Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Reweighting methods for elucidation of conformation ensembles of proteins

Raquel Gama Lima Costa[1] and David Fushman[1,2]

Abstract

Proteins are inherently dynamic macromolecules that exist in equilibrium among multiple conformational states, and motions of protein backbone and side chains are fundamental to biological function. The ability to characterize the conformational landscape is particularly important for intrinsically disordered proteins, multidomain proteins, and weakly bound complexes, where single-structure representations are inadequate. As the focus of structural biology shifts from relatively rigid macromolecules toward larger and more complex systems and molecular assemblies, there is a need for structural approaches that can paint a more realistic picture of such conformationally heterogeneous systems. Here, we review reweighting methods for elucidation of structural ensembles based on experimental data, with the focus on applications to multidomain proteins.

Addresses

[1] Chemical Physics Program, Institute for Physical Sciences and Technology, University of Maryland, College Park, 20742, MD, USA
[2] Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, 20742, MD, USA

Corresponding authors: Fushman, David (fushman@umd.edu); Gama Lima Costa, Raquel (rgama@umd.edu)

Current Opinion in Structural Biology 2022, **77**:102470

This review comes from a themed issue on **Biophysical Methods**

Edited by **David Fushman** and **Dagmar Ringe**

For complete overview of the section, please refer the article collection - Biophysical Methods

Available online 29 September 2022

https://doi.org/10.1016/j.sbi.2022.102470

Introduction

Proteins are inherently dynamic macromolecules that exist in equilibrium among multiple conformational states, and knowledge of the relevant conformers and of the extent of motions is essential for understanding their energy landscapes and the molecular mechanisms underlying key elements of their biological function, including recognition, allostery, and catalysis [1]. Conformational heterogeneity of macromolecules is essential for conformational selection that underlies biomolecular recognition and other events that enable protein function. Aside from conformational heterogeneity of a macromolecule, there is also heterogeneity in assembly. In addition, metabolites and post-translational modifications can shift the population of the conformational ensemble rather than fully lock down a single conformation. Thus, methods that can probe the energy landscapes of macromolecular systems and detect population shifts as a result of interactions or other factors provide valuable tools for structural biology.

Characterization of proteins as structurally heterogeneous systems presents a major challenge because traditional structural biology approaches are geared toward producing a coherent set of similar structures and are generally deficient in treating macromolecules as conformational ensembles. For example, experimental data from solution NMR measurements generally reflect physical characteristics averaged over multiple conformational states of a molecule; yet, the existing software packages for biomolecular structure determination (such as CNS/XPLOR, CYANA, HADDOCK, etc.) [2–4] were originally designed to produce a single-structure snapshot that matches the conformationally averaged constraints in their entirety. This might work well for relatively "rigid" single-domain proteins but could paint an inadequate portrait of inherently flexible macromolecular systems, such as intrinsically disordered proteins (IDP), multidomain proteins, and weakly bound macromolecular assemblies. A conceptually different approach is needed that aims at determining an ensemble of conformers where no single one needs to match the composite experimental data, but instead, a weighted average of a given pertinent physical observable over such an ensemble has to be in agreement with the experiment. This paradigm shift in structural biology, from a single-snapshot picture to a more adequate ensemble representation of biomacromolecules, requires novel computational approaches and tools. Here, we review some of such approaches, with the focus on applications to multidomain proteins in solution.

Determining structural ensembles from experimental data faces a fundamental challenge of solving a mathematically underdetermined system because the number of degrees of freedom associated with dynamic macromolecules generally greatly exceeds the number of experimentally available independent observables. This

renders the direct conversion of experimental data into a representative ensemble an ill-posed problem and can yield an unlimited number of possible solutions. This also implies that an integrated structural approach combining different methods/types of data is needed because no single experimental technique can fully capture all the features of a conformationally heterogeneous macromolecule, and thus has difficulties in "explaining" the composite of the data [5–7].

Experimental techniques that can provide ensemble-related structural information range from those capable of observing or detecting individual molecules (like cryo-electron microscopy (cryo-EM) or single-molecule Förster resonance energy transfer (smFRET)) to methods extracting distance distribution information (smFRET and double electron–electron resonance (DEER)) to those methods where the measured values of the observables are averaged over all molecules in the sample (FRET and small-angle scattering (SAS)) and also over various conformations each individual molecule samples during the characteristic measurement time (solution nuclear magnetic resonance (NMR)). Determining structural ensembles from experimental data obtained by the latter methods is particularly challenging because of the need to deconvolute contributions from various conformational states; therefore, here we mainly focus on those methods.

