Representing Structures of the Multiple Conformational States of Proteins

Theresa A. Ramelot*, Roberto Tejero, and Gaetano T. Montelione*

Dept of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Sciences, Rensselaer Polytechnic Institute, Troy, New York, 12180 USA

Theresa A. Ramelot	orcid: 0000-0002-0335-1573	email: ramelt2@rpi.edu
Roberto Tejero	orcid: 0000-0003-2504-5988	email: roberto.tejero@uv.es
Gaetano T. Montelione	orcid: 0000-0002-9440-3059	email: monteg3@rpi.edu

Highlights

- Improved methods have advanced multi-conformational structural modeling
- Two or more multiple-state conformations often best describe a protein structure
- Single-state representation depicts local model uncertainty on one representative conformer
- Consistent data structures are needed for archiving multiple-state models

Abstract

Biomolecules exhibit dynamic behavior that single-state models of their structures cannot fully capture. We review some recent advances for investigating multiple conformations of biomolecules, including experimental methods, molecular dynamics simulations, and machine learning. We also address the challenges associated with representing single- and multiple-state models in data archives, with a particular focus on NMR structures. Establishing standardized representations and annotations will facilitate effective communication and understanding of these complex models to the broader scientific community.

Abbreviations: AF2 – AlphaFold2 Multimer; BMRB - biological magnetic resonance bank; CEST – chemical exchange by saturation transfer; CPMG – Carr-Purcell-Meiboom-Gill; DEER – double electron-electron resonance; FID - free induction decay NMR data; FRET - Förster resonance energy transfer; LDDT – local-distance difference test; MD – molecular dynamics; ML – machine learning; mmCIF - macromolecular crystallographic information file; MSA – multiple sequence alignment; NMR - nuclear magnetic resonance spectroscopy; PDB - protein data bank; pLDDT - predicted local model confidence score predicted from ML; PRE - paramagnetic relaxation enhancement; RDC – residual dipolar coupling; wwPDB – worldwide PDB.

*Corresponding authors. Email: ramelt2@rpi.edu; monteg3@rpi.edu

Introduction

Biomolecules exhibit dynamic behavior and adopt a distribution of conformations influenced by factors including sequence, temperature, pressure, ligand binding, and solution conditions. Traditionally, structural biology has predominantly focused on single-conformation models. However, it is broadly appreciated that most biomolecules must move to function. One of the earliest experimental demonstrations of protein structure plasticity came from NMR studies of aromatic-ring flips of the small protein bovine pancreatic trypsin inhibitor, where it was observed that conformational fluctuations allow rapid rotations of aromatic rings buried in the hydrophobic core [1]. Recent advances in experimental and computational methods illustrate the importance of multiple-conformation modeling for understanding biomolecule functions. In particular, as the machine learning (ML) methods of AlphaFold2 (AF2) [2], RosettaFold [3], OpenFold [4], ESMFold [5], RaptorX [6], and other advanced techniques have reached the stage where single structure prediction of small proteins is robust and reliable, and a current frontier is multiple-state modeling [7,8]. Establishing consistent ontologies and formats for representing such multiple-state models is crucial for supporting and advancing this important area of structural biology.

This perspective addresses the significance and handling of multiple-conformation models of biomolecules. We begin with some key definitions. Conformers refer to atomic structures capable of interconversion without making or breaking covalent bonds. Conformational "collections" are defined here as sets of these conformers. Note that this is a looser definition than the concept of "thermodynamic statistical ensemble" in statistical mechanics, which describes the Boltzmann distribution of conformations contributing to ensemble-averaged measurable parameters. Conformational “states” are distinct conformational models (or collections of conformational models) that, in principle, can be distinguished experimentally. Structural models can be categorized as “single-state” or “multiple-state” based on the nature of the experimental data or the theoretical inference. Multiple-state models may constitute a pair of conformers, pairs of conformational collections, or collections of many conformers of, for example, disordered polymer chains. The distinction between states, and the enumeration of the number of states, is determined by the interpretation used in modeling the data. Terms used for such multiple-state models in the literature include alternative conformations, multi-conformer models, switched folds, metamorphic states, chameleonic states, conformational ensembles, and conformational excited states. As various methods for interpreting experimental data in terms of multiple-state structural models are evolving, there is a current pressing need for standardized representations of such models and their corresponding data in structural databases and across the structural biology community.

The significance of multiple conformational states

Modeling multiple conformations is fundamental to understanding biomolecule functions, as dynamics determines their ability to carry out these functions. In this perspective, we focus on representing diverse conformational states of proteins, but similar challenges also apply to nucleic acids such as DNA and RNA [9,10]. Conformational dynamics underlie enzyme function [11,12,] and are especially important for membrane protein activities as receptors and transporters of ions, metabolites, and drugs. Protein-protein interfaces may also exhibit multiple conformational states, as observed, for example, in dimers of the

influenza A virus non-structural protein NS1 [13] and between domains of MHC class I molecules [14]. Additionally, the significance of intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) is increasingly recognized in biology [15,16].

Biomolecules often undergo conformational changes when interacting with binding partners or substrates, encompassing both induced fit and conformational selection mechanisms. These structural changes can occur at binding sites or be distributed across the structure. The significant role of allostery in enzyme function, where an allosteric modulator molecule binds to sites distal to the active site, has been recognized for decades [17]. Recent advances combining experimental data with advanced modeling methods have revealed structural details of allosteric mechanisms [17-20]. Evolutionary coupling (EC) based on sequence covariance analysis has also been utilized to enhance enzyme activities by perturbing allosteric networks with mutations distant from their active sites [21].

Structural heterogeneity is also relevant to *de novo* protein design, and has been successfully used to create cyclic chameleon peptides that switch between exposed hydrophobic and hydrophilic surfaces to provide membrane permeability [22], two-state hinge proteins [23], and fold-switching metamorphic proteins [24]. Membrane protein transporters have also been the subject of multiple-state *de novo* design efforts, such as the Zn^{2+}-transporting four-helix bundle transmembrane protein Rocker [25]. Advancements in the controlled design of proteins that switch between alternative conformations are crucial for creating novel protein effectors and catalysts.

