Predicting pose distribution of protein domains connected by flexible linkers is an unsolved problem

Allen C. McBride[†1], Feng Yu[†2], Edward H. Cheng[†3], Aulane Mpouli[†3], Aimee C. Soe[2], Michal Hammel[2], Gaetano T. Montelione[4], Terrence G. Oas[5], Susan E. Tsutakawa[*2], and Bruce R. Donald[*1,3,5,6]

[1]Department of Computer Science, Duke University

[2]Department of Structural Biology, Lawrence Berkeley National Laboratory

[3]Department of Chemistry, Duke University

[4]Department of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Sciences, Rensselaer Polytechnic Institute

[5]Department of Biochemistry, Duke University

[6]Department of Mathematics, Duke University

April 27, 2025

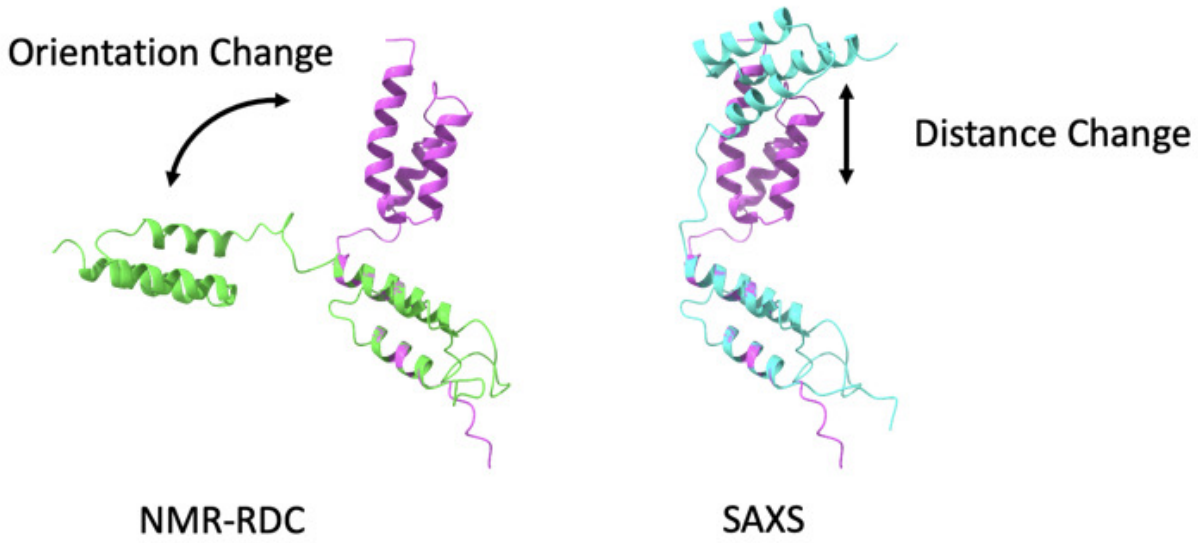
Abstract

In CASP16, we assessed the ability of computational methods to predict the distribution of relative orientations of two domains tethered by a flexible linker. The range of interdomain distances and orientations (poses) of such domain-linker-domain (D-L-D) proteins can play an important role in protein function, allostery, aggregation, and the thermodynamics of binding. The CASP16 Conformational Ensembles Experiment included two challenges to predict the interdomain pose distribution of a Staphylococcal protein A (SpA) D-L-D construct, called ZLBT-C, in which two of SpA's five nearly identical domains are connected by either (1) a six-residue wild-type (WT) linker (KADNKF), or (2) an all-glycine (Gly6) linker. The wild-type linker has a highly conserved sequence and is thought to contribute to the energetic barrier for binding with host antibodies. Ground truth was provided by nuclear magnetic resonance (NMR) residual dipolar coupling (RDC) data on WT protein and small angle X-ray scattering (SAXS) data on both proteins in solution. Twenty-five predictor groups submitted 35 sets of predicted conformational distributions, in the form of population-weighted finite ensembles of discrete structures. Unlike traditional CASP assessments that compare predicted atomic models to experimental atomic models, the accuracy of these predictions was assessed by back-calculating NMR RDCs and

[†]These authors contributed equally to this work

[*]Correspondence to: brd+proteins25@cs.duke.edu, setsutakawa@lbl.gov

SAXS curves from each ensemble of atomic models and comparing these results to respective experimental data. Accuracy was also assessed by using kernelization to compare ensembles to the continuous orientational distributions optimally fit to experimental data. In our assessment, predictions spanned a wide range of accuracy, but none were close fits to the combined NMR and SAXS data. In addition, none were able to recapitulate the observed difference between WT and Gly6 proteins, as observed in the SAXS data. These results, and our analysis, highlighted strengths and weaknesses, plus complementarity of NMR RDC and SAXS analysis.



1 Introduction

Short, flexible sequences of amino acids often link more rigid domains, both in nature and in engineered proteins [24, 14, 51]. These linkers exist in constantly changing conformational ensembles and do not fold into a stable native structure [56]. Although linkers, like intrinsically disordered regions (IDRs) more generally, are considered as random coils to be described by polymer theory [23], they usually have constrained flexibility due to amino acid sequence constraints—both flexible yet not fully random. For such linkers, the size and shape of amino acid sidechains will constrain the conformation of the linker backbone and thereby the pose distributions of linked domains. If this pose distribution has functional significance, such as transcription activation, phase separation, or antibody binding, the sequence is likely conserved, which is atypical among flexible linkers [26, 33, 9, 45].

Characterizing and predicting not only the stiffness of linkers, but also the specific distributions of interdomain poses that they allow, is an important challenge [45, 7, 25, 1, 4]. Differences in these distributions among different linkers is determined by linker sequence as well as by the properties of the domains they connect. A better understanding of the relationship between linker sequence and interdomain pose distribution would contribute to our study of the natural biological systems containing such linkers; it would also allow us to engineer linkers to achieve desired pose distributions for improved binding in therapeutic applications or for optimized synthetic

pathways in biomanufacturing.

Linked multidomain proteins belong to a larger class of proteins and protein regions possessing flexible and highly dynamic regions that do not fold into a stable native structure [56]. Such sequences longer than 30 amino acids are identified as intrinsically disordered proteins (IDPs) or IDRs [60, 63]. At least 40% of human proteins contain IDRs.

The two Critical Assessment of protein Structure Prediction (CASP) targets assessed in this report, T1200 and T1300, represent the first time that ensemble predictions have been assessed on the basis of how well each predicted ensemble, taken as a whole, fits experimental data collected from real ensembles in solution. Prior to CASP15, flexibility-related CASP challenges assessed the ability to predict which regions in a sequence were flexible (IDRs), but not the actual conformational distributions adopted by flexible regions [39, 40, 41, 10, 32, 38]. CASP15 introduced ensemble targets to CASP; however, these targets were assessed on the basis of individual models, not as conformational distributions [34].

T1200 and T1300 are two domain-linker-domain (D-L-D) proteins with the same rigid domains but different 6-amino-acid flexible linker regions. For these targets, a X-ray crystallographic structure of the two rigid domains of each construct is known and published in the Protein Data Bank (PDB) (IDs: 2LR2 [6]; 4NPD [19]). With the given sequence and reference structure of the rigid domain, predictors were instructed to predict how the linker sequence impacts the relative pose of the two well-folded protein domains by submitting an ensemble with population weights. The ensembles were to be assessed in relation to nuclear magnetic resonance (NMR) and small angle X-ray scattering (SAXS) experimental data, as well as to our previously modeled orientational distributions for one of the constructs [45], none of which predictors had access to.

The T1200 and T1300 targets are biologically relevant because they are constructs derived from protein A, which is a virulence factor in *Staphylococcus aureus*, binding to mammalian antibodies and inhibiting opsonization. Understanding the structure and function of protein A in *S. aureus* (SpA) could allow for the design of improved vaccines or therapeutics targeting *S. aureus* virulence. The N-terminal half of SpA (SpA-N) consists of five homologous three-helix domains connected by four homologous flexible linkers. These linkers are highly conserved in evolution, possibly because their constrained flexibility supports binding to antibodies such as to the Fc region of IgG [45].

The T1200 and T1300 targets correspond to two simplified D-L-D systems created to further investigate the role of these linkers in the flexibility and function of protein A, using two experimental methods (NMR and SAXS). The T1200 target is the D-L-D construct ZLBT-C. The C domain is the wild-type C domain of SpA-N; the ZLBT domain is similar to the B domain of SpA-N but includes a lanthanide binding tag (LBT) described by Barb and Subedi [5] and modified from Barb et al. [6] to facilitate partial alignment in the NMR magnetic field for purposes of residual dipolar coupling (RDC) experiments. The flexible linker between the two domains is the wild-type

linker (KADNKF) in three of the four linkers of SpA-N. The T1300 target is identical to T1200 except that the linker sequence GGGGGG is substituted for KADNKF. The intent of this substitution is to increase flexibility [45]. The constructs for T1200 and T1300 are referred to here as "WT" and "Gly6," respectively.

The NMR data used in this assessment comes from our prior work applying RDCs to characterize flexibility [45]. RDCs carry information about the orientation of internuclear bond vectors, such as N-H and C-H, with respect to the external magnetic field in an NMR experiment [12]. As molecules tumble and sample various conformations in solution, the angles between these bond vectors and the magnetic field change. The RDC experiment captures the average angle of these vectors over the experimental timescale, representing both time- and population-averaged molecular dynamics. This averaging reflects the distribution of interdomain orientations in the sample and can be represented as continuous probability distributions. Historically, these distributions of interdomain orientation have been represented implicitly in the form of finite, discrete, atomistic ensembles fit to RDC data [e.g. 27, 36]. In Qi et al. [45], we demonstrated benefits of representing these distributions explicitly using a family of low-dimensional continuous probability distributions fit to RDC data, termed continuous distributions of interdomain orientation (CDIOs). One limitation of RDC data is that its information is strictly orientational in nature and does not contain information on the relative probability of different interdomain translations. This limitation is partially addressed in the present study through the use of SAXS, a complementary technique that studies proteins in physiologically-relevant solution.

SAXS is a technique useful for studying flexible protein systems under relevant biological conditions [28, 11]. SAXS data contains information on the distribution of electron pair distances [44]. As the data is a sum from all molecules in solution, SAXS captures information from all conformations occurring in solution, containing information on both orientation and domain-domain distance, albeit at lower resolution then NMR RDC. For example, we showed in a CASP study that over half of the proteins adopted a different conformation in solution than in the crystal lattice [31], as was similarly observed in an NMR study [2]. This lack of agreement between solution and crystal data confounded earlier CASP attempts to use SAXS to improve prediction accuracy using crystallographic models as the ground truth. SAXS measures the intensity of X-rays scattered by the protein solution as a function of the scattering vector magnitude q, which is related to scattering angle θ and wavelength λ by $q = 4\pi \sin(\theta)/\lambda$. The variable q represents a point in reciprocal space, which is the Fourier transform of real space. By transforming reciprocal space information to real space by Fourier transformation, we can obtain a more intuitive distribution of pairwise distance for protein electrons. Importantly, SAXS reciprocal space curves can be calculated from atomic models and compared to experimental SAXS data, identifying atomic models or ensembles of atomic models that are consistent with protein solution data and can detect even small domain-domain shifts that alter multiple distances [16, 53, 15, 46]. Thus, SAXS is a useful tool in characterizing ensembles of proteins in distinct conformations.