A conformational ensemble can be defined by a set of relevant structures/conformers and their respective populations (relative weights). Conceptually, the ensemble selection methods aim at finding the weights for various members of (typically) an in-silico generated set of structures [8] that provide the best match to experimental data. In this chapter, we review the so-called reweighting methods for ensemble selection. The name "reweighting" reflects that at the beginning, all conformations included in the initial input ensemble are considered possible and with equal a priori probabilities/weights [9]. As the result of analysis, a new weight w_i is assigned to each conformer i, such that the ensemble-averaged predicted data ($\mathbf{d}_{\mathbf{pred}}$) match the experimental data ($\mathbf{d}_{\mathbf{expt}}$) within their errors, see Figure 1. This is achieved by predicting experimental data for each member of the input ensemble and finding the appropriate weights (elements of vector $\mathbf{w}$) by solving the relevant optimization problem, as discussed below. Thus, reweighting methods work in a posterior way: an initial pool of structures is generated, and

Figure 1

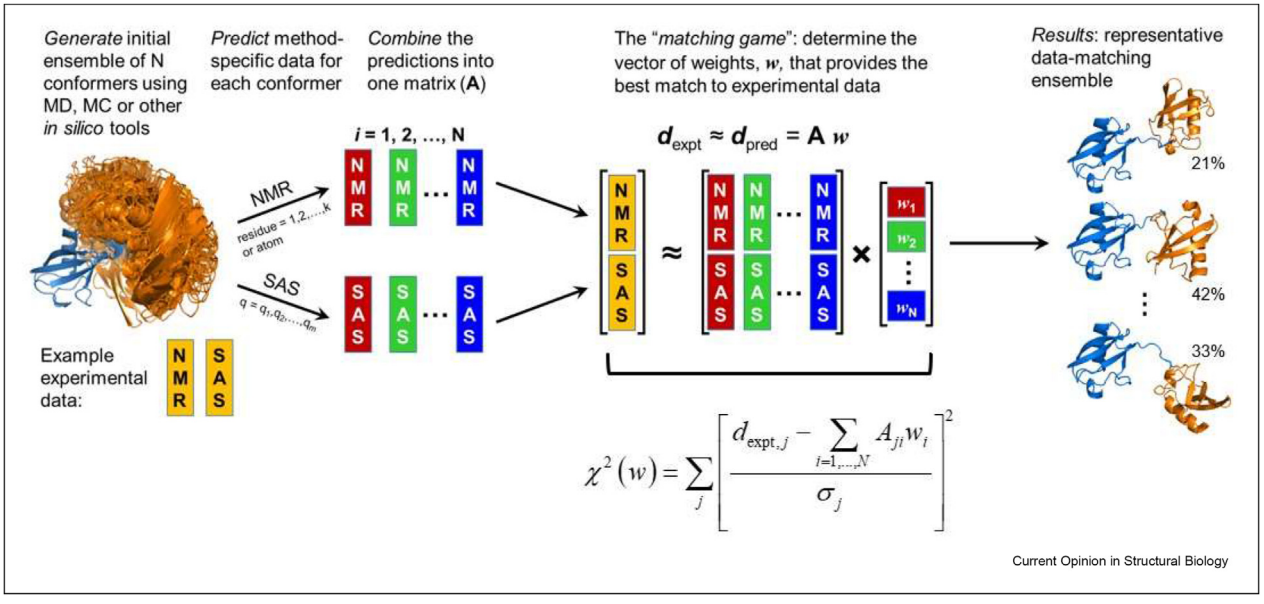

Illustration of the workflow of reweighting methods for determining the conformational ensemble based on experimental data. (From left to right) Step 1: An initial ensemble of N possible conformers is generated. Step 2: Given the type of experimental data available, the corresponding data are predicted for each conformer of the initial ensemble. From the mathematical perspective, the experimental data can be considered as a vector ($\mathbf{d}_{\mathbf{expt}}$), and the predicted data can be combined to form a matrix ($\mathbf{A}$) in which columns 1 through N represent predicted data for each of the N conformers. Step 3: The relative populations/weights for the conformers (vector $\mathbf{w}$ of length N) are determined by solving the optimization problem of matching the experimental data to the conformationally weighted predicted data for the entire ensemble, $\mathbf{d}_{\mathbf{pred}} = \mathbf{Aw}$. The agreement between $\mathbf{d}_{\mathbf{expt}}$ and $\mathbf{d}_{\mathbf{pred}}$ is quantified through χ^2, where σ's represent experimental errors. Finally, step 4: The resulting ensemble is analyzed, and the relevant structures are visualized. The input experimental data can be in a form of various observables (e.g., RDC, PCS, PRE data for various residues/atoms in a protein or SAS data for various values of the scattering vector length q) analyzed separately or a combination of such data analyzed together, as exemplified here for NMR and SAS data [6,11].

experimental data are used to refine the ensemble to a final solution. Note that reweighting methods are different from approaches that include experimental data as restraints when sampling the conformational space using molecular dynamics or Monte Carlo simulations [10].

Reweighting methods

A variety of methods for ensemble selection using reweighting have been developed. They generally range from those searching for a smallest-size ensemble (based on the maximum parsimony or Occam's razor principle) that can reproduce the experimental data to methods searching for solutions that encompass the entire input set of conformers (using the maximum entropy principle). Obviously, these principles reflect two extremes in approaching the complicated problem of treating structurally heterogeneous macromolecular systems. The maximum entropy principle provides an intuitively meaningful approximation of the generally continuous distribution of structures. However, due to the sheer size of the ensemble, the resulting solutions could be difficult to visualize and interpret without performing additional analyses, including clustering/discretization. Furthermore, such solutions often contain numerous conformers having almost negligible weights and might eventually require a further analysis to separate significant from less significant conformers [12]. On the other hand, the typically small discrete set of structures resulting from the maximum parsimony principle, while clearly a simplification that picks conformers making major contributions to the measured data, has the appeal of yielding a solution that often contains an easier visualizable and interpretable number of conformers.

