow a proper separation of the target protein, (ii) the matrix does not aspecifically bind the target protein (a possibility typically ruled out by assessing that SEC elution profile at different ionic strengths remains unvaried). From this calibration curve one can thus infer the R_s of the protein of interest. This experimentally observed value (R_s^{Obs}) is then compared to the value expected for a natively

folded protein (R_s^{NF}), and for a fully unfolded (*i.e.* denatured) protein (R_s^U). The latter can be calculated according to [47]:

$$\log(R_s^{NF}) = 0.357 * (\log MM^{\text{Theo}}) - 0.204 \quad (\text{Eq. 2})$$

$$\log(R_s^U) = 0.521 * (\log MM^{\text{Theo}}) - 0.649 \quad (\text{Eq. 3})$$

where MM^{Theo} is the expected MM calculated from the amino sequence.

The compaction index (CI) allows comparing the degree of compaction between two proteins in a way that is independent of MM. It can be calculated according to [48]:

$$CI_{Rs} = \frac{R_s^U - R_s^{\text{Obs}}}{R_s^U - R_s^{NF}} \quad (\text{Eq. 4})$$

where R_s^{Obs} is the experimental value for a given protein, and R_s^U and R_s^{NF} are the reference values calculated for an unfolded protein, according to equation 3, and for a globular folded protein, according to equation 2. CI increases with increasing compaction and, in principle, varies between 0 and 1. However, values slightly higher than 1 and lower than 0 can be obtained due to the empiric nature of the reference curves that were derived by fitting procedures on experimental results.

When performing estimations of R_s using SEC, it is recommended to check if the elution profile is not (strongly) dependent on pH, composition and ionic strength of the buffer. A certain dependence on ionic strength is however expected, given the typically high content in charged residues of IDPs. By this relatively simple analysis, along with the comparison of the observed R_s with the values expected for an unfolded and a natively folded form, one can draw conclusions about the degree of compaction of a protein under study.

Alternatively, the R_s can be determined by SEC coupled to multiple angle laser light scattering (SEC-MALLS). In this case, the R_s and the MM are directly measured and not inferred from a calibration curve. While SEC relies on fast protein liquid chromatography (FPLC), SEC-MALLS requires the more sophisticated, and hence more expensive, high-pressure liquid chromatography (HPLC) system.

The hydrodynamics properties of proteins can also be estimated *via* dynamic light scattering (DLS). DLS measures the translational diffusion coefficient of the protein, and its R_s is inferred from the Stokes-Einstein equation, which gives the diffusion coefficient of a spherical particle with radius R_s in a medium of viscosity ν and at a temperature T :

$$D = \frac{k_B T}{6\pi\nu R_s} \quad (\text{Eq. 5})$$

where k_B is the Boltzmann's constant.

It should be emphasized, however, that once a protein is shown to possess an R_s much higher than the value expected for a folded protein of the same size, this does not constitute a proof of its conformational disorder. In fact, folded proteins with elongated shape (for instance coiled-coils) can also have larger apparent hydrodynamic radii. Besides, a higher-than-expected R_s could also arise from the presence of structured oligomers, rather than from disordered conformations. To discriminate between these two possibilities, one has to experimentally determine the MM of the protein (which can be done for instance by mass spectrometry (MS) under native conditions, by SEC-MALLS and by analytical

ultracentrifugation, AUC), and/or to determine the content in secondary and tertiary structure (which can be done by circular dichroism (CD) and/or NMR spectroscopy).

3.2. Analytical ultracentrifugation (AUC)

The physics of molecular sedimentation in a centrifuge is very well characterized. Sedimentation consists in a balance between centrifugal, buoyant and friction forces. The resultant sedimentation velocity depends on the difference between protein and medium densities, and on protein MM and friction coefficient. Sedimentation can be described as the evolution of protein concentration c , with time t , and radial position, r .

AUC experiments typically require protein samples to be analyzed at various concentrations (in the 1 – 5 mg range, depending on the absorption coefficient). There are two types of AUC experiments, sedimentation velocity (SV) and sedimentation equilibrium (SE).

In a SV experiment, a moving boundary, separating cushions of protein solutions, is formed upon application of a strong centrifugal field. Typically, data are acquired overnight. The radial concentration of particles over time during sedimentation is given by the Lamn equation, also referred to as the transport equation. For a homogeneous diluted solution of a non-interacting macromolecule in a diluted buffer (*i.e.* an ideal solution):

$$\frac{\partial c}{\partial t} = D \left[\frac{\partial^2 c}{\partial r^2} + \frac{1}{r} \frac{\partial c}{\partial r} \right] - s\omega^2 \left[r \frac{\partial c}{\partial r} + 2c \right] \quad (\text{Eq. 6})$$

where c is the protein concentration at a radial distance r , t is the time, D is the diffusion coefficient, s is the sedimentation coefficient, and ω is the rotor speed. s is defined as the ratio of the macromolecule velocity (cm/s) to the centrifugal field ($\omega^2 r$ in cm/s^2) ratio. s is expressed in Svedberg unit S ($1 S = 10^{-13}$ s).

The Lamn equation has, however, no analytical solution and needs a theoretical model to be assumed. Experimentally measured sedimentograms can be fit on theoretical ones, thereby allowing determination of both D and s . This information allows the determination of the MM of the macromolecules in solution using the Svedberg equation.

$$\frac{s}{D} = \frac{M(1-\rho\nu)}{RT} \quad (\text{Eq. 7})$$

where ν is the partial specific volume of the solute, ρ is the solvent density, M is the solute mass, R is the universal gas constant and T is the temperature in Kelvin. The partial specific volume ν of proteins can be estimated from amino acid composition using the program SEDNTERP (available from <http://www.jphilo.mailway.com/>).

The Svedberg equation also relates s to the R_S and M :

$$s = \frac{M(1-\rho\nu)}{N_A 6\pi\eta R_S} \quad (\text{Eq. 8})$$

where η is the solvent viscosity and N_A is the Avogadro's number. The sedimentation coefficient is generally expressed as $s_{20,w}$ after correction for solvent density and viscosity in relation to the density and viscosity of water at 20°C ($\rho_{20,w} = 0.99832$ g/mL and $\eta_{20,w} = 1.022$ mPa/s). The obtained R_S value can be compared with the minimum theoretical hydrodynamic

radius R_0 compatible with the non-hydrated volume, V , of the particle through the frictional ratio f/f_0 :

$$V = \frac{4}{3}\pi R_0^3 = \frac{Mv}{N_A} \quad (\text{Eq. 9})$$

$$R_S = (f/f_0)R_0 \quad (\text{Eq. 10})$$

The frictional ratio f/f_0 depends on the hydration, surface roughness, shape, and flexibility of the sedimenting species. It is a measure of the degree of asymmetry in a molecule with respect to a sphere of the same mass. Using $v = 0.73$ mL/g as a typical value for proteins, Manon and Ebel have extracted f/f_0 values for globular proteins, random coil (RC)-like IDPs and premolten globule (PMG)-like IDPs [49, 50], where PMGs are forms that are more compact than RCs, but still less compact than globular or molten globule proteins, and conserve some residual transiently populated secondary and/or tertiary structure [35, 51]. IDPs have f/f_0 values significantly larger than those of globular proteins. In addition, while for IDPs the frictional ratio increases with protein size, there is only a very slight increase of f/f_0 with MM in the case of globular proteins. Typical values of f/f_0 for globular proteins are 1.25 ± 0.5 [49]. Note, however, that f/f_0 values higher than the latter ones can be obtained for structured and yet non-globular (*i.e.* elongated) proteins such as coiled-coils, for instance [52, 53].

Because the shape of IDPs is anomalous and because SE allows estimating the MM independently of protein shape, it is a method of choice to assess, in a rigorous way, the MM of IDPs. In a SE experiment, a steady-state concentration gradient develops as the tendency of the molecule to sediment in the centrifugal field is counterbalanced by its tendency to diffuse according to the concentration gradient so established. In this case a smooth gradient is formed instead of distinct boundaries. Equilibrium is typically reached after 24 hours. Usually nine equilibrium profiles are obtained using three different concentrations and three different rotor speeds. Different approaches have been used to derive the equations describing sedimentation equilibrium [54-57]. The final, equilibrium concentration gradient provides accurate estimates of solute mass. SE provides s/D values (and hence M values).

Unfortunately, measuring both SE and SV is not feasible in the same experiment. While equilibrium needs relatively low speed, in order to get a nice resolution on the balance between sedimentation and diffusion, velocity needs higher speed, in order to maximize the friction, which gives information on particle shape. Thus, an AUC experiment commonly consists of two distinct experiments, namely SE and SV. Fortunately, both experiments can be carried out using the same equipment. Finally, the experimental set-up of AUC is very versatile (centrifugation speed, cuvette path length...), which makes this technique well suited for a wide range of systems. Thus, analyzing the sedimentation of protein particles is a powerful approach to assess their hydrodynamics properties.

SV and SE analyses by AUC are also particularly well suited for rapid analyses of IDP supramolecular complexes. In a typical experimental design, SV experiments give information on the shape of the protein and how it changes upon binding, while SE yields the stoichiometry of the interaction and provides good estimates of the interaction strength.

3.3. Small-angle X-ray scattering (SAXS)

X-ray diffraction methods are often used in structural studies, in order to get information on molecule geometry. Since IDPs are highly entropic systems, those methods fail to provide information at the atomic detail. However, the irradiation of particles in solution gives rise to an X-ray scattering pattern containing low-resolution information (around 20 Å) on the geometry of the molecules, even when the latter features high conformational entropy. One of the strong advantages of SAXS, compared to experiments of similar resolution, is the very few restrictions it imposes on sample preparation. Indeed, SAXS is compatible with a wide range of particle size and buffer types, provided that the sample is mono-disperse in those conditions. The main bottleneck of SAXS resides in the need for access to a synchrotron facility or for the availability of an in-house SAXS set-up that includes an X-ray source, a detector, and a 2-3 meters long tube under vacuum, which ensures light scattering between source and detector. The use of synchrotron radiation dramatically speeds up the process of data collection and yields data of better quality compared to SAXS data obtained using laboratory equipment.

In a SAXS experiment, the sample is irradiated by X-rays and diffracted beams diffuse on hundreds of cm (typically from 1 to 3 m) before hitting the detector. The higher the sample-detector distance, the lower the scattering angle (2θ), and hence the lower the resolution. The resulting scattering curve depends on the size and shape of the protein. A typical diffusion pattern is represented by a curve of X-ray intensities (usually represented on a logarithmic (Log) scale) at different values of the scattering vector, q , with $q = 4\pi\sin(\theta)/\lambda$ where λ is the wavelength and 2θ is the scattering angle, *i.e.* the angle between the incident beam and a deviated beam. This two-dimensional curve is obtained by radial integration of X-ray intensities around a central point corresponding to the incident, non-deviated beam. Intensities at high angles (“high” relatively to the angle range used in SAXS) contain information on the shape of the particles in solution. However, the region of the scattering curve at very low angles also contains information on the particles and their global organization in solution. When attempting at inferring structural parameters from a scattering curve, one should keep in mind that SAXS is very sensitive to inter-particle interactions. The latter have the potential of dramatically distorting the structural parameters that can be extracted after processing SAXS curves. IDPs, and especially those belonging to the class of strong-electrolytes, are strongly subjected to inter-particle interactions, which makes their characterization by SAXS very challenging. Reducing inter-particle interactions might be tackled by screening different buffers (salt concentrations, pHs, additives), in order to identify the conditions leading to a mono-disperse solution. Sample heterogeneity, due, for instance, to shape and mass distribution asymmetry, can be revealed by careful analysis of SEC elution peaks using quasi-elastic and multi-angle light scattering techniques. Such features should be considered during SAXS data analysis [58].

Typically, 30-50 µL protein samples are used in a concentration range of 1 to 10 mg/mL. It is recommended to collect data at various protein concentrations, so as to rule out any possible attractive and/or repulsive inter-particle interactions. It is also recommended to use *strictly* the same buffer, in which the protein is prepared when recording background scattering. Experimental scattering curves at low and high concentrations can then be combined to yield a single, *merged curve*. While intensity data collected at the lowest

concentration are assumed to correspond to a monodisperse sample and allow a correct description of the protein at low angles, intensities obtained at high protein concentrations provide a better signal-to-noise ratio at high angles. Thus, the merged curve contains information both at high angles and at low angles, which allows eliminating data possibly suffering from inter-particle interactions while keeping data at the highest possible resolution [59]. However, as this approach assumes that the lowest concentration corresponds to a monodisperse sample, it is recommended to carry out different measurements at different low-medium concentrations. Intensities measured at those concentrations are assumed to correspond to a monodisperse sample, if all the scattering data yield similar R_g and normalized I_0 values, as estimated by the Guinier approximation (see below). Extrapolation of intensities at zero concentration could also be an alternative to the merging procedure to infer I_0 and R_g values not biased by inter-particle interactions. However, this method assumes that intensities evolve linearly with protein concentration and predicts intensities out of the protein concentration range under study. Thus, it is recommended to use intensities that directly result from measurements (*i.e.* merged curve) when data at low concentration meet quality criteria, and rather perform extrapolation when data strongly suffer from inter-particle interactions.

From the scattering curve one can derive the radius of gyration, R_g , and its distribution (for IDPs), the maximal intramolecular distance and the MM. The R_g parameter is the average root-mean-square distance to the center of density in the molecule weighted by the scattering length density. It is therefore different from R_s , which corresponds to the radius of a spherical particle with the same diffusion coefficient as the macromolecule.

Thanks to the Guinier approximation (Eq. 11), the average R_g of the particles in solution can be estimated from the Guinier representation, where $\ln(I)$ is plotted as a function of q^2 . For small q values ($q < 1.3/R_g$ for a spherical particle), the Guinier approximation allows to fit the curve by linear regression. The slope of the line is a function of R_g (slope = $-R_g^2/3$) and the intensity at $q = 0$ is proportional to the mean MM of the scatterer. The R_g and forward intensity at zero angle I_0 can be determined according to the following equation:

$$\ln[I(q)] = \ln[I_0] - \frac{q^2 R_g^2}{3} \quad (\text{Eq. 11})$$

From the mean R_g one can infer information on the degree of compaction of the molecule in solution. The experimentally determined R_g (R_g^{Obs}) can be compared to the values expected for a folded, chemically denatured or natively unfolded form. The theoretical value of R_g (in Å) of a protein composed by N amino acids can be calculated from the following equations describing the behavior of a set of natively folded (R_g^{NF}) [60], chemically denatured (R_g^U) [61] or intrinsically disordered (R_g^{IDP}) [60] proteins of given length:

$$R_g^{NF} = \sqrt{\left(\frac{3}{5}\right)} 4.75 N^{0.29} \quad (\text{Eq. 12})$$

$$\log(R_g^U) = 0.58 \log(N) + 0.40 \quad (\text{Eq. 13})$$

$$R_g^{IDP} = R_0 N^v \quad (\text{Eq. 14})$$

where R_0 is 2.54 ± 0.01 and v is 0.522 ± 0.001 .