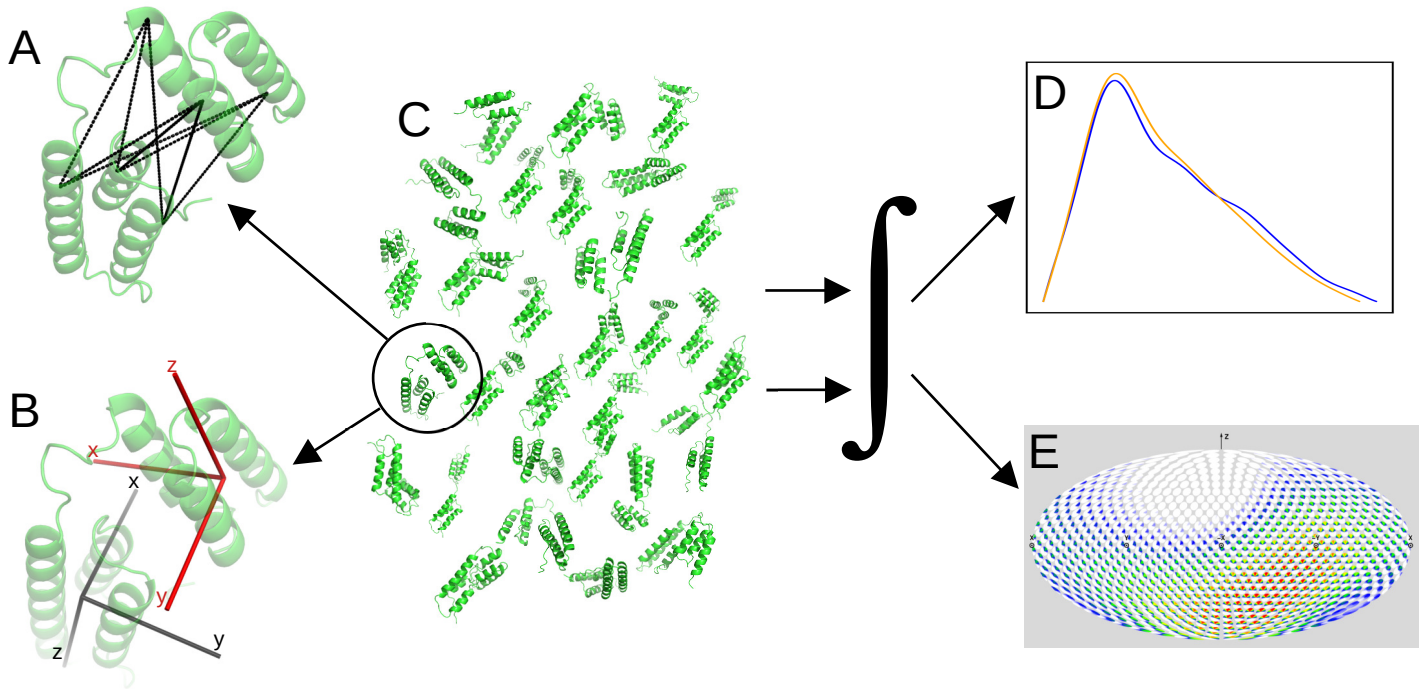


Figure 1: Cartoon of relationship between SAXS and NMR RDC information. For a given conformation in an ensemble (C), SAXS carries information about all pairwise interatomic distances (A), and RDCs carry information related to interdomain orientation (B). This information is integrated over the ensemble and, once analyzed, can yield models of the distributions of interatomic distances (i.e., SAXS $P(r)$ curves, D) and interdomain orientation (i.e., CDIOs, E) for the entire ensemble. (Panels D and E are scaled-down images of Figure 4B and Supplemental Information Figure S2A).

RDC and SAXS information can be viewed as complementary (Figure 1). RDC information strictly relates to orientation in a global frame, whereas SAXS information relates to interatomic distances. SAXS also carries information about the relative orientation of the domains of a D-L-D system, because a change in this relative orientation necessarily changes the distribution of interatomic distances between the domains. However, the information is not residue specific and is convolved with the domain distances. Therefore, combined RDC and SAXS data are an attractive choice for a study whose goal is to assess the accuracy of prediction of interdomain pose distribution, where pose consists of interdomain translation and orientation.

2 Materials and Methods

2.1 NMR Computational Methods

We used NMR RDC experimental data from our previous paper [45], which are deposited at BMRB (accession number 53002) [59]. The amino acid sequence of the wild-type ZLBT-C protein is MVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDYIDTNNDGAYEGDELQSANLLAEAKKLNDAQAPKADNKFNKEQQNAFYEILHLPNLTEEQRNGFIQSLKDDPSVSKEILAEAKKLNDAQAPK. The amino acid sequence for the Gly6 ZLBT-C variant is MVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDYIDTNNDGAYEGDELQSANLLAEAKKLNDAQAPGGGGGGNKEQQNAFYEILHLPNLTEEQRNGFIQSLKDDPSVSKEILAEAKKLNDAQAPK. NMR data collection for Gly6 is ongoing.

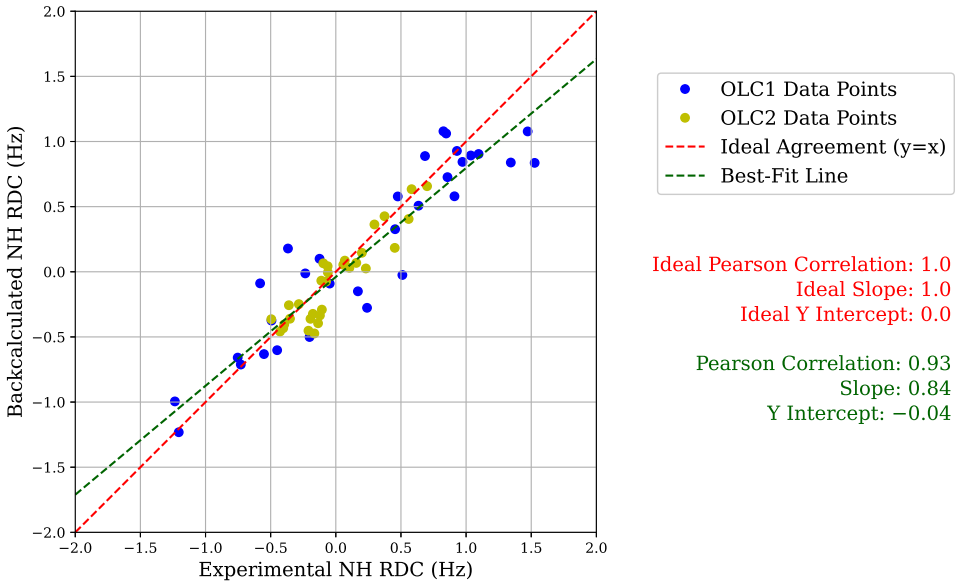
Much of our NMR RDC data analysis was identical to what we described in Qi et al. [45]. Experimental Saupe tensors for the ZLBT and C domains were fit to the RDCs measured in four alignment conditions for each domain, using singular value decomposition (SVD) [37, 45, 3]. Also as in Qi et al. [45], we used SVD to extract two effectively independent alignment conditions from experimental data to obtain orthogonal linear combinations (OLCs) of measured RDCs and corresponding Saupe tensors [48, 45].

A branch-and-bound search algorithm was used to find CDIOs in the form of Bingham distributions that are close fits to these OLC RDCs, as described in Qi et al. [45]. Two distinct solutions were found with similar goodness-of-fit; these are qualitatively similar to the two reported in Qi et al. [45], and are named Solution 1 and Solution 2 accordingly.

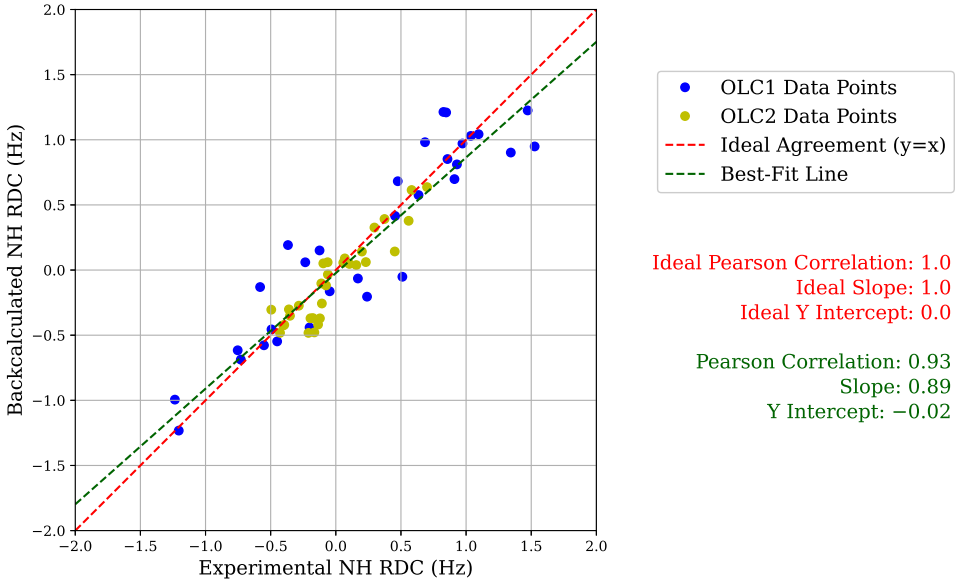
Following Qi et al. [45, SI, Eq. (12)], the 4×4 parameter matrix X for a Bingham distribution can be decomposed as

$$X = M^{\intercal} \Lambda M,$$

where M is a matrix representing a rotation in SO(4), the group of rotations in four-dimensional Euclidean space, and Λ is a diagonal matrix with constant trace specifying the variances along the four principle directions [45, SI].



(A)



(B)

Figure 2: Comparing CDIO solutions to NMR RDC data. Linear regressions of OLC RDCs back-calculated from (A) Solution 1 and (B) Solution 2, which are our CDIOs optimally fit (using the algorithm in Qi et al. [45]) to our experimental NMR data, showing close fits to experimental RDCs.

M has 6 degrees of freedom and Λ has 3, giving X a total of 9 degrees of freedom [45, SI]. For Solution 1, we have

$$M = \begin{bmatrix} 0.35 & 0.34 & -0.85 & 0.20 \\ -0.91 & 0.29 & -0.22 & 0.17 \\ 0.05 & 0.68 & 0.12 & -0.72 \\ 0.20 & 0.59 & 0.46 & 0.64 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 2.37 & 0 & 0 & 0 \\ 0 & -2.47 & 0 & 0 \\ 0 & 0 & -1.23 & 0 \\ 0 & 0 & 0 & 1.33 \end{bmatrix}.$$

For Solution 2, we have

$$M = \begin{bmatrix} -0.08 & 0.66 & 0.05 & -0.74 \\ 0.06 & 0.46 & 0.76 & 0.45 \\ 0.28 & 0.57 & -0.63 & 0.44 \\ -0.95 & 0.14 & -0.15 & 0.22 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 2.67 & 0. & 0. & 0. \\ 0. & -3.51 & 0. & 0. \\ 0. & 0. & -0.43 & 0. \\ 0. & 0. & 0. & 1.27 \end{bmatrix}.$$

OLC RDC data back-calculated from these two solutions are a good fit for the experimental data, as shown in Figure 2. Supplemental Information (SI) Figure S2 shows these two solutions in the form of Disk-on-Sphere (DoS) visualizations, which are described in Qi et al. [45]. (The M matrices above are given in the coordinate system that corresponds to these DoS visualizations.)

If two distributions ϕ_1 and ϕ_2 each fit experimental RDCs, then any convex linear combination of them $\alpha\phi_1 + (1-\alpha)\phi_2$ is also a solution [45]. Therefore we also consider many such solutions for values of α between zero and one, in increments of 0.01. These solutions are not Bingham distributions, but mixtures of two Bingham distributions.

2.2 Assessment methods

A document provided to CASP predictors (see Specifications in SI) specified that backbone atoms of helices 2 and 3 of each domain be within 0.5 Å of reference models. In practice, we computed the atomic root mean square deviation (RMSD) of these backbone atoms for each domain, and then found the mean between these domains. Because the resulting distance was greater than 0.5 Å for 20 out of 35 predictors, we used a cutoff of 0.75 Å instead of 0.5 for each target. Because the goal of these CASP targets was purely to predict linker conformations, this atomic RMSD cutoff was used stricly as a filter to ensure meaningful analysis; it plays no other role in assessing the success of predictions.

2.2.1 SAXS assessment

To evaluate the submitted ensembles, we use fast open-source X-ray scattering (FoXS) to convert the 3D protein structure to a theoretical scattering profile, which requires two parameters for each ensemble, c_1 and c_2, to estimate

the possible water layer surrounding the protein [49, 50]. The water layer should be consistent for a given protein sequence and ensemble. Thus, to estimate the proper water layer parameters for the conversion, we conducted the BilboMD simulation with integrated SAXS fitting on reconstructed atomic model based on WT PDBs for both targets, and used the output parameters to preprocess all submitted ensembles [43]. We further calculated the weighted average based on the submitted population profile to obtain the predicted SAXS profile of each prediction group. Chi-squared (χ^2) value is calculated between experimental and predicted data to evaluate the accuracy of the prediction, as has been described previously [49].