Experimental methods for characterizing multiple conformational states

Recent reviews discuss the experimental methods that provide structural information on multiple conformational states of biomolecules [26-28]. Crystallographic studies, using either X-ray or neutron diffraction, may capture different states in different crystal forms. Electron density can also often be fit to multiple conformations within a single crystal. Room-temperature (or higher temperature) X-ray crystallography may avoid structural bias from cryogenic cooling and reveal motions crucial for catalysis, ligand binding, and allosteric regulation [12,28-30]. Other experimental data types, such as small-angle X-ray scattering (SAXS) [27] and electron microscopy (cryoEM) [31-33] data, can frequently only be fit to multi-conformer models. Additionally, Förster resonance energy transfer (FRET), Double Electron-Electron Resonance (DEER) spectroscopy, and chemical cross-linking data have been used to model multiple conformational states since they can characterize interprobe distance distributions [26]. With all experimental data, multiple-state fitting may be indicated and (cross)-validated only when a single-state model is inadequate.

Nuclear Magnetic Resonance (NMR) spectroscopy is a valuable tool for studying the dynamic behavior of biomolecules. It employs properties such as nuclear relaxation and chemical exchange saturation transfer to determine interconversion rates and populations of conformations [19,34-36]. NMR parameters reflect conformational averaging on parameter-specific timescales: for ^{1}H chemical shift, slow exchange (conformational lifetime >> ~ 1 ms) yields distinct resonances for individual states, fast exchange

(conformational lifetime << ~ 1 ms) results in population-weighted average resonance frequencies, and intermediate exchange leads to characteristic resonance lineshapes. The distinction between slow and fast exchange depends on the difference in the resonance frequencies of the individual states and the relative populations; e.g., for a two-state system with ^{1}H chemical shift differences between 0.1 and 1 ppm at 800 MHz, intermediate exchange corresponds to rates from ca. 10 to 250,000 s^{-1} [37]. Chemical shift refocusing experiments like Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion and $T1_{rho}$ relaxation experiments, and saturation transfer experiments (e.g., chemical exchange by saturation transfer, CEST), can provide quantitative information about conformational exchange on the intermediate or slow chemical shift timescale [34] and can be used to characterize sparsely-populated states that cannot otherwise be observed in NMR spectra [35]. NOESY and residual dipolar coupling (RDC) data can reveal multiple conformations in fast dynamic equilibrium, and may provide structural restraints for modeling each state [38-41]. Paramagnetic effects in metal-containing biomolecules provide ensemble-averaged distance restraints and can also determine ensemble-averaged relative orientations of structural domains [42,43]. In many cases, the structures of the conformations in dynamic exchange are modeled by fitting back-calculated NMR parameters (e.g., chemical shifts) to ensemble-averaged NMR data.

Computational methods for modeling multiple conformational states

For decades, molecular dynamics (MD) simulations and normal mode analysis methods have been utilized to study the various conformational states of biomolecules [44]. Integrating experimental data, such as NMR and X-ray crystallographic data, with MD simulations has led to improved conformational heterogeneity evaluation. MD has been combined with X-ray crystallography data to generate multiple-state models that have improved fit to X-ray data compared to single-structure models [45,46]. However, conventional MD simulations often fall short in capturing slower motions, particularly allosteric conformational changes. Approaches have also been developed to integrate experimental NMR data with MD simulations. By incorporating time-averaged distance restraints from NOE data [19,40], MD simulations can better model conformational distributions consistent with experimental data, as in the case of the DNA-binding loops of *E. coli* tryptophan repressor [47]. Bayesian inference and ensemble fitting approaches, which leverage experimental data alongside MD simulations, can also generate multiple-state models [48,49]. Similarly, chemical shift data have been used to guide or interpret computational methods. For example, NMR chemical shift perturbation analysis using programs like CHESCA has been used to characterize allosteric conformational switching upon ligand binding by chemical shift covariance analysis [50]. Alternative conformational state modeling with Ohm, a structural perturbation propagation method, for a set of ~ 20 allosterically-modulated proteins was observed to provide excellent predictions of CHESCA-based chemical shift changes [20]. Accelerated MD methods have also proven effective in modeling multiple conformational states of proteins, which align well with NMR data [19].

Machine learning (ML) approaches show tremendous promise for modeling conformational dynamics. Although AF2 was not trained to model protein dynamics, in some cases, it can provide information about the individual states in dynamic equilibrium. For example, comparison of AF2 and NMR models for CASP14 target T1027, *Gaussia* luciferase, against NMR data suggested that the AF2 prediction model

corresponds to just one of the multiple conformations in the NMR sample [8]. Subsequently, the proclivity of AF2 to model one of multiple conformational states was also reported for a collection of ~100 apo / holo protein structure pairs [51]. Recently, methods have emerged to extend AF2 and other machine learning networks to model alternative protein conformations explicitly. These extensions involve leveraging ECs that distinguish multiple conformations [52-54], employing multiple templates with diverse conformations [55,56], using shallow multiple-sequence alignments [56-58], or by perturbing the neural network weights [59] to generate conformational diversity. SPEACH_AF utilizes *in silico* mutations as input to AF2 to model conformational switching in soluble and membrane proteins [60]. AlphaFold2-RAVE uses the structural outputs from AF2 for AI-augmented MD, resulting in Boltzmann-ranked collections of conformations [61].

Databases of multiple conformational states

The advancements in experimental methods and modeling techniques for determining multiple conformational states of biomolecules necessitate improved methods for representing and archiving information about conformations in dynamic equilibrium. This can be challenging, as definitions of alternative conformational "states" depend on the timescale of the experimental data and/or the modeling methods used. Apart from the well-known Protein Data Bank (PDB) and Nucleic Acid Database (NDB), several other databases (ACMS, CoDNaS, D3PM, GLOCON, and MultiComp) primarily store and annotate data on alternative conformations obtained from diverse crystal structures, as reported in the PDB. These databases, along with other important structural databases that primarily archive single-state models but can also provide data about multiple conformational states, are listed in Table 1, along with their URLs.