As in the case of R_s , the CI allows comparing the degree of compaction of a given IDP, through comparison of the observed R_g to the reference values expected for a globular or unfolded protein of identical MM. The CI referred to the R_g can be calculated as follows [48]

$$CI_{Rg} = \frac{R_g^U - R_g^{Obs}}{R_g^U - R_g^{NF}} \quad (\text{Eq. 15})$$

where R_g^{Obs} is the experimental value for a given protein, and R_g^U and R_g^{NF} are the reference values calculated for an unfolded protein, according to Eq. 13, and for a globular folded protein, according to Eq. 12. As for the R_s -based CI , this index increases with increasing compaction.

Furthermore, comparison of the R_g with the R_s also allows discrimination between globular proteins, RCs and PMGs. The ratio R_g/R_s should in fact be $\sqrt{3/5}$ for globular proteins, close to 1.0 for PMG-like IDPs and about 1.5 for extended, RC-like IDPs [62].

The characterization of the global flexibility of the system can also be addressed using the Kratky representation ($q^2 \times I(q)$ versus q). Because SAXS intensity is proportional to the MM and the angular scale depends on the protein size, comparison of different curves in Kratky representation should be preferably performed using the dimensionless Kratky plot, where q is multiplied by R_g and $I(q)$ is divided by I_0 [63]. The scattering curve of a typical folded protein in a dimensionless Kratky representation has a typical bell shape and is characterized by an increase of $I(q)/I_0 \times q^2$ from 0 to 3×10^{-1} for q ranging from 0 to $\sqrt{3}$, followed by a decrease of $I(q)/I_0 \times q^2$. Any deviation of this maximum in either $I(q)/I_0 \times q^2$ or q can arise from flexibility or particle anisotropic shape. Fully unfolded proteins and IDPs are characterized by a continuous increase of $I(q)/I_0 \times q^2$ in Kratky representation (**Figure 3A**).

This feature readily allows globular conformations (ordered and molten globule-like) to be distinguished from non-globular ones. However, it is noteworthy that the Kratky plot does not yield only a simple binary information, being able to discriminate IDPs by their extent of compaction (**Figure 3B**). In addition, PMGs can give rise to Kratky plots featuring a small bump followed by a plateau (for an example see [64]).

Combining SEC and SAXS is an appealing alternative, as it enables ensuring that measurements are performed on a non-aggregated, mono-disperse sample [65]. After sample loading on a SEC column, the eluted fractions are directly irradiated with X-rays. Thus, the elution of the protein is followed by both UV absorbance and scattering intensities. Those intensities are then used for subtracting buffer contribution to all measurements made on the column resolving volume. Determination of R_g and I_0 using the Guinier approximation is then computed for all the frames. Protein elution can be followed by both I_0 and absorbance. The evolution of R_g over protein elution is then used in order to ascertain the possible presence of different species eluted with different ratio at different time. Intensities of frames of similar R_g can be averaged and the resulting curve is then suitable for further analysis and modeling. SEC coupling is a powerful way to measure intensities on single species for samples consisting of a mixture of interacting proteins. If the different eluted particles are well resolved, the final scattering curve can be obtained without using any deconvolution methods.

Notably, SAXS can also be used to study complexes of IDPs bound to their partner(s). Deconvolution methods can be useful to characterize systems of well-defined composition [66].

Although SAXS is very well suited to study IDPs [67-70], it should be emphasized that, since IDPs consist of particles of different shape, their scattering curve is an average of the contributions of all the particles. Thus, describing IDPs using a single value of R_g is not appropriate. R_g distributions provide a much more meaningful and realistic description of

IDPs (see section 8 on modeling IDPs as conformational ensembles). In the same vein, one should be aware that attempting at modeling IDPs using *ab-initio* shape reconstruction is a non-sense, as IDPs inherently exist as ensembles of conformers of different shape and they cannot be described by a single conformer. By contrast, envelopes can be calculated for complexes made of IDPs bound to their folded partner(s), provided that the complex possesses at least a folded core (for an example see [71]).

4. Assessing protein secondary structure content

Various methods enable assessing the secondary structure content of proteins. Below, we provide a description of two commonly used methods, *i.e.* CD in the far-ultraviolet (UV) region and infrared (IR) spectroscopy. NMR and site-directed spin-labeling (SDSL) EPR spectroscopy, two powerful techniques to assess the (dis)ordered nature of proteins, are not included in this section, as they are discussed in two distinct reviews within this special issue.

4.1. Circular dichroism (CD) spectroscopy

CD spectroscopy is a quick and simple spectroscopic technique that provides information on the secondary and tertiary structure of proteins, as well as on its temperature dependence [72]. It is well suited to study IDPs [73]. This technique requires protein amounts in the microgram range and relies on the interaction of circularly polarized light with optically active (chiral) compounds. The signal that is measured in CD reflects the differential absorption of left- and right-handed circularly polarized light. As a result of the differential absorption, the light that has crossed the sample is no longer circularly polarized but elliptically polarized. Depending on the spectropolarimeter used, the differential absorption can be given in ellipticity (θ) or in difference between left- and right-handed molar attenuation coefficient ($\Delta\epsilon$). Those two different units both refer to the same phenomenon and are related to each other according to Eq. 16:

$$\theta = 3300 \times \Delta\epsilon \quad (\text{Eq. 16})$$

However, the standard unit used to describe proteins ellipticity is MRE (Molar Residue Ellipticity). The latter can be calculated as follows:

$$MRE = \frac{\theta}{(lcn)} = \frac{(3300\Delta\epsilon)}{(lcn)} \quad (\text{Eq. 17})$$

where the ellipticity θ is given in degrees, $\Delta\epsilon$ is the difference of molar attenuation coefficient between the left- and right-handed radiation, the path length l is given in cm, protein concentration c is expressed in $\text{dmol}\cdot\text{cm}^{-3}$ and the number of residues n has no unit. Thus, the final unit corresponds to $\text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$.

There are two types of optically active chromophores in proteins, side groups of aromatic amino acid residues and peptide bonds [74].

CD spectra in the far-UV region (peptide part of the spectrum, 180-250 nm) reflect the peptide bond environment and indicate the content of secondary structure in the protein molecule. The main contributor to the far-UV CD spectrum of a protein is the excitation of the partially delocalized electrons of the peptide bonds. Isolated amino acids, except glycine, also show a CD signal in this wavelength range. Thus, side chains of amino acids, tryptophan in particular, contribute to the protein CD spectrum, as well [75]. This contribution is

typically much smaller than that of the protein backbone but could become significant in proteins of low helical content. In the far-UV region, the different forms of regular secondary structure found in proteins give rise to characteristic CD spectra. Any CD spectrum can be resolved into components corresponding to main secondary structure types (α -helix, β -sheet, random etc ...). The CD spectrum of an IDP is typically dominated by the random-coil features, with a largely negative ellipticity near 200 nm, a negligible ellipticity at 222 nm and by an ellipticity close to zero at 185 nm (**Figure 4A**). The CD spectra of many unordered polypeptides show a remarkable resemblance to that of poly(L-proline) II conformation (P_{II}), that shows a strong negative band at 206 nm and a weak positive band at 225 nm [76]. Other unordered polypeptides show a strong negative band below 200 nm and a negative rather than positive shoulder in the 220-230 nm wavelength region. This type of spectrum reflects a higher extent of disorder as compared to that of P_{II} -like spectra. Far-UV CD spectroscopy also enables discriminating RCs from PMGs, based on the ratio of the ellipticity values at 200 and 222 nm (**Figure 4B**) [35].

A number of algorithms exist, which use the data from far UV CD spectra to estimate the secondary structure composition of proteins. An online server, DICHROWEB [77, 78], has been developed, which allows CD data to be analyzed by various algorithms with a choice of databases. It is also possible to obtain estimates of the α -helical content of peptides and proteins from more limited data by using the values of the CD signals at 208 nm or at 222 nm. However, it must be emphasized that all these estimates are to be treated with caution, as they do not provide reliable quantitative estimations and are, rather, more appropriate for comparative analyses.

4.2. Fourier transform infrared (FTIR) spectroscopy

FTIR spectroscopy also provides quantitative information on the secondary structure motifs of proteins through decomposition of the Amide I absorption band (1700-1600 cm^{-1} region) by curve fitting procedures. This band originates from the C=O stretching vibration of the peptide group, whose frequency is sensitive to the protein conformation. To resolve the Amide I band constituents, Fourier self-deconvoluted and second-derivative spectra can be calculated from the raw spectrum [79].

Because RC and α -helix contributions largely overlap in H_2O but not in D_2O , spectra should be acquired in both solvents, in order to discriminate between disordered and structured proteins. The absorption spectra of proteins in aqueous solution can be measured despite the high absorption of water in the amide I region, using highly performing FTIR spectrometers and algorithms for the subtraction of the solvent spectrum. Typically, protein solutions at a concentration ranging from 1 to 15 mg/mL are required for measurements in water, while concentrations can be decreased down to 0.3-0.5 mg/mL when measurements are performed in D_2O . However, the total protein amount is about 10–100 μg , since only 10–20 μL are required for a single measurement.

The different structural components can be identified, and their relative intensities can be quantified by Gaussian fitting from the ratio between band area and total Amide-I area (**Figure 5**). In D_2O , peaks at $\sim 1645 \text{ cm}^{-1}$ are indicative of RC structure. β -turns can be found at 1671 cm^{-1} , while 3_{10} helices occur in the range 1645 to 1634 cm^{-1} . Peaks at 1652 cm^{-1} correspond to α -helices. Intramolecular β -sheets display two absorption bands at 1630 and

1679 cm⁻¹ of different intensity.

In contrast to CD, FTIR spectroscopy enables a better deconvolution and quantification of the various structural components. In addition, it also allows discrimination of helices with different extents of flexibility [80]. This information is encoded by minor shifts in band position. In particular, higher wavenumbers characterize short and flexible helices, while lower wavenumbers are associated to long and rigid structures. Like CD, FTIR spectroscopy offers the possibility to explore the temperature-dependence of protein content in secondary structure. When recording spectra in water, the temperature needs to be carefully controlled since the infrared absorption of water displays strong temperature dependence [79].

Notably, FTIR spectroscopy is one of the few techniques that allow studying dry proteins (*i.e.* protein films), obtained by evaporation of protein solutions. This can be achieved through attenuated total reflection (ATR) measurements, a method in which the sample is placed on a support having a refractive index higher than the sample. The infrared beam reaches the ATR-plate/sample-interface at a larger angle than the limit angle of total reflection. Under these conditions, the beam is totally reflected by the interface and penetrates, as an evanescent wave, into the sample where it can be absorbed. The optical path of the penetrating beam, which is of the order of the IR wavelength, depends on the incident wavelength and angle, and on the refractive indices of the sample and ATR support. After one or several reflections, the attenuated beam is collected at the detector. This method requires that very small amounts (*i.e.* µg range) of the protein under study are placed on the ATR support and allowed to evaporate (for an example see [43]). An interesting advantage of ATR measurements resides in the possibility of studying the interactions, and hence possible ensuing conformational changes, of proteins with surface and synthetic membranes.

5. Assessing protein tertiary structure

5.1. Near-UV CD spectroscopy

CD spectra in the near-UV region (250-350 nm, also called the “aromatic region”) reflect the environment of aromatic amino acid residues and, consequently, characterize the tertiary structure of a folded protein. Each aromatic residue has its own characteristic dichroic bands. Signals in the 250-270 nm region arise from phenylalanine residues, while signals in the 270-290 nm and 280-300 nm regions arise from tyrosine and tryptophan residues, respectively. Disulfide bonds give rise to broad, weak bands throughout the near-UV spectrum [81]. The near-UV CD spectrum, therefore, provides a detailed fingerprint of the tertiary structure around these specific residues. Proteins with rigid tertiary structure are typically characterized by intense near-UV CD spectra, which reflects the restricted orientation of the individual aromatic side chains. By contrast, the near-UV CD spectra of IDPs generally show low intensity and low complexity, reflecting their large conformational freedom (**Figure 6**). Although near-UV CD spectroscopy requires larger amounts of sample compared to far-UV CD (typically ten times more), it remains an affordable technique, also considering that samples can be recovered and used for further analyses. Although IDPs are typically depleted in aromatic residues, near-UV CD has been used to probe conformational transitions of this class of proteins (as an example see [82]).

5.2. Differential scanning calorimetry (DSC)

Protein structures that undergo thermally-induced conformational changes present a redistribution of non-covalent bonds accompanied by the absorption of heat. Differential scanning calorimetry (DSC) is an analytical technique that measures this heat uptake in terms of molar heat capacity of sample, as a function of temperature. DSC experiments are carried out on protein solutions heated at a constant rate. DSC instruments can operate with very small quantities of sample, approximately 100 μL or 100 μg . In a DSC experiment, the thermal unfolding results in a temperature profile of the heat capacity of the system, which serves as a structural “fingerprint” of the protein [83, 84]. DSC provides measures of unfolding enthalpy (ΔH), changes in heat capacity (ΔC_p) of denaturation, and thermal transition midpoint (T_m). In the case of ordered proteins undergoing a two-state transition from the native to fully unfolded state, the thermal profile shows a characteristic peak indicative of cooperative transitions involving its secondary and tertiary structure [85]. Helix-to-coil transitions in peptides also give rise to heat capacity peaks. By contrast, the lack of a heat absorption peak in a calorimetric curve (*i.e.* a flat profile) is indicative of the absence of ordered structure [86-88]. Such a gradual, non-cooperative denaturation transition, or “uphill” unfolding transitions, is typical of IDPs, although it can also be found in very small/simple structured proteins [89]. DSC not only allows recognizing IDPs from their typically flat profile, but also enables recognizing the presence of residual (e.g. intrinsic) structure in IDPs. This is well illustrated in the case of *Bacillus pasteurii* UreG, whose smooth DSC thermogram is consistent with a continuous and non-cooperative loss of residual secondary structure upon increasing temperature (**Figure 7**) [90].