2.2.2 Assessing back-calculated NMR-RDC values

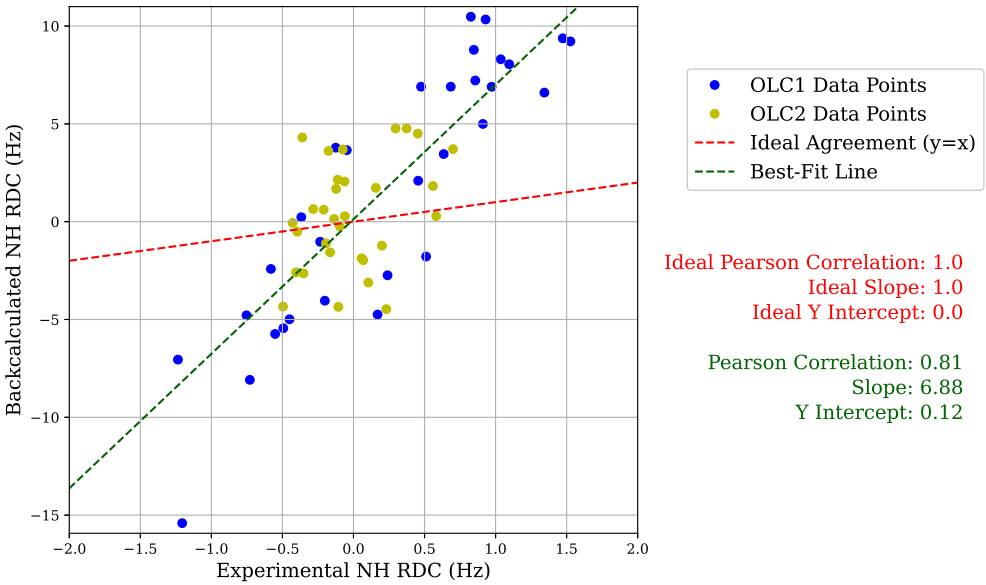
Our NMR-based assessments fall into two categories: those directly comparing against experimental measurements, and those involving ground-truth CDIO models. We address each in turn.

OLC RDCs (Section 2.1) were compared to corresponding RDCs back-calculated from each predictor ensemble. To back-calculate RDCs for a given predictor ensemble, each structure in the ensemble was considered in turn. First, the orientation of each predicted domain was defined by attaching a coordinate frame to it as described in Qi et al. [45] and the Algorithm document in the SI. Then, the two reference domains (for this purpose, PDB IDs 1Q2N [64] and 4NPE [19]) were reoriented such that the relative orientation between them matched the relative orientation between the two domains of the predicted structure. That is, if the frames attached to the two domains in the predicted structure are represented as column vectors in matrices F_{PC} and F_{PZ} and are related by $F_{PC} = RF_{PZ}$ for some rotation matrix R, then the two reference structures were reoriented such that $F_{RC} = RF_{RZ}$, where F_{RC} and F_{RZ} are, similarly, the domain-attached coordinate frames for the two reference structures. Once the reference structures were oriented in this manner, RDCs were back-calculated based on

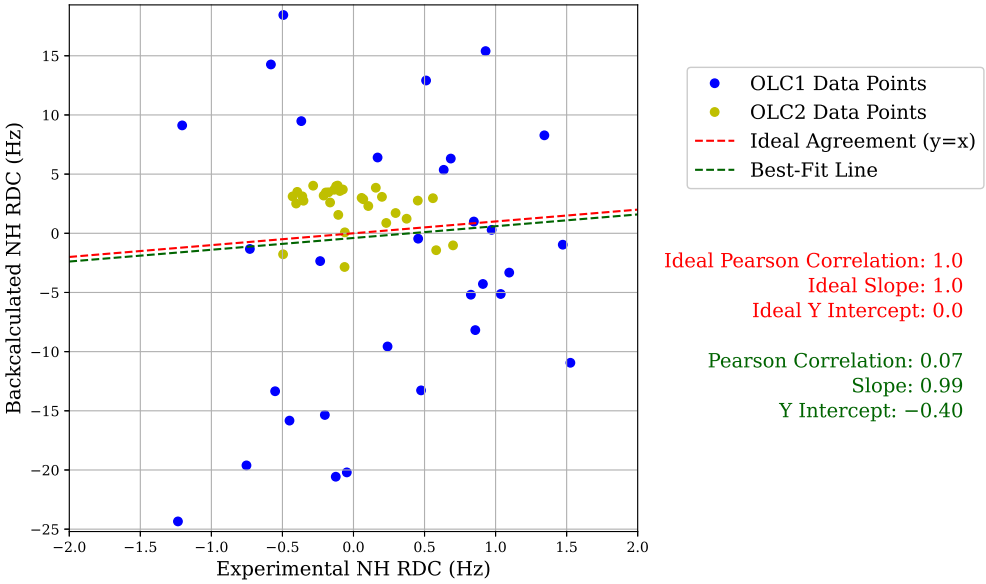
$$D = \frac{k}{2} \sum_{R \in \Xi \subset \mathrm{SO}(3)} \mathbf{v}^{\intercal} R^{\intercal} \mathbf{S} R \mathbf{v}\, p(R), \tag{1}$$

where k is the dipolar coupling constant (SI Section S2), Ξ is a finite sampling of SO(3), $p : \Xi \to \mathbb{R}$ is a discrete probability density function (PDF), $\mathbf{v}$ is the unit bond vector, R is the rotation matrix defined above, and $\mathbf{S}$ is each of the two OLC Saupe tensors. $p(R)$ is the population weight provided by predictors. When predictors did not provide population weights we assumed them to be uniform. This finite sum is a discrete approximation to the integral in SI Section S2 (SI Equation S7), in which the continuous PDF is replaced by population weights on a discrete set of orientations.

After back-calculating RDCs from a given predictor ensemble, we performed linear regression predicting back-calculated RDCs from experimental RDCs. If an ensemble were to perfectly reproduce the experimental data, then the slope of the best-fit line would be 1, the y-intercept would be 0, and the Pearson correlation would be 1.



(A) Predictor 331



(B) Predictor 015

Figure 3: Examples of comparing experimental RDCs to RDCs back-calculated from ensemble predictions. (A) For ensemble 331, back-calculated RDCs plotted against experimental RDCs, along with the best-fit linear regression for these back-calculated RDCs as well as for the ideal case when the experimental and predicted RDCs are equal. (B) Same as above, for ensemble 015.

We considered each of these quantities separately, because each represents a different aspect in which back-calculated RDCs can be similar or dissimilar from experimental. For example, the back-calculated RDCs for predictor 331 (MULTICOM_AI) had the strongest Pearson correlation with the experimental RDCs of any predictor ($R = 0.81$); however, the slope of the best-fit line was 6.88, much higher than the slope of 1 that would result from a perfect match (Figure 3A). By contrast, the corresponding best-fit line for predictor 15 (PEZYFoldings) had a slope of 0.99, the closest to 1.0 of any predictor. However, the Pearson correlation for their back-calculated RDCs was close to 0 ($R = 0.07$), indicating the near-unity slope of the best-fit line are not meaningful in the sense of signifying back-calculated RDCs that fit the experimental RDCs (Figure 3B).

2.2.3 Comparing predictions to CDIOs

Although predictors had the option to submit parameterized continuous distributions, all chose to submit discrete ensembles of structural predictions. To compare predicted ensembles to ground-truth CDIOs, it is necessary to convert these discrete ensembles to continuous probability density functions. This goal is achieved through kernel density estimation, taking a weighted sum of kernel distributions, one for each structure in a discrete ensemble. A narrow kernel function was chosen as described in SI Section S3.

To visualize these kernelized predictor ensembles in comparison with our solution CDIOs, we used the disk-on-sphere (DoS) approach we introduced in Qi et al. [45]. Instead of using a perspective projection, which requires two views per distribution (for front and rear views) and makes it hard to see densities on the edges of the projection, we used the Mollweide projection from cartography [52], originally introduced by Karl B. Mollweide in 1805. The correspondence between these projections is illustrated in SI Figure S2. This projection shows the entire sphere in one shape, and is also equal-area, simplifying the visual interpretation of different amounts of color in different parts of the sphere.

Each kernelized predictor distribution was visualized in comparison to one of the α values considered as convex combinations of Solution 1 and Solution 2 (Section 2.1, SI Figure S2). For visualization purposes, α was chosen such that, if the CDIO corresponding to that α were correct, the free energy difference between that solution and the given kernelized predictor distribution would be minimized. Free energy difference calculations are described in Qi et al. [45] and in SI Section S4. The distribution of these best-fit α values is shown in SI Figure S1.

A central concern of our work on the ZLBT-C system is to characterize how flexible the linker is, that is, in terms of relative interdomain orientation, how close to or far from isotropic its true CDIO is. In keeping with this emphasis, we assessed the accuracy of each predicted ensemble in this regard by defining "anisotropicity" and comparing the anistropicity of each predictor ensemble with the range of anisotropicities associated with our solution CDIOs for α values between 0 and 1. We define anisotropicity to be thermodynamically interpretable: We calculated the free energy difference (SI Section S4) between a given orientational distribution and the uni-

form distribution, under the provisional assumption for this purpose that the uniform distribution reflects the comparison energy field.

As our α parameter ranges from zero to one, this anisotropicity ranges from 0.09 to 0.42 kcal/mol. If a kernelized predictor distribution was within this range, this predictor was treated as correctly predicting isotropicity. Otherwise, each predictor was assessed based on the absolute value of the difference between their predicted anisotropicity and the nearest endpoint of this range.

2.2.4 Combining NMR-based assessments

Our assessments resulted in four values relating to NMR RDC data. Three of these relate to linear regressions of back-calculated on experimental RDCs: Pearson correlation coefficient, slope of best-fit line, and y-intercept of best-fit line. The fourth is our anisotropicity quantity, which relates to a free energy difference between a kernelized predicted ensemble and a uniform distribution. These four quantities must be combined for purposes of assessment. We chose to first scale each quantity according to the range of possible predicted ensembles, and to then combine them using Chebyshev distance [20]. Chebyshev distance treats the distance between two vectors as simply the maximum of the component-wise distances, so in this case, it corresponds to assessing each predicted ensemble based on its worst performance among the four quantities summarized above. This approach has two advantages: First, after scaling according to the worst that these quantities can be, comparing one quantity to another may be more meaningful. Second, because the final Chebyshev distance directly reflects only one of the underlying quantities, it does not implicitly assume that the quantities are independent. [1]

This approach requires computing the worst that a hypothetical predictor could perform with respect to each of these quantities. Determining the worst possible Pearson correlation is not empirically straightforward, so we assume that a perfect anticorrelation (i.e., a Pearson correlation of -1) is possible. By contrast, the most extreme possible slopes and y-intercepts are straightforward to determine empirically. These occur when the magnitude of back-calculated RDCs is maximized, which in turn occurs when an ensemble consists of a single structure, resulting in no averaging of back-calculated RDCs. We back-calculated RDCs for single-structure ensembles whose interdomain orientations were each of 294,912 grid points on SO(3) [62]. Performing linear regression on each back-calculated RDC set against the experimental RDCs, the slope furthest from 1.0 was -14.53, and the y-intercept furthest from zero was 8.80.

[1] They are not independent; for example, as Pearson correlation gets close to zero, slope becomes less meaningful.

3 Results

3.1 CASP experiment

The goal for these two CASP targets was to evaluate the prediction of the orientation and translation between the two rigid domains generated by different computational methods, as driven by the structural distribution of the six-residue flexible linker domain. The structures of the globular domains are deposited in the PDB as PDB IDs 2LR2 [6] and 4NPD [19]; predictors were given this information, in keeping with the CASP challenge to predict conformational diversity in the flexible linker, not in the rigid domains. Predictions were made using a variety of methods, mostly involving either molecular dynamics or deep learning (Table 1).

Predictors were invited to submit predictions either in the form of discrete ensembles of PDB-format structures with accompanying population weights or in the form of Bingham orientational distributions [8, 35]. All predictors chose to submit discrete ensembles, suggesting this is more familiar for prediction algorithms. The number of discrete structures per ensemble ranged from 3 to 10,000.

Predictors were instructed (see Specifications in SI) to ensure that atomic backbone RMSDs should match the reference structures within 0.5 Å, because deviations would call into question the validity of back-calculated SAXS and RDC data, as discussed in Sections 2.2.1 and 2.2.2. However, 20 out of the 35 predicted ensembles had RMSDs above this 0.5 Å, so we filtered according to a cutoff of 0.75 Å, which reduced the number of ensembles excluded from consideration to 8, for both T1200 and T1300 targets. Excluded ensembles were numbers 2, 3, 4, 5, 84, 88, 100, and 139, with average RMSDs ranging from 0.96 to 1.75 Å (SI Table S1).