Clearly, multiple conformational state information is important in biology, and there is a need for consistent representation of such information in databases. In the following sections, we discuss some challenges in representing collections of molecular models derived from NMR data. While some points are specific to NMR structures and data, most are relevant for representing biomacromolecule structures obtained via various experimental or predictive modeling techniques.

Representation of biomolecular NMR structures: single-state models

Solution NMR structures are typically represented as collections of coordinate sets, where each model in the collection is independently generated by fitting experimental data to a single conformer. This is done multiple times under different initialization conditions, generating a collection of conformers. The single-state model representation commonly used for NMR structures encodes information about which regions of the protein structure are "well-defined" by the NMR data and which regions are not. Less-well-defined segments of the structure often (but not always) correspond to regions undergoing conformational dynamics. In the simplest case, the coordinates for the collection of conformer models are deposited to the PDB, accompanied by restraints, while the chemical shifts, peaks lists, and raw FID data are deposited to the Biological Magnetic Resonance Bank (BMRB).

This single-state model representation can be confusing for scientists using NMR structures. The individual conformer models in the single-state representation are not meant to describe actual conformers contributing to the Boltzmann distribution of states present in the sample; rather the conformer collection provides information about the consistency of modeling. Although each conformer is considered to be a good fit to the data, the coordinate uncertainty in different regions can only be assessed by analyzing the conformer collection as a whole. The single-state model representation also does not provide a statistically reliable estimate of the atomic coordinate precision based on experimental measurement uncertainties, although Bayesian methods have been proposed for this purpose [48]. Despite its limitations, the prevailing convention for biomolecular structure modeling using NMR data continues to be the single-state model representation. When using such models, it is crucial to distinguish regions that are well-defined from the experimental data versus less precise regions where atomic positions are highly uncertain. This distinction is critical for the correct application of structure validation methods, which generally apply only to well-defined regions [62]. Accordingly, it is important that the single-state model representation is conveyed in a simple way to users of NMR structures.

In X-ray crystallography, single-state models use B-factors (which include various data quality effects) to describe the uncertainty of atomic positions. In a similar way, a single-state conformer collections generated from NMR data can be represented by a *single representative conformer*, along with information about coordinate uncertainty (Figure 1). The wwPDB uses the medoid conformer as the representative structure, defined as the single conformer in the collection most like all the other conformers [62,63]. Tools like Dihedral Angle Order Parameter (DAOP) [63,64], FindCore [65,66], and CYRANGE [67] assess well-defined and not-well-defined residue ranges across the NMR conformer collection. The cutoffs used by these tools are based on standardized conventions. Presently, the wwPDB has adopted CYRANGE conventions to annotate well-defined residue ranges in the NMR structure validation report. PDBStat [63] also uses these tools and writes information about well-ordered residues as well as atom-specific coordinate variances into a mmCIF (or conventional PDB format) coordinate file, allowing graphical rendering of this information onto a single representative conformer (e.g., the medoid conformer) (Figure 1). After aligning the models with respect to a core atom set, coordinate uncertainties are converted using the Debye-Waller equation to effective "NMR B-factors" [66], indicating the variability and uncertainty in atomic positions across the conformer collection. It is unfortunate that these annotations are not more widely used compared to the widespread adoption of the analogous concept of predicted LDDT (pLDDT) scores reported for AlphaFold2 models [2]. These annotations are essential for the informed use of NMR structures.

Representation of biomolecular NMR structures: multiple-state models

Solution NMR can provide valuable insights into multiple conformational states in dynamic equilibrium. Different modeling approaches are employed depending on the timescale of conformational averaging, such as slow or fast chemical shift exchange (Figure 2). These include generating single-state or multi-state models for each set of slowly exchanging resonances [22], deconvoluting information in spectra of rapidly-exchanging systems [39], fitting to spectral features of intermediate exchange [34,37], and matching chemical shifts of slowly-exchanging states to chemical shifts predicted from known structures

[68]. Various methods of data interpretation will generate different numbers of chemical shift lists, restraint lists, and coordinate sets, which need to be accounted for in creating the archived data structure.

The community does not yet have a consensus scheme for deposition of multiple-state models and supporting data (e.g., FID, chemical shift, and restraint data). Figure 2B illustrates examples of multiple-state models in the PDB. Some were deposited as pairs of separate PDB files, while in other cases, the multiple-state models were concatenated in a single PDB file. For the separate entries, although comments describing the relationships of these pairs of PDB files may be included in their header files, it is not always clear if these multiple conformations were derived from a single or multiple sets of experimental data, and information on relative populations may be lost. Conformational collections for three of the multiple-state models illustrated in Figure 2B were deposited as a single PBD file. In two other cases (5tm0 and 2lwa), the sets (2 or 3, respectively) of structures were reported with separate chain IDs within a single PDB file. In contrast, the individual multiple-state models reported in the single PDB file 7r95 were not distinguished by any specific designator. In X-ray crystal structures, alternative local conformations are often represented with distinct 'AltLoc IDs', the alternative location indicators, where atoms are assigned unique letters to represent different conformations [28,32,33]. These AltLocs can range from single atoms to sets of connected or non-connected residues, and have the benefit of including relative populations of conformers as well as the positions for each atom with alternative coordinates. Despite progress in automating the assignment and creation of PDB files with AltLoc annotations, challenges persist regarding ease of interpretation and how these annotations are used. For some software, only one of the alternative conformations is used, and there are compatibility issues with other analysis software. In NMR studies, like those shown in Figure 2B, multiple-state models are usually refined using multiple complete copies of the entire molecule for each state. This does not easily align with the AltLoc ID usage, which also does not currently support association with the corresponding alternative chemical shift assignments.

