



An Integrative Approach to Determine 3D Protein Structures Using Sparse Paramagnetic NMR Data and Physical Modeling

Kari Gaalswyk^{1†}, Zhihong Liu^{2†}, Hans J. Vogel² and Justin L. MacCallum^{1*}

¹Department of Chemistry, University of Calgary, Calgary, AB, Canada, ²Department of Biological Sciences, University of Calgary, Calgary, AB, Canada

OPEN ACCESS

Edited by:

Massimiliano Bonomi,
Institut Pasteur, France

Reviewed by:

Enrico Ravera,
University of Florence, Italy
Alexandre M. J. J. Borvin,
Utrecht University, Netherlands

*Correspondence:

Justin L. MacCallum
justin.maccallum@ucalgary.ca

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Biological Modeling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 04 March 2021

Accepted: 29 July 2021

Published: 12 August 2021

Citation:

Gaalswyk K, Liu Z, Vogel HJ and
MacCallum JL (2021) An Integrative
Approach to Determine 3D Protein
Structures Using Sparse
Paramagnetic NMR Data and
Physical Modeling.
Front. Mol. Biosci. 8:676268.
doi: 10.3389/fmolb.2021.676268

Paramagnetic nuclear magnetic resonance (NMR) methods have emerged as powerful tools for structure determination of large, sparsely protonated proteins. However traditional applications face several challenges, including a need for large datasets to offset the sparsity of restraints, the difficulty in accounting for the conformational heterogeneity of the spin-label, and noisy experimental data. Here we propose an integrative approach to structure determination combining sparse paramagnetic NMR with physical modelling to infer approximate protein structural ensembles. We use calmodulin in complex with the smooth muscle myosin light chain kinase peptide as a model system. Despite acquiring data from samples labeled only at the backbone amide positions, we are able to produce an ensemble with an average RMSD of ~ 2.8 Å from a reference X-ray crystal structure. Our approach requires only backbone chemical shifts and measurements of the paramagnetic relaxation enhancement and residual dipolar couplings that can be obtained from sparsely labeled samples.

Keywords: paramagnetic relaxation enhancement, NMR, modeling, protein structure, integrative structural biology, calmodulin

INTRODUCTION

Protein nuclear magnetic resonance (NMR) spectroscopy has played an important role in biomolecular structure determination. To date more than 13,000 NMR structures have been deposited in the Protein Data Bank [PDB (Berman et al., 2000)], accounting for about 7.5 percent of all available protein structures (PDB Statistics, 2021). The vast majority of the deposited NMR solution structures are determined for smaller proteins or independently-folded isolated protein domains (Tugarinov et al., 2004). Without special stable isotopic labelling techniques, NMR methods struggle with structure determination of proteins larger than ~ 25 kDa as the slow molecular tumbling results in rapid relaxation, leading to poor resolution and spectral quality (Kay, 2016). The most commonly used method to overcome this challenge is to combine transverse relaxation optimized spectroscopy (TROSY) (Pervushin et al., 1997) with site-specific protonation in an otherwise perdeuterated background (Tugarinov et al., 2006), although there are many alternatives, e.g., Tugarinov et al. (2004) and Ruschak and Kay (2010). While such site-specific isotope labelling can dramatically increase the spectral quality and interpretability, the overall perdeuteration results in protons being sparsely distributed within the structure.

The overwhelming majority of solution NMR structures in the PDB are based around Nuclear Overhauser Effect Spectroscopy (NOESY), which provides information about through-space interactions between protons that can be used to derive distance restraints for 3D structure determination. For the homonuclear NOE to be detectable, however, the protons must be within about 6 Å or closer, which can pose a substantial challenge for sparsely-labelled samples due to the lack of proton pairs that are in close proximity within the folded protein structure (Gardner and Kay, 1998).

This phenomenon is particularly acute for samples that are labelled with protons only on the exchangeable backbone amide positions, as the amide protons within an alpha helix are typically too far away from amide protons in other secondary structure elements to produce a detectable NOE. Consequently, labelling of only the amide protons of alpha-helical proteins leads to a restraint network that is too sparse to calculate a 3D structure. Other site-specific labelling schemes can supplement amide labelling, leading to a denser restraint network (Goto and Kay, 2000). Site-specific labelling of the terminal methyl groups of isoleucine, leucine, and valine (ILV-labeling) is particularly common (Tugarinov et al., 2006), but several alternatives and complementary labeling methods exist, e.g., Kainosho and Güntert (2009), Otten et al. (2010), and Gifford et al. (2011). While these additional labelling schemes can increase the density of the restraint network, they often come at the cost of increased complexity and the need to synthesize or purchase expensive precursors that are required to generate the isotope-labelled samples.

Paramagnetic NMR methods have emerged as potentially viable alternatives, capable of providing valuable information about electron-nucleus distances up to ~20–30 Å (Koehler and Meiler, 2011; Pilla et al., 2017). Paramagnetic relaxation enhancement (PRE) experiments have been performed with native metalloproteins and proteins modified with covalent paramagnetic tags such as nitroxide spin labels and metal chelates (Bertini et al., 2005; Clore and Iwahara, 2009; Keizers and Ubink, 2011). These techniques can be used to extend the scope of NMR methods to larger, more complex systems by providing long-range distances when short-range NOEs are unavailable or limited. Due to the long-range nature of paramagnetic relaxation enhancement, PRE experiments can provide valuable distance restraints even in sparsely labelled perdeuterated protein samples.

The utility of a distance restraint generally depends on two factors (Sullivan et al., 2003; Sullivan and Kuntz, 2004). First, restraints with short spatial distances are more valuable than those with long spatial distances because there are many more ways for two particles to be far apart than close together. Thus, a short-distance restraint provides more information than a long one. Second, this effect is more substantial for restraints involving residues that are more distant in the sequence. Thus, the most valuable restraints involve residues distant in sequence but close together in space. NOESY experiments provide powerful short spatial distance restraints (<6 Å) but can miss many crucial long sequence distance restraints due to distribution of the isotope labels. In contrast, PRE experiments will yield more distance

restraints due to the paramagnetic relaxation enhancement effect's long-range nature. Many of these restraints will be of limited utility due to their long spatial distances; however, the collective effect of all of these long spatial distance restraints with the remaining short spatial distance (<12 Å) restraints can still be potent. As an aside, PRE methods have also become a popular approach to explore lowly populated transient protein states (Iwahara and Clore, 2006).

Using PRE data for 3D protein structure determination presents several challenges. First, each experiment only provides information about the spatial proximity of a given proton to a single site labelled with a paramagnetic tag. Adequately determining the 3D structure requires multiple experiments with different tag locations, increasing both experimental time and cost. Second, each experiment provides only a limited amount of information (Battiste and Wagner, 2000; Gottstein et al., 2012). Although a single experiment provides information about the spatial proximity of each residue to the paramagnetic tag, much of this information is redundant. For example, if a residue is close to the tag, then neighbouring residues in the sequence are also likely to be close. Furthermore, information that a residue is close to the tag provides a far more powerful structural constraint than information that a residue is distant from the tag, but the latter occurrence is far more frequent. Third, the derived distances can be imprecise due to intermolecular interactions, secondary metal-binding sites, and diamagnetic contamination (Clore and Iwahara, 2009). Fourth, heterogeneity and dynamics (Clore et al., 1990; Ryabov and Fushman, 2007; Bertini et al., 2012) can complicate the interpretation of PRE data. Relaxation can be strongly affected by conformational heterogeneity due to the inverse sixth power relationship between the PRE and the electron-nucleus distance; i.e., a minor structural population with a strong paramagnetic effect can have a significant impact on the measured data (Clore and Iwahara, 2009). Finally, the effects of spin diffusion due to dipole-dipole coupling can limit the accuracy of the measured PRE data (Vlasie et al., 2007; Vlasie et al., 2008; Bellomo et al., 2021). These challenges have slowed the widespread adoption of PRE-based methods for structure determination in favour of traditional NOE-based approaches.