3.2 SAXS

The Size-Exclusion-Coupled (SEC)-SAXS data were collected on the T1200 and T1300 systems at the SIBYLS beamline 12.3.1 at the Advanced Light Source Synchrotron [47] (details to be published in an experimental paper). The columns were equilibrated in the same buffer used for the NMR samples except for the addition of glycerol: 25 mM 3-(N-morpholino)propanesulfonic acid (MOPS) pH 7.2, 100 mM KCl, and 1% glycerol. The proteins were stoichiometrically monodisperse, based on co-collected multi-angle light scattering (MALS), SAXS-calculated molecular weight, and linear Guinier curves.

The T1200 (WT) and T1300 (Gly6) targets were similar to each other in terms of SAXS results, with the SAXS data in reciprocal and real space indicating that T1200 was more globular and T1300 more extended (Figure 4). Tellingly, the real space analysis that describes the pairwise distance distribution of protein residues revealed an important difference: the $P(r)$ curve of T1200 had visible shoulders after the main peak, which were not as prominent in the curve for the T1300 sample (Figure 4B, arrows). While the main peak represents the intradomain distances, the peaks at a longer distance represent interdomain distances. The visible shoulders in

Pred. Name	Pred. ID	# of Models (T1200)	# of Models (T1300)	Category	Ensemble Diversity Methods
Cool-PSP	014	6	6	DL (MSA)	AlphaFold 2-based model with custom MSA
PEZYFoldings	015	4	8	DL (MSA)	AlphaFold 2 with PLM-based MSA
Zheng-Server	019	3	5	Other	D-I-TASSER (REMC-based)
MULTICOM	051	1000	1000	DL (MSA)	AlphaFold 2 & 3 with various MSA
JFK-THG-IDPCONFGEN	097	1000	1000	Other	IDPConformerGenerator
MIEnsembles-Server Zheng-Multimer Zheng	110 147 462	27	32	DL (MSA)	DMFold
Lindorff-LarsenCLVDS	135	1000	1000	MD	CALVADOS
Lindorff-LarsenM3PPS	136	1000	1000	MD	Martini 3
Lindorff-LarsenM3PWS	137	1000	1000	MD	Martini 3
GuijunLab-Complex	148	162	173	DL (MSA)	AlphaFold-Multimer and DeepAssembly with various MSA
McGuffin	164	626	366	MD	ReFOLD4
OpenComplex	167	10000	10000	DL	OpenComplex-2
aicb	177	71	89	DL	OpenFold
CSSB_FAKER CSSB_experimental CSSB_Human	221 286 419	1000	1200	DL (MSA)	Faker and UltimateMSA
elofsson	241	500	500	DL	AlphaFold 3
UNRES	261	100	100	MD	UNRES (MREMD-based)
GuijunLab-Human	264	30	60	DL	DeepAssembly
KiharaLab	294	1000	1000	MD	Desmond
AF3-server	304	100	100	DL	AlphaFold 3
MoMAteam1	308	918	753	Other	Statistical methods, Amber
MULTICOM_LLM	319	1000	1000	DL (MSA)	AlphaFold 2 & 3 with various MSA
MULTICOM_AI	331	1000	1000	DL (MSA)	AlphaFold 2 & 3 with various MSA
MULTICOM_human	345	2000	2000	DL (MSA)	AlphaFold 2 & 3 with various MSA
pert-plddt	384	56	82	DL	OpenFold
MULTICOM_GATE	425	1000	1000	DL (MSA)	AlphaFold 2 & 3 with various MSA
OpenComplex_Server	450	10000	10000	DL	OpenComplex-2
Wallner	465	17	21	DL (MSA)	AFsample2

Table 1: Summary of predicted ensembles meeting RMSD cutoff, including the number of atomic structural models in each ensemble and key prediction methods. This table omits information on refinement methods applied to each predicted structure in an ensemble; instead, it focuses on a key question relating to the overall shape of an ensemble: What methods were most important in generating a pool of decoys ("Ensemble diversity methods")? For qualitative analysis purposes, we categorized predictions according to methods used to generate a pool of decoys: "DL" for deep learning with default MSA, "DL (MSA)" for deep learning with custom MSA, "MD" for molecular dynamics, and "Other" for other methods such as Monte Carlo.

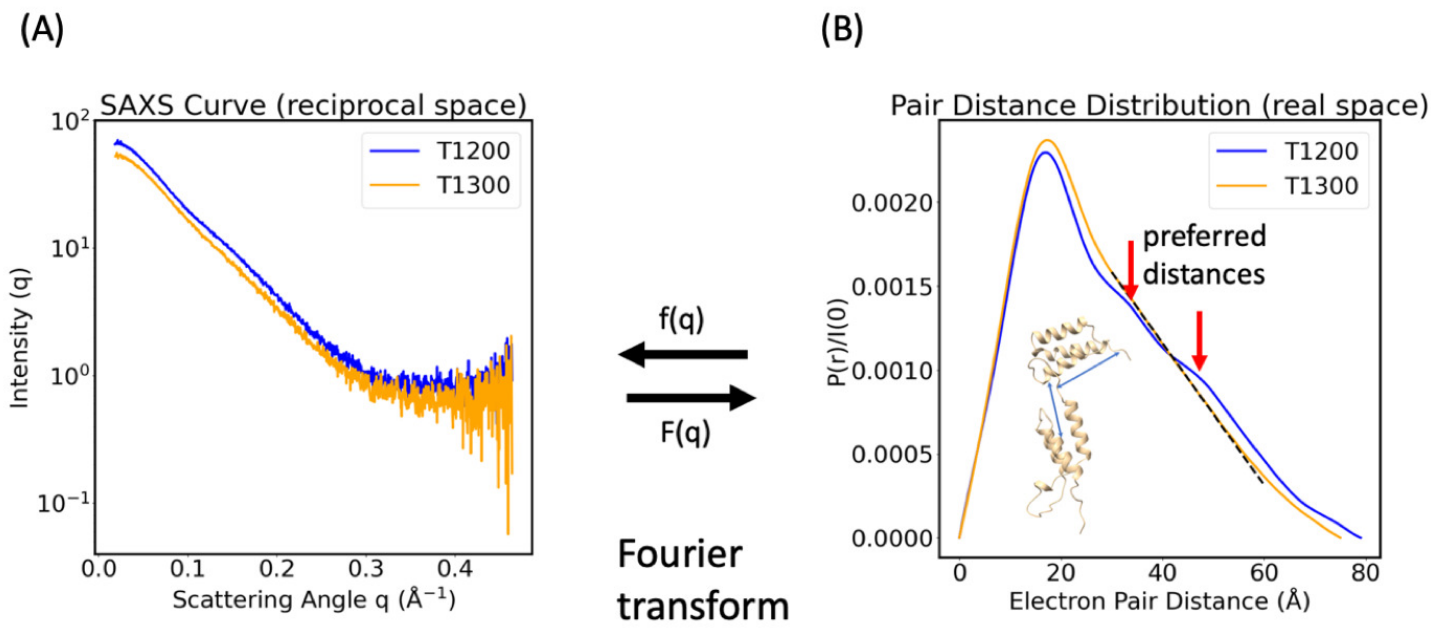


Figure 4: SAXS experimental data for targets T1200 and T1300. (A) The reciprocal space data for T1200 and T1300. (B) The real space data for T1200 and T1300, calculated from the reciprocal space data. The Y-axis represents the normalized probability of the pair distance distribution.

the experimental data for T1200 suggest a population with longer preferred interdomain distances while the lack of shoulders is consistent with the loss of preferred distances in T1300. This finding supports the model that the conserved WT linker in T1200 constrains the conformational space, while the Gly6 replacement in T1300 reduces that constraint. This difference was important for assessing the ensemble predictions of each predictor.

Based on these SAXS data, we conducted a two-step evaluation of the predictors. First, we evaluated the overall quality of the predictors' models using the χ^2 metric in reciprocal space and filtered out models that diverged from the experimental data. The χ^2 metric provides a rapid assessment of the overall structure. However, the χ^2 metric can obscure subtle yet significant differences in the scattering. So, to detect whether those models with reasonable χ^2 values actually captured the structural preferences of the protein ensemble, we compared the reciprocal and real space curves of the top predictors with corresponding experimental data, as a function of the scattering vector magnitude q.

To assess the overall quality of predictions, we calculated SAXS curves from the ensembles of atomic models provided by the predictors using FoXS and compared them to the experimental data using a standard workflow (Figure 5). The commonly-used χ^2 metric was employed as a measure of accuracy; note that some of our predictors may have calibrated/trained their models with other SAXS data using this metric. χ^2 evaluates the difference between model and experimental curves and is thus dominated by low q and interdomain distances [30]. Typically, a χ^2 value between 1 and 2 is considered reasonable, with the caveat that χ^2 is a function of the experimental error in the denominator. Models or model ensembles for the same protein based on the same experiment and

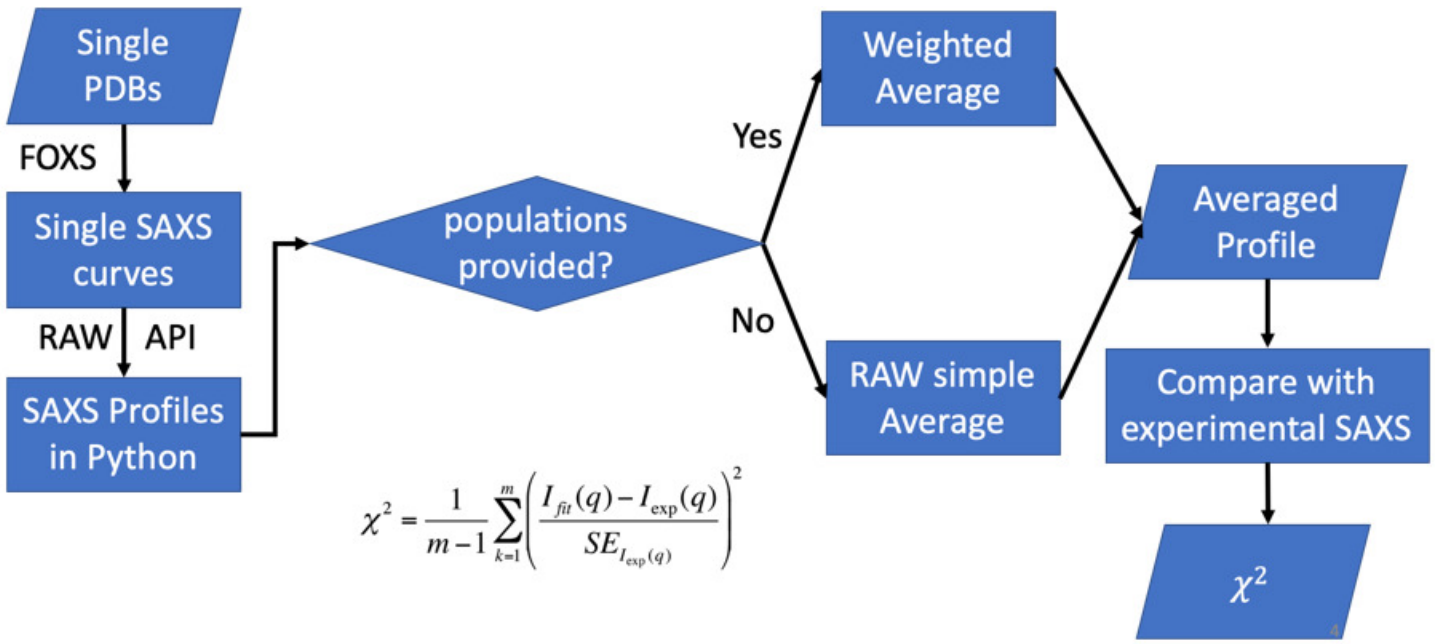


Figure 5: SAXS evaluation workflow. The single PDBs of each group were converted to single SAXS curves with FoXS for preprocessing (Left). The converted SAXS curves were averaged based on the population file provided with the PDBs (Right). The same evaluation workflow is applied on both targets. In the χ^2 difference comparison between experimental data and reciprocal space curves calculated from predicted ensemble atomic models, m is the number of experimental data points, $I_{\text{fit}}(q)$ is the predicted intensity of each group's ensembles models, $I_{\exp}(q)$ is the experimental intensity, $SE_{I_{\exp}(q)}$ is the standard error of each experimental data point, q is the scattering vector magnitude.

analyzed using the same experimental error are comparable; however, χ^2 values cannot be compared between different experiments. In this evaluation, 16 of 27 predictors had at least one prediction with a χ^2 value between 3 and 10 (Figure 6, panels A, B, and C). The largest χ^2 value exceeded 200 for the worst single PDB among all the conformations submitted, which is surprising given that the domain structures are known and only the structural conformations of 6 residues are unknown. As a reference, the ranking in the plot is sorted by the root mean square average of the χ^2 values for T1200 and T1300.