At the heart of finding the weights for various conformers is the "matching game" of minimizing the difference between the experimental and ensemble-averaged predicted data, quantified as $\chi^2(\mathbf{w})$ (see Figure 1). This has been generally approached from two different directions: (*A*) solving the following minimization problem:

$$\mathbf{w} = \arg\min_{\mathbf{w}\geq 0}\{\chi^2(\mathbf{w}) + F(\mathbf{w})\} \quad (1)$$

where $F(\mathbf{w})$ is a regularization term included to prevent overfitting (see below); or (*B*) by maximizing the probability $p(\mathbf{w}|\mathbf{d}_{\mathbf{expt}}, E)$ of finding the proper combination of weights $\mathbf{w}$, given the experimental data $\mathbf{d}_{\mathbf{expt}}$ and ensemble E:

$$\mathbf{w} = \arg\max_{\mathbf{w}\geq 0}\{p(\mathbf{w}|\mathbf{d}_{\mathbf{expt}}, E)\} \quad (2)$$

where the solution with the highest probability is the optimal solution. To solve the latter problem, Bayesian inference-based methods are used, and the model is often assumed to be a Gaussian-type function that depends on $(\mathbf{d}_{\mathbf{pred}} - \mathbf{d}_{\mathbf{expt}})$, such that by maximizing the probability, minimization of $\chi^2(\mathbf{w})$ is achieved under regularization functions that depend on the data and implementation.

Maximum parsimony

In the search for the smallest number of conformers necessary to "explain" experimental data, constraints that limit the resulting ensemble size have to be imposed depending on the approach. When minimizing χ^2 using Eq. (1), regularization is directly imposed by finding solutions for a fixed size (M) of the resulting ensemble (i.e., the number of nonzero elements in $\mathbf{w}$), and screening various M values to determine the smallest M that provides a match between $\mathbf{d}_{\mathbf{expt}}$ and $\mathbf{d}_{\mathbf{pred}}$ within experimental errors (see, e.g., [12,13]). Finding the right size solutions can be tricky, and L-curve-based methods [12,14], initial guesses, etc., have been used to achieve that. For an initial ensemble of N conformers, testing all possible $N!/(M!(N-M)!)$ combinations for a given solution size M could be intractable even for N as small as $\sim$ 100, and greedy-type algorithms (e.g., Ref. [12]) reducing the computational complexity are used to efficiently select possible relevant combinations of conformers, while minimizing the risk of missing proper solutions.

In the probabilistic approach (Eq. (2)), the ensemble size reduction is used as a means to simplify the probability $p(\mathbf{w}|\mathbf{d}_{\mathbf{expt}}, E)$, such that convergence of the assumed probability with the true probability is achieved [11,15]. The weights are optimized by reducing the ensemble size iteratively using a cutoff threshold for w_i until an optimal size is reached.

Examples of maximum parsimony-based methods using χ^2 minimization include sparse ensemble selection (SES) [12], sample and select [16], a selection tool for ensemble representations of intrinsically disordered states (ASTEROIDS) [7], minimum ensemble search (MES) [17], minimal ensemble solutions to multiple experimental restraints (MESMER) [18], and ensemble optimization method (EOM) [19,20], while Bayesian inference-based methods (maximizing the probability $p(\mathbf{w}|\mathbf{d}_{\mathbf{expt}}, E)$) include Bayesian weighting (BW) [21] and Bayesian inference of electron microscopy (BioEM) [22].

A different approach called maximum occurrence (MaxOcc) [23] has also been developed. Instead of finding an ensemble solution, the focus is on determining the maximum possible weight a conformer from a predefined set can have as part of an ensemble. After finding the conformers with the highest possible weights, the method can be combined with the maximum and minimum occurrence of a region (MaxOR and MinOR) [24] to zoom on respective regions of the conformational space that provide a match to experimental data.

It should be mentioned here that maximum parsimony-based methods could produce multiple solutions with comparable values of the target function [12], and care should be exercised to validate them by comparison with other experimental data (e.g., Refs. [25,26]) as well as with the outcomes of the maximum entropy-based analysis (see e.g., Ref. [12]).

Maximum entropy

In this approach, when minimizing $\chi^2(\mathbf{w})$ that contains contributions from the entire input ensemble, a relative entropy term, $F(\mathbf{w}) = \lambda\sum_{i=1}^{N} w_i \log\left(\frac{w_i}{p_i}\right)$ is included as the regularizer in Eq. (1), where $\lambda > 0$ is a regularization parameter that can be obtained using an L-curve method [27], and p_i is a prior probability ($p_i = 1/N$ for a uniform distribution).

When solving the problem by maximizing the probability $p(\mathbf{w}|\mathbf{d}_{\mathbf{expt}}, E)$, the Bayesian inference principle is applied, and regularization is achieved by using relative entropy as a measure of deviation between the probability density and a reference distribution [28].

In both cases, the reweighting is performed for the entire initial ensemble. The implementation of these methods has been done using various approaches (e.g., Ref. [29]) to facilitate the solution convergence. Examples of methods using the maximum entropy principle to minimize $\chi^2(\mathbf{w})$ include ensemble-refinement of SAXS (EROS) [30], convex optimization for ensemble reweighting (COPER) [31], ENSEMBLE [32], and maximum entropy (MaxEnt) [12]. Methods maximizing $p(\mathbf{w}|\mathbf{d}_{\mathbf{expt}}, E)$ are used in Bayesian ensemble SAXS (BE-SAXS) [33], Bayesian energy landscape tilting (BELT) [34], Bayesian ensemble refinement [28], Bayesian/maximum entropy (BME) [35], and its iterative version, iBME [36].