Residual dipolar coupling (RDC) measurements are a common supplemental data source to PRE and NOE-based experiments. RDC measurements are carried out on systems where the protein is weakly aligned relative to the external magnetic field. Rather than reporting on distances, RDCs report on the angles between bonded atoms (typically backbone N-H bond vectors) and the external magnetic field, which provides valuable orientational information that complements distance information from PRE or NOE experiments (Prestegard et al., 2004). RDCs have been used for structure refinement and as restraints in *de novo* structure prediction software (Banci et al., 1998; Raman et al., 2010; Prestegard et al., 2014). While many protein structures based on RDC measurements have been reported, molecular modeling and low temperature annealing procedures are often used to derive and refine the 3D structures (Chou et al., 2000; Lipsitz and Tjandra, 2004; Huang and Vogel, 2012). Clearly there is room for

more unbiased approaches to incorporate such RDC data into protein structure calculations.

Integrative approaches to structure determination (Ward et al., 2013) have emerged as practical tools for converting NMR and other experimental data into useful structural models. For example, PRE and RDC measurements have been used to drive molecular docking studies (van Dijk et al., 2005; Gelis et al., 2007; Gochin et al., 2011), as restraints in molecular dynamics simulations to generate ensembles of conformers (Dedmon et al., 2005; Ascianto et al., 2011), or they have been incorporated into Rosetta scoring functions (Lange et al., 2012; Kuenze et al., 2019). We recently demonstrated the structure determination of a small protein using PRE measurements in solid-state NMR (Perez et al., 2019). However, integrative methods are not without their own set of challenges. Even the most sophisticated methods can still struggle as the data becomes sparse, ambiguous, or unreliable, and considerable method development is often required to treat a new type of experimental data in order to correctly account for its characteristics, e.g., the conformational heterogeneity of spin-labels in PRE measurements (Iwahara et al., 2004; Anthis et al., 2011; Andrałojć et al., 2017).

Here, we show that these challenges can be overcome by using a sophisticated integrative structural biology approach called Modeling Employing Limited Data (MELD) (MacCallum et al., 2015). MELD combines experimental data from multiple sources with physical modelling to overcome the challenges of sparse, ambiguous, and difficult to interpret experimental data to infer accurate protein structural ensembles. We combine PRE and RDC measurements with secondary structure predictions based on backbone chemical shifts. We use MELD to infer the structure of Calmodulin in complex with the 20-residue smooth muscle myosin light chain kinase peptide (169 residues total). Calmodulin was selected for this exploratory work as it has an almost completely helical structure where the absence of inter-helical close contacts between amide protons makes 3D structure determination by NOE-based approaches difficult. Calmodulin-peptide complex have previously been used as models for integrative approaches using sparse NMR data (Andrałojć et al., 2014; Carlon et al., 2019). We show that MELD can identify conformations within 3 Å of a reference X-ray crystal structure using only sparse paramagnetic NMR restraints and RDCs from amide protons in combination with backbone chemical shifts, while successfully addressing conformational heterogeneity and noise in the NMR data.

EXPERIMENTAL METHODS

Calmodulin–Peptide Complex as a Model System

In this work, we illustrate our approach for the protein calmodulin in complex with the smooth muscle Myosin Light Chain Kinase (smMLCK) peptide. Throughout, we will use a previously solved crystal structure of this complex [PDB ID: 1cdl (Meador et al., 1992)] as a reference.

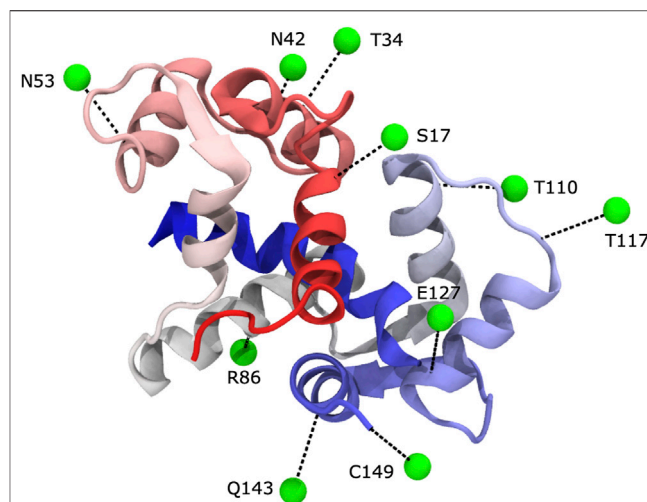


FIGURE 1 | We carried out PRE experiments with ten different label sites. In each experiment, Calmodulin was MTSL-labeled at a different position. Spin labels were generally located in predicted surface-exposed sites within secondary structures. Spin labels are shown in green as virtual sites (see text), and their corresponding cysteine mutation linkage site is shown in black (PDB 1cdl).

Overview of Labeling Strategy

In previous studies, specific nitroxide spin-labeled target peptides that bind to calmodulin were used; in this manner it was possible to map out the orientation of the peptide with respect to the protein (Zhang et al., 1995; Yuan et al., 2004). In this work, we collected PRE data for a total of ten spin-labelled protein sites (Figure 1). Nine of the sites on the protein were chosen to be solvent-exposed and within secondary structure elements by manual inspection of predicted secondary structure and solvent exposure. The remaining site, C149, was a single-residue extension of the C-terminus. To better simulate the process for a system without a previously determined structure, the protein's known structure **was not used** in choosing the spin-labelled sites. Indeed, we learned later that several of the selected sites provided little useful information because they are either distant from the rest of the protein or they provided information that is mostly redundant with that obtained from other labeling sites.

To simulate the limited availability of isotopically labelled peptide, either due to cost or difficulty of production, only four of the ten spin-label data sets (chosen randomly) were collected with isotopically labelled peptide. The remaining six data sets were collected without labelled peptide, which results in the peptide being present but unlabeled and undetectable in the ^1H , ^{15}N HSQC NMR experiments.

Protein Production

The ten single-cysteine point calmodulin (CaM) mutants (S17C, T34C, N42C, N53C, R86C, T110C, T117C, E127C, Q143C, 149C) were made by standard site-directed mutagenesis methods for attachment of the thiol-specific nitroxide spin label (1-oxyl-2,2,5,5-tetramethyl- δ -3-pyrroline-3-methyl)

methanethiosulfonate (MTSL, Toronto Research Chemicals). Correctness of the mutations was confirmed by DNA sequencing. Calmodulin contains no Cys residues, so highly site-specific labeling can be obtained in this manner. MTSL is a relatively compact, yet highly reactive molecule compared to other commercially available nitroxide spin-labels that have been used to modify Cys residues; its shorter more rigid structure would be an advantage for the PRE studies [for discussion see for example (Guo et al., 2008; Fawzi et al., 2011)]. ^{13}C and ^{15}N -labeled CaM was expressed in M9 minimal medium with 99.9% $^{15}\text{NH}_4\text{Cl}$ and $^{13}\text{C}_6$ -glucose (0.5 gr/L and 3 gr/L, respectively; Cambridge Isotopes Laboratories) as isotope sources. Proteins were expressed and purified as described previously (Liu and Vogel, 2012; Ishida et al., 2016).

We followed a standard protocol for attaching the nitroxide spin-label to each single-cysteine CaM mutants with the spin-labelling reagent MTSL (Battiste and Wagner, 2000). To prepare the CaM/smMLCK complex sample, a 1.2-fold excess of either unlabeled or labeled peptide was mixed with each CaM mutant protein. All preparations were divided into two NMR samples. One sample was reduced to inactivate the spin-label by adding a 3-fold excess of ascorbic acid.

