Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Structural Biology

Small-angle scattering studies of intrinsically disordered proteins and their complexes

Tiago N Cordeiro[1], Fátima Herranz-Trillo[1,3], Annika Urbanek[1], Alejandro Estaña[1,2], Juan Cortés[2], Nathalie Sibille[1] and Pau Bernadó[1]

Intrinsically Disordered Proteins (IDPs) perform a broad range of biological functions. Their relevance has motivated intense research activity seeking to characterize their sequence/structure/function relationships. However, the conformational plasticity of these molecules hampers the application of traditional structural approaches, and new tools and concepts are being developed to address the challenges they pose. Small-Angle Scattering (SAS) is a structural biology technique that probes the size and shape of disordered proteins and their complexes with other biomolecules. The low-resolution nature of SAS can be compensated with specially designed computational tools and its combined interpretation with complementary structural information. In this review, we describe recent advances in the application of SAS to disordered proteins and highly flexible complexes and discuss current challenges.

Addresses
[1] Centre de Biochimie Structurale, INSERM U1054, CNRS UMR 5048, Université de Montpellier, 29, rue de Navacelles, 34090 Montpellier, France
[2] LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France
[3] Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen, Denmark

Corresponding author: Bernadó, Pau (pau.bernado@cbs.cnrs.fr)

Current Opinion in Structural Biology 2016, **42**:15–23

This review comes from a themed issue on **Proteins: bridging theory and experiment**

Edited by **Igor N Berezovsky** and **Ugo Bastolla**

For a complete overview see the Issue and the Editorial

Available online 26th October 2016

http://dx.doi.org/10.1016/j.sbi.2016.10.011

Introduction

In the last two decades, Intrinsically Disordered Proteins or Regions (IDPs/IDRs) have emerged as fundamental molecules in a broad range of crucial biological functions such as cell signaling, regulation, and homeostasis [1,2,3••]. Due to their lack of a permanent secondary and tertiary structure, IDPs and IDRs are highly plastic and have the capacity to perform specialized functions that complement those of their globular (folded) counterparts [4]. Disordered regions, which can finely adapt to the structural and chemical features of their partners, are very well suited for protein–protein interactions and are thus abundant in hub positions of interactomes [5–7].

The importance of disordered proteins in a multitude of biological processes has fostered intense research efforts that seek to unravel the structural bases of their function. Nuclear Magnetic Resonance (NMR) has been the main structural biology technique used to characterize the conformational preferences at residue level, and, therefore, to localize partially structured elements [8,9]. However, a number of structural features related to the overall size and shape of IDPs or their complexes remain elusive to NMR. To study these properties, thereby complementing NMR residue-specific information, Small-Angle Scattering (SAS) of X-rays (SAXS) or Neutrons (SANS) is the most appropriate technique [10–12]. Although SAS is a low-resolution technique, the data obtained is sensitive to large-scale protein fluctuations and the presence of multiple species and/or conformations in solution [13–15]. However, the conversion of SAS properties into structural restraints is challenging due to the enormous conformational variability of IDPs and the ensemble-averaged nature of the experimental data [16]. The quantitative analysis of these data in terms of structure has prompted the development of computational approaches to both model disordered proteins and to use ensembles of conformations to describe the experimental data. Here we highlight the most relevant developments and applications of SAS to IDPs and IDRs, with a special emphasis on the computational strategies required to fully exploit the data in order to achieve biologically insightful information.

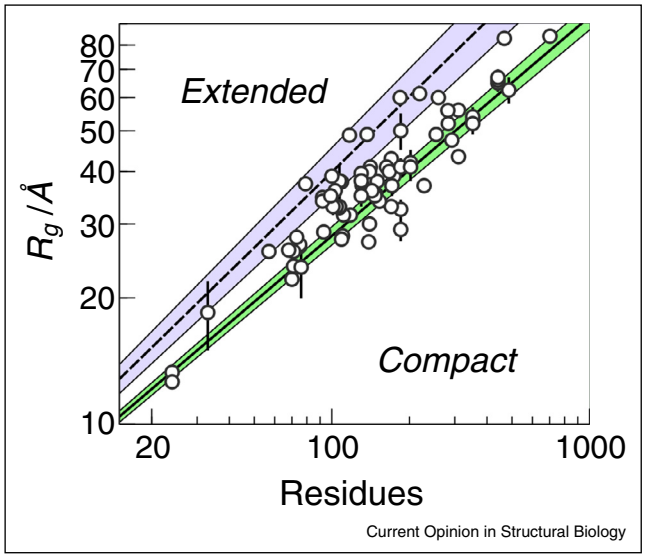
Structural models of IDPs and their experimental validation