DSC is especially convenient for proteins devoid of aromatic amino acids, for which information on the lack of tertiary structure cannot be obtained by other spectroscopic methods such as near-UV CD and intrinsic fluorescence [91]).

5.3. Fluorescence spectroscopy

Fluorescence spectroscopy is a widely used technique to study protein structural transitions. Aromatic residues are intrinsic fluorescent chromophores with different quantum yield (tryptophan > tyrosine > phenylalanine). The aromatic residue most commonly exploited for fluorescence spectroscopy studies is tryptophan, since the quantum yield of phenylalanine fluorescence is extremely low and tyrosine fluorescence is strongly quenched mostly due to its ionization, or location near polar groups (such as amide or carboxyl groups), or energy transfer to tryptophan [92]. Tryptophan has an absorption maximum close to 280 nm and an emission maximum that is highly dependent on the polarity of its environment. In particular, a tryptophan in a non-polar environment (*i.e.* buried in the protein core) has an emission maximum close to 320 nm, while in a polar environment (*i.e.* solvent exposed) has an emission maximum close to 350 nm, a phenomenon known as solvatochromism [93]. In addition, the intensity of tryptophan fluorescence is dependent on interactions with neighboring groups, which can lead to transfer of the excitation energy. This results in reduction of fluorescence quantum yield and in smaller emission intensity, a phenomenon known as fluorescence quenching [93]. Quenching is the result of a variety of molecular interactions requiring molecular contact between the fluorophore and the quencher [93]. Therefore, the solvatochromic and the quenching effects can be exploited to probe how

variations in experimental conditions affect polarity/hydrophobicity of tryptophan environment and its overall accessibility/permeability to quenchers. As such, variations in fluorescence intensity under varying experimental conditions (such as the presence or absence of ligands) provide hints on the gain/loss of ordered structure in the vicinity of the fluorophore(s) (**Figure 8**).

Beyond intrinsic fluorescence, it is also possible to use extrinsic fluorescence probes, such as 1-anilino-8-naphthalenesulfonic acid (ANS), whose fluorescence in water is almost negligible, while strongly enhanced in hydrophobic environments. These compounds can enter hydrophobic cavities of protein structures and strongly emit fluorescence when bound to a non-polar surface. Thanks to this property, they have been widely used to study protein conformational changes (see [94] and references therein cited). In particular, extrinsic fluorescent probes can be used to assess the presence of hydrophobic pockets within globular or partially folded proteins, such as molten globules and PMGs, and to discriminate between them and extended forms (*i.e.* RCs).

A related approach is differential scanning fluorimetry (DSF). DSF monitors thermal unfolding of proteins in the presence of a fluorescent dye that has the same properties as ANS. When the fluorescence intensity is plotted as a function of the temperature, the apparent melting temperature (T_m) of the protein can be derived from the inflection point of the resulting sigmoidal curve. This method, also known as thermal shift assay (TSA), is routinely used for structured proteins to identify the best (in terms of protein stability) storage buffer. In striking contrast, IDPs, which do not possess well-defined hydrophobic cavities, are characterized by a fluorescence profile that is rather flat and temperature-independent. This behavior therefore enables IDPs to be readily distinguished from structured proteins.

5.4. Electrospray ionization (ESI) mass spectrometry (MS)

ESI-MS relies on the generation of gas-phase ions through nebulization of a liquid sample in the presence of a strong electric field and accurate determination of the mass-to-charge ratio (m/z) for each detected ion [95, 96]. Instrumental progress and, most importantly, development of nano-ESI equipments, has enabled maintenance of non-covalent interactions, such as those responsible of protein folding and binding, under ESI conditions, driving growth of an entirely new field named “native-MS” [97]. By this approach, structural investigation of macromolecules is conjugated with the exceptional analytical power of MS in dissecting complex samples, without averaging over the molecular population. This aspect makes native-MS a very valuable tool for the investigation of heterogeneous systems and conformational ensembles [98]. One technical limitation of ESI-MS is its poor solvent compatibility, requiring volatile buffers and accurate desalting. Aqueous solutions of ammonium acetate are typically used for native MS. Proteins prone to precipitation are sometime challenging to analyze under these conditions. However, highly diluted samples can be used (*i.e.* μM protein concentrations), reducing aggregation risk and sample consumption.

A large amount of evidence suggests that the extent of protein ionization under electrospray conditions is strongly affected by structural compactness at the moment of transfer from solution to the gas phase [99]. Thus, protein charge-state distributions (CSDs) by native-MS can deliver important structural information. The higher the structural compactness, the lower the average net charge will be for each protein conformer populating a

given molecular ensemble [100] (**Figure 9A**). This observation has been rationalized by the influence of native-like intramolecular interactions on the protonation propensity of ionizable sites [101, 102]. These structural features, including zwitterionic configurations, contribute lowering net charge of folded proteins in the gas-phase and preserving native-like structures. Simulations support this conclusion even when proton mobility is allowed for [103, 104].

CSDs can be deconvoluted by Gaussian-fitting procedures, upon transformation from $x=m/z$ to $x=z$ of the abscissa axis (**Figure 9B**) [105]. The average charge state of each component yields an estimate of the solvent-accessible surface area (SASA) for each resolved protein conformer (**Figure 9C**) [106-109]. Such a charge-to-SASA relation seems to hold regardless of protein size and protein conformational state, offering a link to a quantitative structural parameter that can be used as a constraint for computational modeling of conformational ensembles [110, 111]. A SASA-based compaction index has been developed to compare the extent of compaction in polypeptide chains of different length and to provide a continuous, quantitative parameter for IDP classification [112].

Ion mobility (IM) spectrometry is a separation technique that is increasingly used in combination with mass spectrometry (IM-MS) to analyze conformation of desolvated analyte ions, including macromolecules [113, 114]. The mobility of an ion is a measure of its velocity in a neutral buffer gas under the influence of a weak electric field. The number of collisions with the gas, together with the charge and shape of the ion, leads to an arrival time distribution, which can be used to determine the rotationally averaged collision cross-section (CCS) of the selected ion. The CCS serves as an indicator of the shape of the molecule in the gas phase [113]. Thus, this technique adds a second dimension to the ion sorting process inside a mass spectrometer, increasing resolution and yielding the two different descriptors m/z and CCS, for each isolated ion. ESI-IM-MS allows resolving possible conformational heterogeneity even within a given charge state.

ESI-IM-MS is a useful tool for the structural investigation of IDPs and their non-covalent complexes [113]. Although normally folded proteins have been shown to retain solution-like structures upon ESI, IDPs can undergo even dramatic structural rearrangements in the gas phase [115-118]. Nonetheless, folding-upon-binding events have been detected, showing good consistency with CSD analysis and solution spectroscopy data. One example is provided by α -synuclein binding to dopamine and epigallocatechin-3-gallate, two small molecules inhibiting the fibrillation process through opposite effects in terms of protein compaction [119]. Analogously, IM-MS allowed documenting conformational transitions towards more compact states in p53 and MDM2 upon interaction with small ligands [120].

The measles virus N_{TAIL}-XD complex exemplifies a challenging type of assembly for ESI-MS, as its interface is characterized by almost exclusively hydrophobic interactions [111]. Although only a fraction of the complex could be preserved, complex-specific peaks could be identified and their CSD was found to be bimodal, indicating the coexistence of at least two distinct conformations in the bound state differing in the extent of structural compaction. ESI-IM-MS identified distinct components, providing CCS values to filter structural models of the complex obtained by computational simulations. The results highlighted a large conformational freedom of the N_{TAIL} moiety when bound to XD and unveiled that not only the main interface, but also the neighboring, disordered appendages are dominated by hydrophobic interactions [111].

ESI-IM-MS combined with molecular dynamic simulations has also been applied to investigate the conformational heterogeneity of isolated domains from yeast and human cyclin-dependent kinase inhibitors (Sic1 and p27, respectively). Molecular-dynamics simulations in water gave results consistent with the distribution of SASA values derived from CSDs for the Sic1 KID domain [110] but failed to sample compact conformers for the p27 QT domain detected by IM, unless step-wise water evaporation was introduced in the simulation protocol [117]. These results suggest that even major structural rearrangements of highly disordered polypeptides can take place upon desolvation and indicate the need to account for these effects while interpreting IM data of IDPs. Nonetheless, consistent effects of charge patterning in IDPs have been captured by IM and solution techniques (such as partial proteolysis, CD, SEC and SAXS) showing increasing compaction along with increasing charge clustering in the protein sequence [117].

5.5. Hydrogen-deuterium exchange (HDX) reaction

The HDX reaction gives information on the protection of exchangeable hydrogens exerted by the protein structure [121]. Exchangeable hydrogens are those bound to electronegative atoms (not carbon), being continuously shuffled between protein and solvent. Protection can derive from shielding or hydrogen bonding. The detection of HDX reactions can be exploited to probe secondary and tertiary structural properties by measuring the rate of exchange of protonated proteins in D₂O. Side-chain exchangeable hydrogens typically exchange too fast to be structurally informative. Amide protons instead can be effectively monitored. Those located in α -helices, β -sheets, or buried in the hydrophobic core of a protein exchange slowly, due to hydrogen bonding [122]. In contrast, hydrogens located in “floppy loops” on the protein surface exchange much faster [123]. Therefore, fast exchange rates reflect local lack of ordered structure. Accordingly, amide hydrogens can be used as a structural probe to discriminate IDRs from regions with inherent structural features. Several instrumental techniques can be used to detect or to map exchangeable hydrogens. Among these we will consider NMR, FTIR spectroscopy and MS spectrometry.

NMR can give information about the rates of exchange of both amide and side chain hydrogens, provided that these have been assigned in the spectrum [124]. This is possible because ²H nuclei are silent, *i.e.* they do not give rise to a signal, in contrast to ¹H. FTIR spectroscopy can also be used to monitor HDX by following variations of the vibrational frequency of the carbonyl C=O group of the peptide bonds involved in hydrogen bonds [79].

Given the difference in mass between hydrogen and deuterium, MS is an alternative and convenient approach to monitor HDX reactions. HDX-MS not only allows recognizing IDPs and IDRs from their typically high exchange rates [125, 126], but also enables monitoring induced folding [125, 127] and aggregation processes [128, 129]. Sequence-specific information can be obtained by bottom-up approaches, *i.e.* digesting the protein by a protease and analyzing the exchange at the peptide level, or by top-down approaches, *i.e.* by subjecting the whole protein to fragmentation and analyzing exchange products at single-residue resolution [130-132]. Both approaches allow identifying regions of the IDPs that are engaged in intrinsic structure [133], in folding induced by binding to ligands [31, 134] (**Figure 10A**) or partner proteins [135] (**Figure 10B**), or in amyloid aggregation [128]. HDX-

MS also enables the peculiar modality of gas-phase isotopic exchange, which can provide insightful information on the role of solvent on IDP structure [136].

6. Approaches relying on protein labeling and/or site-directed mutagenesis

Once the prevalently disordered nature of a protein has been ascertained using the above-described approaches, additional insights into structure and dynamics of IDPs can be obtained using more demanding approaches that require protein labeling and/or site directed mutagenesis. Among these, heteronuclear NMR, which relies on isotopic labeling, is an extremely powerful approach. It gives access to parameters (such as chemical shifts, secondary structure propensity, relaxation rates, and residual dipolar couplings) that collectively enable a detailed, quantitative description of the structural and dynamic properties of IDPs [137]. SDSL EPR spectroscopy and fluorescence resonance energy transfer (FRET) are two commonly used approaches for the characterization of IDPs. While the former relies on the covalent grafting of a nitroxide spin label onto a cysteine residue [138, 139], FRET typically (although not exclusively) relies on the grafting of fluorophores onto selected protein sites. Less common, though promising approaches are vibrational spectroscopy of cyanylated cysteines and tryptophan triplet cysteine quenching (TCQ). Both methods require the presence of a cysteine residue at a desired protein site, and thus entail the need for site-directed mutagenesis, unless a unique cysteine residue already natively occurs at the proper site. In addition, TCQ also requires a tryptophan residue to be introduced into a unique protein site. Consequently, these two approaches, as all those based on labeled proteins, are quite demanding in terms of experimental set-up and better suited for an in-depth characterization of IDPs rather than for assessing their disordered state.

6.1. Fluorescence resonance energy transfer (FRET)

Fluorescence resonance energy transfer (FRET) is defined as the non-radiative process whereby an excited-state donor fluorophore (D) transfers energy to a proximal ground-state acceptor (A) *via* long-range dipole coupling. The phenomenon arises from the ability of the electromagnetic field of the excited oscillator to induce oscillation with the same frequency in the non-excited oscillator [92, 140]. FRET experiments can be designed by using either intrinsic chromophores (tyrosines or tryptophans) as donors and extrinsic chromophores as acceptors, or by using only extrinsic chromophores. Chromophores of the latter type emit light in the visible region and are covalently grafted to the protein, generally onto cysteine side chains.

The transfer of excitation energy between the donor and the acceptor occurs only provided that: *i*) the absorption (excitation) spectrum of the acceptor overlaps with the emission (luminescence) spectrum of the donor; *ii*) the donor and the acceptor are in close proximity of each other (*i.e.* the distance must not exceed a few dozens of Å); *iii*) the donor has a sufficiently high quantum yield; *iv*) the donor and the acceptor have a mutual spatial orientation compatible with an effective energy transfer.

FRET can be used as a “molecular ruler” to measure distances between the donor and acceptor. In fact, the efficiency of energy transfer, E , from the excited donor, D , to the non-

excited acceptor, A , located at a distance r from the donor is determined by the Förster equation [140]:

$$E(r) = \frac{R_0^6}{R_0^6 + r^6} \quad (\text{Eq. 18})$$

where R_0 is the Förster radius, *i.e.* the characteristic donor-acceptor distance that results in a transfer efficiency of 50%. The mean distance between dyes can be estimated from the mean FRET efficiency. If the donor and the acceptor are grafted both on the same polypeptide chain, then intra-molecular FRET can be measured.

From the above description it is quite obvious that FRET is a rather complex technique that requires the use of labeled proteins and dedicated equipment.