Peptide Production

A construct with a 6xHis-KSI (D38A) fusion-protein tag was generated for smMLCK peptide expression in *Escherichia coli* (Jaroniec et al., 2005; Ishida and Vogel, 2010). The ketosteroid isomerase (KSI) coding sequence generates an insoluble protein, and this directs the protein-peptide fusion directly into inclusion bodies, where they are protected from proteolytic cleavage (Hwang et al., 2014). A linkage sequence “GGGGSSDP” with the Asp-Pro acid cleavage site was designed between the KSI protein and the sequence of the smMLCK peptide. The entire 6xHis-KSI-GGGGSSDP-smMLCKp gene sequence was inserted between the NdeI and XhoI sites of the pET15b(+) plasmid (Novagen), which was subsequently transferred into BL21(DE3) *E. coli* cells for protein expression. The cells were grown in either LB media (for unlabeled peptide) or minimal M9 media (containing $^{13}\text{C}_6$ -glucose and $^{15}\text{NH}_4\text{Cl}$ isotope to produce isotope-labeled peptide) and they were induced at $\text{OD}_{600} = 0.6$ with 1 mM IPTG for 4 h at 37°C. A cell lysate was prepared as previously described. The insoluble fusion protein was separated after one hour of centrifugation (18,000 rpm) and then resuspended in 6 M guanidine hydrochloride. Impurities were removed before the insoluble proteins can be extracted with metal chelate chromatography on a nickel affinity column. After extensive dialysis with double distilled H_2O , the precipitated insoluble protein was collected and the Asp-Pro bond was cleaved in 10% formic acid at 80°C for 90 min (Hwang et al., 2014). The protein-peptide mixture was flash frozen with liquid nitrogen and lyophilized. Insoluble proteins and other impurities were removed after the lyophilized mixture was resuspended in a 20 mM Tris-HCl buffer (pH = 8.0). Finally, the unlabeled and isotope-labeled smMLCK peptides were purified with reverse-phase HPLC (COSMOSIL 5C₁₈-AR-300, Nacalai United States). All purified peptides were lyophilized and stored at -20°C for further use. The final peptide sequence after cleaving is

PARRKWQKTGHAVRAIGRLSS. The N-terminal proline is not observable in the NMR experiments and was not included in modeling with MELD.

Chemical Shift Assignments

All NMR experiments were carried out on a 600 MHz Bruker AVANCE spectrometer with a field strength of 14.1 T. Backbone resonance assignments for the protein and the bound peptide were confirmed with the following 3D experiments: HNCO, HNCA, HNCOCA, HNCACB, and CBCA(CO)NH, as described previously (Liu and Vogel, 2012). All data were processed using NMRPipe (Delaglio et al., 1995) and analyzed with the program NMRView (Johnson and Blevins, 1994). All chemical shifts from these experiments were used to obtain backbone torsion angles from the program TALOS+ (Cornilescu et al., 1998; Shen et al., 2009). Secondary structure elements as identified through the assigned chemical shifts were as expected based on the known structure.

Paramagnetic Relaxation Enhancement Measurements

Two ^1H , ^{15}N HSQC spectra were obtained for each spin-label construct. Each system contained each ^{15}N -labeled protein and either unlabeled or ^{15}N -labeled peptide, depending on the spin-label site (S17C, N53C, T127C, and I49C had isotopically labeled peptide). One HSQC was collected with active spin-label, whereas the other HSQC was collected with reduced, inactivated spin-label. The distances between the spin-label and the affected nuclei were calculated using the two-time point method (Iwahara et al., 2007).

Residual Dipolar Coupling Measurements

Finally, to supplement the PRE experiments, we obtained RDC measurements for the amide groups in the complex with a sample where both the protein and peptide are isotopically labeled. Residual dipolar couplings (RDC) were measured for the CaM/smMLCK complex sample in a partially aligned media, which contains 2 mM bis-Tris (pH = 7.0), 300 mM KCl and 16 mg/ml Pf1 bacteriophage (Asla Biotech Ltd.). The IPAP-HSQC experiment was used for the RDC measurements (Ottiger et al., 1998). In these experiments the effects of dipole-dipole cross-correlated relaxation can impact the accuracy of $^1\text{J}_{\text{NH}}$ splitting measured from the spectra introducing a small residual bias in the RDCs. While these systematic errors can be eliminated by using a selectively-decoupled sequence (Yao et al., 2009), the errors are small relative to the magnitude of the measured RDCs and are expected to have a minimal effect on structure determination (Yao et al., 2009). Our work uses only a single RDC alignment. Notably, a mutant of Calmodulin is capable of selective binding to lanthanides, which provides a strategy for the measurement of multiple RDCs (Bertini et al., 2009). A quantitative assessment of protein mobility/heterogeneity by RDC would require the use of multiple alignments (Barbieri et al., 2002; Tolman, 2002; Bouvignies et al., 2006; Higman et al., 2011; Guerry et al., 2013; Andraščić et al., 2015). However, as discussed further

below, it is not currently possible to conduct such an analysis with the MELD approach, as MELD compares individual structures, rather than ensembles of structures, to the experimental data.

COMPUTATIONAL APPROACH

Overview of Modeling Employing Limited Data Approach

Here, we employ MELD, a physics-based Bayesian approach for structural determination to infer the ensemble of structures most consistent with the known physics of protein structure and experimental data (MacCallum et al., 2015; Perez et al., 2015). MELD uses a Bayesian framework to combine a physics-based prior distribution with a data likelihood function to make statistically consistent inferences about conformations that explain the experimental data.

MELD uses Bayes' theorem:

$$p(x|D) \propto p(x)p(D|x), \quad (1)$$

where x represents the atomic coordinates and D represents the data. The physics-based prior, $p(x)$, specifies which structures are more likely *a priori* and determines the distribution of structures in the absence of data. In the present study the physics-based prior is given by the Amber ff14SB force field (Maier et al., 2015) with a grid-based torsion potential (Perez et al., 2015) and the OBC generalized-Born implicit solvent model (Onufriev et al., 2004). The likelihood function, $p(D|x)$, captures the compatibility between the data and some structure x . In MELD, the likelihood function takes the form of a unique restraint function (MacCallum et al., 2015), explained in more detail below. The goal of Bayesian inference is to compute the posterior distribution, $p(x|D)$, which is the most statistically consistent inference given the prior, likelihood, and data.

As discussed in the Results section below, the term *ensemble* is highly overloaded in structural biology and care is required in interpretation. MELD belongs to the class of methods where a single structure, rather than entire ensemble of structures, is considered in the likelihood function, such that each member of the ensemble individually agrees with the experimental data. Any conformational heterogeneity (e.g., flexible loops) may represent true intrinsic heterogeneity, but may also simply reflect a lack of data. As such, MELD produces a form of *uncertainty ensemble* in the terminology of Bonomi et al., 2017.

Overview of Experimental Data

The input to our approach is: 1) the protein sequence, 2) TALOS+ secondary structure predictions derived from backbone chemical shifts (Cornilescu et al., 1998; Shen et al., 2009; MacCallum et al., 2015), 3) distance restraints derived from PRE measurements, and 4) orientational restraints derived from RDC measurements. We have

recently demonstrated the success of a similar approach for PRE measurements in solid-state NMR (Perez et al., 2019).

PRE data is often both noisy and sparse (Kim et al., 2014), which makes structural inference challenging. Despite collecting data for ten spin-label positions, we can derive only a few distance restraints that are short in spatial distance (in this case, we define short as $<12 \text{ \AA}$) (Figure 2). Of these short spatial distances, only a small number correspond to residues that are distant in sequence, which would provide the most information about folding (Gottstein et al., 2012). Furthermore, as stated previously, to simulate the limited availability of isotopically labelled peptide, only four of ten datasets (S17C, N53C, T127C, and 149C) had labelled peptide, and there are no short distance PREs between the peptide and the protein. This leaves the peptide's correct placement to be dictated by longer, less informative spatial distance restraints and the physical model, which makes accurate inference more challenging.

Deriving Distances From Paramagnetic Relaxation Enhancement Data

Our first step was to develop a consistent method to convert ensemble-averaged PRE measurements into distance restraints. PRE data were turned into approximate distances using the Solomon-Bloembergen equations following the standard approach (Battiste and Wagner, 2000; Iwahara et al., 2007).