In the methods discussed here reweighting is primarily used to select conformers from an initial unbiased ensemble of structures. However, reweighting can also be implemented when generating ensembles using molecular dynamics simulations [37,38].

Types of experimental data

Below, we briefly review various types of experimental data typically used for ensemble determination.

Nuclear magnetic resonance spectroscopy

NMR spectroscopy is a versatile technique capable of detecting and providing information on every magnetically active nucleus in a molecule. It allows studies of proteins in their native milieu and usually with no need for chemical modifications. Solution NMR experiments provide various types of data/constraints for structure/ensemble analysis. These include information on interatomic distances (nuclear Overhauser effect (NOE)), bond orientations (residual dipolar coupling (RDC)) [25,26,39], and dynamics (spin-relaxation rates [40–42]), as well as distances between protein atoms and a paramagnetic moiety (metal ion or radical) in a form of pseudo-contact shift (PCS) [43–47] and paramagnetic relaxation enhancement (PRE) [40,44,45]. RDCs can be caused by weak molecular alignment in anisotropic media (such as liquid crystalline media, stretched gels, etc.) or the anisotropy of magnetic susceptibility of the paramagnetic moiety (Ref. [13] compares RDCs caused by steric versus paramagnetic alignment and PCSs as constraints for ensemble selection).

The interconversion among various conformational states of a protein is typically fast on the time scale of NMR experiment, resulting in the measured values of the observables being averaged over the conformational ensemble. Several dual-domain proteins have been used as "toy" systems for developing tools for ensemble analysis using various NMR data, including calmodulin [11,13] and covalently linked ubiquitin dimers [12,25,26,48] among others.

Small-angle scattering: SAXS and SANS

Data from SAS of X-rays (SAXS) and neutrons (SANS) in solution are also commonly used for ensemble selection [25,36,49,50]. Although SAS measurements do not yield atomic-resolution structures, this technique is capable of providing information on the overall size and shape of the molecule [51]. The measured SAS data come from all molecules in the sample, thus reflecting all possible conformational states. The scattering profile or the reconstructed pair distance distribution function can then be used for ensemble selection either on their own or combined with other types of data, for example, from NMR measurements [5,6,25,52,53]. SAS data have also been used for validation and ranking of the NMR-derived ensemble solutions [5,25,53].

Förster resonance energy transfer

FRET measures distance-dependent energy transfer between the donor and acceptor fluorophores attached to specific sites on a molecule [54,55]. Bulk/ensemble FRET data are averaged over a large number of molecules, while single-molecule FRET (smFRET) can detect those distances in individual molecules, thus uniquely sensing the conformational states for each molecule at both spatial and temporal resolution. This enables smFRET to directly sample the conformational ensemble of a macromolecule (e.g., Ref. [55]). However, this capability is limited by the time resolution of the measurement, and for flexible molecules, like IDPs, that interconvert among different states on a faster time scale, the measured smFRET efficiency becomes averaged over a range of distances/conformations [7]. Therefore, ensemble FRET and smFRET can be used as input in ensemble selection methods and also to

validate ensemble solutions derived from SAS, NMR, or other data.

Double electron–electron resonance spectroscopy

The DEER (aka PELDOR) method measures the distance between two paramagnetic labels attached to select sites in a protein [56,57]. It utilizes magnetic dipole–dipole interaction between their electron spins and allows extraction of the distance distribution profile directly from the observed modulation of the signal. Unlike smFRET, DEER measurements are performed on frozen samples, and the results are averaged over all molecules/conformations. DEER has recently been used to characterize conformational ensembles of structurally heterogeneous protein systems [58,59].

Cryo-electron microscopy

Cryo-EM is a powerful technique that has recently revolutionized structural biology. It takes 2D snapshots of single particles frozen in various conformations and therefore could be ideally suited for direct observation and reconstruction of the structural ensembles of macromolecules at unprecedented details. However, converting the low-contrast 2D images into high-resolution 3D density maps/structures of the individual states requires the ability to detect numerous molecules trapped in the same conformational state. This works well for relatively rigid macromolecular systems (e.g., the proteasome) or for highly populated states of a dynamic system, but the reconstruction of less populated conformers or transition states still presents a significant challenge. Addressing this problem requires integrative approaches using ensemble selection methods [60,61].

We would like to mention here that the accuracy of ensemble reweighting methods depends on the ability to generate realistic predicted values, which is more challenging for some experimental techniques relative to others. For example, simplified physical models (disks, rods) used to simulate steric alignment can limit the accuracy of the predicted RDCs [62,63], and prediction of paramagnetic effects (PCSs, PREs, RDCs, and DEER), while based on exact equations, can be affected by the intrinsic conformational heterogeneity of some paramagnetic tags/moieties [44]. In addition, data predicted for multidomain structures generated using rigid-body Monte Carlo-based methods might not faithfully account for internal motions (of loops, side chains, etc.) within segments that were kept rigid.

Outlook: progress and challenges

Past reviews of the subject from 2015 to 2017 [64,65] proposed several community goals for further development of ensemble determination methods. Below, we highlight recent progress toward these goals as well as some challenges.