As already discussed for data provided by SAXS, it should be kept in mind that, since IDPs are extremely heterogeneous systems, describing them by a single, averaged parameter is poorly informative. Single-molecule approaches are convenient tools to resolve heterogeneity and to avoid the complications arising from ensemble averaging [141, 142]. Single-molecule FRET is an attractive method for probing long-range distances and dynamics in IDPs, because its sensitivity range (2 to 10 nm) corresponds well to the typical intramolecular distances of polypeptide segments from tens to hundreds of residues [143]. Thanks to the fact that the molecules are observed individually, structural heterogeneity can be resolved provided that the various conformations differ in transfer efficiency and that their interconversion dynamics are sufficiently slow, compared to the time resolution of the measurement. Therefore it is possible to single out the properties of folded and unfolded proteins co-existing and freely diffusing in the same solution, provided that the distinct subpopulations differ in their transfer efficiency and that they interconvert on a timescale slower than approximately a millisecond, the typical diffusion time through the confocal volume [143]. In this case, the unfolded state can be investigated without overlap from the signal of folded molecules. To further solve the conformational heterogeneity within the unfolded state ensemble, polymer-physical models (such as Gaussian chain, worm-like chain, excluded volume chain or a weighted Flory-Fisk distribution), which are meant to reflect the underlying chain statistics, can be used to obtain appropriate distance distributions.

The combination of FRET with nanosecond fluorescence correlation spectroscopy (nsFCS) allows probing the reconfiguration dynamics of IDPs. From the measured decay time of the fluorescence correlation curves, the diffusion coefficient, D , which determines the relaxation time, can be obtained. This relaxation time corresponds to the reconfiguration time, τ_r , of the chain segment between the two dyes.

Single-molecule FRET measurements can be carried out under a variety of experimental conditions. The changes in transfer efficiency induced by variations in the experimental conditions can be used to extract information on the corresponding changes in chain conformations. Using polymer-physical models, one can infer the underlying distance distributions, and the latter can be used to quantify the dimensions of the polypeptide chains in terms of either mean-squared intramolecular distances or R_g of the segments separating the donor and acceptor fluorophores. Typically, in IDPs the FRET efficiency increases in the presence of crowding agents, ligands and partner proteins [144] (**Figure 11**).

In FRET measurements, the donor and the acceptor can also be grafted individually

onto two distinct polypeptide chains, thereby allowing inter-molecular distances to be measured. In this case, it is possible to monitor binding events. Borgia and co-workers have recently used this approach to investigate the interaction between Prothymosin α and histone H1, two prototypic small IDPs of opposite electrostatic charge [145]. Through the generation of a set of variants of both proteins bearing each a single fluorophore grafted at different positions along the IDP sequence, they could probe various combinations of interacting sites on the two proteins and came up with a picture whereby the two proteins form an extremely disordered, though ultrahigh-affinity, complex [145].

FRET studies based on extrinsic fluorescent probes suffer however from possible labeling heterogeneity, structural alteration of the proteins resulting from the labeling *per se* (including protein compaction [146]), and perturbation of the binding process, due to the large size and hydrophobic nature of the fluorescent labels typically used in these studies. This can lead to possible discrepancies with respect to measurements that do not rely on the grafting of probes (for an example see [147]).

The combined use of tryptophan as donor and of nitrated tyrosine (which has maximal absorbance at ~ 350 nm, and does not emit light) as an energy acceptor in FRET allows overcoming some, if not all, these limitations. For these experiments, tyrosine residues have to be converted into their nitro-form, Tyr(NO₂), by reaction with tetranitromethane. The decrease of tryptophan fluorescence in the presence of Tyr(NO₂) provides a measure of the average distance between these residues. The approach can be used both to investigate the unfolded state of proteins [148] as well as protein-protein interactions [149].

6.2. Vibrational spectroscopy of cyanylated cysteines

Because the IR radiation has a high frequency and because the period of most molecular vibrations is quite short (tens to hundreds of fs), vibrational spectroscopy should be particularly useful in providing an accurate description of the conformational distribution of a dynamic protein. Unfortunately, the overlap in width and frequency of most naturally occurring protein functional groups makes IR spectroscopy only useful as a probe of global secondary structure without any site-specific information. The controlled incorporation of artificial functional groups that absorb in the clear region of aqueous biomolecular IR spectra, between 1900 and 2700 cm⁻¹, provides an excellent means to overcome this limitation.

A simple approach to a vibrationally active artificial amino acid is the cyanylation of cysteine residues, which converts the cysteine thiol group to a covalently attached thiocyanate (**Figure 12A**). Since IDPs are generally depleted in cysteines, the approach often requires the introduction of a cysteine at selected site by site-directed mutagenesis. Cyanylation of cysteine residues is achieved through reaction with Ellman's reagent (**Figure 12A**). This method enables the introduction of a site-specific vibrational probe in IDP systems of arbitrary size (for a review see [150]).

The side chain of a cyanylated cysteine contains a CN group, whose stretching vibration absorbs at 2155–2164 cm⁻¹ in typical protein environments (**Figure 12B**) [151]. The frequency of the SCN stretching vibration depends on the hydrogen-bonding character of its local environment. The frequency undergoes a blue-shift upon accepting a hydrogen bond at the nitrogen atom [151]. Water-exposed SCN bands are centered at ~ 2163 cm⁻¹, while SCN bands in hydrophobic solvents appear at lower frequencies (**Figure 12C**). Thus, the frequency

can be used to discriminate between water-exposed and water-excluded environments. As such, it can be used to document binding events whereby the SCN group becomes buried at an interface with a partner or a lipid membrane [152, 153]. The SCN linewidth depends on the picosecond dynamics of its local environment [151], where “local” means a length scale of only several Å, making this vibration a very site-specific probe. The structural dependence of the linewidth and the sensitivity to water exposure of the frequency make cyanylated cysteines an ideal probe of structural transitions in IDPs.

Accordingly, the CN stretching band of cyanylated cysteine introduced in an IDP was shown to successfully report partner-induced structural transitions in two ways: *i*) *via* frequency shift, due to changes in water exclusion, and *ii*) *via* lineshape change, due to dynamical changes associated with structure formation (**Figure 12D**) [154].

Since the S-bound nitrile moiety is easy to incorporate and sensitive to structural changes (including order/disorder transitions), and since it has a small size compared to any other covalently attached probe causing negligible perturbations [153], site-specific introduction of cyanylated cysteines is expected to become a widely-used approach to map binding-induced structural transitions in IDPs/IDRs.

6.3. Tryptophan triplet cysteine quenching (TCQ)

TCQ monitors the quenching dynamics of the tryptophan triplet state by a cysteine (or cystine) located within a polypeptide chain. Although this approach requires the introduction of a cysteine and of a tryptophan residue, usually achieved by site-directed mutagenesis, it offers multiple experimental advantages. It utilizes tryptophan phosphorescence instead of requiring the introduction of large extrinsic fluorophores such as those used in FRET. Furthermore, the experimentally observed lifetime simultaneously yields two basic parameters of IDPs. The static tryptophan-cysteine reaction rate constant k_R is closely related to the end-to-end distribution of the polypeptide chain, enabling comparisons with various theoretical polymer models, such as the random Gaussian or the wormlike chain. In addition, TCQ provides information about chain dynamics *via* the diffusive rate constant k_D , by which the ends come into contact. The tryptophan triplet state is quenched by close contact with cysteine and disulfides (at distances <0.4 nm). After nanosecond ultraviolet excitation of the unique tryptophan residue to the triplet state, this can contact the cysteine (or cystine) at a rate k_{D+} (by diffusion), and either be quenched with a rate q or diffuse away with a rate k_{D-} (**Figure 13**). The observed lifetime of the tryptophan triplet state, $1/k_{obs}$, is

$$\frac{1}{k_{obs}} = \frac{k_{D-} + q}{qk_{D+}} = \frac{1}{k_R} + \frac{1}{k_{D+}} \quad (\text{Eq. 19})$$

where $k_R = q(k_{D+}/k_{D-})$ is the reaction-limited rate and depends on the probability of forming a contact at equilibrium, k_{D+}/k_{D-} , and on the quenching rate upon contact, q . Because k_R is independent of solvent viscosity (η) while k_{D+} is a function of viscosity, k_R and k_{D+} can be experimentally obtained by measuring k_{obs} as a function of η . Plotting $1/k_{obs}$ as a function of viscosity therefore yields $1/k_R$ as the intercept and $1/\eta k_{D+}$ as the slope.

This experimental approach has been used to compare the mean size and internal chain dynamics of calcitonin-gene-related peptide (CGRP) and amylin (islet amyloid polypeptide, IAPP), two closely related IDPs of the calcitonin peptide family [155]. End-to-end contact formation and end-to-end distance data obtained in that study not only contribute to better

describe these two IDPs but are also extremely useful to validate results from molecular dynamics simulations of calcitonin family peptides in aqueous solution. TCQ holds promises for obtaining information on both intra- and intermolecular interactions. Its main drawback resides in the fact that it is rather demanding in terms of protein amounts (in the mg range). In addition, experiments have to be carried out in anoxic conditions so as to avoid intermolecular disulfide formation.

7. *In-vivo* assessment of disorder

One of the major challenges of structural biology pertains the possibility of exploring biomolecules in the context of their natural intracellular environment. The viscosity of the cytoplasm, the macromolecular crowding, as well as the presence of a huge number of potential interactors can all in fact exert a large impact on protein structure and dynamics. Characterizing proteins in living cells is even more relevant for IDPs, as one can question to which extent structural features determined *in vitro* can be reliably extrapolated to a crowded cellular environment. Do IDPs, which lack a stable 3D structure in isolation, fold in a cellular environment? Do interactions of IDPs with partners/ligands characterized under *in vitro* conditions bear resemblance to those occurring in the cellular milieu? Do post-translational modifications cause significant changes in the conformation, dynamics and partner recognition of IDPs? The recent advent of in-cell spectroscopic techniques has made it possible to start addressing these important questions.

In-cell NMR spectroscopy has gained popularity since it provides a means to analyze the conformational and functional properties of proteins inside living cells at atomic resolution [156]. This emerging technique has been pioneered by the Gary Pielak's and Philipp Selenko's groups (for a review see [157]). In-cell NMR was originally established in bacterial cells and relied on protein over-expression under conditions of isotopic labeling, followed by sample analysis in bacterial intact cells [158]. This approach, however, suffers from the inherent drawback related to the typically high expression levels achieved in bacteria, which might introduce biases and hence lead to erroneous conclusions. Studies relying on the introduction, *via* injection and/or electroporation, of labeled IDPs in *Xenopus laevis* oocytes [159, 160] or in other eukaryotic cell types [161] offer an attractive alternative to overcome this limitation.

The Ben Shuler's group has recently demonstrated the feasibility of probing protein dynamics from ms down to the ns regime in live eukaryotic cells using confocal single-molecule FRET spectroscopy [162]. Following this approach, the authors could determine the dimensions and submicrosecond chain dynamics of prothymosin α (ProT α), used as a model IDP. The study relied on the use of recombinant proteins, labeled *via* the thiol-maleimide chemistry by organic fluorophores in combination with three different acceptors. Microinjection was chosen for the introduction of the labeled proteins into cultured eukaryotic cells. Compared to transfection techniques, such as electroporation and cell-penetrating peptides, microinjection offered the advantage of enabling reproducible introduction of sample concentrations in the picomolar to low nanomolar range, with cell remaining viable beyond the required observation times. In addition, microinjection *i*) enabled the shortest delay between protein application and measurement, *ii*) was compatible with fluorescence

detection directly after injection, *iii*) offered the opportunity to inject molecules selectively into different cellular compartments, and *iv*) preserved cell adherence, thereby avoiding problems with positional drift of the cells during measurements. The authors successfully overcome the major technical limitations, namely design and reproducible delivery of labeled molecules and background (e.g. cellular) autofluorescence. This study can therefore be regarded as an asset for an expectedly broad use of in-cell smFRET in a next future.

Karthikeyan and co-workers have recently reported a study of in-cell EPR spectroscopy, relying on the use of a bioresistant nitroxide spin label [163]. Until recently, the use of nitroxide-based spin labels for in-cell studies has been hampered by their very short persistence in the cellular context and by the propensity of the protein–label bond to be cleaved by endogenous thiols. In this study, the authors designed and synthesized a maleimido-proxyl-based spin label (M-TETPO) (**Figure 14A**) that was found to be much more resistant towards reduction than its tetramethyl analogue, maleimido-proxyl (M-PROX) (**Figure 14B**). M-TETPO was found to probe protein dynamics as efficiently as M-PROX. Most importantly, the extended lifetime of M-TETPO enabled the study of structural features of a bacterial chaperone in the absence and presence of its binding partner at endogenous concentration directly inside *X. laevis* oocytes (**Figure 14C**). This pioneering study paves the way towards a broader use of this powerful technique to investigate the behavior of IDPs inside living cells.

8. Modeling IDPs as conformational ensembles

One of the most challenging aspects in the field of “unstructural biology” [164] is to improve experimental and computational methods to describe the structural and dynamic features of IDPs. The representation of IDPs requires the use of models that take the high entropy of the system into account. The current state of computational methods for studying IDP structure and dynamics, as well as the major challenges faced in this field were recently reviewed [165]. Overall, an efficient representation of IDPs can be addressed by building conformational ensembles. A conformational ensemble consists in a pool of multiple “static” structures that, together, represent the conformational sampling of the protein under study. Although an IDP shouldn’t be represented by a fixed number of conformers - as there is probably infinity of them - conformational ensembles made of a “reasonable” number of conformers represent the minimum conformational diversity necessary and sufficient to explain experimental observables. Thus, they shouldn’t be regarded as a precise description of the conformer population and only as possible, and non-unique, solutions that satisfy the experimental data. Indeed, describing IDPs as conformational ensembles is a typical ill-defined problem, in that the number of experimental observables used as constraints is much lower than the number of degrees of freedom of the system.

The challenge of ensemble approaches is to demonstrate that they provide an accurate representation of the range of conformations explored by a given protein during its thermal fluctuations. One of the most important issues is to define the minimal number of conformers for such ensembles to be relevant. Although much caution should be taken when attempting generalizations, in most cases 50–100 structures in the final ensemble were shown to suffice in describing experimental observations [166].

The second limitation comes from the so-called “non-uniqueness” problem. Different ensembles can fit experimental data equally well because of the huge discrepancy between the number of observables and the number of degrees of freedom of a typical IDP. This lack of information can be however at least partially compensated by combining data from different and complementary techniques such as SAXS and NMR [167].

Two different approaches are commonly used for generating disordered state ensembles. The first uses *in silico* simulations, referred to as *ensemble-restrained molecular dynamics*, and the second starts from a pre-defined conformational library that is reduced to a smaller ensemble.