For nitroxides, Curie-spin relaxation is negligible and the transverse relaxation enhancement, Γ_2 , is dominated by direct dipole-dipole interactions (Clore and Iwahara, 2009). In this case, Γ_2 is related to the distance between the paramagnetic center and the observed nucleus, r :

$$\Gamma_2 = \frac{K}{r^6} \left[4\tau_c + \frac{3\tau_c}{1 + \omega_H^2 \tau_c^2} \right]$$

where ω_H is the Larmor frequency of the proton, and τ_c is the correlation time for the electron-nuclear interaction defined as $\tau_c = (\tau_r^{-1} + \tau_s^{-1})^{-1}$ where τ_r is the rotational correlation time, and τ_s is the electron relaxation time. Previous experiments (Lee et al., 2002) have determined that $\tau_c \approx 9.5 \text{ ns}$ for Calmodulin in complex with smMLCK peptide at 25°C. For nitroxides, the electron relaxation time is long (Iwahara et al., 2007) ($\tau_s > 10^{-7} \text{ s}$), so the rotational correlation time dominates and $\tau_c \approx \tau_r \approx 9.5 \text{ ns}$. K is given by:

$$K = \frac{1}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \gamma_1^2 g^2 \mu_B^2 S(S+1)$$

where μ_0 is the permeability of vacuum, γ_1 is the nuclear gyromagnetic ratio, g is the electronic g factor, μ_B is the Bohr magneton, and S is the electron spin quantum number. Γ_2 can be estimated using a two-point time measurement (Iwahara et al., 2007):

$$\Gamma_2 = R_{2,para} - R_{2,dia} = \frac{1}{\Delta T} \ln \frac{I_{dia}(T_b)I_{para}(T_a)}{I_{dia}(T_a)I_{para}(T_b)}$$

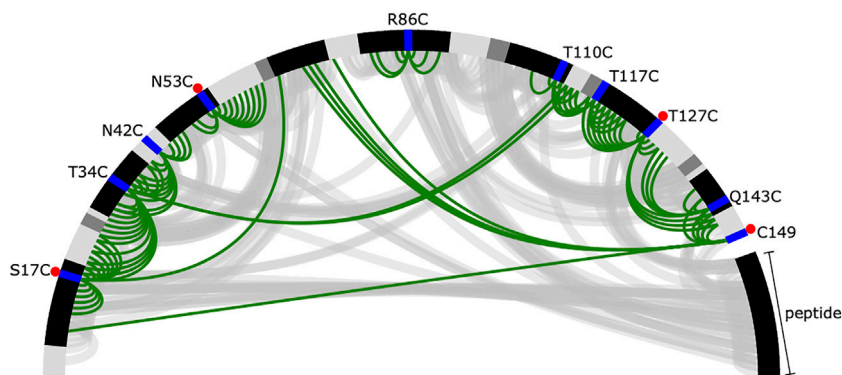


FIGURE 2 | Summary of short distances from the reference crystal structure and inferred from the PRE data. The protein and peptide backbones are represented in a half-circle where the colour depends on the secondary structural element (black—helix, dark grey—extended, light grey—loop). The spin-label locations are shown in blue. A red dot indicates the experiment was performed with labelled peptide. Short distances [(i, j) pairs where $|i - j| > 4$ and $r_{ij}^{Ca} < 7.6 \text{ \AA}$] derived from the crystal structure are shown as grey arcs. Short distances ($r_{ij}^{\text{label-NH}} < 12 \text{ \AA}$) derived from the PRE data are shown in green. Note that C149 is a single-residue extension of calmodulin, which is 148 residues long.

TABLE 1 | Distance bounds for calculated PRE distances, r .

	PRE distance range (Å)	Restraint upper bound (Å)	Restraint lower bound (Å)
Short	$r \leq 12$	17	0.0
Medium	$12 < r < 20$	$r + 5$	$r - 5$
Long	$r \geq 20$	∞	15

Ranges are chosen based on the nature of the PRE and include a $\pm 5 \text{ \AA}$ buffer to account for heterogeneity and flexibility.

where ΔT is a time delay chosen to minimize the error in Γ_2 , and I_{dia} and I_{para} are the peak intensities for the diamagnetic and paramagnetic samples, respectively (Iwahara et al., 2007). In this work, we use $\Delta T = 20 \text{ ms}$.

Incorporating Paramagnetic Relaxation Enhancement Information Into Modeling Employing Limited Data Calculations

The distances derived from PRE data correspond to ensemble averages with an r^{-6} weighting, but in MELD (and most other structure determination software), restraints are applied to single structures rather than ensembles. To account for conformational heterogeneity of both the protein and the flexible spin-label, the PRE-derived distances are turned into flat bottomed harmonic restraints that allow for a range of distances without penalty. This approach is a tradeoff that ensures that individual structures are not erroneously over-restrained but this can allow discrepancies between the measured and modelled ensemble averages. Our aim is to produce an approximately correct ensemble starting from an extended chain. If desired, the resulting ensemble can be further refined using a variety of ensemble approaches (Boomsma et al., 2014; Hummer and Köfinger, 2015; Bonomi et al., 2016; Gaalswyk et al., 2018).

We divided the data into *short*, *medium*, and *long* distances with corresponding upper and lower bounds (Table 1). *Short* and *long* distances are difficult to quantify with precision because the peak is either completely broadened for residues close to the spin-label or

barely changes intensity for those that are far away. These distances are turned into broad restraints that either start from zero or extend to infinity for *short* and *long*, respectively. *Medium* distances correspond to peak intensity changes that can be quantified more precisely and are turned into restraints centered around the predicted value. All distances include a buffer of $\pm 5 \text{ \AA}$ of the measured distance to account for the flexibility of the spin-label and noise in the experimental data (Perez et al., 2019).

Due to noise in the experimental data, partially overlapping peaks, and instantaneous fluctuations in both the protein structure and the position of the spin label, we observed that restraints are sometimes violated even with a $\pm 5 \text{ \AA}$ buffer. To mitigate this issue, we used MELD's unique ability to require that only a certain fraction of the restraints must be satisfied by each structure. We set this *active fraction* to 0.9. Essentially, as long as 90 percent of the restraints are satisfied, the resulting restraint energy will be zero. We treat the remaining restraints as being derived from spurious data, so they are entirely ignored. Every timestep, MELD decides which restraints are active based on the current structure. Further details can be found in the SI and in MacCallum et al. (2015).

In our approach, the various hyperparameters (boundaries between *short/medium/long*, size of buffer, active fraction) are fixed. One potential improvement would be to place a hyperprior on these values and infer them using an extended Bayesian approach like Inferential Structure Determination (Rieping et al., 2005). This would allow the data and physical model to determine the most likely values of these hyperparameters, rather

than requiring their specification *a priori*. As MELD does not currently support inference of hyperparameters, we chose the simpler approach of setting a wide buffer and lower active fraction, which potentially sacrifices a small amount of information.

The spin-label was modeled using virtual sites (Banci et al., 1996) following the approach of Islam and Roux (Islam et al., 2013; Islam and Roux, 2015). These virtual sites represent the spin-label as a non-interacting dummy particle to simplify the simulation without losing relevant information for structural refinement. These simplified dummy nitroxide spin-labels are parameterized to match the spin-labels' 3D spatial distribution and dynamics in all-atom simulations. The virtual sites are non-interacting, allowing us to account for all ten spin labels in a single simulation without the risk of interactions between them.

Secondary structure restraints were derived from TALOS+ (Cornilescu et al., 1998; Shen et al., 2009) and used to restraint MELD simulations as previously described (MacCallum et al., 2015). Our approach works by first breaking the protein into overlapping 5-residue fragments. If 4/5 of the residues in the fragment are predicted to be helical or extended, then the fragment is restrained using a combination of torsion and distance restraints (MacCallum et al., 2015). All secondary structure restraints are then combined into a collection with an active fraction of 0.95, which allows 5 percent of fragments to differ from their predicted secondary structure.