Validation of the derived structural ensembles is essential, and recent studies [6,7,49] exhibit advances in this direction. An encouraging trend is to use more than one category of ensemble selection methods, for example, one based on maximum entropy and one using maximum parsimony. Another approach to validation is by using different types of experimental data, for example, comparing predicted data for ensembles derived using NMR and SAS with experimental smFRET and/or DEER data. The use of the probabilistic/Bayesian approach has significantly enhanced the capability of ensemble selection methods by allowing different types of data to be combined together and new types of data (e.g., cryo-EM) to be incorporated. Coverage of the experimentally relevant conformational space by the initial ensemble is essential for the success of reweighting methods. However, finding proper means to assess and visualize the completeness of conformational coverage has proven challenging. An advance in visual and quantitative evaluation of conformational coverage was made in a 2017 study [8] that used density plots obtained by dividing the 3D space into voxels to count the number of occupied voxels as a function of the generated ensemble size.

Visual representation of the ensemble solutions is another challenge, especially when considering the full ensemble (maximum entropy), due to the dimensionality of the conformational distribution space. Even a simple dual-domain protein requires six degrees of freedom to specify the orientation and spatial location of one domain relative to another. Thus, there is a need for reduced-dimensionality representations. An example of such an approach, proposed in a 2018 paper [66], utilizes a disk-on-sphere model to visualize probability distributions of interdomain orientations in a dual-domain protein, and enables representation of continuous distributions.

Ensemble determination methods thus far have focused on soluble proteins and RNA (not discussed here). However, tools are much needed to elucidate the structure and conformational ensembles of membrane proteins which, despite constituting about a quarter of all proteins, remain largely unexplored structurally [67,68]. These systems are particularly difficult to study for a number of reasons, including flexibility, instability, and the need to extract them from cell membrane while maintaining or closely mimicking their native environments in order to preserve their structure and function. Developing experimental and computational approaches to tackle membrane proteins should be one of the next frontiers for structural biology.

Several web portals are available at no cost to users. SASSIE [69], URL: https://sassie-web.chem.utk.edu/

sassie2, provides efficient tools for generating structural ensembles and for prediction and analysis of SAS data; ATSAS online [70], URL: https://www.embl-hamburg.de/biosaxs/atsas-online/, implements RanCh and GAJOE in the EOM method [19] to generate an ensemble of structures and perform ensemble selection. As for ensemble selection; NMRsuite, URL: https://nmrsuite.genapp.rocks/, enables ensemble determination using MaxEnt and SES methods; WeNMR, URL: http://py-enmr.cerm.unifi.it/access/index/maxocc, performs the MaxOcc analysis [23]. Other resources and tutorials are also available online.

As the recent publications demonstrate [7,20,35,36], various methods for ensemble determination have been developed in the recent years. These are exciting advancements in the methodology, and it is important that new implementations do not stay hidden within a specific research group, otherwise combining these tools could become a challenge. Therefore, although the potential of ensemble selection methods is vast and multiple developments are already available, new ideas still need to reach their potential of impact.

Outside the method development community, ensemble selection methods have yet to reach the broad community of potential users. Published applications often use "toy" models for the purpose of testing new methods, although encouraging recent extensions to other protein systems are found in Refs. [6,7,49,50]. Scientists studying structurally heterogeneous protein systems are generally focused on directly using data as restraints and are often unaware that reweighting methods can help answer their questions at a lesser computational cost. Making the reweighting methods known and accessible to the scientific community requires presenting a clear picture of what each method does, what it is best used for, and what the limitations are.

Disclosure statement

Given his role as Guest Editor, David Fushman had no involvement in the peer review of the article and has no access to information regarding its peer-review. Full responsibility for the editorial process of this article was delegated to Dagmar Ringe.

Conflict of interest statement

Nothing declared.

Acknowledgements

This work was supported by NIH grant GM065334 to D.F. D.F. thanks Drs. Emre H. Brookes, Joseph E. Curtis, and Susan Krueger for insightful discussions.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Boehr DD, Nussinov R, Wright PE: **The role of dynamic conformational ensembles in biomolecular recognition**. *Nat Chem Biol* 2009, **5**:789–796.

2. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, *et al.*: **Crystallography & NMR system: a new software suite for macromolecular structure determination**. *Acta Crystallogr D* 1998, **54**:905–921.

3. Güntert P: **Automated NMR structure calculation with CYANA**. In *Protein NMR Techniques*. Edited by Downing AK, Humana Press; 2004:353–378.

4. Dominguez C, Boelens R, Bonvin AMJJ, HADDOCK: **A protein-protein docking approach based on biochemical or biophysical information**. *J Am Chem Soc* 2003, **125**:1731–1737.

5. * Boughton AJ, Krueger S, Fushman D: **Branching via K11 and K48 bestows ubiquitin chains with a unique interdomain interface and enhanced affinity for proteasomal subunit Rpn1**. *Structure* 2020, **28**:29–43.e6.
Structural studies combining X-ray crystallography, solution NMR, and SANS revealed a novel interdomain interface in K11/K48-linked branched tri-ubiquitin. However, the crystal and NMR-derived structures of this compact state of the tri-ubiquitin were significantly different from each other and neither agreed with SANS data. A multi-state ensemble comprising compact and open conformers was needed to achieve agreement with SANS data.