To address these inconsistencies and improve data organization, it is crucial to establish standardized conventions for archiving multiple-state models and their relative populations in the PDB and other structural databases. Additionally, it is essential to ensure that the underlying experimental data, including raw FID data, are archived along with the model coordinates [69]. This will allow for future reinterpretation of data and regeneration of models as methods improve. As illustrated in Figure 2C, for cases of multiple coordinate sets derived from NMR data on fast-exchanging systems, there are (i) a set of raw data, (ii) a single set of chemical shifts, but potentially (iii) two (or more) sets of restraints, and (iv) two (or more) sets of atomic coordinates. In the case of multiple coordinate sets derived from slowly exchanging systems, there is again (i) a single set of data, but (ii) multiple sets of chemical shifts, one associated with each member of the slowly exchanging system, as well as potentially (iii) two (or more) sets of restraints, and (iv) two (or more) sets of atomic coordinates. This data organization required for representing multiple-state NMR structures is not currently supported by public biomolecular structure databases.

Generalization to other experimentally- and computationally-generated multiple-state models

The issues of representing single-state and multiple-state models also impact the representation of biomacromolecular structures based on cryoEM, X-ray crystallography, FRET, chemical cross-linking data, and other experimental methods. Issues of data structures needed to represent these models and data are analogous to those discussed above for solution NMR data. However, their details are beyond the scope of this perspective. Ensuring consistency in representing multiple conformational states modeled from various experimental and computational methods is critical for developing integrated structural biology methods and advancing dynamic modeling techniques.

Future outlook

The representation and management of multiple conformational models of biomolecules is very important for understanding their dynamic behavior and how these dynamics influence their biochemical functions. Standardized ontologies and formats for representing conformational collections, thermodynamic ensembles, and multiple-state models are crucial for effective communication and integration of structural data across the scientific community. Of course, the format of data deposition is inextricably related to the viewing capabilities of the software tools used to display such structures. Recent advances in experimental and computational methods, including machine learning (ML), provide exciting new opportunities for modeling and characterizing multiple conformations. Experimental distance restraint data can also be used as input for training ML-based structure prediction methods [70] and will certainly impact ML-based methods for modeling multiple conformational states. As combined experimental and computational methods develop, models of the multiple conformational states of proteins and nucleic acids will enable biochemical, biophysical, and biological studies. The ability to consistently represent and archive information about conformations in dynamic equilibrium will facilitate research and enhance our understanding of biomolecular function.

Acknowledgments

We thank Drs. K. Baskaran, J. Hoch, D. Snyder, and S. Valenkar, as well as N. Dube, K. Fraga, Y.J. Huang, B. Shurina, and G.V.T Swapna for helpful discussions. This research was supported by grant R35-GM141818 (to GTM) from the National Institutes of Health.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT3.5 to improve manuscript readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of Interests

GTM is a founder of Nexomics Biosciences, Inc. This does not represent a conflict of interest for this study.

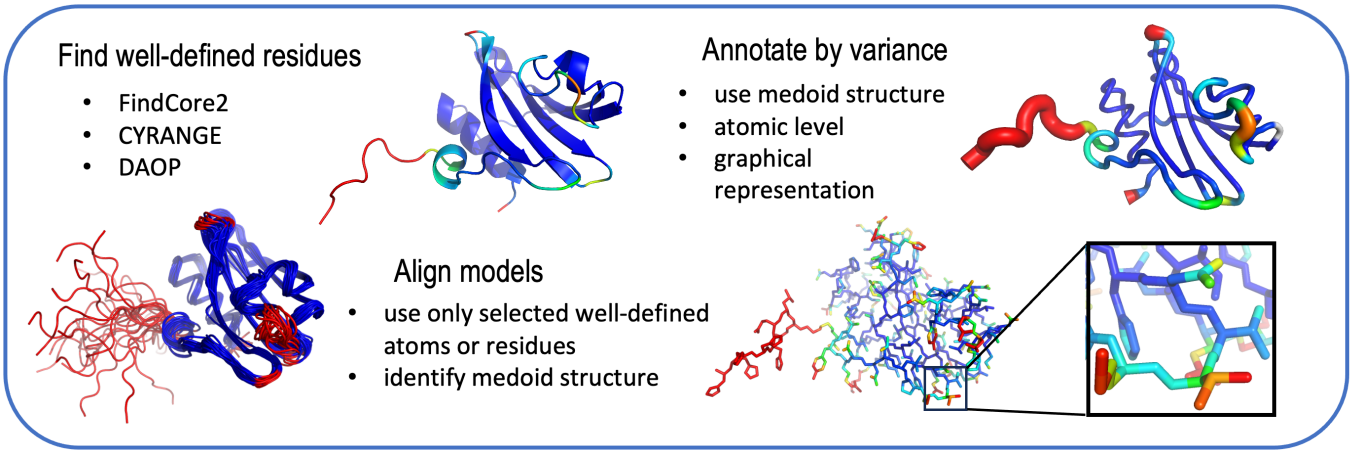

Fig. 1. Schematic representations of conformational collections for "single-state" models. PDB entries for NMR structures typically consist of a collection of ~20 conformers obtained through restrained modeling with NMR data. "Well-defined" parts of the structure can be determined using conventions encoded in programs like CYRANGE [67], FindCore2 [66], and dihedral angle order parameter DAOP [63,64] software, and colored to indicate well-defined (blue) and not-well-defined (red) residues (left side, pdb 2kcd). The PDBStat program [63] provides tools to create an image of the protein annotated with this information about model convergence. The medoid structure, determined by aligning models using well-defined heavy (or backbone N, Cα, C') atoms, is the first entry in a new file containing coordinates for the conformer collection in mmCIF (or conventional) PDB format. This file includes per-residue tags (q=1 for well-defined residues, q=0 for others). In addition, atom-specific coordinate variances are determined from the average atomic root mean-squared fluctuation (RMSF) across the conformer collection [65,66,71], and reported as effective "NMR B-factors", representing positional uncertainties across the collection. These mmCIF (or conventional) PDB format files can then be used to visualize well-defined regions (left) or atomic variances (shown schematically in three ways, top and right side) using programs like PyMOL. Representations showing atomic variances by coloring or by scaling the size of the ribbon are shown. For multiple domain structures, variance matrix analysis is used to parse the coordinates into well-defined units, which are then analyzed separately [65-67].