Incorporation of Residual Dipolar Coupling Information Into Modeling Employing Limited Data

The traditional approach to incorporate RDCs into simulations is based on solving for the optimal alignment tensor, which requires solving a system of equations every time step using singular value decomposition (SVD) or related methods, which can be computationally intensive (Losonczi et al., 1999). We found this to be particularly problematic in the GPU-accelerated framework of MELD, where this traditional approach led to a 300 percent increase in run time (data not shown), primarily due to the extreme speed of the rest of the force/energy calculations and the challenge of efficiently parallelizing SVDs for small systems of equations on a GPU. To mitigate this issue, we instead followed the approach in Habeck, Nilges, and Riepling (Habeck et al., 2008), which we implemented using an OpenMM CustomCentroidBondForce (Eastman et al., 2017). In our implementation, the alignment tensor elements are encoded in two non-interacting dummy particles coupled to the rest of the system through an additional energy term. This approach has two benefits. First, it is dramatically faster than the standard approach on GPUs, with negligible cost compared to the calculation of the non-bonded forces. Second, this approach accounts for uncertainty and produces a joint distribution of alignment tensors and structures, providing a Bayesian posterior estimate of the conformational ensemble that better reflects uncertainty. A full explanation of our implementation can be found in the SI. To account for uncertainty in the experimental data and to avoid erroneously over-restraining individual structures to the ensemble average data, we use a flat-bottomed restraint where

the energy is zero if the computed RDC is within 1.5 Hz of the measured value. Another approach that avoids the need to solve for the alignment tensor is given in Camilloni and Vendruscolo (2015).

RESULTS AND DISCUSSION

Interpretation of Modeling Employing Limited Data-Computed Ensembles

The term “ensemble” is highly overloaded in structural biology, with a variety of meanings in different contexts (Bonomi et al., 2017; Andralojc and Ravera, 2018; Gaalswyk et al., 2018). Care must be taken to ensure correct interpretation.

MELD samples from a well-defined conformational ensemble (Gaalswyk et al., 2018), specifically the Bayesian posterior distribution given by Eq. 1. Interpretation of this ensemble is straightforward: it is the statistically consistent posterior distribution inferred from the prior, likelihood, and data. How should one select or report structures from this ensemble? The approach we take here is simply to report all structures, as this fully captures the heterogeneity of the distribution. If there is a limit to the number of structures reported, one simple, correct approach is to select a subset of structures at random. Alternatively, one could cluster the structures and report the cluster medoids and populations along with some measure of the variance of structures within the cluster. A variety of approaches are supported by the PDB-Dev archival system which is being developed for structural models obtained using integrative modeling (Vallat et al., 2018). However, we note that since MELD samples structures with the correct posterior probabilities, it is incorrect to further select structures based on other criteria, such as selecting the lowest energy structures.

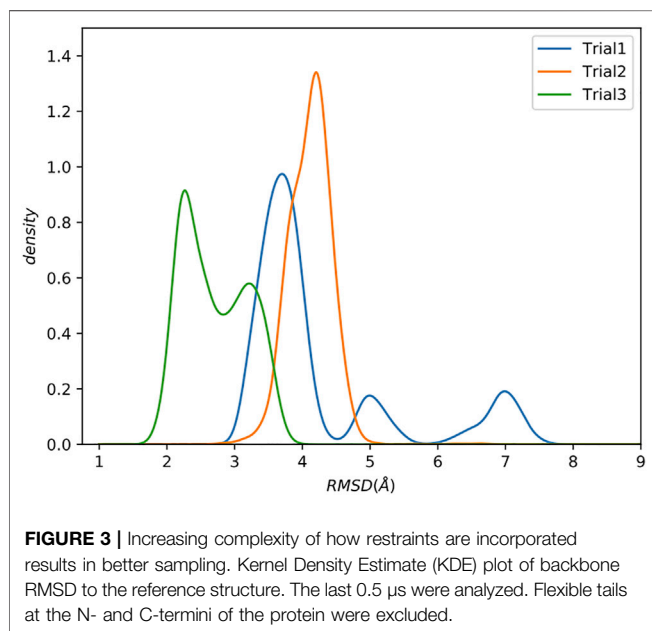
A second consideration in ensemble interpretation is the nature of the likelihood function. The experimental measurements are averages over a thermodynamic ensemble. The most correct modeling approach is to use a likelihood function that considers an entire ensemble of models, ensuring that the predicted average quantities match their corresponding experimental measurements (Bonomi et al., 2017; Andralojc and Ravera, 2018; Gaalswyk et al., 2018). This is an ill-defined inverse problem (Bonomi et al., 2017; Andralojc and Ravera, 2018), so regularization is required, typically in the form of physical modelling and entropy maximization (Bonomi et al., 2017; Andralojc and Ravera, 2018; Gaalswyk et al., 2018). While conceptually appealing, ensemble likelihood methods are complex with high computational requirements. Alternatively, most methods in structural biology, including the MELD approach described here, use single-structure likelihoods (Boomsma et al., 2014). These methods are overly restrictive, as they require each member of the ensemble to be consistent with the data to within some tolerance. In the current approach, we use relatively wide tolerances, but this still does not guarantee that that the computed ensemble accurately models the true distribution.

The primary issue is that for a given set of experimental measurements, there are many possible ensembles that could produce it. The ensemble that MELD generates ensures that each

TABLE 2 | Grouping of restraints for simulations and description of individual trials.

Trial	Number of collections	Description
Trial 1	1	All PRE restraints in a single collection
Trial 2	$3 \times 10 = 30$	PRE restraints are combined by spin-label position and distance (Short, Medium Long) into 30 collections.
Trial 3	$3 \times 10 = 30$	PRE restraints are combined by spin-label position and distance (short, medium long) into 30 collections. RDC restraints are included.
Trial 4	$3 \times 10 = 30$	PRE restraints are combined by spin-label position and distance (short, medium long) into 30 collections. RDC restraints are included. Simulation starts from native crystal structure.

PRE restraints have a force constant of $250 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. RDC restraints have a force constant of $0.5 \text{ kJ mol}^{-1} \text{ Hz}^{-2}$.



structure is in reasonable agreement with the data and allows for a reasonable degree of flexibility. However, the MELD average might not precisely match the experimental measurement due to the use of wide tolerances. Furthermore, the true ensemble could be “broader” than the one generated by MELD—the true ensemble could have many structures that are individually in poor agreement with the data, while still having the same ensemble average, see **Figure 1** of Gaalswyk et al. (2018) for a simple illustration. In the terminology of Bonomi et al. (2017), MELD produces an *uncertainty ensemble*, where heterogeneity in the calculated ensemble could represent true heterogeneity in the system or could simply reflect a lack of data for some part of the protein. The single-structure approach is likely reasonable when the true ensemble has only a modest amount of heterogeneity, e.g., small fluctuations around an average structure, but could be expected to break down for highly heterogeneous systems, e.g., systems containing intrinsically disordered regions.

Although we do not pursue it here, a promising approach would be to use a method like MELD to compute an initial approximate ensemble that could then be used as a starting point for ensemble approaches (Boomsma et al., 2014; Hummer and Köfinger, 2015; Bonomi et al., 2016; Gaalswyk et al., 2018).

The Accuracy of Inference Depends on the Protocol Used

To determine how the experimental data should be incorporated, we performed several simulations varying in their set up (Trial1–Trial4). We explored various ways of combining the restraints into *collections* (**Table 2**). In MELD, at every timestep, the restraints in a collection are sorted by energy, and the *active fraction* with the lowest energy are “active” and contribute their forces and energy to the system, while the remainder are “inactive” and ignored. The division of restraints into collections matters because it determines how MELD decides which restraints are active and which are ignored. For example, Trial1 combines all of the restraints into a single collection. In this case with an active fraction of 0.9, MELD can freely ignore any 10 percent of the restraints, which could be, for example, ignoring one of the ten spin labels entirely. Trial2 separates the restraints by spin-label and into *short*, *medium*, and *long-distance* ranges, resulting in 30 collections. Now MELD can only ignore 10 percent from each spin label/distance combination, while the remaining 90% will be active. Trial3 extends Trial2 by adding the RDC restraints. Trial1–Trail3 start from an extended conformation generated by the tleap tool from the AmberTools suite (Case et al., 2021). Trial4 follows the same protocol as Trial3 but starts from the reference crystal structure as a control.