Ensemble-restrained molecular dynamics simulation consists in a molecular dynamics simulation in which restraints determined from measurements are used to bias the trajectories. Thus, the simulated structures converge to conformations that satisfy experimental constraints. Multiple simulations are carried out in parallel and will converge to a solution according to biases calculated on averages of all replicas. Molecular dynamics simulations can provide an accurate representation of IDP conformation and dynamics, provided that methods ensuring exploration of a broad conformational space are implemented. Metadynamics is a powerful simulation method that enables enhanced sampling of multidimensional free-energy landscapes [168]. The complexity of the system (*i.e.* its degrees of freedom) is reduced by selecting a few collective variables for its description. Furthermore, a history-dependent bias is introduced to the potential to prevent resampling of previously visited states and, therefore, forcing the system to leave local minima and sampling low-probability states. Metadynamics is an adaptive method, automatically biasing the employed potential to encourage sampling configurations away from the most visited ones. This approach has the advantage of not requiring an initial estimate of the energy landscape to be explored. Its intrinsic limitation resides in the difficulty of choosing appropriate collective variables for the simulation of complex systems, such as IDPs in solution. Furthermore, the parameters for the potential correction should be chosen carefully in order to ensure that the statistics of the population do not depend on them. Enhanced sampling simulations have been successfully used to describe IDPs as conformational ensembles with good agreement with experimental data (for examples see [169, 170]).

Although enhanced sampling simulations ensures a broad conformational sampling, *ensemble-restrained molecular dynamics* suffers from the high number of restraints that have to be integrated in order to drive the simulation in the typically large conformational landscape of IDPs. The minimal and acceptable set of constraints might include both inter-atomic distances (calculated for example from EPR, NMR, FRET...) and observables describing the shape of the protein (calculated from SAXS, ESI-MS...). While the amount of constraints required to achieve model convergence can be decreased by reducing model complexity, using for example coarse-grained simulation [171], determination of the adequate number and type of required restraints remains obscure. Thus, *ensemble-restrained molecular dynamics* modeling gives no guarantee to achieve a consistent model for any given set of constraints.

Another option consists in modeling conformational ensembles by selecting conformers from a pre-defined conformational library sampling an as large conformational space as possible. The initial pool of structures can be defined from an already existing

library, computed by molecular-dynamic simulation, or modeled using statistical coil modeling approaches such as Flexible-Meccano [172], RANCH-EOM [68, 173] or TraDES [174, 175]. Once the initial pool is generated, a subset of conformers is selected to build ensembles that minimize the differences between the theoretical data back-calculated from the ensembles and experimental data. This selection of conformers involves a stochastic search using either Monte Carlo algorithms, as implemented in ENSEMBLE or Sample and Select (SAS) [176], or genetic algorithms, such as those used by ASTEROIDS [177] or GAJOE (EOM) [68, 173].

For SAXS data, programs such as CRY SOL [178] and FOXS [179] are available for the calculation of theoretical scattering curves for each individual conformer. The back-calculated scattering curves are then compared to measurements. In this way, it is possible to select a subset of conformers that, together, fit the actual experimental SAXS data. The fitting between theoretical and measured values is evaluated using χ^2 . The χ^2 is given by:

$$\chi^2 = \frac{1}{n-1} \sum_{i=1}^n \left[\frac{I_{th}(s_i) - I_{obs}(s_i)}{\sigma(s_i)} \right]^2 \quad (\text{Eq. 20})$$

where n is the number of experimental points, $\sigma(s)$ are standard deviations, $I_{th}(s_i)$ and $I_{obs}(s_i)$ are the back-calculated and observed intensities at scattering vector s_i , respectively. Modeling based on SAXS data usually uses χ^2 as minimization function. For values greater than 1, the lower is χ^2 the more relevant is the model. However, χ^2 values lower than 1 only indicate that the models are equally meaningful, according to both standard deviations on intensities and SAXS resolution.

Modeling of IDPs as conformational ensembles is required to assess a statistical description of the system. One can derive many parameters to describe IDPs in the form of density functions, as long as the experimental technique used to guide modeling is sensitive to the parameter of interest. While SAXS is sensitive to the R_g and can thus be used to study the density of this parameter, it fails to describe secondary structure propensities, as the technique is not sensitive to this parameter. Therefore, even if calculations can be performed on all parameters, it is extremely important to pay attention to the kind of data used for the generation of the ensembles, so as to be able to critically estimate whether the parameter of interest suffers from over-fitting, *i.e.* if it has been modeled independently from data restraints.

9. Assessing binding events

Many IDPs undergo a (partial) disorder-to-order transition upon binding to their partner(s), a process known as “folding coupled to binding”. In many cases the gain of structure is partial and limited to a region of the IDP that therefore remains conspicuously disordered after binding, thus leading to a so-called “fuzzy complex” [180]. The extent of *fuzziness* can be investigated using a combination of methods [181]. Moreover, several small molecules have been identified to modulate the behavior of IDPs and most commonly used techniques to probe these types of interactions were recently reviewed [182]. Although the detailed description of the various methods enabling the assessment of binding (and folding) events of IDPs is beyond the scope of this review, we nevertheless would like to briefly

mention the arsenal of approaches that can be used.

The partner-induced folding of an IDP, irrespective of the underlying mechanisms (*i.e.* folding *after* or *before* binding), can be detected using any of the different spectroscopic methods that we have described above. Some techniques, such as far- and near-UV CD, FTIR SAXS, and fluorescence spectroscopy, allow documenting an overall structural transition through comparison of measured signals in the absence and in the presence of small ligands, [31, 64, 183-185] (**Figure 4A**). When such experiments are carried out in the presence of a protein partner, the latter contributes to the observed signal. Hence, structural transitions can be better inferred from comparison between the experimentally measured signal and the theoretical (*i.e.* calculated) average spectrum, where the latter corresponds to the spectrum expected in case no transition takes place. In all cases, the obtained information is global, with no indication about the specific IDP region(s) involved. The major difficulty lies in extracting the signal of the folded IDP from that arising from the partner and from the unbound form of the IDP. The presence of unbound IDP is not necessarily an issue, as the amount of the uncomplexed form can be minimized (or even zeroed) by increasing the concentration of the two proteins well above the equilibrium dissociation constant (K_D) (provided that this is compatible with the requirements of the given technique), as well as by performing titrations. By contrast, the problem related to the contribution of the partner to the measured signal persists. Methods that allow only the IDP to be monitored (while the partner remains invisible) are of course much more powerful. Examples of such methods are heteronuclear NMR [186], site-directed spin-labeling EPR spectroscopy [187, 188], and HDX-MS [135]. These approaches, in addition, provide site-resolved information and are thus ideally suited to identify the IDP region(s) involved in structural transitions.

Isothermal titration calorimetry (ITC) is an additional method that can be used not only to document binding events but also folding coupled to binding processes [189]. Under specific experimental conditions, ITC allows estimating the extent of folding upon binding as well as to distinguish between binding mechanisms (conformational selection *vs* induced fit) (for an example see [190]).

Regions of IDPs involved in binding can also be identified by generating truncation constructs and then assessing their ability to interact with the partner. These interactions can be documented using methods such as isothermal titration calorimetry, surface plasmon resonance, microscale thermophoresis, and biolayer interferometry that all rely on the use of purified proteins, although this approach implicitly assumes a mono-dimensional description of the protein, which may fail to offer an adequate representation of binding sites. A valuable and attractive alternative to detect protein interactions is provided by protein complementation assays such as those based on reassembly of split-GFP [6, 181, 191-194] and on proximity ligation [195-197]. The advantage of these assays resides in the fact that protein interactions are directly detected *in cellula*, without requiring prior protein purification.

10. Conclusions

Conformational heterogeneity is an inherent property of IDPs that renders their characterization extremely challenging owing to the inherent multiplicity of species, which entails problems of resolution, spectral broadening and averaging.

Here, we have reviewed a variety of techniques enabling recognition of IDPs that differ in terms of the information they convey and ease of implementation. In-depth characterization of the conformational landscape of IDPs can be achieved by combining computational methods and experimental data, where the latter are used as constraints. Some issues, such as averaging over collective motions with different timescales, have still to be solved. Single-molecule approaches are ideally suited to study IDPs as they circumvent the inherent averaging problem from which all the other approaches suffer [145]. The combination of different techniques, such as NMR and SAXS for instance, is of course an additional step toward a more realistic description of IDPs as conformational ensembles, as it allows limiting data over fitting thanks to the enlarged number of experimental constraints used to compute the ensemble.

The number of experimentally characterized IDPs is continuously expanding, as are the methods to investigate them. Among emerging approaches, high-speed single-molecule atomic force microscopy has attracted attention as a powerful approach for the characterization of IDPs [198, 199] that also holds promises for their site-resolved conformational analysis. Any novel original approach will enlarge the repertoire of available means to depict the structure of these ubiquitous and elusive proteins, thus shedding light on their peculiar mode of action. We hope that the next years will see the rise of increasingly more performant and sensitive methods.

Acknowledgments

We wish to thank Sébastien Brier (Institut Pasteur, Unité de Spectrométrie de Masse Structurale et Protéomique, Paris, France) for generating Figure 10. We are also grateful to Barbara Zambelli (Bioinorganic Chemistry Lab., Dept. Pharmacy and Biotechnology, University of Bologna, Bologna, Italy) and Elisabetta Mileo (BIP Lab., CNRS and Aix-Marseille University, France) for kindly providing us with the data used in Figure 7 and Figure 14, respectively. A.S. is a recipient of a Ph.D. fellowship from the French Ministry of National Education, Research and Technology allotted to the Ecole Doctorale des Sciences de la Vie et de la Santé (Aix-Marseille Université). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Figure Legends

Figure 1. General suggested procedure for the experimental assessment of protein disorder. The different approaches that can be used to assess protein disorder are herein summarized. The different techniques are listed from top to bottom according to the suggested work-flow (see arrows) and from left to right according to increasing complexity (in terms of both protein and technical requirements). The molecular information provided by each approach is indicated, as is the information common to all the approaches grouped within the same box.

Figure 2. Anomalous electrophoretic mobility and protease sensitivity assay. (A) SDS-PAGE analysis of the intrinsically disordered C-terminal domain of the measles virus (MeV) nucleoprotein (N_{TAIL}). M: protein molecular-mass standards, *l* and *h* stand for *low-κ* and *high-κ* variants, respectively. The three variants display a different electrophoretic migration although they possess the same net charge. In addition, none of the variants migrates at the expected molecular mass. Data were taken from [20]. (B) SDS-PAGE analysis of thermolysin digestion of Nipah virus N_{TAIL} at different time intervals (0, 1 h and 24 h) showing that the IDP is readily fully digested after one hour of incubation. The protein also displays an aberrant migration (expected molecular mass: 15 kDa). M: protein molecular mass markers. Data were taken from [45].

Figure 3. Dimensionless Kratky plots. (A) Three proteins illustrative of three different conformations. Unfolded protein: TtASR1 [31]; partially folded protein: Hendra virus V protein [200]; folded protein: Hendra virus XD [201]. (B) MeV N_{TAIL} charge permutant variants, where *l* and *h* stand for *low-κ* and *high-κ* variants, respectively. Modified from [20].

Figure 4. Far-UV CD spectroscopy analysis. (A) Far-UV CD spectra of an IDP (e.g. MeV N_{TAIL} Q499E variant) (at 3.5 μM) in the presence of a 2-fold molar excess of its folded partner (XD) (violet). The CD spectra of the IDP alone (red) or the folded partner alone (blue), as well as the theoretical average curve (dark pink) calculated by assuming that no structural variations occur are also shown. The deviation of the spectrum of the mixture (violet) from the calculated average spectrum (dark pink) indicates that a structural transition (*i.e.* a gain of α -helicity) takes place. Modified from [202]. MRE: mean residue ellipticity (B) 222–200 ellipticity plot modified from [35]. The mean residue ellipticity (MRE) at 222 nm of a set of well-characterized unfolded, RC-like or PMG-like proteins (from [35]) has been plotted against the MRE at 200 nm.

Figure 5. Fourier transform infrared (FTIR) spectroscopy analysis. Amide-I absorption spectrum (continuous line) and curve-fitted spectrum (dashed line) of the intrinsically disordered Sic1^{Δ214} truncated variant in D₂O solution. The curve fitting of the spectrum as a linear combination of Gaussian components (dot-dashed lines) is superimposed to the measured spectrum. The Fourier self-deconvoluted (dotted line) and the second derivative (continuous bold line) spectra are also reported. Modified from [48].

Figure 6. Near-UV CD spectroscopy analysis. CD spectra in the near-UV wavelength range of green fluorescent protein (GFP), of three IDPs (e.g. MeV N_{TAIL}, MeV PNT and Sic1) and of their respective GFP fusions. Modified from [203].

Figure 7. Differential scanning calorimetry (DSC) analysis. Temperature dependence of heat capacity of *Bacillus pasteurii* UreG (57 μ M), followed by DSC in the 10–90°C temperature range. The theoretical profile for the fully unfolded state is also shown. Modified from [90].

Figure 8. Intrinsic fluorescence spectroscopy analysis. Fluorescence emission spectra of HvASR1, an IDP of the ASR family from barley, at 20 °C. The protein concentration was 1 μ M in 10 mM sodium phosphate buffer pH 7 supplemented with various concentrations of ZnSO₄. The spectra show a decrease in the intensity of the emission peak with increasing concentrations of ZnSO₄ reflecting a metal-induced folding process leading to tryptophan burying and decreased quantum yield. Modified from [31].

Figure 9. Native mass spectrometry (MS) analysis. (A) ESI-MS spectrum of an IDP acquired under non-denaturing conditions. The most intense peak of each CSD is labeled by the corresponding charge state. (B) Deconvolution of the peaks envelope of panel (A) by multi-Gaussian fitting. A cartoon representative of the conformational compactness of the conformer labels each component. (C) SASA vs Z_{av} relationship under ESI conditions allows the estimation of SASA for each protein conformer.