6. ** Gomes GNW, Krzeminski M, Namini A, Martin EW, Mittag T, Head- Gordon T, Forman-Kay JD, Gradinaru CC: **Conformational ensembles of an intrinsically disordered protein consistent with NMR, SAXS, and single-molecule FRET**. *J Am Chem Soc* 2020, **142**:15697–15710.
An integrative approach combining modeling with SAXS, NMR, and smFRET data is developed and applied to characterize the intrinsically disordered N-terminal region of protein Sic1 and (phosphorylated) pSic1. When analyzed separately, SAXS and smFRET data, yield discrepant inferences of Sic1 and pSic1 global dimensions. However, a joint refinement of SAXS and NMR/PRE data produced conformational ensembles that are consistent with smFRET and other experimental data, demonstrating the utility of integrating diverse experimental data for characterization of IDPs.

7. ** Naudi-Fabra S, Tengo M, Jensen MR, Blackledge M, Milles S: **Quantitative description of intrinsically disordered proteins using single-molecule FRET, NMR, and SAXS**. *J Am Chem Soc* 2021, **143**:20109–20121.
An integrated approach using a combination of FRET, NMR, and SAXS data in conjunction with the ASTEROIDS selection method is developed and applied to characterize the conformational landscape of an intrinsically disordered N-terminal region of the measles virus phosphoprotein. This detailed study shows how to properly generate predicted smFRET data for the ensemble of an IDP and how combining with PRE data helps overcome smFRET limitations, to obtain conformational ensembles that agree with all these data.

8. Zhang W, Howell SC, Wright DW, Heindel A, Qiu X, Chen J, Curtis JE: **Combined Monte Carlo/torsion-angle molecular dynamics for ensemble modeling of proteins, nucleic acids and carbohydrates**. *J Mol Graph Model* 2017, **73**:179–190.

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

10. Kaynak BT, Krieger JM, Dudas B, Dahmani ZL, Costa MGS, Balog E, Scott AL, Doruker P, Perahia D, Bahar I: **Sampling of protein conformational space using hybrid simulations: a critical assessment of recent methods**. *Front Mol Biosci* 2022, **9**, 832847.

11. Potrzebowski W, Trewhella J, Andre I: **Bayesian inference of protein conformational ensembles from limited structural data**. *PLoS Comput Biol* 2018, **14**, e1006641.

12. Berlin K, Castañeda CA, Schneidman-Duhovny D, Sali A, Nava-Tudela A, Fushman D: **Recovering a representative conformational ensemble from underdetermined macromolecular structural data**. *J Am Chem Soc* 2013, **135**: 16595–16609.

13. Andralojć W, Berlin K, Fushman D, Luchinat C, Parigi G, Ravera E, Sgheri L: **Information content of long-range NMR data for the characterization of conformational heterogeneity**. *J Biomol NMR* 2015, **62**:353–371.

14. Hansen PC: **The L-curve and its use in the numerical treatment of inverse problems**. In *Computational Inverse Problems in Electrocardiology*. Edited by Johnston P, WIT Press; 2000: 119–142.

15. Fisher CK, Ullman O, Stultz CM: **Efficient construction of disordered protein ensembles in a Bayesian framework with optimal selection of conformations**. *Pac Symp Biocomput* 2012:82–93.

16. Chen Y, Campbell SL, Dokholyan NV: **Deciphering protein dynamics from NMR data using explicit structure sampling and selection**. *Biophys J* 2007, **93**:2300–2306.

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

18. Ihms EC, Foster MP: **MESMER: minimal ensemble solutions to multiple experimental restraints**. *Bioinformatics* 2015, **31**:1951.

19. Bernadó 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**: 5656–5664.

20. Sagar A, Jeffries CM, Petoukhov MV, Svergun DI, Bernadó P: **Comment on the optimal parameters to derive intrinsically disordered protein conformational ensembles from small-angle X-ray scattering data using the ensemble optimization method**. *J Chem Theor Comput* 2021, **17**:2014–2021.

21. Fisher CK, Huang A, Stultz CM: **Modeling intrinsically disordered proteins with Bayesian statistics**. *J Am Chem Soc* 2010, **132**:14919–14927.

22. Cossio P, Hummer G: **Bayesian analysis of individual electron microscopy images: towards structures of dynamic and heterogeneous biomolecular assemblies**. *J Struct Biol* 2013, **184**:427–437.

23. Bertini I, Ferella L, Luchinat C, Parigi G, Petoukhov M, Ravera E, Rosato A, Svergun DI: **MaxOcc: a web portal for maximum occurrence analysis**. *J Biomol NMR* 2012, **53**:271–280.

24. Andralojć W, Luchinat C, Parigi G, Ravera E: **Exploring regions of conformational space occupied by two-domain proteins**. *J Phys Chem B* 2014, **118**:10576–10587.

25. Castañeda CA, Chaturvedi A, Camara CM, Curtis JE, Krueger S, Fushman D: **Linkage-specific conformational ensembles of non-canonical polyubiquitin chains**. *Phys Chem Chem Phys* 2016, **18**:5771–5788.

26. Castañeda CA, Dixon EK, Walker O, Chaturvedi A, Nakasone MA, Curtis JE, Reed MR, Krueger S, Cropp TA, Fushman D: **Linkage via K27 bestows ubiquitin chains with unique properties among polyubiquitins**. *Structure* 2016, **24**:423–436.