For each trial, we ran a 2.5 μs replica exchange simulation using 48 replicas. The temperature and the force constant for each restraint collection varied across replicas (see SI for details). The last 0.5 μs of the lowest replica was used for analysis.

Using Only Paramagnetic Relaxation Enhancement-Derived Information Leads to Modest Structural Quality

We compare the trials using kernel density estimation plots (KDE; see Supporting Information for details) of the backbone root mean square deviation (RMSD) to the reference structure [PDB: 1cdl (Meador et al., 1992)], excluding the flexible tails at the N and C terminals of the protein which are not present in the reference (**Figure 3**).

Trial1 is the most straightforward approach and combines all of the data into a single collection. Many of the structures have relatively low RMSD to the reference ($<4\text{Å}$), which is

promising considering the rather limited experimental data, but there are also structures with much higher RMSDs of ~ 5 Å and ~ 7.5 Å RMSD. There are various explanations for these high RMSD conformations, but perhaps the simplest is that this way of grouping all restraints into a single collection allows MELD to ignore short spatial distance restraints that would otherwise eliminate these conformations. As previously stated, the utility of a restraint depends on its spatial distance. Shorter distances provide highly constraining information. However, this highly constraining nature means that these restraints are more difficult to form, leading MELD ignore them in favour of more easily satisfied restraints.

To test this hypothesis, in Trial2, we further subdivided the restraints by separating the *short*, *medium*, and *long* restraints from each dataset into separate collections, resulting in 30 total collections.

The resulting RMSD distribution is centered at a modest RMSD of ~ 4 Å, which is slightly worse than the mode from Trial1. However, this method of combining restraints into collections has wholly eliminated the high RMSD conformations.

Although the RMSDs obtained are only modest (~ 4 Å), these results were obtained with a very sparse dataset with only one spin-label per 17 amino acids. This equates to 6.4 total restraints per residue, and only 0.8 short-distance restraints per residue. For context, NOE-based structures from fully protonated samples typically have >15 NOE restraints per residue, all with short distances.

Based on visual examination, several of our spin-label sites appear to give restraints that are largely uninformative, either because they are far from the remainder of the protein (e.g., R86C) or because they are mostly redundant with other spin-label positions (e.g., T110C), see **Figures 1, 2**. Our results could be improved with a more judicious choice of the 10 label sites, but it is unclear how to do this without pre-existing knowledge of the structure. The results could likely be improved further by adding additional spin-label sites using the calculated structural ensemble to optimize probe location, although we do not pursue this here. Such an iterative strategy could be a viable approach to improve model accuracy but comes at an additional experimental cost. More rigid spin-labels (Fawzi et al., 2011) could also improve results, as MTSL still displays significant conformational heterogeneity that results in less precise distance restraints.

Residual Dipolar Couplings Provide Complementary Information That Improves Accuracy

Despite collecting data for 10 different spin-label sites, few yielded informative short spatial distance, high sequence distance restraints (**Figure 2**), limiting the models' achievable accuracy to relatively modest RMSDs of around 4 Å. Rather than collect additional PRE data, we instead chose to explore the utility of combining PRE information with residual dipolar couplings (RDCs) measured for the amide groups.



FIGURE 4 | Superposition of a typical model (green) from the Trial3 ensemble with the reference structure (white). Peptide is shown in dark green and grey respectively. The superposition was over residues 4–146 of calmodulin plus the peptide. The backbone RMSD of this structure is 2.8 Å, which is near the mean of the ensemble. Structures with RMSDs as low as 1.6 Å are sampled.

Residual dipolar couplings provide information about the orientation of amide NH bonds complementary to the distance information from PRE experiments. In Trial3, we combined PRE information (using the same strategy as Trial2) with RDC data. The inclusion of RDC data led to a substantial improvement in the RMSD (**Figure 3**). The RMSD ranges from approximately 1.6–4.0 Å with an average RMSD of 2.8 Å, including both Calmodulin and the smMLCK peptide (**Figure 4**). This improvement of RMSD upon inclusion of RDC data is consistent with previous studies showing that RDC data provides valuable information on the relative orientation of the two lobes of calmodulin (Mal et al., 2002; Gifford et al., 2011).

RDCs provide information about how the amides are oriented, which, when combined with secondary structure restraints (derived from the measured backbone chemical shifts) and distance restraints (derived from PREs), serves to dramatically limit the possible structures that simultaneously agree with the experimental data and the physical model.

To assess the potential quality of sidechain packing, we examined the single best structure obtained during our simulations, which has an RMSD of 1.6 Å (**Supplementary Figure S1**). For this “best” structure, the RMSD for sidechain heavy atoms is 2.1 Å for all sidechains and 1.4 Å for core sidechains. This is notable as there are no restraints on the side chains themselves, only between the spin labels and the backbone amide protons. This packing phenomenon with MELD has been noted previously and can be attributed to the accuracy of the physical model (Perez et al., 2019). However, we note that the

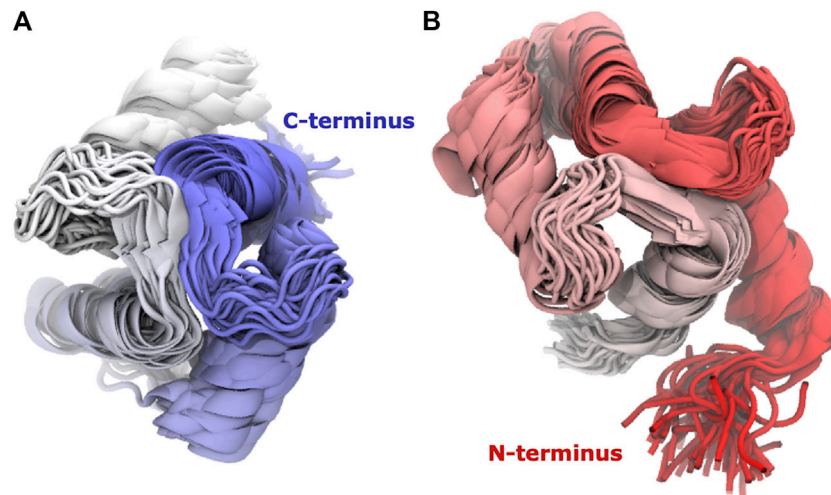


FIGURE 5 | The domains have tightly clustered ensembles. Superpositions of **(A)** C-lobe (residues 82–149) and **(B)** N-lobe (residues 1–76) of Calmodulin for Trial3. Every 100th frame from the last 0.5 microseconds is shown, coloured from N-terminus (red) to C-terminus (blue).

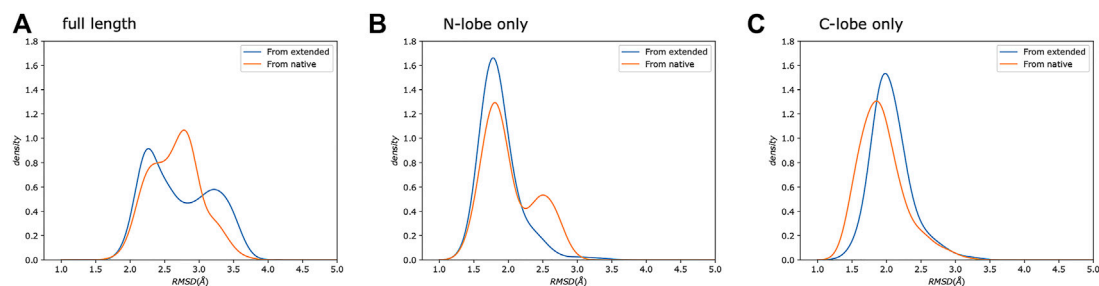


FIGURE 6 | Simulations from extended and native show similar distributions. We show a comparison of the same protocol started from either an extended chain (Trial3, blue) or from the reference crystal structure (Trial4, orange). Each panel shows the backbone RMSD compared to the reference crystal structure. We compare: **(A)** the full-length protein, as well as the **(B)** N-, and **(C)** C-lobes.

sidechain and backbone RMSDs are generally correlated, and this structure has a lower backbone RMSD than average, so the average side chain RMSDs will be higher than these figures.