Figure 10. Hydrogen-Deuterium exchange mass spectrometry (HDX-MS) analysis. (A) HDX-MS analysis of HvASR1 in the presence of 10% TFE and 2 mM ZnSO₄. The relative fractional uptake was plotted as a function of peptide position for the full-length ASR1 protein, where each color and dot represents a specific time point. The region encompassing residues 81-115 shows a dynamic behavior, *i.e.* increasing uptake with time, indicating that this region is not fully exposed to the solvent and reflecting metal-induced gain of structure occurring therein. Data were taken from [31]. (B) HDX-MS analysis of the catalytic domain of the adenylate cyclase toxin from *Bordetella pertussis*. The relative fractional uptake was plotted as a function of peptide position for the full-length protein, where each color and dot represents a specific time point. Regions displaying a dynamic HDX-MS behavior, indicative of the presence of structured elements, are colored in green, whereas non-dynamic regions are colored in gray. Data were taken from [135].

Figure 11. Single-molecule FRET spectroscopy analysis and quantification of the compaction of IDPs by molecular crowding. Representative FRET efficiency histograms at different concentrations of polyethylene glycol (PEG) 6000 for four IDPs, the N- and C-terminal segments of human prothymosin- α (ProT α -N and -C), the binding domain of the activator for thyroid hormones and retinoid receptors (ACTR), and the N-terminal domain of HIV-1 integrase (IN). Histograms of ACTR and IN in the presence of their respective interaction partners, NCBD and Zn²⁺, are shown for comparison. The solid line in all panels corresponds to fits. The transfer efficiency peaks from molecules lacking an active acceptor

dye are shaded in gray. The data show an increase in FRET efficiency with increasing crowder concentration, except for IN that exhibits only a small crowding-induced population at the transfer efficiency of the folded state. Modified from [144].

Figure 12. Vibrational spectroscopy of cyanylated cysteines. (A) Reaction scheme for the cyanylation of cysteines. (B) IR absorption spectrum of methyl thiocyanate in water, showing the CN stretching band of thiocyanate on top of the collective absorption band of H₂O. (C) IR absorption bands in the CN stretching region for methyl thiocyanate in varying mixtures of tetrahydrofuran (THF) with water illustrating the red shift with increasing hydrophobicity. (D) Infrared spectra in the CN stretching region for the cyanylated MeV N_{TAIL} S491C variant in the absence (black) and presence (red) of the folded partner XD. The addition of the partner causes an increase in the line width reflecting formation of α -helical structure in the vicinity of residue 491. Data were taken from [154].

Figure 13. Kinetic scheme of tryptophan triplet quenching by cystine (TCQ) experiment. The tryptophan is shown in red and the cystine in yellow. The asterisk denotes the excited form of tryptophan. Taken from [155].

Figure 14. *In-vitro* and *in-cell* EPR spectroscopy. (A) Chemical structure of the new maleimido-proxyl-based spin label (M-TETPO) and of its tetramethyl analogue, maleimido-proxyl (M-PROX). (B) Reduction profiles at room temperature of M-TETPO (filled circles) and M-PROX (empty circles) grafted on a cysteine variant of NarJT (Glu119Cys) at 50 μ M in the presence of 4 mM ascorbate in aqueous solution. The EPR amplitude of the central peak (h_0), as observed in CW X-band EPR spectra recorded at room temperature, is plotted as a function of time. Results show the much higher resistance of the new spin label to reduction. (C) Inter-label distances distribution for NarJT variant H21C/Q104C labeled with M-TETPO and M-PROX as inferred from double electron electron resonance (DEER) experiments with a protein solution at 120 μ M (top and middle panels) or with injected *X. laevis* oocytes. The inter-spin distance profile obtained in the whole cells is similar to that obtained *in vitro*, indicating that the protein retains its dynamic nature in the cytosol and demonstrating that M-TETPO is efficient in determining inter-spin distances at endogenous concentration inside cells. Data were taken from [163].

References

- [1] Dunker AK, Babu MM, Barbar E, Blackledge M, Bondos SE, Dosztányi Z, et al. What's in a name? Why these proteins are intrinsically disordered. *Intrinsically Disordered Proteins*. 2013;1:e24157.
- [2] Habchi J, Tompa P, Longhi S, Uversky VN. Introducing Protein Intrinsic Disorder. *Chem Rev*. 2014;114:6561-88.
- [3] Schad E, Ficho E, Pancsa R, Simon I, Dosztanyi Z, Meszaros B. DIBS: a repository of disordered binding sites mediating interactions with ordered proteins. *Bioinformatics*. 2018;34:535-7.
- [4] Miskei M, Antal C, Fuxreiter M. FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. *Nucleic Acids Res*. 2017;45:D228-D35.
- [5] Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics*. 2008;9 Suppl 1:S1.
- [6] Bignon C, Troilo F, Gianni S, Longhi S. Partner-mediated polymorphism of an intrinsically disordered protein. *J Mol Biol*. 2018;430:2493-507.
- [7] Contreras-Martos S, Nguyen HH, Nguyen PN, Hristozova N, Macossay-Castillo M, Kovacs D, et al. Quantification of Intrinsically Disordered Proteins: A Problem Not Fully Appreciated. *Front Mol Biosci*. 2018;5:83.
- [8] Juhasova A, Baliova M, Jursky F. A Dynamic Interaction of Coomassie Dye with the Glycine Transporters N-termini. *Protein J*. 2016;35:371-8.
- [9] Jakob U, Kriwacki R, Uversky VN. Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chem Rev*. 2014;114:6779-805.
- [10] Reichmann D, Xu Y, Cremers CM, Ilbert M, Mittelman R, Fitzgerald MC, et al. Order out of disorder: working cycle of an intrinsically unfolded chaperone. *Cell*. 2012;148:947-57.
- [11] Fraga H, Pujols J, Gil-Garcia M, Roque A, Bernardo-Seisdedos G, Santambrogio C, et al. Disulfide driven folding for a conditionally disordered protein. *Sci Rep*. 2017;7:16994.
- [12] Gontero B, Maberly SC. An intrinsically disordered protein, CP12: jack of all trades and master of the Calvin cycle. *Biochem Soc Trans*. 2012;40:995-9.
- [13] Meszaros B, Erdos G, Dosztanyi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res*. 2018;46:W329-W37.
- [14] Tompa P. Intrinsically unstructured proteins. *Trends Biochem Sci*. 2002;27:527-33.
- [15] Westerhuis WH, Sturgis JN, Niederman RA. Reevaluation of the electrophoretic migration behavior of soluble globular proteins in the native and detergent-denatured states in polyacrylamide gels. *Anal Biochem*. 2000;284:143-52.
- [16] Iakoucheva L, Kimzey A, Masselon C, Smith R, Dunker A, Ackerman E. Aberrant mobility phenomena of the DNA repair protein XPA. *Protein Sci*. 2001;10:1353-62.
- [17] Kirkland TN, Finley F, Orsborn KI, Galgiani JN. Evaluation of the proline-rich antigen of *Coccidioides immitis* as a vaccine candidate in mice. *Infect Immun*. 1998;66:3519-22.
- [18] Blocquel D, Habchi J, Gruet A, Blangy S, Longhi S. Compaction and binding properties of the intrinsically disordered C-terminal domain of Henipavirus nucleoprotein as unveiled by deletion studies. *Mol Biosyst*. 2012;8:392-410.
- [19] Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc Natl Acad Sci U S A*. 2013;110:13392-7.

- [20] Tedeschi G, Salladini E, Santambrogio C, Grandori R, Longhi S, Brocca S. Conformational response to charge clustering in synthetic intrinsically disordered proteins. *Biochim Biophys Acta Gen Subj*. 2018;1862:2204-14.
- [21] Reynolds JA, Tanford C. Binding of dodecyl sulfate to proteins at high binding ratios. Possible implications for the state of proteins in biological membranes. *Proc Natl Acad Sci U S A*. 1970;66:1002-7.
- [22] Fontana A, Fassina G, Vita C, Dalzoppo D, Zamai M, Zambonin M. Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry*. 1986;25:1847-51.
- [23] Hubbard SJ, Campbell SF, Thornton JM. Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J Mol Biol*. 1991;220:507-30.
- [24] Hubbard SJ, Eisenmenger F, Thornton JM. Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci*. 1994;3:757-68.
- [25] Fontana A, Polverino de Laureto P, De Filippis V, Scaramella E, Zambonin M. Probing the partly folded states of proteins by limited proteolysis. *Fold Des*. 1997;2:R17-26.
- [26] Fontana A, Polverino de Laureto P, Spolaore B, Frare E, Zambonin M. Detecting disordered regions in proteins by limited proteolysis. In: Uversky V, Longhi S, editors. *Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation*. Hoboken, New Jersey: John Wiley and Sons; 2010. p. 569-626.
- [27] Fontana A, de Laureto P, Spolaore B, Frare E. Identifying disordered regions in proteins by limited proteolysis. In: Uversky V, Dunker AK, editors. *Intrinsically disordered protein analysis*. New York: Springer; 2012. p. 297-318.
- [28] Carey J. A systematic and general proteolytic method for defining structural and functional domains of proteins. *Methods Enzymol*. 2000;328:499-514.
- [29] Fontana A, Zambonin M, De Filippis V, Bosco M, Polverino de Laureto P. Limited proteolysis of cytochrome c in trifluoroethanol. *FEBS Lett*. 1995;362:266-70.
- [30] Karlin D, Longhi S, Receveur V, Canard B. The N-terminal domain of the phosphoprotein of morbilliviruses belongs to the natively unfolded class of proteins. *Virology*. 2002;296:251-62.
- [31] Hamdi K, Salladini E, O'Brien DP, Brier S, Chenal A, Yacoubi I, et al. Structural disorder and induced folding within two cereal, ABA stress and ripening (ASR) proteins. *Sci Rep*. 2017;7:15544.
- [32] Kalthoff C. A novel strategy for the purification of recombinantly expressed unstructured protein domains. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2003;786:247-54.
- [33] Livernois AM, Hnatchuk DJ, Findlater EE, Graether SP. Obtaining highly purified intrinsically disordered protein by boiling lysis and single step ion exchange. *Anal Biochem*. 2009;392:70-6.
- [34] Brocca S, Samalikova M, Uversky VN, Lotti M, Vanoni M, Alberghina L, et al. Order propensity of an intrinsically disordered protein, the cyclin-dependent-kinase inhibitor Sic1. *Proteins*. 2009;76:731-46.
- [35] Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*. 2002;11:739-56.
- [36] Kumar N, Shukla S, Kumar S, Suryawanshi A, Chaudhry U, Ramachandran S, et al. Intrinsically disordered protein from a pathogenic mesophile *Mycobacterium tuberculosis* adopts structured conformation at high temperature. *Proteins*. 2008;71:1123-33.
- [37] Zhou HX, Pang X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem Rev*. 2018;118:1691-741.
- [38] Dill KA, Shortle D. Denatured states of proteins. *Annu Rev Biochem*. 1991;60:795-825.

- [39] Neumann S, Matthey U, Kaim G, Dimroth P. Purification and properties of the F1F0 ATPase of *Ilyobacter tartaricus*, a sodium ion pump. *J Bacteriol.* 1998;180:3312-6.
- [40] Haines DC, Sevrioukova IF, Peterson JA. The FMN-binding domain of cytochrome P450BM-3: resolution, reconstitution, and flavin analogue substitution. *Biochemistry.* 2000;39:9419-29.
- [41] Cortese MS, Baird JP, Uversky VN, Dunker AK. Uncovering the unfoldome: Enriching extracts for unstructured proteins by acid treatment. *J Proteome Res.* 2005;In press.
- [42] Uversky VN. Intrinsically disordered proteins and their “mysterious” (meta)physics. *Frontiers in Physics.* 2019;7:1-18.
- [43] Tedeschi G, Mangiagalli M, Chmielewska S, Lotti M, Natalello A, Brocca S. Aggregation properties of a disordered protein are tunable by pH and depend on its net charge per residue. *Biochim Biophys Acta Gen Subj.* 2017;1861:2543-50.
- [44] Fonin AV, Stepanenko OV, Sitdikova AK, Antifeeva IA, Kostyleva EI, Polyanichko AM, et al. Folding of poly-amino acids and intrinsically disordered proteins in overcrowded milieu induced by pH change. *Int J Biol Macromol.* 2019;125:244-55.
- [45] Habchi J, Mamelli L, Darbon H, Longhi S. Structural Disorder within Henipavirus Nucleoprotein and Phosphoprotein: From Predictions to Experimental Assessment. *PLoS ONE.* 2010;5:e11684.
- [46] Uversky VN. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry.* 1993;32:13288-98.
- [47] Uversky VN. What does it mean to be natively unfolded? *Eur J Biochem.* 2002;269:2-12.
- [48] Brocca S, Testa L, Sobott F, Samalikova M, Natalello A, Papaleo E, et al. Compaction properties of an intrinsically disordered protein: sic1 and its kinase-inhibitor domain. *Biophys J.* 2011;100:2243-52.
- [49] Manon F, Ebel C. Analytical ultracentrifugation, a useful tool to probe intrinsically disordered proteins. In: Uversky VN, Longhi S, editors. *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation.* Hoboken (NJ): John Wiley and Sons; 2010. p. 433-49.
- [50] Salvay AG, Communie G, Ebel C. Sedimentation velocity analytical ultracentrifugation for intrinsically disordered proteins. In: Uversky V, Dunker AK, editors. *Intrinsically disordered protein analysis.* New York: Springer; 2012. p. 91-105.
- [51] Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. *J Mol Graph Model.* 2001;19:26-59.
- [52] Blocquel D, Beltrandi M, Erales J, Barbier P, Longhi S. Biochemical and structural studies of the oligomerization domain of the Nipah virus phosphoprotein: Evidence for an elongated coiled-coil homotrimer. *Virology.* 2013;446:162-72.
- [53] Bloyet LM, Schramm A, Lazert C, Raynal B, Hologne M, Walker O, et al. Regulation of measles virus gene expression by P protein coiled-coil properties. *Sci Adv.* 2019;5:eaaw3702.
- [54] Casassa EF, Eisenberg H. Thermodynamic Analysis of Multicomponent Solutions. *Adv Protein Chem.* 1964;19:287-395.
- [55] Fujita H. *Foundations of Ultracentrifugal Analysis.* New York: Wiley; 1975.
- [56] Goldberg RJ. Sedimentation in the ultracentrifuge. *J Phys Chem.* 1953;57:194-202.
- [57] Williams JW, Van Holde KE, Baldwin RL, Fujita H. The theory of sedimentation analysis. *Chem Rev.* 1958;58.
- [58] Rambo RP. Considerations for sample preparation using size-exclusion chromatography for home and synchrotron sources. In: Chaudhuri B, Muñoz I, Qian S, Urban V, editors. *Biological Small Angle Scattering: Techniques, Strategies and Tips.* Singapore: Springer; 2017. p. 31-45.