For disordered proteins, the structural insights gained from overall SAS parameters, such as the radius of gyration, R_g, the pairwise intramolecular distance distribution, $p(r)$, and the maximum intramolecular distance, D_{max}, are limited. Neither these parameters nor the traditional Kratky representation, $I(s)s^2$ versus s where $I(s)$ represents the scattering intensity and s the momentum transfer, which qualitatively report on the compactness of biomolecules in solution, directly account for the ensemble

nature of disordered proteins. In order to fully exploit the structural and dynamic information encoded in SAS data, it is necessary to use realistic three-dimensional (3D) models. However, the generation of conformational ensembles of disordered proteins is extremely challenging, mainly because of the flat energy landscape and the large number of local minima separated by low-energy barriers [17]. The most popular methods to generate 3D models of IDPs are based on residue-specific conformational landscapes derived from large databases of crystallographic structures [18,19,20•]. However, the main limitation of these approaches is the absence of sequence context information, thereby precluding the prediction of transiently formed secondary structure elements or the presence of long-range interactions between distant regions of the protein. Accurate energy models (force-fields) accounting for the interactions within the chain and with the solvent are required to describe these features. The development of specific force-fields to study conformational fluctuations in disordered proteins is a very active field of research [21–24]. Molecular Dynamics (MD) or Monte-Carlo (MC) simulations, when an appropriate energy description is provided, are suitable methods to correctly sample the conformational space of IDPs. However, the high-dimensionality and the breadth of the energy landscape hamper exhaustive exploration of this space. Replica Exchange MD (REMD) [25,26], which exchanges conformations between parallel simulations running at multiple temperatures, or Multiscale Enhanced Sampling (MSES) [27], which couples temperature and Hamiltonian replica exchange, have been proposed to enhance the conformational exploration of MD methods. The performance of MD-based methods can also be improved by the inclusion of experimental data to delimit the exploration to the most relevant regions of the conformational space [28–30].

The quality of computational models of disordered proteins is normally validated using experimental data. The R_g derived from the low-angle region of SAXS curves or from the $p(r)$ function is an excellent probe of the overall size of a particle in solution. R_g compilations have been extensively used to validate models of denatured and natively disordered proteins through Flory's relationship, which correlates the R_g observed with the number residues of the chain [31,14]. The compilation of the R_gs from 76 IDPs (Figure 1) reveals that these proteins are more compact than chemically denatured ones. It has been shown that denatured proteins present an enhanced sampling of extended conformations, probably due to the interaction of the protein with chemical agents [32]. Importantly, deviations from the expected R_g values for canonical random-coil behavior, which is represented by the green line in Figure 1, indicate the presence of structural features that modify the overall size of the particle in solution towards more extended or more compact (Figure 1). The extendedness detected using this analysis for several Tau protein constructs has been linked to the presence of secondary structural elements probed by NMR [29]. These structural properties can be more thoroughly examined when the complete SAXS curve is used to validate the ensemble models of peptides [33] or proteins [19,34,35].

Figure 1



R_g values from 76 IDPs as a function of the number of residues of the protein are plotted in Log–Log scale. Only proteins lacking a permanent secondary or tertiary structure were considered for the compilation. Proteins with ordered domains, molten globules, or denatured proteins were not considered. Straight lines correspond to Flory's relationships parametrized for denatured proteins using experimental data (purple-dashed) [31] and IDPs using computational ensembles calculated with Flexible-Meccano (green-solid) [32]. Colored bands correspond to uncertainty of the parametrization for both models. Some IDPs contain local structural features and consequently they are globally more extended or more compact than expected for a random coil. These structural features, even if transient, can be manifested in the experimental R_g.

Ensemble approaches

In the last decade, ensemble methods have become highly popular to structurally characterize disordered proteins. Guided by experimental data, these methods aim to derive accurate ensemble models of flexible proteins. Several strategies that apply these methods to SAS data have been reported: Ensemble Optimization Method (EOM) [36,37]; Minimal Ensemble Search (MES) [38]; Basis-Set Supported SAXS (BSS-SAXS) [39]; Maximum Occurrence (MAX-Occ) [40]; Ensemble Refinement of SAXS (EROS) [41]; Broad Ensemble Generator with Re-weighting (BEGR) [42]; and Bayesian Ensemble SAXS (BE-SAXS) [43]. These methods share a common strategy that consists of the following three consecutive steps: (i) computational generation of a large ensemble that describes the conformational landscape of