27. Chiang YW, Borbat PP, Freed JH: **Maximum entropy: a complement to Tikhonov regularization for determination of pair distance distributions by pulsed ESR**. *J Magn Reson* 2005, **177**:184–196.

28. Hummer G, Köfinger J: **Bayesian ensemble refinement by replica simulations and reweighting**. *J Chem Phys* 2015, **143**, 243150.

29. Byrd R, Hribar M, Nocedal J: **An interior point Algorithm for large-scale nonlinear programming**. *SIAM J Optim* 1999, **9**: 877–900.

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

31. Leung HTA, Bignucolo O, Aregger R, Dames SA, Mazur A, Bernéche S, Grzesiek S: **A rigorous and efficient method to reweight very large conformational ensembles using average experimental data and to determine their relative information content**. *J Chem Theor Comput* 2016, **12**:383–394.

32. Choy WY, Forman-Kay JD: **Calculation of ensembles of structures representing the unfolded state of an SH3 domain**. *J Mol Biol* 2001, **308**:1011–1032.

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

34. Beauchamp KA, Pande VS, Das R: **Bayesian energy landscape tilting: towards concordant models of molecular ensembles**. *Biophys J* 2014, **106**:1381–1390.

35. Bottaro S, Bengtsen T, Lindorff-Larsen K: **Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach**. In *Structural Bioinformatics*. Gáspári Z. Springer US; 2020:219–240.

36. * Pesce F, Lindorff-Larsen K: **Refining conformational ensembles of flexible proteins against small-angle x-ray scattering data**. *Biophys J* 2021, **120**:5124–5135.

A method is presented for obtaining conformational ensembles of flexible proteins by combining SAXS data with computer modeling using Bayesian Maximum Entropy (BME method) approach and improved prediction of SAXS data. The method is applied to several IDPs and a multi-domain protein containing flexible linkers. iBME (iterative version of BME) is pro- posed where minimization of χ^2 is used first to rescale the predicted data, then BME is used to optimize the weights and calculate a new solution.

37. Rangan R, Bonomi M, Heller GT, Cesari A, Bussi G, Vendruscolo M: **Determination of structural ensembles of proteins: restraining vs reweighting**. *J Chem Theor Comput* 2018, **14**:6632–6641.

38. Cesari A, Reißer S, Bussi G: **Using the maximum entropy principle to combine simulations and solution experiments**. *Computation* 2018, **6**:15.

39. Tjandra N, Bax A: **Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium**. *Science* 1997, **278**:1111–1114.

40. Fushman D, Varadan R, Assfalg M, Walker O: **Determining domain orientation in macromolecules by using spin-relaxation and residual dipolar coupling measurements**. *Prog Nucl Magn Reson Spectrosc* 2004, **44**:189–214.

41. Ryabov YE, Fushman D: **A model of interdomain mobility in a multidomain protein**. *J Am Chem Soc* 2007, **129**:3315–3327.

42. * Kauffmann C, Zawadzka-Kazimierczuk A, Kontaxis G, Konrat R: **Using cross-correlated spin relaxation to characterize backbone dihedral angle distributions of flexible protein segments**. *ChemPhysChem* 2021, **22**:18–28.

The methodology is developed for use of cross-correlated spin-relaxation effects (CCR) in conjunction with Maximum Entropy-based ensemble reweighting approach to characterize the distribution of backbone dihedral angles in a protein. Ubiquitin is used as an example where both flexible and rigid regions are analyzed. This study also demonstrates the utility of the less commonly used CCR data as experimental constraints to guide ensemble analysis.

43. Parigi G, Ravera E, Luchinat C: **Magnetic susceptibility and paramagnetism-based NMR**. *Prog Nucl Magn Reson Spectrosc* 2019, **114–115**:211–236.

44. Miao Q, Nitsche C, Orton H, Overhand M, Otting G, Ubbink M: **Paramagnetic chemical probes for studying biological macromolecules**. *Chem Rev* 2022, **122**:9571–9642.

45. * Orton HW, Huber T, Otting G: **Paramagpy: software for fitting magnetic susceptibility tensors using paramagnetic effects measured in NMR spectra**. *Magn Reson* 2020, **1**:1–12.

A new publicly available software package for analysis of paramagnetic NMR data (PCS, PRE, RDC and CCR) is presented. Given experimental PCS data and molecular structure, the software determines the

magnetic susceptibility tensor and predicts related NMR data that can be used in ensemble selection methods.

46. Müntener T, Joss D, Häussinger D, Hiller S: **Pseudocontact shifts in biomolecular NMR spectroscopy**. *Chem Rev* 2022, **122**:9422–9467.

47. Pilla KB, Leman JK, Otting G, Huber T: **Capturing conformational states in proteins using sparse paramagnetic NMR data**. *PLoS One* 2015, **10**:1–16.

48. ** Hou XN, Sekiyama N, Ohtani Y, Yang F, Miyanoiri Y, Akagi Ki, Su XC, Tochio H: **Conformational space sampled by domain reorientation of linear diubiquitin reflected in its binding mode for target proteins**. *ChemPhysChem* 2021, **22**: 1505–1517.
Paramagnetic NMR data were utilized to characterize the conformations of linear (head-to-tail linked) di-ubiquitin. Two sets of PCS and RDC data resulting from lanthanide tagging at two different positions were analyzed together using ensemble selection methods MESMER and MaxOcc. The results are in agreement with published SAXS and smFRET data. Comparison with crystal structures of di-ubiquitin in complexes with binding partners suggests existence of conformers prealigned for ligand binding.