- [59] Kikhney AG, Svergun DI. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* 2015;589:2570-7.
- [60] Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry.* 1999;38:16424-31.
- [61] Bernado P, Blackledge M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophys J.* 2009;97:2839-45.
- [62] Gast K, Damaschun H, Misselwitz R, Muller-Frohne M, Zirwer D, Damaschun G. Compactness of protein molten globules: temperature-induced structural changes of the apomyoglobin folding intermediate. *Eur Biophys J.* 1994;23:297-305.
- [63] Receveur-Brechot V, Durand D. How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr Protein Pept Sci.* 2012;13:55-75.
- [64] Longhi S, Receveur-Brechot V, Karlin D, Johansson K, Darbon H, Bhella D, et al. 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.* 2003;278:18638-48.
- [65] Perez J, Vachette P. A successful combination: coupling SE-HPLC with SAXS. In: Chaudhuri B, Muñoz I, Qian S, Urban V, editors. *Biological Small Angle Scattering: Techniques, Strategies and Tips.* Singapore: Springer; 2017. p. 183-99.
- [66] Cordeiro TN, Chen PC, De Biasio A, Sibille N, Blanco FJ, Hub JS, et al. Disentangling polydispersity in the PCNA-p15PAF complex, a disordered, transient and multivalent macromolecular assembly. *Nucleic Acids Res.* 2017;45:1501-15.
- [67] Bernado P, Svergun DI. Structural Insights into Intrinsically Disordered Proteins by Small-Angle X-ray Scattering. In: Uversky VN, Longhi S, editors. *Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation.* Hoboken, New Jersey John Wiley and Sons; 2010. p. 451-76.
- [68] Bernado P, Svergun DI. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol Biosyst.* 2012;8:151-67.
- [69] Cordeiro TN, Herranz-Trillo F, Urbanek A, Estana A, Cortes J, Sibille N, et al. Structural Characterization of Highly Flexible Proteins by Small-Angle Scattering. *Adv Exp Med Biol.* 2017;1009:107-29.
- [70] Cordeiro TN, Herranz-Trillo F, Urbanek A, Estana A, Cortes J, Sibille N, et al. Small-angle scattering studies of intrinsically disordered proteins and their complexes. *Curr Opin Struct Biol.* 2017;42:15-23.
- [71] Bourhis JM, Receveur-Bréchet V, Oglesbee M, Zhang X, Buccellato M, Darbon H, et al. 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.* 2005;14:1975-92.
- [72] Wallace BA. The role of circular dichroism spectroscopy in the era of integrative structural biology. *Curr Opin Struct Biol.* 2019.
- [73] Chemes LB, Alonso LG, Noval MG, de Prat-Gay G. Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. *Methods Mol Biol.* 2012;895:387-404.
- [74] Kelly S, Jess T, Price N. How to study proteins by circular dichroism. *Biochim Biophys Acta.* 2005;1751:119-39.
- [75] Woody RW. Contributions of tryptophan side chains to the far-ultraviolet circular dichroism of proteins. *Eur Biophys J.* 1994;23:253-62.

- [76] Woody RW. Circular dichroism of intrinsically disordered proteins. In: Uversky VN, Longhi S, editors. *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation*. Hoboken (NJ): John Wiley and Sons; 2010. p. 303-21.
- [77] Whitmore L, Wallace BA. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res*. 2004;32:W668-73.
- [78] Whitmore L, Wallace BA. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*. 2008;89:392-400.
- [79] Natalello A, Doglia SM. Intrinsically disordered proteins and induced folding studied by Fourier transform infrared spectroscopy In: Uversky V, Longhi S, editors. *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation*. Hoboken (NJ): John Wiley and Sons; 2010. p. 225-52.
- [80] Natalello A, Ami D, Brocca S, Lotti M, Doglia S. Secondary structure, conformational stability and glycosylation of a recombinant *Candida rugosa* lipase studied by Fourier transform infrared spectroscopy. *Biochem J*. 2004;385:511-7.
- [81] Kelly SM, Price NC. The use of circular dichroism in the investigation of protein structure and function. *Curr Protein Pept Sci*. 2000;1:349-84.
- [82] Szasz CS, Alexa A, Toth K, Rakacs M, Langowski J, Tompa P. Protein disorder prevails under crowded conditions. *Biochemistry*. 2011;50:5834-44.
- [83] Ibarra-Molero B, Naganathan AN, Sanchez-Ruiz JM, Munoz V. Modern Analysis of Protein Folding by Differential Scanning Calorimetry. *Methods Enzymol*. 2016;567:281-318.
- [84] Johnson CM. Differential scanning calorimetry as a tool for protein folding and stability. *Arch Biochem Biophys*. 2013;531:100-9.
- [85] Farber P, Darmawan H, Sprules T, Mittermaier A. Analyzing protein folding cooperativity by differential scanning calorimetry and NMR spectroscopy. *J Am Chem Soc*. 2010;132:6214-22.
- [86] Privalov PL. Stability of proteins: small globular proteins. *Adv Protein Chem*. 1979;33:167-241.
- [87] Ptitsyn OB. Molten globule and protein folding. *Adv Protein Chem*. 1995;47:83-229.
- [88] Uversky VN. A multiparametric approach to studies of self-organization of globular proteins. *Biochemistry (Mosc)*. 1999;64:250-66.
- [89] Malhotra P, Udgaonkar JB. How cooperative are protein folding and unfolding transitions? *Protein Sci*. 2016;25:1924-41.
- [90] Zambelli B, Cremades N, Neyroz P, Turano P, Uversky VN, Ciurli S. Insights in the (un)structural organization of *Bacillus pasteurii* UreG, an intrinsically disordered GTPase enzyme. *Mol Biosyst*. 2012;8:220-8.
- [91] Aphasizheva IY, Dolgikh DA, Abdullaev ZK, Uversky VN, Kirpichnikov MP, Ptitsyn OB. Can grafting of an octapeptide improve the structure of a de novo protein? *FEBS Lett*. 1998;425:101-4.
- [92] Lackowicz J. *Principles of Fluorescence Spectroscopy*. New York: Kluwer Academic/Plenum Publishers; 1999.
- [93] Lakowitz J. *Principles of fluorescence spectroscopy*. New York: Plenum Press; 1983.
- [94] Receveur-Bréchet V, Bourhis JM, Uversky VN, Canard B, Longhi S. Assessing protein disorder and induced folding. *Proteins: Structure, Function and Bioinformatics*. 2006;62:24-45.
- [95] Grandori R, Santambrogio C, Brocca S, Invernizzi G, Lotti M. Electrospray-ionization mass spectrometry as a tool for fast screening of protein structural properties. *Biotechnol J*. 2009;4:73-87.

- [96] Kebarle P, Verkerk U. On the Mechanism of Electrospray Ionization Mass Spectrometry (ESIMS). In: Cole R, editor. *Electrospray and MALDI Mass Spectrometry: Fundamentals, Instrumentation, Practicalities, and Biological Applications*: Wiley and Sons; 2010.
- [97] Chandler SA, Benesch JL. Mass spectrometry beyond the native state. *Curr Opin Chem Biol*. 2018;42:130-7.
- [98] Natalello A, Santambrogio C, Grandori R. Are Charge-State Distributions a Reliable Tool Describing Molecular Ensembles of Intrinsically Disordered Proteins by Native MS? *J Am Soc Mass Spectrom*. 2017;28:21-8.
- [99] Chowdhury S, Katta V, Chait B. Probing conformational changes in proteins by mass spectrometry. *J Am Chem Soc*. 1990;112:9012-3.
- [100] Kaltashov IA, Bobst CE, Abzalimov RR. Mass spectrometry-based methods to study protein architecture and dynamics. *Protein Sci*. 2013;22:530-44.
- [101] Marchese R, Grandori R, Carloni P, Rauegi S. On the zwitterionic nature of gas-phase peptides and protein ions. *PLoS Comput Biol*. 2010;6:e1000775.
- [102] Marchese R, Grandori R, Carloni P, Rauegi S. A computational model for protein ionization by electrospray based on gas-phase basicity. *J Am Soc Mass Spectrom*. 2012;23:1903-10.
- [103] Bakhtiari M, Konermann L. Protein Ions Generated by Native Electrospray Ionization: Comparison of Gas Phase, Solution, and Crystal Structures. *J Phys Chem B*. 2019;123:1784-96.
- [104] Li J, Lyu W, Rossetti G, Konijnenberg A, Natalello A, Ippoliti E, et al. Proton Dynamics in Protein Mass Spectrometry. *J Phys Chem Lett*. 2017;8:1105-12.
- [105] Abzalimov RR, Frimpong AK, Kaltashov IA. Detection and characterization of large-scale protein conformational transitions in solution using charge-state distribution analysis in ESI-MS. *Methods Mol Biol*. 2012;896:365-73.
- [106] Kaltashov IA, Mohimen A. Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Anal Chem*. 2005;77:5370-9.
- [107] Testa L, Brocca S, Grandori R. Charge-surface correlation in electrospray ionization of folded and unfolded proteins. *Anal Chem*. 2011;83:6459-63.
- [108] Testa L, Brocca S, Santambrogio C, D'Urzo A, Habchi J, Longhi S, et al. Extracting structural information from charge-state distributions of intrinsically disordered proteins by non-denaturing electrospray-ionization mass spectrometry. *Intrinsically Disordered Proteins* 2013;1:e25068.
- [109] Li J, Santambrogio C, Brocca S, Rossetti G, Carloni P, Grandori R. Conformational effects in protein electrospray-ionization mass spectrometry. *Mass Spectrom Rev*. 2016;35:111-22.
- [110] Lambrugh M, Papaleo E, Testa L, Brocca S, De Gioia L, Grandori R. Intramolecular interactions stabilizing compact conformations of the intrinsically disordered kinase-inhibitor domain of Sic1: a molecular dynamics investigation. *Front Physiol*. 2012;3:435.
- [111] D'Urzo A, Konijnenberg A, Rossetti G, Habchi J, Li J, Carloni P, et al. Molecular Basis for Structural Heterogeneity of an Intrinsically Disordered Protein Bound to a Partner by Combined ESI-IM-MS and Modeling. *J Am Soc Mass Spectrom*. 2015;26:472-81.
- [112] Santambrogio C, Natalello A, Brocca S, Ponzini E, Grandori R. Conformational Characterization and Classification of Intrinsically Disordered Proteins by Native Mass Spectrometry and Charge-State Distribution Analysis. *Proteomics*. 2019;19:e1800060.
- [113] Stuchfield D, Barran P. Unique insights to intrinsically disordered proteins provided by ion mobility mass spectrometry. *Curr Opin Chem Biol*. 2018;42:177-85.
- [114] Ben-Nissan G, Sharon M. The application of ion-mobility mass spectrometry for structure/function investigation of protein complexes. *Curr Opin Chem Biol*. 2018;42:25-33.

- [115] Beveridge R, Chappuis Q, Macphee C, Barran P. Mass spectrometry methods for intrinsically disordered proteins. *Analyst*. 2013;138:32-42.
- [116] Beveridge R, Phillips AS, Denbigh L, Saleem HM, MacPhee CE, Barran PE. Relating gas phase to solution conformations: Lessons from disordered proteins. *Proteomics*. 2015;15:2872-83.
- [117] Beveridge R, Migas LG, Das RK, Pappu RV, Kriwacki RW, Barran PE. Ion Mobility Mass Spectrometry Uncovers the Impact of the Patterning of Oppositely Charged Residues on the Conformational Distributions of Intrinsically Disordered Proteins. *J Am Chem Soc*. 2019;141:4908-18.
- [118] Han JY, Choi TS, Heo CE, Son MK, Kim HI. Gas-phase conformations of intrinsically disordered proteins and their complexes with ligands: Kinetically trapped states during transfer from solution to the gas phase. *Mass Spectrom Rev*. 2019.
- [119] Konijnenberg A, Ranica S, Narkiewicz J, Legname G, Grandori R, Sobott F, et al. Opposite Structural Effects of Epigallocatechin-3-gallate and Dopamine Binding to alpha-Synuclein. *Anal Chem*. 2016;88:8468-75.
- [120] Dickinson ER, Jurneczko E, Nicholson J, Hupp TR, Zawacka-Pankau J, Selivanova G, et al. The use of ion mobility mass spectrometry to probe modulation of the structure of p53 and of MDM2 by small molecule inhibitors. *Front Mol Biosci*. 2015;2:39.
- [121] Oganessian I, Lento C, Wilson DJ. Contemporary hydrogen deuterium exchange mass spectrometry. *Methods*. 2018;144:27-42.
- [122] Wales TE, Engen JR. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom Rev*. 2006;25:158-70.
- [123] Engen JR, Smith DL. Investigating protein structure and dynamics by hydrogen exchange MS. *Anal Chem*. 2001;73:256A-65A.
- [124] Nishimura C, Dyson HJ, Wright PE. Enhanced picture of protein-folding intermediates using organic solvents in H/D exchange and quench-flow experiments. *Proc Natl Acad Sci U S A*. 2005;102:4765-70.
- [125] Balasubramaniam D, Komives EA. Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins. *Biochim Biophys Acta*. 2013;1834:1202-9.
- [126] Al-Naqshabandi MA, Weis DD. Quantifying Protection in Disordered Proteins Using Millisecond Hydrogen Exchange-Mass Spectrometry and Peptic Reference Peptides. *Biochemistry*. 2017;56:4064-72.
- [127] Keppel TR, Howard BA, Weis DD. Mapping unstructured regions and synergistic folding in intrinsically disordered proteins with amide H/D exchange mass spectrometry. *Biochemistry*. 2011;50:8722-32.
- [128] Przygonska K, Poznanski J, Mistarz UH, Rand KD, Dadlez M. Side-chain moieties from the N-terminal region of Abeta are Involved in an oligomer-stabilizing network of interactions. *PLoS One*. 2018;13:e0201761.
- [129] Kheterpal I, Wetzel R. Hydrogen/deuterium exchange mass spectrometry--a window into amyloid structure. *Acc Chem Res*. 2006;39:584-93.
- [130] Rand KD, Zehl M, Jorgensen TJ. Measuring the hydrogen/deuterium exchange of proteins at high spatial resolution by mass spectrometry: overcoming gas-phase hydrogen/deuterium scrambling. *Acc Chem Res*. 2014;47:3018-27.
- [131] Pan J, Zhang S, Borchers CH. Comparative higher-order structure analysis of antibody biosimilars using combined bottom-up and top-down hydrogen-deuterium exchange mass spectrometry. *Biochim Biophys Acta*. 2016;1864:1801-8.
- [132] Kan ZY, Ye X, Skinner JJ, Mayne L, Englander SW. ExMS2: An Integrated Solution for Hydrogen-Deuterium Exchange Mass Spectrometry Data Analysis. *Anal Chem*. 2019.