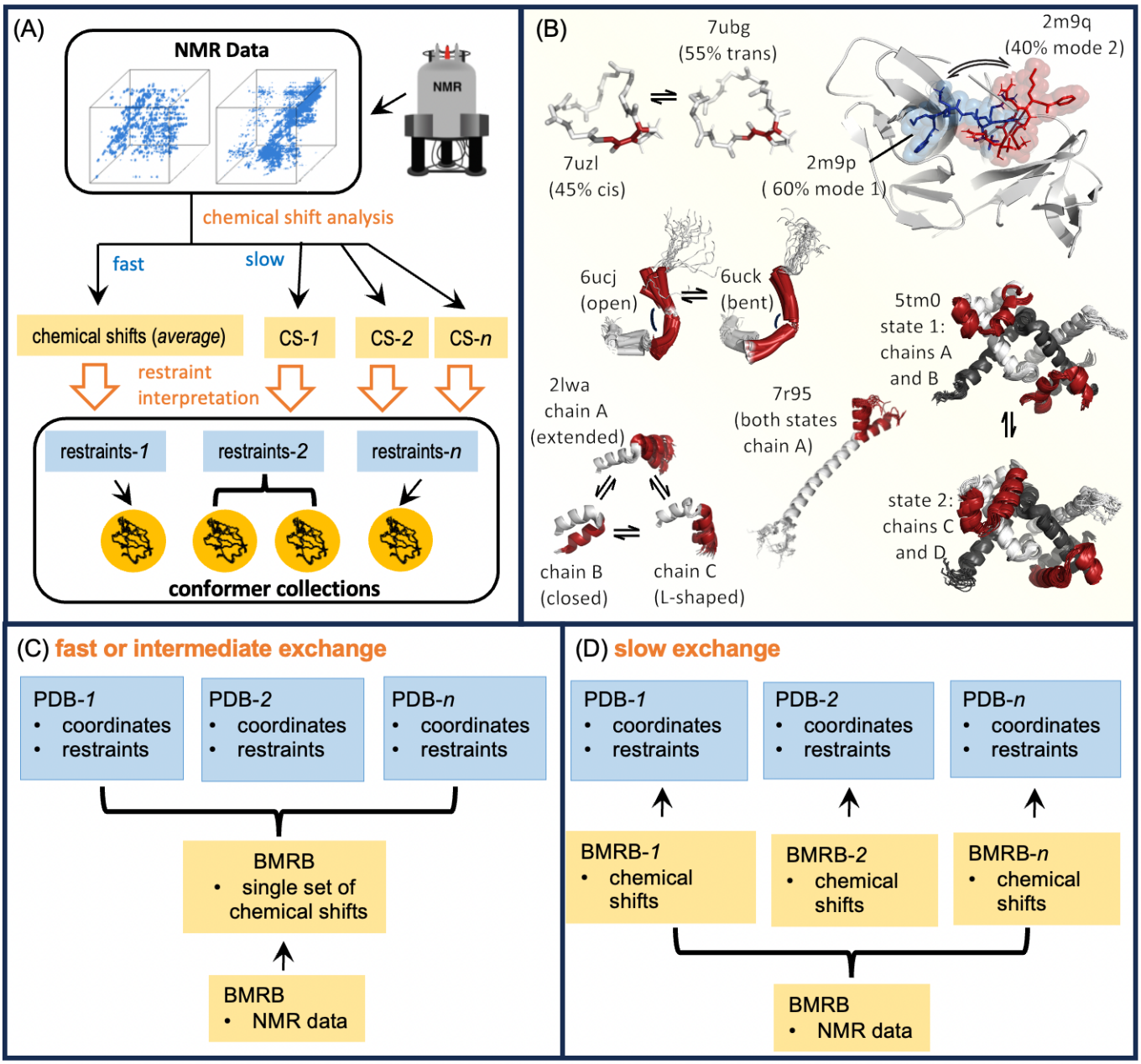

Fig. 2. Schematic of NMR data deposition pipeline for proteins reported to have multiple-state structures

(A) The process involves data collection and chemical shift analysis resulting in one or more sets of chemical shifts. NMR data interpretation is then used to derive one or more sets of restraints used to model one or more conformational states. (B) Representative examples of two-state models including: 9-residue Rosetta-designed cyclic peptide D9.16 (pdb IDs 7ubg and 7uzl) [22], inhibitor-bound dengue virus NS2B/NS3 protease (two binding modes, 2m9p and 2m9q) [39], pro-islet amyloid polypeptide in detergent micelles (6ucj and 6uck) [72], *E. coli* tryptophan repressor (two states combined as pdb ID 5tm0) [47], and the membrane-bound SARS-CoV-2 spike protein HR1 ectodomain (two states combined as 7r95) [73]. Also shown is the three-state model of influenza hemagglutinin fusion peptide A (combined

as 2lwa) [74] (C and D). Schematic representations for the data organization required for deposition of multiple-state models into the PDB and BMRB: (C) Fast or intermediate exchange between the conformers yields population-averaged chemical shifts, resulting in one or more sets of restraints and corresponding PDB coordinate sets. (D) Slow exchange between conformers leads to distinct chemical shifts, with multiple chemical shift entries from a single NMR dataset that are then used to calculate multiple sets of PDB coordinates. As in the fast/intermediate exchange case, a single set of chemical shifts arising from one slow-exchange ensemble may generate multiple restraints, leading to multiple coordinate sets, depending on the data analysis method. Multiple-state models may also be generated from a single restraint set in certain cases.