To see if any predictor model can reveal the difference between the two targets, T1200 and T1300, we plotted the χ^2 for the two targets for groups that have χ^2 lower than 10 for at least one target (Figure 6D). The data points for groups that perform similarly on both targets would fall near the diagonal. About half of the predictors performed better on the WT compared to the Gly6 mutant. This suggests that prediction methods may be more accurate in modeling endogenous linkers similar to those found in nature as opposed to artificial constructs such as Gly6 where there are no evolutionary relationships.

Because χ^2 is a single metric for a three-dimensional comparison, we looked at the reciprocal and real space curves for predicted vs. experimental data for the two best-scoring predictors 135 (Lindorff-LarsenCLVDS) and 308 (MOMAteam1) (Figure 7). More informative than visual inspection of the overlain curves, in terms of accuracy, is the clear bias in the residual difference between predicted and experimental data in reciprocal space ($\Delta I(q)/\sigma(q)$)

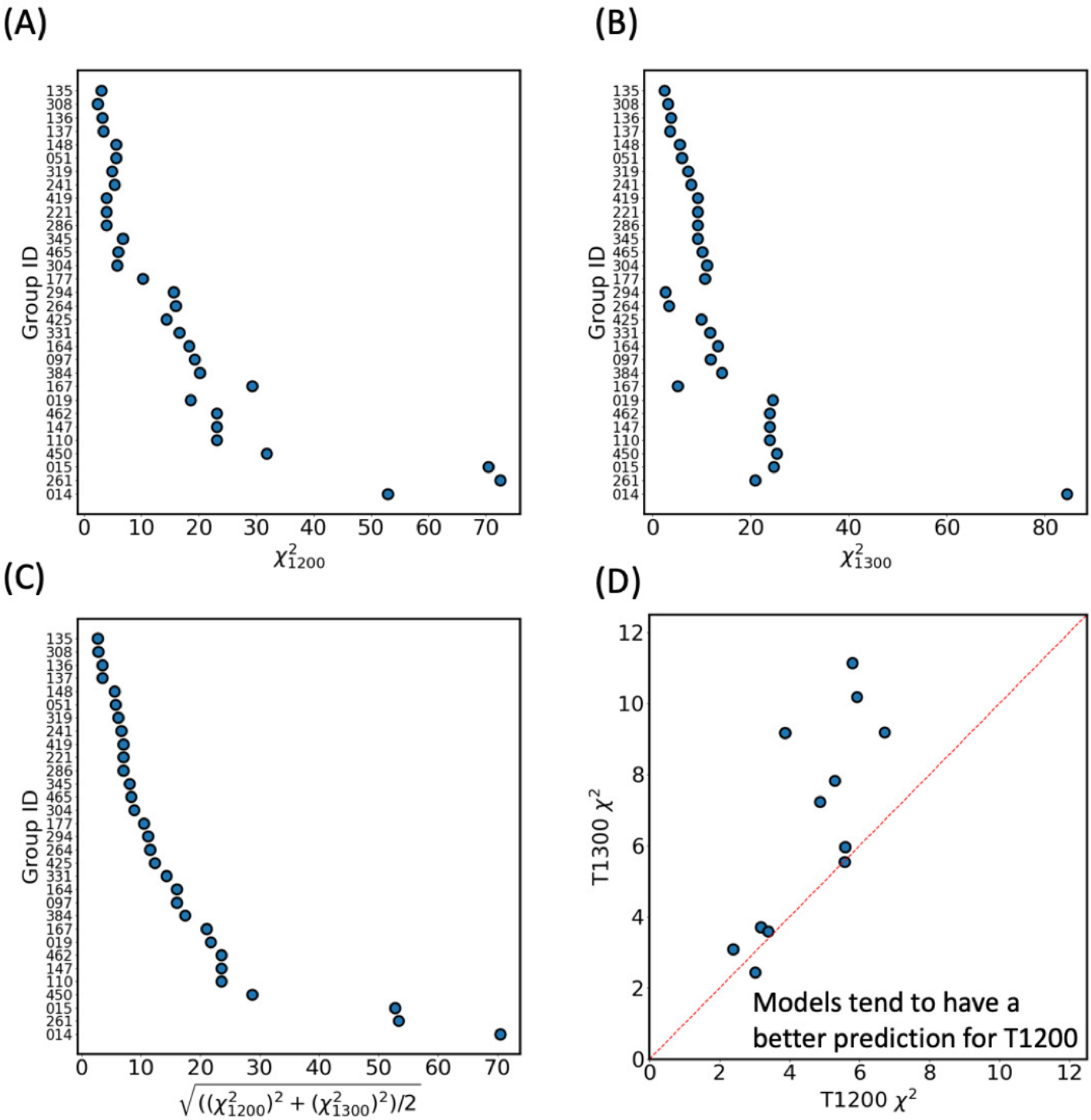


Figure 6: Comparison of ensemble-based predictions of SAXS scattering curves with ground truth using the χ^2 metric (Section 2.2.2). (A) Average χ^2 for T1200 (WT). (B) Average χ^2 for T1300 (Gly6). (C) Root mean square average of the χ^2 values for T1200 and T1300. (D) Comparison of predictor performance on WT vs. mutants for the predictions whose T1300 χ^2 values were less than 12. Each point represents a predictor group. The y-axis shows the χ^2 value for T1300, the Gly6 mutant. The x-axis shows the χ^2 value for T1200, the WT protein. The dashed red line represents the diagonal, indicating the same performance on both targets.

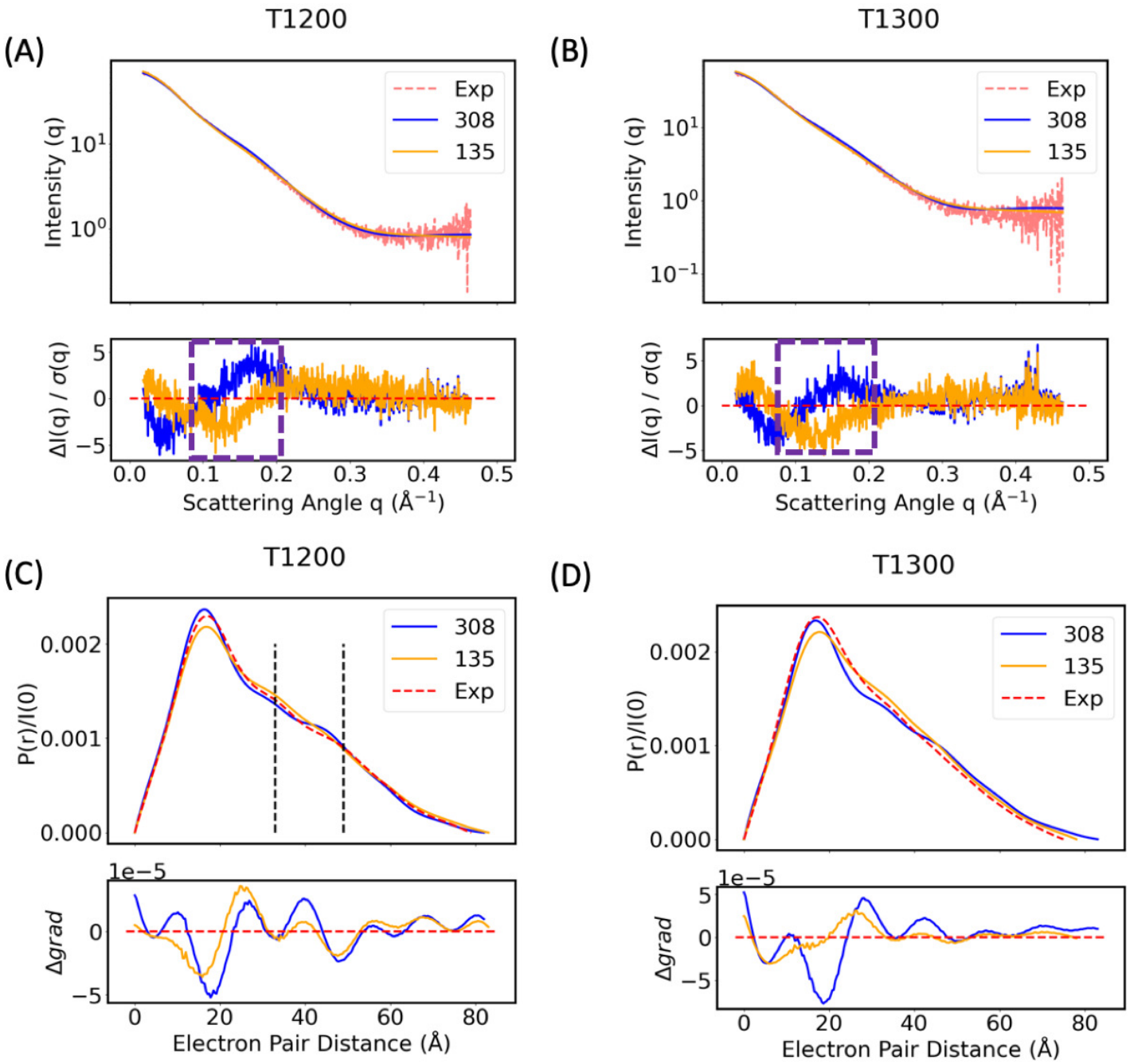


Figure 7: Comparison of two predicted ensembles (135 and 308) with experimental SAXS data for T1200 and T1300 showing biased differences. SAXS curves were calculated from predictor's ensemble of atomic models and overlaid with experimental data in reciprocal (A,B) and real (C,D) space for T1200 (A,C) and T1300 (B,D). Residual graphs between predictions and experimental are shown below each SAXS plot.

Quantity	Minimum among predictors	Maximum among predictors	Ideal value	Worst possible value
Slope for RDC regression	−3.0	7.6	1.0	−14.5
Y-intercept for RDC regression	−3.7	1.7	0.0	−8.8
Pearson correlation for RDCs	−0.76	0.81	1.0	−1.0
Anisotropicity (kcal/mol)	0.40	2.45	0.09 to 0.42	2.50

Table 2: Quantities used to compute Chebyshev distances. The first two columns are the minimum and maximum values of each quantity among all the predictions. The ideal value represents the quantity for perfect agreement with ground truth. Calculation of worst possible values is described in Section 2.2.4. Chebyshev distances were computed by scaling the prediction's values for each quantity by the range from ideal to worst possible. The overall distance is the highest of these four values. SI Table S3 lists each of these values and the overall distance for each prediction.

(Figure 7A,B). If the predictions were accurate, the residuals would fall with equal probability above and below the diagonal. For top scoring predictors 135 and 308, there are large residuals falling in the low-q region, which indicates an inaccurate prediction of the interdomain distance. Since the structures of the rigid domains were given, the mid- to high-q regions show relatively good predictions of the intradomain distance.

We further plotted the change of the derivative in the $P(r)$ curve to see if predictors captured the semi-stabilized domain-domain orientations indicated by the shoulders at 33 and 49 Å in the real space data (Figure 4B). These shoulders represent the preferred population, especially in the WT. For the WT predictions, we observed shoulders (Figure 7C), but they were at incorrect distances. The Gly6 predictions also had shoulders. However, shoulders were not present in the experimental data, indicating that the loss of preferred distance was not captured in these predictions (Figure 7D). No metrics that can quantitatively measure bias or preferred distances have been developed for SAXS analysis, hence our analysis here, based on SAXS data alone, can only qualitatively conclude that the predictors captured neither the preferred distance of the WT protein nor the loss of this preferred distance in the Gly6 mutant.