49. ** Sicorello A, Różycki B, Konarev PV, Svergun DI, Pastore A: **Capturing the conformational ensemble of the mixed folded polyglutamine protein ataxin-3**. *Structure* 2021, **29**:70–81.e5.
Ensemble selection methods are applied to study protein Ataxin-3 using NMR and SAXS data in conjunction with EROS and EOM methods. The study explores how the conformational plasticity of Ataxin-3 affects its function as it relates to the disease.

50. Hammel M, Amlanjyoti D, Reyes FE, Chen JH, Parpana R, Tang HYH, Larabell CA, Tainer JA, Adhya S: **HU multimerization shift controls nucleoid compaction**. *Sci Adv* 2016, **2**, e1600650.

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

52. Sterckx YGJ, Volkov AN, Vranken WF, Kragelj J, Jensen MR, Buts L, Garcia-Pino A, Jové T, Van Melderen L, Blackledge M, *et al.*: **Small-angle X-ray scattering- and nuclear magnetic resonance-derived conformational ensemble of the highly flexible antitoxin PaaA2**. *Structure* 2014, **22**:854–865.

53. Castañeda CA, Kashyap TR, Nakasone MA, Krueger S, Fushman D: **Unique structural, dynamical, and functional properties of K11-linked polyubiquitin chains**. *Structure* 2013, **21**:1168–1181.

54. Hellenkamp B, Schmid S, Doroshenko O, Opanasyuk O, Kühnemuth R, Adariani SR, Ambrose B, Aznauryan M, Barth A, Birkedal V, *et al.*: **Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study**. *Nat Methods* 2018, **15**:669–676.

55. Ye Y, Blaser G, Horrocks MH, Ruedas-Rama MJ, Ibrahim S, Zhukov AA, Orte A, Klenerman D, Jackson SE, Komander D: **Ubiquitin chain conformation regulates recognition and activity of interacting proteins**. *Nature* 2012, **492**:266–270.

56. Schiemann O, Heubach CA, Abdullin D, Ackermann K, Azarkh M, Bagryanskaya EG, Drescher M, Endeward B, Freed JH, Galazzo L, *et al.*: **Benchmark test and guidelines for DEER/PELDOR experiments on nitroxide-labeled biomolecules**. *J Am Chem Soc* 2021, **143**:17875–17890.

57. Pannier M, Veit S, Godt A, Jeschke G, Spiess HW: **Dead-time free measurement of dipole-dipole interactions between electron spins**. *J Magn Reson* 2000, **142**:331–340.

58. Sweger SR, Pribitzer S, Stoll S: **Bayesian probabilistic analysis of DEER spectroscopy data using parametric distance distribution models**. *J Phys Chem A* 2020, **124**:6193–6202.

59. Kniss A, Schuetz D, Kazemi S, Pluska L, Spindler PE, Rogov VV, Husnjak K, Dikic I, Güntert P, Sommer T, *et al.*: **Chain assembly and disassembly processes differently affect the conformational space of ubiquitin chains**. *Structure* 2018, **26**: 249–258.e4.

60. Bonomi M, Pellarin R, Vendruscolo M: **Simultaneous determination of protein structure and dynamics using cryo-electron microscopy**. *Biophys J* 2018, **114**:1604–1613.

61. Bonomi M, Vendruscolo M: **Determination of protein structural ensembles using cryo-electron microscopy**. *Curr Opin Struct Biol* 2019, **56**:37–45.

62. Zweckstetter M: **NMR: prediction of molecular alignment from structure using the PALES software**. *Nat Protoc* 2008, **3**: 679–690.

63. Berlin K, O'Leary DP, Fushman D: **Improvement and analysis of computational methods for prediction of residual dipolar couplings**. *J Magn Reson* 2009, **201**:25–33.

64. Bonomi M, Heller GT, Camilloni C, Vendruscolo M: **Principles of protein structural ensemble determination**. *Curr Opin Struct Biol* 2017, **42**:106–116.

65. Ravera E, Sgheri L, Parigi G, Luchinat C: **A critical assessment of methods to recover information from averaged data**. *Phys Chem Chem Phys* 2016, **18**:5686–5701.

66. Qi Y, Martin JW, Barb AW, Thélot F, Yan AK, Donald BR, Oas TG: **Continuous interdomain orientation distributions reveal components of binding thermodynamics**. *J Mol Biol* 2018, **430**: 3412–3426.

67. Carpenter EP, Beis K, Cameron AD, Iwata S: **Overcoming the challenges of membrane protein crystallography**. *Curr Opin Struct Biol* 2008, **18**:581.

68. Liang B, Tamm LK: **NMR as a tool to investigate the structure, dynamics and function of membrane proteins**. *Nat Struct Mol Biol* 2016, **23**:468–474.

69. Curtis JE, Raghunandan S, Nanda H, Krueger S, SASSIE: **A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints**. *Comput Phys Commun* 2012, **183**: 382–389.

70. Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HDT, Konarev PV, Svergun DI: **New developments in the ATSAS program package for small-angle scattering data analysis**. *J Appl Crystallogr* 2012, **45**:342.