359 **Table 1. Databases of multiple conformational states of proteins and nucleic acids, and related**
360 **resources**

Database	url
ACMS: Provides a detailed description of the alternate conformations of various residues for more than 60,000 high-resolution crystal structures.	http://iris.physics.iisc.ac.in/acms
AlphaFold Protein Structure Database: Provides over 200 million protein structure predictions.	https://alphafold.ebi.ac.uk
Binding MOAD - Mother of All Databases: A subset of the Protein Data Bank (PDB), with many high-quality structures of ligand-protein complexes.	http://www.bindingmoad.org
Biological Magnetic Resonance Bank (BMRB): Archive of biological NMR data.	https://bmrb.io
CoDNaS 2.0: A comprehensive database of protein conformational diversity.	http://ufq.unq.edu.ar/codnas
D3PM: A comprehensive database for protein motions, including changes with ligand binding.	http://www.d3pharma.com/D3PM/index.php
DNAproDB: Web-based database and structural analysis tool designed to access and visualize structural data of DNA–protein complexes.	https://dnaprodb.usc.edu
Electron Microscopy Data Bank (EMDB): A public repository for electron cryo-electron microscopy maps and tomograms of macromolecular complexes and subcellular structures.	https://www.ebi.ac.uk/emdb
EM Data Resource: A unified data resource for 3-Dimensional Electron Microscopy (3DEM) structure data archiving and retrieval.	https://www.emdataresource.org
ESM Metagenomic Atlas: Atlas of 772 million predicted metagenomic protein structures.	https://esmatlas.com
GLOCON: Clusters of protein chains based on a GLObal CONformation difference score, computed from pairwise C-alpha distances.	https://doi.org/10.1101/2023.07.13.545008
MultiComp: A database for exploring multiple conformations of membrane proteins.	https://multicomp.nibiohn.go.jp
Nucleic Acid Data Bank (NDB) and Nucleic Acid Knowledgebase: A data resource for experimentally determined structures containing DNA and RNA nucleic acid polymers and their biological assemblies.	http://nakb.org

wwPDB worldwide Protein Data Bank: Primary international data repository for protein and nucleic structures.	https://www.wwpdb.org

361

References

1. Wuthrich K, Wagner G: **NMR investigations of the dynamics of the aromatic amino acid residues in the basic pancreatic trypsin inhibitor**. *FEBS Lett* 1975, **50**:265-268.
2. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, et al.: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583-589.
3. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, et al.: **Accurate prediction of protein structures and interactions using a three-track neural network**. *Science* 2021, **373**:871-876.
4. Ahdritz G, Bouatta N, Kadyan S, Xia Q, Gerecke W, O'Donnell TJ, Berenberg D, Fisk I, Zanichelli N, Zhang B, et al.: **OpenFold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization**. *bioRxiv* 2022:2022.2011.2020.517210.
5. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, Smetanin N, Verkuil R, Kabeli O, Shmueli Y, et al.: **Evolutionary-scale prediction of atomic-level protein structure with a language model**. *Science* 2023, **379**:1123-1130.
6. Wu F, Jing X, Luo X, Xu J: **Improving protein structure prediction using templates and sequence embedding**. *Bioinformatics* 2023, **39**.
7. Lane TJ: **Protein structure prediction has reached the single-structure frontier**. *Nature Methods* 2023, **20**:170-173.
8. Huang YJ, Zhang N, Bersch B, Fidelis K, Inouye M, Ishida Y, Kryshtafovych A, Kobayashi N, Kuroda Y, Liu G: **Assessment of prediction methods for protein structures determined by NMR in CASP14: impact of AlphaFold2**. *Proteins-Structure, Function and Bioinformatics* 2021.
9. Nikolova EN, Kim E, Wise AA, O'Brien PJ, Andricioaei I, Al-Hashimi HM: **Transient Hoogsteen base pairs in canonical duplex DNA**. *Nature* 2011, **470**:498-502.
10. Liu B, Shi H, Al-Hashimi HM: **Developments in solution-state NMR yield broader and deeper views of the dynamic ensembles of nucleic acids**. *Curr Opin Struct Biol* 2021, **70**:16-25.
11. Palmer AG, 3rd: **Enzyme dynamics from NMR spectroscopy**. *Acc Chem Res* 2015, **48**:457-465.
12. Otten R, Padua RAP, Bunzel HA, Nguyen V, Pitsawong W, Patterson M, Sui S, Perry SL, Cohen AE, Hilvert D, et al.: **How directed evolution reshapes the energy landscape in an enzyme to boost catalysis**. *Science* 2020, **370**:1442-1446.
13. Aramini JM, Hamilton K, Ma LC, Swapna GVT, Leonard PG, Ladbury JE, Krug RM, Montelione GT: **(19)F NMR reveals multiple conformations at the dimer interface of the nonstructural protein 1 effector domain from influenza A virus**. *Structure* 2014, **22**:515-525.
14. Beerbaum M, Ballaschk M, Erdmann N, Schnick C, Diehl A, Uchanska-Ziegler B, Ziegler A, Schmieder P: **NMR spectroscopy reveals unexpected structural variation at the protein–protein interface in MHC class I molecules**. *Journal of Biomolecular NMR* 2013, **57**:167-178.
15. Camacho-Zarco AR, Schnapka V, Guseva S, Abyzov A, Adamski W, Milles S, Jensen MR, Zidek L, Salvi N, Blackledge M: **NMR Provides Unique Insight into the Functional Dynamics and Interactions of Intrinsically Disordered Proteins**. *Chem Rev* 2022, **122**:9331-9356.
16. Abyzov A, Blackledge M, Zweckstetter M: **Conformational Dynamics of Intrinsically Disordered Proteins Regulate Biomolecular Condensate Chemistry**. *Chem Rev* 2022, **122**:6719-6748.
17. Nussinov R, Zhang M, Liu Y, Jang H: **AlphaFold, Artificial Intelligence (AI), and allostery**. *J Phys Chem B* 2022, **126**:6372-6383.
18. Schueler-Furman O, Wodak SJ: **Computational approaches to investigating allostery**. *Current Opinion in Structural Biology* 2016, **41**:159-171.