Despite Limited Data, the Peptide Is Routinely Placed Correctly

As noted previously, the experimental data contained no short distance PREs to the peptide, so placement of the peptide is dependent on a combination of *medium* and *long* restraints with the physical model. Furthermore, only 4 of 10 experiments contained labelled peptide, with the peptide undetected in the remaining experiments. Nevertheless, the combination of available data and the physical model was still able to routinely position the peptide correctly (the peptide is included in the RMSD calculations shown in **Figure 3**). The structure of calmodulin depends on the peptide and its binding (Barbato et al., 1992; Ikura et al., 1992), so correct placement of the peptide is critical.

The Individual Lobes Are Better Defined Than the Complex

Calmodulin consists of two lobes connected by a flexible linker that becomes structured upon peptide binding (Barbato et al., 1992). Examination of each lobe individually shows that our modeled ensembles are tightly clustered (**Figure 5**), indicating that most of the heterogeneity in our calculated ensemble arises from the relative motion of the two lobes. If we consider the RMSD of each lobe to the reference individually, the results are consistently lower than for the whole protein (**Figure 6**). The RMSD of the C-lobe to the reference structure is ~ 2.1 Å (**Figure 6B**), which is consistent with typical RMSDs for small globular proteins seen in MD simulations. The results for the N-lobe are similar. The resulting heterogeneity in the relative orientation of the two domains should be interpreted with caution, due to the use of a single-structure likelihood, as discussed above.

A previous study (Carlon et al., 2019) examining the joint X-ray/NMR refinement of Calmodulin in complex with the

Death-Associated Protein kinase (DAPk) peptide revealed poor agreement between X-ray and NMR data due to large interprotein contacts in the crystal that stabilize a conformation that is in poor agreement with the solution NMR data. The crystal structure of the full-length DAPk protein in complex with Calmodulin lacks these contacts and is in much better agreement with the NMR data. These results highlight the need for caution when comparing structures determined by X-ray crystallography and NMR, particularly in cases where flexibility can be expected.

A study of 109 pairs of NMR and crystal structures (Sikic et al., 2010) showed that typical C α RMSDs range from ~0.5 to 4 Å with a mean of ~2.0 Å when using the DALI (Holm and Sander, 1993) alignment. The typical variability of models within a given NMR ensemble was similar (Sikic et al., 2010). Our results for the individual lobes of calmodulin give similar average RMSDs, indicating that our approach is producing results comparable to typical NMR structures using NOEs and fully protonated samples despite the substantial sparsity in our data. Our results for the full-length complex produce a slightly higher average RMSD, which reflects heterogeneity in the exact relative placement of the two lobes.

To further test our predictions' quality, we also ran calculations using the same protocol as Trial3 but starting from the reference crystal structure rather than from an extended chain (Figure 6), which sets a bound on the possible accuracy that could be obtained. The resulting RMSD distributions are similar to our predictions. This indicates that given: 1) the available experimental data, 2) potential limitations of the physical model used, 3) the use of a single-structure rather than ensemble likelihood, and 4) the challenges of comparing with a static crystal structure, the results obtained using MELD are essentially as good as they could be.

Computational Requirements

Each calculation was over 48 replicas for 2.5 μ s, which required approximately 6 days on 48 GTX 1080Ti GPUs. However, examination of the RMSD over time (Supplementary Figure S2) shows that the simulations appear to be converged after ~500 ns. In hindsight, the simulation length could have been reduced to 1 μ s without a loss in quality, which would reduce simulation time to 2.5 days. While computationally expensive, our approach is readily feasible with access to advanced research computing or cloud computing resources.

CONCLUSION

Our approach can generate accurate protein structures starting from an extended chain using backbone chemical shifts in combination with PRE and RDC measurements from backbone amide labeled samples. We demonstrate this on a relatively large, complex system

with only one spin label per 17 residues. This gives an average of 6.4 PRE restraints per residue of which less than 0.8 per residue are short-distance, compared to the >15 short-distance restraints per residue that are typical in NOE-based structure determination. Our approach is able to routinely identify dominant conformations within 3 Å of the reference crystal structure for calmodulin in complex with a peptide and correctly places the peptide despite a lack of information relating the peptide to the protein. These results approach the quality of gold-standard, fully protonated NMR structures based on NOEs, but were obtained from a far sparser dataset using methods that are more applicable to large proteins. The inclusion of RDCs highlights their value in structure determination with minimal PRE-derived distance restraints. These results showcase the importance of spin label location and the effect it has on the value of the resulting restraints. We show that MELD can accurately account for challenges related to conformational heterogeneity and noise and achieve moderate side chain packing. These results also highlight the capabilities of integrative approaches when experimental information is limited.

DATA AVAILABILITY STATEMENT

Restraint files, run scripts, and analysis scripts can be found on our github repository (https://github.com/maccallumlab/calmodulin_pre_paper) and are archived with Zenodo (DOI: 10.5281/zenodo.5071079).

AUTHOR CONTRIBUTIONS

KG, ZL, HV, and JM contributed to the design of the study, analysis of the data, and wrote the manuscript. ZL performed the NMR experiments. KG performed the MELD calculations.

FUNDING

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, Canada Foundation for Innovation, and Compute Canada. JM is a Canada Research Chair.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.676268/full#supplementary-material>