3.3 NMR

Predictions were compared to NMR RDC data both directly and through comparison with models as described in Materials and Methods. Briefly, a linear regression of back-calculated RDC values on experimental RDC values was computed for each predictor, and the slope and y-intercept of this regression were considered, along with the corresponding Pearson correlation. In addition, kernelized predicted ensembles were compared to uniform distributions in terms of free energy as a measure of anisotropicity. The ranges of these four quantities across predictors are shown in Table 2, along with ideal and worst possible values or range of values for that quantity, and Table S2 shows these four quantities for all predictors. The four quantities were normalized according to their worst possible values and then combined using Chebyshev distance, selecting for each predictor their worst

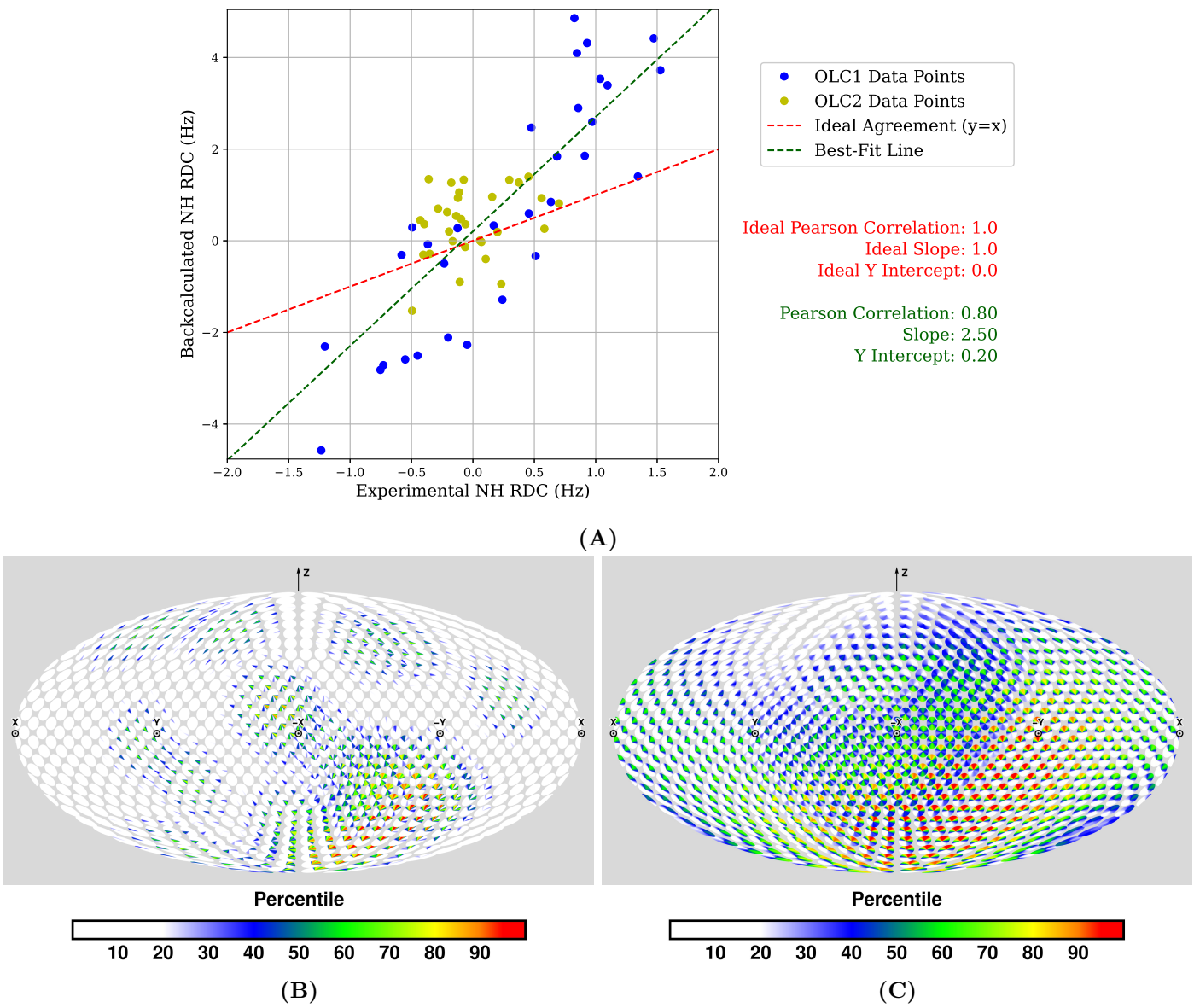


Figure 8: Representative data and assessment for predictor 264. (A) back-calculated RDCs plotted vs. experimental RDCs. *Green*: best-fit linear regression for back-calculated RDCs. *Red*: ideal regression line when the experimental and predicted RDCs are in perfect agreement (equal). (B) kernelized distribution of 264's interdomain orientations. (C) mixture of 2 ground truth solution CDIOs (SI Figure S2, [45]) corresponding to $\alpha = 0.73$, the value of α minimizing the free energy difference between the resulting α-mixture CDIO vs. the kernelized predictor distribution. (B,C) are disk-on-sphere (DoS) figures [45], in the Mollweide cartographic projection [52].

performing score among the four (Table S3).

Results of analysis for predictor 264 (GuijunLab-Human), which had the smallest Chebyshev distance from ideal, are shown in Figure 8, with similar figures for every predictor in SI Section S5. Figure 8A shows the set of back-calculated RDCs plotted against experimental RDCs and with information shown regarding the linear regression. Figure 8B shows this predictor's kernelized distribution in DoS form, along with the DoS for the solution CDIO corresponding to $\alpha = 0.73$ (Figure 8C), which is the value of α for which the corresponding solution CDIO had the smallest free energy difference with predictor 264's kernelized distribution. The DoS illustrates that the top two modes of ensemble 264 are each similar to the modes of Solution 1 and 2 (and therefore to the two modes of a mixed solution such as $\alpha = 0.73$). It is also clear from the DoS's that this prediction is considerably less isotropic than the corresponding $\alpha = 0.73$ solution CDIO, and it is anisotropicity that limits this predictor's Chebyshev distance score, which is 0.23.

Chebyshev distances range from 0.23 for predictor 264 (GuijunLab-Human) to 0.98 for predictor 014 (CoolPSP). This latter ensemble's Chebyshev distance is limited by its near-maximal anisotropicity; it consists of six structures, all similarly oriented.

3.4 Joint evaluation of NMR and SAXS

An accurate prediction would be consistent with both SAXS and NMR RDC experimental data. These two types of data provide complementary information: SAXS data contain residue-nonspecific distance and orientational information of the entire protein, while NMR RDC provides residue-specific orientational information between the two domains. Both SAXS and NMR RDC data reflect distance and/or orientational information integrated, both temporally and spatially, over the conformational variation within an ensemble. NMR RDC data has relatively little noise, whereas SAXS data has relatively high noise.

SI Figure S2 shows that the CDIOs best fit to NMR RDC data demonstrate preference for some orientations over others. Similarly, the WT $P(r)$ curve from SAXS in Figure 4B shows that there are populations in the WT ensemble demonstrating preferred interdomain distances; no such interdomain preferences were visually detectable in the $P(r)$ curve for the Gly6 mutant. These observations confirm our previous descriptions of the WT linker as neither rigid nor completely flexible, and are consistent with the hypothesis that the conserved sequence of the WT linker encodes constrained flexibility. Thus, an ideal prediction consistent with data from both experiments would not describe an isotropic ensemble with uniform orientational and distance distributions. Instead, it would describe a somewhat anisotropic ensemble exhibiting these structural preferences in terms of both distance and orientation.

Figure 9 shows Chebyshev distances reflecting NMR-based analysis plotted against χ^2 statistics reflecting SAXS analysis. We colored the dot for each prediction according to a broad categorization of the methods used

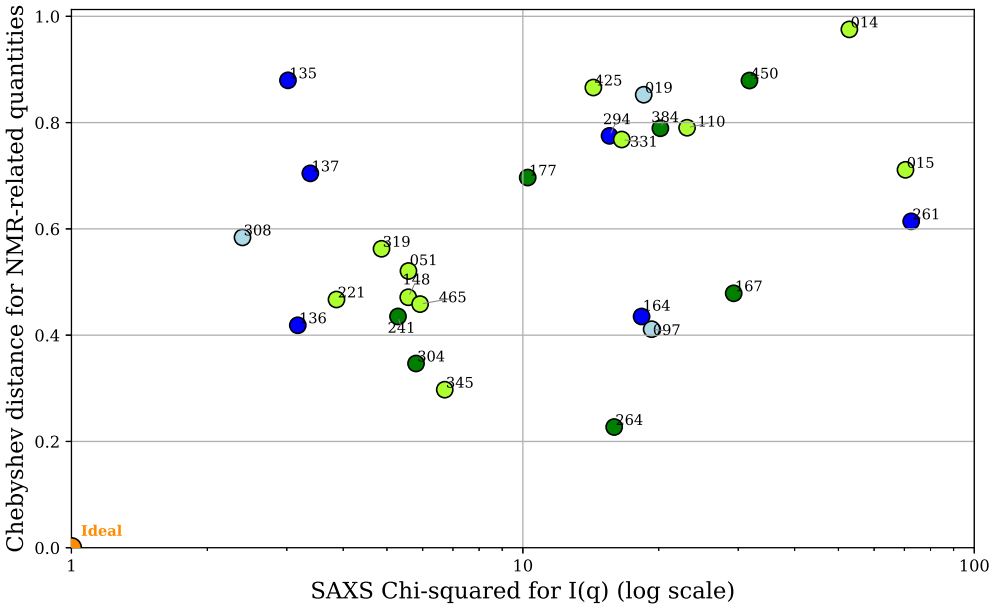


Figure 9: For each predictor, Chebyshev distances reflecting NMR-based analysis vs. χ^2 statistics reflecting SAXS analysis. An ideal prediction would lie at the lower-left corner of the plot, with a Chebyshev distance of zero and a χ^2 of one. χ^2 statistics are plotted on a log scale to reflect that values greater than 10 are considered a poor fit to data. Dot color for each predictor reflects a broad categorization of methods used to predict the diversity of conformations for each ensemble (*blue*: molecular dynamics; *green*: deep learning; *yellow-green*: deep learning with custom MSA; *light blue*: other) and orange indicates the ideal values for a perfect prediction.

to predict the diversity of linker conformations for each ensemble, as shown in Table 1. Where possible, we categorized these methods as being primarily based on molecular dynamics (blue), deep learning with default multi-sequence alignment (MSA) (green), and deep learning with other MSA methods (yellow-green). Light blue indicates a method that did not fall into these three categories; specifically, the D-I-TASSER workflow used by predictor 019 (Zheng-Server), which uses replica exchange Monte Carlo (REMC) as the primary method for generating diverse conformations [65], IDPConformerGenerator, used by predictor 097 (JFK-THG-IDPCONFGEN), which predicts ensembles of IDPs by sampling torsion angles found in the RCSB Protein Data Bank [54], and a workflow combining molecular dynamics with a statistical method used to predict conformations of IDPs [22].

For each predictor, panels summarizing the various analyses described in this section are shown in SI Section S5.

4 Discussion

4.1 Comparison with other CASP challenges

In a typical CASP assessment, predictors have submitted one or more structures, each of which was compared individually to atomistic models. Even ensemble targets have taken this approach, for example, by comparing each predicted structure one-on-one with each member of a ground-truth ensemble. Our assessment is the first to assess each predicted population-weighted ensemble as an interconverting temporal or molecular thermodynamic

ensemble. We assessed each predicted ensemble against both experimental SAXS and NMR RDC data as well as to ensemble models fit to experimental data.

In comparing predictions to NMR data, our assessment is somewhat similar to a previous CASP experiment that evaluated prediction models against NMR NOESY peak lists, backbone chemical shifts, and RDC data [29]. Unlike that experiment, which compared singleton atomistic structures to NMR observables, we treat the submitted set of population-weighted structures as a thermodynamic ensemble. Our assessment also differs by incorporating SAXS data as well as NMR.

Other past CASP challenges have used SAXS data [31, 42]. However, these were not blind assessments; rather, SAXS data were provided to predictors in advance, and comparisons were made largely on the basis of fit of predicted atomistic models to experimentally determined atomistic models from X-ray crystallography. These past CASP experiments demonstrated possible pitfalls of non-blind prediction: when initial structures with incorrect protein topology were adjusted to fit SAXS data, good fits could be achieved through distortion of the structure without correction of topological errors. By assessing data-blind predictions, we increased the likelihood that predictions that are good fits to data reflect more realistic models, as opposed to overfitting.

Our assessment's use of back-calculated data allowed for comparisons of entire ensembles vs. ensemble-averaged NMR RDC and SAXS experimental data. Such whole-ensemble comparisons are necessary for our goal of assessing predictions of continuously varying interdomain pose. This flexible interdomain structure is not naturally described by fixed, finite numbers of discrete conformations but rather by continuous probability distributions. Therefore, discrete, pairwise comparisons between atomic structural models are not adequate to assess these predictions.

4.2 Evaluation of predictions

For this assessment, a good prediction must be close in terms of back-calculated SAXS curves, back-calculated RDCs, and anisotropicity. None of the predicted ensembles were close fits in all these measures. This lack of clear success in prediction resembles the first years of CASP.