19. East KW, Skeens E, Cui JY, Belato HB, Mitchell B, Hsu R, Batista VS, Palermo G, Lisi GP: **NMR and computational methods for molecular resolution of allosteric pathways in enzyme complexes**. *Biophys Rev* 2020, **12**:155-174.
20. Wang J, Jain A, McDonald LR, Gambogi C, Lee AL, Dokholyan NV: **Mapping allosteric communications within individual proteins**. *Nature Communications* 2020, **11**:3862.
21. Wang X, Jing X, Deng Y, Nie Y, Xu F, Xu Y, Zhao YL, Hunt JF, Montelione GT, Szyperski T: **Evolutionary coupling saturation mutagenesis: Coevolution-guided identification of distant sites influencing Bacillus naganoensis pullulanase activity**. *FEBS letters* 2020, **594**:799-812.
22. Bhardwaj G, O'Connor J, Rettie S, Huang YH, Ramelot TA, Mulligan VK, Alpkilic GG, Palmer J, Bera AK, Bick MJ, et al.: **Accurate de novo design of membrane-traversing macrocycles**. *Cell* 2022, **185**:3520-3532 e3526.
23. Praetorius F, Leung PJY, Tessmer MH, Broerman A, Demakis C, Dishman AF, Pillai A, Idris A, Juergens D, Dauparas J, et al.: **Design of stimulus-responsive two-state hinge proteins**. Edited by: Cold Spring Harbor Laboratory; 2023.
24. Dishman AF, Volkman BF: **Design and discovery of metamorphic proteins**. *Current Opinion in Structural Biology* 2022, **74**:102380.
25. Joh NH, Grigoryan G, Wu Y, DeGrado WF: **Design of self-assembling transmembrane helical bundles to elucidate principles required for membrane protein folding and ion transport**. *Philos Trans R Soc Lond B Biol Sci* 2017, **372**.
26. Gama Lima Costa R, Fushman D: **Reweighting methods for elucidation of conformation ensembles of proteins**. *Curr Opin Struct Biol* 2022, **77**:102470.
27. Chinnam NB, Syed A, Hura GL, Hammel M, Tainer JA, Tsutakawa SE: **Combining small angle X-ray scattering (SAXS) with protein structure predictions to characterize conformations in solution**. *Methods Enzymol* 2023, **678**:351-376.
28. Du S, Wankowicz SA, Yabukarski F, Doukov T, Herschlag D, Fraser JS: **Refinement of Multiconformer Ensemble Models from Multi-temperature X-ray Diffraction Data**. *bioRxiv* 2023.
29. van den Bedem H, Dhanik A, Latombe JC, Deacon AM: **Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers**. *Acta Crystallogr D Biol Crystallogr* 2009, **65**:1107-1117.
30. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T: **Accessing protein conformational ensembles using room-temperature X-ray crystallography**. *Proc Natl Acad Sci U S A* 2011, **108**:16247-16252.
31. Bonomi M, Pellarin R, Vendruscolo M: **Simultaneous Determination of Protein Structure and Dynamics Using Cryo-Electron Microscopy**. *Biophys J* 2018, **114**:1604-1613.
32. Riley BT, Wankowicz SA, de Oliveira SHP, van Zundert GCP, Hogan DW, Fraser JS, Keedy DA, van den Bedem H: **qFit 3: Protein and ligand multiconformer modeling for X-ray crystallographic and single-particle cryo-EM density maps**. *Protein Sci* 2021, **30**:270-285.
33. Wankowicz SA, Ravikumar A, Sharma S, Riley BT, Raju A, Hogan DW, van den Bedem H, Keedy DA, Fraser JS: **Uncovering Protein Ensembles: Automated Multiconformer Model Building for X-ray Crystallography and Cryo-EM**. *bioRxiv* 2023.
34. Palmer AG: **Chemical exchange in biomacromolecules: Past, present, and future**. *Journal of Magnetic Resonance* 2014, **241**:3-17.
35. Alderson TR, Kay LE: **Unveiling invisible protein states with NMR spectroscopy**. *Curr Opin Struct Biol* 2020, **60**:39-49.