- [133] Rusinga FI, Weis DD. Soft interactions and volume exclusion by polymeric crowders can stabilize or destabilize transient structure in disordered proteins depending on polymer concentration. *Proteins*. 2017;85:1468-79.
- [134] O'Brien DP, Hernandez B, Durand D, Hourdel V, Sotomayor-Perez AC, Vachette P, et al. Structural models of intrinsically disordered and calcium-bound folded states of a protein adapted for secretion. *Sci Rep*. 2015;5:14223.
- [135] O'Brien DP, Durand D, Voegelé A, Hourdel V, Davi M, Chamot-Rooke J, et al. Calmodulin fishing with a structurally disordered bait triggers CyaA catalysis. *PLoS Biol*. 2017;15:e2004486.
- [136] Mistarz UH, Rand KD. Installation, validation, and application examples of two instrumental setups for gas-phase HDX-MS analysis of peptides and proteins. *Methods*. 2018;144:113-24.
- [137] Felli IC, Pierattelli R. *Intrinsically Disordered Proteins studied by NMR spectroscopy*. Printforce, The Netherlands: Springer; 2015.
- [138] Belle V, Rouger S, Costanzo S, Longhi S, Fournel A. Site-directed spin labeling EPR spectroscopy. In: Uversky VN, Longhi S, editors. *Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation*. Hoboken, New Jersey John Wiley and Sons; 2010.
- [139] Martinho M, Fournier E, Le Breton N, Mileo E, Belle V. Nitroxide spin labels: fabulous spy spins for biostructural EPR applications. *Electron Paramag Reson*. 2019;26:66-88.
- [140] Förster T. Intermolecular energy migration and fluorescence. *Ann Physik*. 1948;2:55-75.
- [141] Gomes GN, Gradinaru CC. Insights into the conformations and dynamics of intrinsically disordered proteins using single-molecule fluorescence. *Biochim Biophys Acta Proteins Proteom*. 2017;1865:1696-706.
- [142] LeBlanc SJ, Kulkarni P, Weninger KR. Single Molecule FRET: A Powerful Tool to Study Intrinsically Disordered Proteins. *Biomolecules*. 2018;8.
- [143] Schuler B, Soranno A, Hofmann H, Nettels D. Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins. *Annu Rev Biophys*. 2016;45:207-31.
- [144] Soranno A, Koenig I, Borgia MB, Hofmann H, Zosel F, Nettels D, et al. Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *Proc Natl Acad Sci U S A*. 2014;111:4874-9.
- [145] Borgia A, Borgia MB, Bugge K, Kissling VM, Heidarsson PO, Fernandes CB, et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature*. 2018;555:61-6.
- [146] Fuertes G, Banterle N, Ruff KM, Chowdhury A, Mercadante D, Koehler C, et al. Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. *Proc Natl Acad Sci U S A*. 2017;114:E6342-E51.
- [147] Feng H, Zhou BR, Bai Y. Binding Affinity and Function of the Extremely Disordered Protein Complex Containing Human Linker Histone H1.0 and Its Chaperone ProTalpha. *Biochemistry*. 2018;57:6645-8.
- [148] Tcherkasskaya O, Uversky VN. Denatured collapsed states in protein folding: example of apomyoglobin. *Proteins*. 2001;44:244-54.
- [149] De Filippis V, Frasson R, Fontana A. 3-Nitrotyrosine as a spectroscopic probe for investigating protein protein interactions. *Protein Sci*. 2006;15:976-86.
- [150] Yang H, Habchi J, Longhi S, Londergan CH. Monitoring structural transitions in IDPs by vibrational spectroscopy of cyanylated cysteine. *Methods Mol Biol*. 2012;895:245-70.

- [151] Maienschein-Cline MG, Londergan CH. The CN stretching band of aliphatic thiocyanate is sensitive to solvent dynamics and specific solvation. *J Phys Chem A*. 2007;111:10020-5.
- [152] Alfieri KN, Vienneau AR, Londergan CH. Using infrared spectroscopy of cyanylated cysteine to map the membrane binding structure and orientation of the hybrid antimicrobial peptide CM15. *Biochemistry*. 2011;50:11097-108.
- [153] Dalton SR, Vienneau AR, Burstein SR, Xu RJ, Linse S, Londergan CH. Cyanylated Cysteine Reports Site-Specific Changes at Protein-Protein-Binding Interfaces Without Perturbation. *Biochemistry*. 2018;57:3702-12.
- [154] Bischak CG, Longhi S, Snead DM, Costanzo S, Terrer E, Londergan CH. Probing structural transitions in the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by vibrational spectroscopy of cyanylated cysteines. *Biophys J*. 2010;99:1676-83.
- [155] Sizemore SM, Cope SM, Roy A, Ghirlanda G, Vaiana SM. Slow Internal Dynamics and Charge Expansion in the Disordered Protein CGRP: A Comparison with Amylin. *Biophys J*. 2015;109:1038-48.
- [156] Plitzko JM, Schuler B, Selenko P. Structural Biology outside the box-inside the cell. *Curr Opin Struct Biol*. 2017;46:110-21.
- [157] Theillet FX, Binolfi A, Frembgen-Kesner T, Hingorani K, Sarkar M, Kyne C, et al. Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chem Rev*. 2014;114:6661-714.
- [158] Serber Z, Keatinge-Clay AT, Ledwidge R, Kelly AE, Miller SM, Dotsch V. High-resolution macromolecular NMR spectroscopy inside living cells. *J Am Chem Soc*. 2001;123:2446-7.
- [159] Freedberg DI, Selenko P. Live cell NMR. *Annu Rev Biophys*. 2014;43:171-92.
- [160] Selenko P, Wagner G. Looking into live cells with in-cell NMR spectroscopy. *J Struct Biol*. 2007;158:244-53.
- [161] Stadtmiller SS, Pielak GJ. The Expanding Zoo of In-Cell Protein NMR. *Biophys J*. 2018;115:1628-9.
- [162] Konig I, Zarrine-Afsar A, Aznauryan M, Soranno A, Wunderlich B, Dingfelder F, et al. Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat Methods*. 2015;12:773-9.
- [163] Karthikeyan G, Bonucci A, Casano G, Gerbaud G, Abel S, Thome V, et al. A Bioresistant Nitroxide Spin Label for In-Cell EPR Spectroscopy: In Vitro and In Oocytes Protein Structural Dynamics Studies. *Angew Chem Int Ed Engl*. 2018;57:1366-70.
- [164] Tompa P. Unstructural biology coming of age. *Curr Opin Struct Biol*. 2011;21:419-25.
- [165] Bhattacharya S, Lin X. Recent Advances in Computational Protocols Addressing Intrinsically Disordered Proteins. *Biomolecules*. 2019;9.
- [166] Jensen MR, Markwick PR, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, et al. Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure*. 2009;17:1169-85.
- [167] Mertens HDT, Svergun DI. Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics. *Arch Biochem Biophys*. 2017;628:33-41.
- [168] Laio A, Parrinello M. Escaping free-energy minima. *Proc Natl Acad Sci U S A*. 2002;99:12562-6.
- [169] Rossetti G, Musiani F, Abad E, Dibenedetto D, Mouhib H, Fernandez CO, et al. Conformational ensemble of human alpha-synuclein physiological form predicted by molecular simulations. *Phys Chem Chem Phys*. 2016;18:5702-6.

- [170] Bernetti M, Masetti M, Pietrucci F, Blackledge M, Jensen MR, Recanatini M, et al. Structural and Kinetic Characterization of the Intrinsically Disordered Protein SeV NTAIL through Enhanced Sampling Simulations. *J Phys Chem B*. 2017;121:9572-82.
- [171] Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-Grained Protein Models and Their Applications. *Chem Rev*. 2016;116:7898-936.
- [172] Ozenne V, Bauer F, Salmon L, Huang JR, Jensen MR, Segard S, et al. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*. 2012;28:1463-70.
- [173] Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc*. 2007;129:5656-64.
- [174] Feldman HJ, Hogue CW. A fast method to sample real protein conformational space. *Proteins*. 2000;39:112-31.
- [175] Feldman HJ, Hogue CW. Probabilistic sampling of protein conformations: new hope for brute force? *Proteins*. 2002;46:8-23.
- [176] Chen Y, Campbell SL, Dokholyan NV. Deciphering protein dynamics from NMR data using explicit structure sampling and selection. *Biophys J*. 2007;93:2300-6.
- [177] Nodet G, Salmon L, Ozenne V, Meier S, Jensen MR, Blackledge M. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J Am Chem Soc*. 2009;131:17908-18.
- [178] Svergun D, Barberato C, Koch MHJ. CRY SOL, a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst*. 1995;28:768-73.
- [179] Schneidman-Duhovny D, Hammel M, Sali A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res*. 2010;38:W540-4.
- [180] Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci*. 2008;33:2-8.
- [181] Troilo F, Bignon C, Gianni S, Fuxreiter M, Longhi S. Experimental characterization of fuzzy protein assemblies: interactions of paramyxoviral NTAIL domains with their functional partners. *Methods Enzymol*. 2018;611:137-92.
- [182] Heller GT, Aprile FA, Vendruscolo M. Methods of probing the interactions between small molecules and disordered proteins. *Cell Mol Life Sci*. 2017;74:3225-43.
- [183] Johansson K, Bourhis JM, Campanacci V, Cambillau C, Canard B, Longhi S. Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J Biol Chem*. 2003;278:44567-73.
- [184] Permyakov SE, Millett IS, Doniach S, Permyakov EA, Uversky VN. Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin. *Proteins*. 2003;53:855-62.
- [185] Nyarko A, Hare M, Hays TS, Barbar E. The intermediate chain of cytoplasmic dynein is partially disordered and gains structure upon binding to light-chain LC8. *Biochemistry*. 2004;43:15595-603.
- [186] Brutscher B, Felli IC, Gil-Caballero S, Hosek T, Kummerle R, Piai A, et al. NMR Methods for the Study of Intrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. *Adv Exp Med Biol*. 2015;870:49-122.
- [187] Habchi J, Martinho M, Gruet A, Guigliarelli B, Longhi S, Belle V. Monitoring structural transitions in IDPs by site-directed spin labeling EPR spectroscopy. *Methods Mol Biol*. 2012;895:361-86.

- [188] Le Breton N, Martinho M, Mileo E, Etienne E, Gerbaud G, Guigliarelli B, et al. Exploring intrinsically disordered proteins using site-directed spin labeling electron paramagnetic resonance spectroscopy. *Front Mol Biosci.* 2015;2:21.
- [189] Sahu D, Bastidas M, Lawrence CW, Noid WG, Showalter SA. Assessing Coupled Protein Folding and Binding Through Temperature-Dependent Isothermal Titration Calorimetry. *Methods Enzymol.* 2016;567:23-45.
- [190] Vega S, Abian O, Velazquez-Campoy A. On the link between conformational changes, ligand binding and heat capacity. *Biochim Biophys Acta.* 2016;1860:868-78.
- [191] Magliery TJ, Wilson CG, Pan W, Mishler D, Ghosh I, Hamilton AD, et al. Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *J Am Chem Soc.* 2005;127:146-57.
- [192] Gruet A, Dosnon M, Vassena A, Lombard V, Gerlier D, Bignon C, et al. Dissecting partner recognition by an intrinsically disordered protein using descriptive random mutagenesis. *J Mol Biol.* 2013;425:3495-509.
- [193] Gruet A, Dosnon M, Blocquel D, Brunel J, Gerlier D, Das RK, et al. Fuzzy regions in an intrinsically disordered protein impair protein-protein interactions. *FEBS J.* 2016;283:576-94.
- [194] Bignon C, Troilo F, Gianni S, Longhi S. Modulation of Measles Virus NTAIL Interactions through Fuzziness and Sequence Features of Disordered Binding Sites. *Biomolecules.* 2018;9.
- [195] Soderberg O, Gullberg M, Jarvius M, Ridderstrale K, Leuchowius KJ, Jarvius J, et al. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat Methods.* 2006;3:995-1000.
- [196] Ulke-Lemee A, Turner SR, MacDonald JA. In situ analysis of smoothelin-like 1 and calmodulin interactions in smooth muscle cells by proximity ligation. *J Cell Biochem.* 2015;116:2667-75.
- [197] Neira JL, Bintz J, Arruebo M, Rizzuti B, Bonacci T, Vega S, et al. Identification of a Drug Targeting an Intrinsically Disordered Protein Involved in Pancreatic Adenocarcinoma. *Sci Rep.* 2017;7:39732.
- [198] Miyagi A, Tsunaka Y, Uchihashi T, Mayanagi K, Hirose S, Morikawa K, et al. Visualization of intrinsically disordered regions of proteins by high-speed atomic force microscopy. *Chemphyschem.* 2008;9:1859-66.
- [199] Hashimoto M, Kodera N, Tsunaka Y, Oda M, Tanimoto M, Ando T, et al. Phosphorylation-coupled intramolecular dynamics of unstructured regions in chromatin remodeler FACT. *Biophys J.* 2013;104:2222-34.
- [200] Salladini E, Delauzun V, Longhi S. The Henipavirus V protein is a prevalently unfolded protein with a zinc-finger domain involved in binding to DDB1. *Mol Biosyst.* 2017;13:2254-67.
- [201] Erales J, Beltrandi M, Roche J, Maté M, Longhi S. Insights into the Hendra virus NTAIL-XD complex: evidence for a parallel organization of the helical MoRE at the XD surface stabilized by a combination of hydrophobic and polar interactions *Biochimica and Biophysical Acta* 2015;1854:1038-53.
- [202] Shu Y, Habchi J, Costanzo S, Padilla A, Brunel J, Gerlier D, et al. Plasticity in structural and functional interactions between the phosphoprotein and nucleoprotein of measles virus. *J Biol Chem.* 2012;287:11951-67.
- [203] Sambì I, Gatti-Lafranconi P, Longhi S, Lotti M. How disorder influences order and vice versa--mutual effects in fusion proteins containing an intrinsically disordered and a globular protein. *FEBS J.* 2010;277:4438-51.

Early detection of disorder

SDS-PAGE and limited proteolysis

Electrophoretic behavior & protease sensitivity



