Ensemble 264 (GuijunLab-Human) is an example of a prediction that performed well by some measures but not others. The kernelized distribution based on this ensemble is shown in Figure 8 in comparison to the interpolation between Solutions 1 and 2 that is closest to it, in free energy terms ($\alpha = 0.73$). Although the general shape of the distributions in Figure 8B and Figure 8C look similar, the comparison makes clear that ensemble 264 is too narrow; that is, it is less isotropic than either Solution 1 or 2 or any interpolation between them. Some qualitative aspects of the prediction can be observed in the best-fit regression for this ensemble's back-calculated RDCs. In particular, the relatively good prediction of modes are reflected in a high Pearson correlation, and the prediction's overly-narrow spread is reflected in an overly steep slope. This steep slope derives from the overly large magnitudes of the back-calculated RDCs, which often results from a lack of averaging of RDCs corresponding

to different orientations (Equation 1).

Our results show that it remains challenging to predict, from sequence alone, ensembles of linker conformation (and, more generally, D-L-D protein structure) that are consistent with experimental data. This difficulty persisted even though predictors were given high resolution models of the two globular domains based on X-ray crystallography. The remaining challenge was thus to predict the conformational distribution of the 6 linker residues or, alternatively, the relative translation and orientation between the two relatively rigid domains. It is striking that a 6-residue structure prediction problem remains challenging for current methods. It is much easier to create ensembles consistent with data when RDC and/or SAXS data, at the level of ensembles or individual structures, are available prior to prediction for fitting and/or comparison.

Figure 9 summarizes our assessments, with a broad categorization of prediction methods indicated by different colors of dots. The small number of predictors in each category, combined with the diversity of methods within each category, preclude meaningful statistical analysis. If a clear pattern were qualitatively visible in the locations of differently-colored dots, we might be able to draw preliminary conclusions about strengths and weaknesses of different prediction methods, with respect to the relative orientation and translation of the two domains of the D-L-D system. No such clear patterns emerged in the data overall; however, we make some qualitative observations based on the top predictors with respect to the NMR RDC and SAXS axes of this figure.

Four of the nine non-deep-learning-based predictions were the best-performing predictions in terms of fitting SAXS data. These four predictions all used either energy functions tuned to SAXS data [57, 13] or methods known to produce predictions with good agreement to SAXS data [22]. In addition, these four predictions were also among the most isotropic of all ensembles considered, reflecting an enigmatic relationship between the anisotropicity of predictor ensembles and their closeness of fit to the SAXS data (SI Figure S3).

More broadly, neither molecular dynamics nor deep learning methods seemed to dominate the best predictions with respect to both SAXS and NMR data. The top four predictions with respect to SAXS data were based primarily or partly on molecular dynamics and were not among the best predictions with respect to NMR data. Conversely, the top three predictions with respect to NMR data were based on deep learning and were not among the best predictions with respect to SAXS data.

We have deliberately eschewed any single combined scoring for both SAXS χ^2 statistics with NMR-based Chebyshev distances, because these two quantities reflect performance with respect to two completely different experimental modalities. Whenever entities are scored using two incommensurable values, it is possible to define a Pareto frontier of those entities with the property that every other entity is worse than it on at least one of the two scores. Figure 9 makes such a Pareto frontier clearly visible; it contains predictions 308, 136, 304, 345, and 264.

Before we decided to use Chebyshev distance to combine three different quantities associated with best fits

between back-calculated RDCs and experimental RDCs, we considered using more conventional measures of RDC closeness such as RDC RMSD [58, 21], Q1 [55, 18], and Q2 [55, 17] (all of which are, in the context of fixed experimental data, constant multiples of each other). If predictions had been closer fits to experimental data, these quantities may have been more useful measures of that fit. However, no predictors' back-calculated RDC sets were qualitatively close to the OLC RDCs. Those predictors whose back-calculated RDC sets were of similar magnitude to the OLC RDCs did not correlate well with the latter, and vice versa. As a result, variation in RDC RMSD, Q1, or Q2 was dominated by the magnitude of back-calculated RDCs rather than more meaningful closeness between the back-calculated and OLC RDC sets. Because our experimental RDCs are relatively small, the dominance of RDC magnitude in these measures had the practical effect of heavy bias toward more isotropic predicted ensembles, in which averaging resulted in back-calculated RDCs that were smaller in magnitude but generally uncorrelated with experimental RDCs.

4.3 Value of continuous distributions

Ensembles of discrete structures are commonly used to express various kinds of flexibility in protein structure, including that of flexible linkers. Such ensembles have the advantage that they are compatible with various computational tools designed to work with atomic models such as those deposited in the PDB. However, ensemble models fail to express the fundamentally continuous nature of protein flexibility or explicitly give the range of possible motion. In addition, they can express flexibility with many redundant degrees of freedom, specifying the positions of every atom when the relevant information may be expressible, for example, by a much smaller number of torsion angles. Some ensembles are designed to fit NMR and/or SAXS data with a minimal number of conformational models; such minimal ensembles may be easier for a human to understand [43], but sacrifice the ability to express flexibility or entropy in detail.

Continuous probability distributions offer descriptions of protein flexibility that can be at once precise, compact, and relatively simple. Such distributions are consistent with the assumption that flexibility is continuous in reality. In addition, choosing a parameterized family of distributions (such as Binghams over SO(3)) to represent potential solutions naturally regularizes the problem, potentially making proposed distributions more parsimonious by allowing for fewer degrees of freedom compared to discrete ensembles. Finally, as our study also demonstrates, there are practical analytical advantages to considering protein flexibility in terms of continuous distributions. For example, modes are well-defined on continuous distributions, and two continuous distributions can be compared in terms of their free energy difference via integration, given a known energy field (or the assumption that one of the distributions reflects this field). There remains the issue of comparing experimental SAXS data to probability distributions that include translational and rotational preferences, requiring further work in this area.

4.4 Value of predicting/understanding flexibility

In general, sequence conservation in flexible regions of proteins suggests a role in biological function, such as involvement in an interface. In the case of SpA-N, the linker sequences studied here are likely to be conserved because the constraints they place on flexibility help determine functionally significant interdomain pose distributions. However, more research is needed to clarify the relationship between sequence, flexibility, and function even in this particularly well-studied system.

The ability to predict interdomain pose distribution from sequence would be an important step toward understanding these natural systems, and the results of this study demonstrate that such predictive ability is still an open problem. In addition to contributing to our understanding of natural systems, such predictive ability would be beneficial in designing engineered systems. For example, therapeutics could be designed to have flexibility matching known pose distributions of target proteins or for optimized orientation of active sites to pass intermediates from one protein to another [61].

Our present study points the way toward improving our ability to predict pose distribution from sequence by demonstrating that experimental methods and robust biophysical measurements exist to empirically ground flexibility predictions. We have shown that both NMR RDCs and SAXS scattering curves can be back-calculated from ensemble predictions and used to assess the quality of flexible ensemble predictions. We expect hybrid collaborations combining these experimental methods with computational models to continue to play an important role in the advancement of predicting protein flexibility.

Further work is needed to solve the reverse problem of flexible structure determination based on NMR and SAXS data. In Qi et al. [45], we introduced a powerful method for determining orientational distributions from NMR RDC data. It would be better to have a method for determining not just an orientational distribution (over SO(3)) but a pose distribution (over SE(3)); such a pose distribution would represent a joint distribution of orientation and translation. In principle, pseudocontact shifts caused by a paramagnetic lanthanide would contain both orientational and distance information; however, we see these shifts only in the ZLBT domain and none in the C domain. The idea of combining orientational information from NMR RDCs with distance information from SAXS to fit a pose distribution to both data sources simultaneously is intuitively appealing; mathematical insights and algorithmic advances are needed to achieve this goal.

5 Conclusion

CASP16 targets T1200 and T1300 asked participants to predict the conformational distribution of two flexible linkers, each part of a D-L-D construct designed to explore the flexibility of SpA-N. These targets represent the first blind challenge to predict population-weighted ensembles of structures to be compared both to experimental

data as well as to ensemble models. The introduction of such targets to CASP recognizes the importance of predicting distributions of conformations realized in solution.

None of the predictions were close fits either for the experimental data or for our optimally-fit orientational distributions. These results demonstrate that predicting the distribution of poses of rigid domains connected by flexible linkers remains an unsolved problem. We conclude that NMR RDC and SAXS data work well together as complementary experimental modalities reflecting different aspects of the pose distribution of D-L-D systems. Further work is needed to fit continuous, low-dimensional pose distributions to SAXS and NMR RDC information.

Given the success of AlphaFold and similar models in predicting static structures, the prediction of protein conformational dynamics driving biological function is an important frontier in structural biology. We hope that future CASPs continue to build on the new directions our experiment has established in addressing this challenge.

Data availability

Both predictions and SAXS data will be uploaded to the CASP website before publication. NMR RDC data are available at BMRB (accession number 53002) [59]. We are currently preparing data processing scripts to be uploaded to https://github.com/donaldlab/CASP16.

Competing interests

Bruce R. Donald is a founder of Ten63 Therapeutics, Inc. No other authors have competing interests to declare.

Funding acknowledgments

This work was supported by the National Institutes of Health (R35 GM-144042 to Bruce R. Donald and NIGMS R35 GM141818 to Gaetano T. Montelione). In addition, this work was supported by the U.S. Department of Energy, Offices of Biological and Environmental Research; Basic Energy Sciences and Advanced Scientific Computing Research through the Biopreparednesss Research Virtual Environment (BRaVE) under contract number DE-AC02-O5CH11231 to Taskforce 5 (DOE-BRAVET5) and to Integrated Diffraction Analysis Technologies (IDAT).

References

[1] Sriram Aiyer et al. "A Common Binding Motif in the ET Domain of BRD3 Forms Polymorphic Structural Interfaces with Host and Viral Proteins". In: *Structure* 29.8 (Aug. 5, 2021), 886–898.e6.

[2] Michael Andrec et al. "A Large Data Set Comparison of Protein Structures Determined by Crystallography and NMR: Statistical Test for Structural Differences and the Effect of Crystal Packing". In: *Proteins: Structure, Function, and Bioinformatics* 69.3 (2007), pp. 449–465.

[3] Anthony K. Yan and Bruce R. Donald. "Principal Components Analysis, Residual Dipolar Couplings, and Their Relationship in NMR Structural Biology". In: Bruce R. Donald. *Algorithms in Structural Molecular Biology*. Computational Molecular Biology. Cambridge, Mass: The MIT Press, 2011, pp. 27–45.

[4] Pooja Arora, Gordon G. Hammes, and Terrence G. Oas. "Folding Mechanism of a Multiple Independently-Folding Domain Protein: Double B Domain of Protein A". In: *Biochemistry* 45.40 (Oct. 2006), pp. 12312–12324.

[5] Adam W. Barb and Ganesh P. Subedi. "An Encodable Lanthanide Binding Tag with Reduced Size and Flexibility for Measuring Residual Dipolar Couplings and Pseudocontact Shifts in Large Proteins". In: *Journal of biomolecular NMR* 64.1 (Jan. 2016), pp. 75–85.

[6] Adam W. Barb et al. "Lanthanide Binding and IgG Affinity Construct: Potential Applications in Solution NMR, MRI, and Luminescence Microscopy". In: *Protein Science* 21.10 (2012), pp. 1456–1466.

[7] Ivano Bertini et al. "Experimentally Exploring the Conformational Space Sampled by Domain Reorientation in Calmodulin". In: *Proceedings of the National Academy of Sciences* 101.18 (May 4, 2004), pp. 6841–6846.

[8] Christopher Bingham. "An antipodally symmetric distribution on the sphere". In: *The Annals of Statistics* (1974), pp. 1201–1225.

[9] Wade Borcherds et al. "How Do Intrinsically Disordered Protein Regions Encode a Driving Force for Liquid–Liquid Phase Separation?" In: *Current Opinion in Structural Biology*. Theory and Simulation/Computational Methods · Macromolecular Assemblies 67 (Apr. 1, 2021), pp. 41–50.

[10] Lorenza Bordoli, Florian Kiefer, and Torsten Schwede. "Assessment of Disorder Predictions in CASP7". In: *Proteins: Structure, Function, and Bioinformatics* 69.S8 (2007), pp. 129–136.

[11] Chris A. Brosey and John A. Tainer. "Evolving SAXS Versatility: Solution X-ray Scattering for Macromolecular Architecture, Functional Landscapes, and Integrative Structural Biology". In: *Current Opinion in Structural Biology* 58 (Oct. 2019), pp. 197–213.