36. Jain S, Sekhar A: **Elucidating the mechanisms underlying protein conformational switching using NMR spectroscopy**. *Journal of Magnetic Resonance Open* 2022, **10-11**:100034.
37. Waudby CA, Alfonso I: **An introduction to one- and two-dimensional lineshape analysis of chemically exchanging systems**. *Journal of Magnetic Resonance Open* 2023, **16-17**:100102.
38. Vogeli B: **The nuclear Overhauser effect from a quantitative perspective**. *Prog Nucl Magn Reson Spectrosc* 2014, **78**:1-46.
39. Gibbs AC, Steele R, Liu G, Tounge BA, Montelione GT: **Inhibitor Bound Dengue NS2B-NS3pro Reveals Multiple Dynamic Binding Modes**. *Biochemistry* 2018, **57**:1591-1602.
40. Papaleo E, Camilloni C, Teilum K, Vendruscolo M, Lindorff-Larsen K: **Molecular dynamics ensemble refinement of the heterogeneous native state of NCBD using chemical shifts and NOEs**. *PeerJ* 2018, **6**:e5125.
41. Omar H, Hein A, Cole CA, Valafar H: **Concurrent Identification and Characterization of Protein Structure and Continuous Internal Dynamics with REDCRAFT**. *Front Mol Biosci* 2022, **9**:806584.
42. Parigi G, Ravera E, Piccioli M, Luchinat C: **Paramagnetic NMR restraints for the characterization of protein structural rearrangements**. *Curr Opin Struct Biol* 2023, **80**:102595.
43. Karschin N, Becker S, Griesinger C: **Interdomain Dynamics via Paramagnetic NMR on the Highly Flexible Complex Calmodulin/Munc13-1**. *J Am Chem Soc* 2022, **144**:17041-17053.
44. Banerjee A, Saha S, Tvedt NC, Yang LW, Bahar I: **Mutually beneficial confluence of structure-based modeling of protein dynamics and machine learning methods**. *Curr Opin Struct Biol* 2023, **78**:102517.
45. Gros P, van Gunsteren WF, Hol WG: **Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics**. *Science* 1990, **249**:1149-1152.
46. Burnley BT, Afonine PV, Adams PD, Gros P: **Modelling dynamics in protein crystal structures by ensemble refinement**. *Elife* 2012, **1**:e00311.
47. Harish B, Swapna GV, Kornhaber GJ, Montelione GT, Carey J: **Multiple helical conformations of the helix-turn-helix region revealed by NOE-restrained MD simulations of tryptophan aporepressor, TrpR**. *Proteins* 2017, **85**:731-740.
48. Rieping W, Habeck M, Nilges M: **Inferential structure determination**. *Science* 2005, **309**:303-306.
49. Kofinger J, Stelzl LS, Reuter K, Allande C, Reichel K, Hummer G: **Efficient Ensemble Refinement by Reweighting**. *J Chem Theory Comput* 2019, **15**:3390-3401.
50. Boulton S, Selvaratnam R, Ahmed R, Melacini G: **Implementation of the NMR CHEmical Shift Covariance Analysis (CHESCA): A Chemical Biologist's Approach to Allostery**. *Methods Mol Biol* 2018, **1688**:391-405.
51. Saldano T, Escobedo N, Marchetti J, Zea DJ, Mac Donagh J, Velez Rueda AJ, Gonik E, Garcia Melani A, Novomisky Nechcoff J, Salas MN, et al.: **Impact of protein conformational diversity on AlphaFold predictions**. *Bioinformatics* 2022, **38**:2742-2748.
52. Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS: **Three-dimensional structures of membrane proteins from genomic sequencing**. *Cell* 2012, **149**:1607-1621.
53. Zheng S, Sham LT, Rubino FA, Brock KP, Robins WP, Mekalanos JJ, Marks DS, Bernhardt TG, Kruse AC: **Structure and mutagenic analysis of the lipid II flippase MurJ from Escherichia coli**. *Proc Natl Acad Sci U S A* 2018, **115**:6709-6714.
54. Karamanos TK: **Chasing long-range evolutionary couplings in the AlphaFold era**. *Biopolymers* 2023, **114**:e23530.
55. Heo L, Feig M: **Multi-state modeling of G-protein coupled receptors at experimental accuracy**. *Proteins* 2022, **90**:1873-1885.

56. Sala D, Hildebrand PW, Meiler J: **Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties**. *Front Mol Biosci* 2023, **10**:1121962.
57. Del Alamo D, Sala D, McHaourab HS, Meiler J: **Sampling alternative conformational states of transporters and receptors with AlphaFold2**. *Elife* 2022, **11**.
58. Wayment-Steele HK, Ovchinnikov S, Colwell L, Kern D: **Prediction of multiple conformational states by combining sequence clustering with AlphaFold2**. Edited by: Cold Spring Harbor Laboratory; 2022.
59. Johansson-Akhe I, Wallner B: **Improving peptide-protein docking with AlphaFold-Multimer using forced sampling**. *Front Bioinform* 2022, **2**:959160.
60. Stein RA, McHaourab HS: **SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2**. *PLOS Computational Biology* 2022, **18**:e1010483.
61. Vani BP, Aranganathan A, Wang D, Tiwary P: **AlphaFold2-RAVE: From sequence to Boltzmann ensemble**. Edited by: Cold Spring Harbor Laboratory; 2022.
62. Montelione GT, Nilges M, Bax A, Guntert P, Herrmann T, Richardson JS, Schwieters CD, Vranken WF, Vuister GW, Wishart DS, et al.: **Recommendations of the wwPDB NMR Validation Task Force**. *Structure* 2013, **21**:1563-1570.
63. Tejero R, Snyder D, Mao B, Aramini JM, Montelione GT: **PDBStat: a universal restraint converter and restraint analysis software package for protein NMR**. *J Biomol NMR* 2013, **56**:337-351.
64. Hyberts SG, Goldberg MS, Havel TF, Wagner G: **The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures**. *Protein Sci* 1992, **1**:736-751.
65. Snyder DA, Montelione GT: **Clustering algorithms for identifying core atom sets and for assessing the precision of protein structure ensembles**. *Proteins* 2005, **59**:673-686.
66. Snyder DA, Grullon J, Huang YJ, Tejero R, Montelione GT: **The expanded FindCore method for identification of a core atom set for assessment of protein structure prediction**. *Proteins* 2014, **82 Suppl 2**:219-230.
67. Kirchner DK, Guntert P: **Objective identification of residue ranges for the superposition of protein structures**. *BMC Bioinformatics* 2011, **12**:170.
68. Xie T, Saleh T, Rossi P, Kalodimos CG: **Conformational states dynamically populated by a kinase determine its function**. *Science* 2020, **370**.
69. Fraga KJ, Huang YJ, Ramelot TA, Swapna GVT, Lashawn Anak Kendary A, Li E, Korf I, Montelione GT: **SpecDB: A relational database for archiving biomolecular NMR spectral data**. *J Magn Reson* 2022, **342**:107268.
70. Stahl K, Graziadei A, Dau T, Brock O, Rappsilber J: **Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning**. *Nat Biotechnol* 2023.
71. Kuzmanic A, Zagrovic B: **Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors**. *Biophys J* 2010, **98**:861-871.
72. DeLisle CF, Malooley AL, Banerjee I, Lorieau JL: **Pro-islet amyloid polypeptide in micelles contains a helical prohormone segment**. *FEBS J* 2020, **287**:4440-4457.
73. Chiliveri SC, Louis JM, Ghirlando R, Bax A: **Transient lipid-bound states of spike protein heptad repeats provide insights into SARS-CoV-2 membrane fusion**. *Science Advances* 2021, **7**:eabk2226.
74. Lorieau JL, Louis JM, Schwieters CD, Bax A: **pH-triggered, activated-state conformations of the influenza hemagglutinin fusion peptide revealed by NMR**. *Proc Natl Acad Sci U S A* 2012, **109**:19994-19999.