[12] Bruce R. Donald. *Algorithms in Structural Molecular Biology*. Computational Molecular Biology. Cambridge, Mass: The MIT Press, 2011.

[13] Fan Cao et al. "A Coarse-Grained Model for Disordered and Multi-Domain Proteins". In: *Protein Science* 33.11 (2024), e5172.

[14] Xiaoying Chen, Jennica Zaro, and Wei-Chiang Shen. "Fusion Protein Linkers: Property, Design and Functionality". In: *Advanced drug delivery reviews* 65.10 (Oct. 15, 2013), pp. 1357–1369.

[15] Naga Babu Chinnam et al. "ASCC1 Structures and Bioinformatics Reveal a Novel Helix-Clasp-Helix RNA-binding Motif Linked to a Two-Histidine Phosphodiesterase". In: *The Journal of Biological Chemistry* 300.6 (June 2024), p. 107368.

[16] Naga Babu Chinnam et al. "Combining Small Angle X-ray Scattering (SAXS) with Protein Structure Predictions to Characterize Conformations in Solution". In: *Methods in Enzymology* 678 (2023), pp. 351–376.

[17] G. Marius Clore and Daniel S. Garrett. "R-Factor, Free R, and Complete Cross-Validation for Dipolar Coupling Refinement of NMR Structures". In: *Journal of the American Chemical Society* 121.39 (Oct. 1, 1999), pp. 9008–9012.

[18] Gabriel Cornilescu et al. "Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase". In: *Journal of the American Chemical Society* 120.27 (June 13, 1998), pp. 6836–6837.

[19] Lindsay N. Deis et al. "Multiscale Conformational Heterogeneity in Staphylococcal Protein A: Possible Determinant of Functional Plasticity". In: *Structure* 22.10 (Oct. 7, 2014), pp. 1467–1477.

[20] Michel Marie Deza and Elena Deza. *Encyclopedia of Distances*. Berlin, Heidelberg: Springer, 2013.

[21] Bruce R. Donald and Jeffrey Martin. "Automated NMR Assignment and Protein Structure Determination Using Sparse Dipolar Coupling Constraints". In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 55.2 (Aug. 1, 2009), pp. 101–127.

[22] Alejandro Estaña et al. "Realistic Ensemble Models of Intrinsically Disordered Proteins Using a Structure-Encoding Coil Database". In: *Structure* 27.2 (Feb. 5, 2019), 381–391.e2.

[23] P. J. Flory. "Statistical Thermodynamics of Random Networks". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 351.1666 (Nov. 19, 1976), pp. 351–380.

[24] Richard A. George and Jaap Heringa. "An Analysis of Protein Domain Linkers: Their Classification and Role in Protein Folding". In: *Protein Engineering, Design and Selection* 15.11 (Nov. 1, 2002), pp. 871–879.

[25] M. Gerstein, A. M. Lesk, and C. Chothia. "Structural Mechanisms for Domain Movements in Proteins". In: *Biochemistry* 33.22 (June 7, 1994), pp. 6739–6749.

[26] Gregory-Neal W. Gomes et al. "Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and Single-Molecule FRET". In: *Journal of the American Chemical Society* 142.37 (Sept. 16, 2020), pp. 15697–15710.

[27] Alexander Grishaev et al. "Refinement of Multidomain Protein Structures by Combination of Solution Small-Angle X-ray Scattering and NMR Data". In: *Journal of the American Chemical Society* 127.47 (Nov. 30, 2005), pp. 16621–16628.

[28] Michal Hammel and John A. Tainer. "X-Ray Scattering Reveals Disordered Linkers and Dynamic Interfaces in Complexes and Mechanisms for DNA Double-Strand Break Repair Impacting Cell and Cancer Biology". In: *Protein Science: A Publication of the Protein Society* 30.9 (Sept. 2021), pp. 1735–1756.

[29] Yuanpeng Janet Huang et al. "Assessment of Prediction Methods for Protein Structures Determined by NMR in CASP14: Impact of AlphaFold2". In: *Proteins* 89.12 (Dec. 2021), pp. 1959–1976.

[30] Greg L. Hura et al. "Comprehensive Macromolecular Conformations Mapped by Quantitative SAXS Analyses". In: *Nature Methods* 10.6 (June 2013), pp. 453–454.

[31] Greg L. Hura et al. "Small Angle X-ray Scattering-Assisted Protein Structure Prediction in CASP13 and Emergence of Solution Structure Differences". In: *Proteins: Structure, Function, and Bioinformatics* 87.12 (2019), pp. 1298–1314.

[32] Yumi Jin and Roland L. Dunbrack Jr. "Assessment of Disorder Predictions in CASP6". In: *Proteins: Structure, Function, and Bioinformatics* 61.S7 (2005), pp. 167–175.

[33] Felix Jonas, Yoav Navon, and Naama Barkai. "Intrinsically Disordered Regions as Facilitators of the Transcription Factor Target Search". In: *Nature Reviews Genetics* (Feb. 21, 2025), pp. 1–12.

[34] Andriy Kryshtafovych et al. "Breaking the Conformational Ensemble Barrier: Ensemble Structure Modeling Challenges in CASP15". In: *Proteins: Structure, Function, and Bioinformatics* 91.12 (2023), pp. 1903–1911.

[35] Karsten Kunze and Helmut Schaeben. "The Bingham Distribution of Quaternions and Its Spherical Radon Transform in Texture Analysis". In: *Mathematical Geology* 36.8 (2004), pp. 917–943.

[36] Alexander Lemak et al. "Structural Characterization of a Flexible Two-Domain Protein in Solution Using Small Angle X-Ray Scattering and NMR Data". In: *Structure* 22.12 (Dec. 2, 2014), pp. 1862–1874.

[37] Judit A Losonczi et al. "Order Matrix Analysis of Residual Dipolar Couplings Using Singular Value Decomposition". In: *Journal of Magnetic Resonance* 138.2 (June 1, 1999), pp. 334–342.

[38] Eugene Melamud and John Moult. "Evaluation of Disorder Predictions in CASP5". In: *Proteins: Structure, Function, and Bioinformatics* 53.S6 (2003), pp. 561–565.

[39] Bohdan Monastyrskyy et al. "Assessment of Protein Disorder Region Predictions in CASP10". In: *Proteins* 82 Suppl 2 (0 2 Feb. 2014), pp. 127–137.

[40] Bohdan Monastyrskyy et al. "Evaluation of Disorder Predictions in CASP9". In: *Proteins* 79 Suppl 10.S10 (2011), pp. 107–118.

[41] Orly Noivirt-Brik, Jaime Prilusky, and Joel L. Sussman. "Assessment of Disorder Predictions in CASP8". In: *Proteins: Structure, Function, and Bioinformatics* 77.S9 (2009), pp. 210–216.

[42] Tadeusz L. Ogorzalek et al. "Small Angle X-ray Scattering and Cross-Linking for Data Assisted Protein Structure Prediction in CASP 12 with Prospects for Improved Accuracy". In: *Proteins* 86 Suppl 1 (Suppl 1 Mar. 2018), pp. 202–214.

[43] Martin Pelikan, Greg L. Hura, and Michal Hammel. "Structure and Flexibility within Proteins as Identified through Small Angle X-ray Scattering". In: *General Physiology and Biophysics* 28.2 (June 2009), pp. 174–189.

[44] Christopher D. Putnam et al. "X-Ray Solution Scattering (SAXS) Combined with Crystallography and Computation: Defining Accurate Macromolecular Structures, Conformations and Assemblies in Solution". In: *Quarterly Reviews of Biophysics* 40.3 (Aug. 2007), pp. 191–285.

[45] Yang Qi et al. "Continuous Interdomain Orientation Distributions Reveal Components of Binding Thermodynamics". In: *Journal of Molecular Biology* 430.18 (Sept. 2018), pp. 3412–3426.

[46] Soumya G. Remesh et al. "Conformational Plasticity of the Immunoglobulin Fc Domain in Solution". In: *Structure (London, England: 1993)* 26.7 (July 3, 2018), 1007–1014.e2.

[47] Daniel J. Rosenberg, Greg L. Hura, and Michal Hammel. "Size Exclusion Chromatography Coupled Small Angle X-ray Scattering with Tandem Multiangle Light Scattering at the SIBYLS Beamline". In: *Methods in Enzymology* 677 (2022), pp. 191–219.

[48] Ke Ruan and Joel R. Tolman. "Composite Alignment Media for the Measurement of Independent Sets of NMR Residual Dipolar Couplings". In: *Journal of the American Chemical Society* 127.43 (Nov. 1, 2005), pp. 15032–15033.

[49] Dina Schneidman-Duhovny et al. "Accurate SAXS Profile Computation and Its Assessment by Contrast Variation Experiments". In: *Biophysical Journal* 105.4 (Aug. 20, 2013), pp. 962–974.

[50] Dina Schneidman-Duhovny et al. "FoXS, FoXSDock and MultiFoXS: Single-state and Multi-State Structural Modeling of Proteins and Their Complexes Based on SAXS Profiles". In: *Nucleic Acids Research* 44.W1 (July 8, 2016), W424–429.

[51] Maad Shatnawi, Nazar Zaki, and Paul D. Yoo. "Protein Inter-Domain Linker Prediction Using Random Forest and Amino Acid Physiochemical Properties". In: *BMC Bioinformatics* 15.16 (Dec. 8, 2014), S8.

[52] John P. Snyder. *Map Projections: A Working Manual.* USGS Professional Paper 1395. Washington, D.C.: U.S. Geological Survey, 1987.

[53] Lindsey Spiegelman et al. "Strengthening of Enterococcal Biofilms by Esp". In: *PLoS pathogens* 18.9 (Sept. 2022), e1010829.

[54] João M. C. Teixeira et al. “IDPConformerGenerator: A Flexible Software Suite for Sampling the Conformational Space of Disordered Protein States”. In: *The Journal of Physical Chemistry. A* 126.35 (Sept. 8, 2022), pp. 5985–6003.

[55] Roberto Tejero et al. “AlphaFold Models of Small Proteins Rival the Accuracy of Solution NMR Structures”. In: *Frontiers in Molecular Biosciences* 9 (June 13, 2022).

[56] F. Emil Thomasen and Kresten Lindorff-Larsen. “Conformational Ensembles of Intrinsically Disordered Proteins and Flexible Multidomain Proteins”. In: *Biochemical Society Transactions* 50.1 (Feb. 7, 2022), pp. 541–554.

[57] F. Emil Thomasen et al. “Rescaling Protein-Protein Interactions Improves Martini 3 for Flexible Proteins in Solution”. In: *Nature Communications* 15.1 (Aug. 5, 2024), p. 6645.

[58] Chittaranjan Tripathy et al. “Protein Loop Closure Using Orientational Restraints from NMR Data”. In: *Proteins: Structure, Function, and Bioinformatics* 80.2 (Feb. 2012), pp. 433–453.

[59] E. L. Ulrich et al. “BioMagResBank”. In: *Nucleic Acids Research* 36.Database issue (2008), pp. D402–D408.

[60] Robin van der Lee et al. “Classification of Intrinsically Disordered Regions and Proteins”. In: *Chemical Reviews* 114.13 (July 9, 2014), pp. 6589–6631.

[61] Xi Wang et al. “Efficient Production of Oxidized Terpenoids via Engineering Fusion Proteins of Terpene Synthase and Cytochrome P450”. In: *Metabolic Engineering* 64 (Mar. 2021), pp. 41–51.

[62] Anna Yershova et al. “Generating Uniform Incremental Grids on SO(3) Using the Hopf Fibration”. In: *The International journal of robotics research* 29.7 (June 1, 2010), pp. 801–812.

[63] Feng Yu and Shahar Sukenik. “Structural Preferences Shape the Entropic Force of Disordered Protein Ensembles”. In: *The Journal of Physical Chemistry B* 127.19 (May 18, 2023), pp. 4235–4244.

[64] Deyou Zheng, James M. Aramini, and Gaetano T. Montelione. “Validation of Helical Tilt Angles in the Solution NMR Structure of the Z Domain of Staphylococcal Protein A by Combined Analysis of Residual Dipolar Coupling and NOE Data”. In: *Protein Science* 13.2 (2004), pp. 549–554.

[65] Wei Zheng et al. “Integrating Deep Learning, Threading Alignments, and a Multi-MSA Strategy for High-Quality Protein Monomer and Complex Structure Prediction in CASP15”. In: *Proteins* 91.12 (Dec. 2023), pp. 1684–1703.