REFERENCES

- Andralojc, W., and Ravera, E. (2018). "Chapter 4. Treating Biomacromolecular Conformational Variability," in *Paramagnetism in Experimental Biomolecular NMR*, 107–133. doi:10.1039/9781788013291-00107
- Andralojc, W., Berlin, K., Fushman, D., Luchinat, C., Parigi, G., Ravera, E., et al. (2015). Information Content of Long-Range NMR Data for the Characterization of Conformational Heterogeneity. *J. Biomol. NMR* 62 (3), 353–371. doi:10.1007/s10858-015-9951-6
- Andralojc, W., Hiruma, Y., Liu, W.-M., Ravera, E., Nojiri, M., Parigi, G., et al. (2017). Identification of Productive and Futile Encounters in an Electron Transfer Protein Complex. *Proc. Natl. Acad. Sci. USA* 114 (10), E1840–E1847. doi:10.1073/pnas.1616813114
- Andralojc, W., Luchinat, C., Parigi, G., and Ravera, E. (2014). Exploring Regions of Conformational Space Occupied by Two-Domain Proteins. *J. Phys. Chem. B* 118 (36), 10576–10587. doi:10.1021/jp504820w
- Anthis, N. J., Doucleff, M., and Clore, G. M. (2011). Transient, Sparsely Populated Compact States of Apo and Calcium-Loaded Calmodulin Probed by Paramagnetic Relaxation Enhancement: Interplay of Conformational Selection and Induced Fit. *J. Am. Chem. Soc.* 133 (46), 18966–18974. doi:10.1021/ja2082813
- Asciutto, E. K., Dang, M., Pochapsky, S. S., Madura, J. D., and Pochapsky, T. C. (2011). Experimentally Restrained Molecular Dynamics Simulations for Characterizing the Open States of Cytochrome P450cam. *Biochemistry* 50 (10), 1664–1671. doi:10.1021/bi101820d
- Banci, L., Bertini, I., Bren, K. L., Cremonini, M. A., Gray, H. B., Luchinat, C., et al. (1996). The Use of Pseudocontact Shifts to Refine Solution Structures of Paramagnetic Metalloproteins: Met80Ala Cyano-Cytochrome C as an Example. *J. Biol. Inorg. Chem.* 1 (2), 117–126. doi:10.1007/s007750050030
- Banci, L., Bertini, I., Huber, J. G., Luchinat, C., and Rosato, A. (1998). Partial Orientation of Oxidized and Reduced Cytochrome5at High Magnetic Fields: Magnetic Susceptibility Anisotropy Contributions and Consequences for Protein Solution Structure Determination. *J. Am. Chem. Soc.* 120 (49), 12903–12909. doi:10.1021/ja981791w
- Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992). Backbone Dynamics of Calmodulin Studied by Nitrogen-15 Relaxation Using Inverse Detected Two-Dimensional NMR Spectroscopy: The Central Helix Is Flexible. *Biochemistry* 31 (23), 5269–5278. doi:10.1021/bi00138a005
- Barbieri, R., Bertini, I., Cavallaro, G., Lee, Y.-M., Luchinat, C., and Rosato, A. (2002). Paramagnetically Induced Residual Dipolar Couplings for Solution Structure Determination of Lanthanide Binding Proteins. *J. Am. Chem. Soc.* 124 (19), 5581–5587. doi:10.1021/ja025528d
- Battiste, J. L., and Wagner, G. (2000). Utilization of Site-Directed Spin Labeling and High-Resolution Heteronuclear Nuclear Magnetic Resonance for Global Fold Determination of Large Proteins with Limited Nuclear Overhauser Effect Data†. *Biochemistry* 39 (18), 5355–5365. doi:10.1021/bi000060h
- Bellomo, G., Ravera, E., Calderone, V., Botta, M., Fragai, M., Parigi, G., et al. (2021). Revisiting Paramagnetic Relaxation Enhancements in Slowly Rotating Systems: How Long Is the Long Range? *Magn. Reson.* 2 (1), 25–31. doi:10.5194/mr-2-25-2021
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28 (1), 235–242. doi:10.1093/nar/28.1.235
- Bertini, I., Kursula, P., Luchinat, C., Parigi, G., Vahokoski, J., Wilmanns, M., et al. (2009). Accurate Solution Structures of Proteins from X-ray Data and a Minimal Set of NMR Data: Calmodulin–Peptide Complexes as Examples. *J. Am. Chem. Soc.* 131 (14), 5134–5144. doi:10.1021/ja8080764
- Bertini, I., Luchinat, C., Nagulapalli, M., Parigi, G., and Ravera, E. (2012). Paramagnetic Relaxation Enhancement for the Characterization of the Conformational Heterogeneity in Two-Domain Proteins. *Phys. Chem. Chem. Phys.* 14 (25), 9149–9156. doi:10.1039/C2CP40139H
- Bertini, I., Luchinat, C., Parigi, G., and Pierattelli, R. (2005). NMR Spectroscopy of Paramagnetic Metalloproteins. *ChemBioChem* 6 (9), 1536–1549. doi:10.1002/cbic.200500124
- Bonomi, M., Camilloni, C., and Vendruscolo, M. (2016). Metadynamic Metainference: Enhanced Sampling of the Metainference Ensemble Using Metadynamics. *Sci. Rep.* 6, 31232. doi:10.1038/srep31232
- Bonomi, M., Heller, G. T., Camilloni, C., and Vendruscolo, M. (2017). Principles of Protein Structural Ensemble Determination. *Curr. Opin. Struct. Biol.* 42, 106–116. doi:10.1016/j.sbi.2016.12.004
- Boomsma, W., Ferkinghoff-Borg, J., and Lindorff-Larsen, K. (2014). Combining Experiments and Simulations Using the Maximum Entropy Principle. *PLOS Comput. Biol.* 10 (2), e1003406. doi:10.1371/journal.pcbi.1003406
- Bouvignies, G., Markwick, P., Brüschweiler, R., and Blackledge, M. (2006). Simultaneous Determination of Protein Backbone Structure and Dynamics from Residual Dipolar Couplings. *J. Am. Chem. Soc.* 128 (47), 15100–15101. doi:10.1021/ja066704b
- Camilloni, C., and Vendruscolo, M. (2015). A Tensor-free Method for the Structural and Dynamical Refinement of Proteins Using Residual Dipolar Couplings. *J. Phys. Chem. B* 119 (3), 653–661. doi:10.1021/jp5021824
- Carlson, A., Ravera, E., Parigi, G., Murshudov, G. N., and Luchinat, C. (2019). Joint X-Ray/NMR Structure Refinement of Multidomain/Multisubunit Systems. *J. Biomol. NMR* 73 (6), 265–278. doi:10.1007/s10858-018-0212-3
- Case, D. A., Aktulga, H. M., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., et al. (2021). *Amber 2021*. San Francisco: University of California.
- Chou, J. J., Li, S., and Bax, A. (2000). Study of Conformational Rearrangement and Refinement of Structural Homology Models by the Use of Heteronuclear Dipolar Couplings. *J. Biomol. NMR* 18 (3), 217–227. doi:10.1023/A:1026563923774
- Clore, G. M., and Iwahara, J. (2009). Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chem. Rev.* 109 (9), 4108–4139. doi:10.1021/cr900033p
- Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990). Deviations from the Simple Two-Parameter Model-free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins. *J. Am. Chem. Soc.* 112 (12), 4989–4991. doi:10.1021/ja00168a070
- Cornilescu, G., Marquardt, J. L., Ottiger, M., and Bax, A. (1998). Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase. *J. Am. Chem. Soc.* 120 (27), 6836–6837. doi:10.1021/ja9812610
- Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M., and Dobson, C. M. (2005). Mapping Long-Range Interactions in α -Synuclein Using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* 127 (2), 476–477. doi:10.1021/ja044834j
- Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* 6 (3), 277–293. doi:10.1007/BF00197809
- Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., et al. (2017). OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLOS Comput. Biol.* 13 (7), e1005659. doi:10.1371/journal.pcbi.1005659
- Fawzi, N. L., Fleissner, M. R., Anthis, N. J., Kálai, T., Hideg, K., Hubbell, W. L., et al. (2011). A Rigid Disulfide-Linked Nitroxide Side Chain Simplifies the Quantitative Analysis of PRE Data. *J. Biomol. NMR* 51 (1), 105–114. doi:10.1007/s10858-011-9545-x
- Gaalswyk, K., Muniyat, M. I., and MacCallum, J. L. (2018). The Emerging Role of Physical Modeling in the Future of Structure Determination. *Curr. Opin. Struct. Biol.* 49, 145–153. doi:10.1016/j.sbi.2018.03.005
- Gardner, K. H., and Kay, L. E. (1998). The Use of ^2H , ^{13}C , ^{15}N Multidimensional Nmr Gto Study the Structure and Dynamics of Proteins. *Annu. Rev. Biophys. Biomol. Struct.* 27, 357–406. doi:10.1146/annurev.biophys.27.1.357
- Gelis, I., Bonvin, A. M. J. J., Karamisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., et al. (2007). Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR. *Cell* 131 (4), 756–769. doi:10.1016/j.cell.2007.09.039
- Gifford, J. L., Ishida, H., and Vogel, H. J. (2011). Fast Methionine-Based Solution Structure Determination of Calcium-Calmodulin Complexes. *J. Biomol. NMR* 50 (1), 71–81. doi:10.1007/s10858-011-9495-3
- Gochin, M., Zhou, G., and Phillips, A. H. (2011). Paramagnetic Relaxation Assisted Docking of a Small Indole Compound in the HIV-1 Gp41 Hydrophobic Pocket. *ACS Chem. Biol.* 6 (3), 267–274. doi:10.1021/cb100368d
- Goto, N. K., and Kay, L. E. (2000). New Developments in Isotope Labeling Strategies for Protein Solution NMR Spectroscopy. *Curr. Opin. Struct. Biol.* 10 (5), 585–592. doi:10.1016/S0959-440X(00)00135-4

- Gottstein, D., Reckel, S., Dötsch, V., and Güntert, P. (2012). Requirements on Paramagnetic Relaxation Enhancement Data for Membrane Protein Structure Determination by NMR. *Structure* 20 (6), 1019–1027. doi:10.1016/j.str.2012.03.010
- Guerry, P., Salmon, L., Mollica, L., Ortega Roldan, J. L., Markwick, P., van Nuland, N. A. J., et al. (2013). Mapping the Population of Protein Conformational Energy Sub-States from NMR Dipolar Couplings. *Angew. Chem. Int. Ed.* 52 (11), 3181–3185. doi:10.1002/anie.201209669
- Guo, Z., Cascio, D., Hideg, K., and Hubbell, W. L. (2008). Structural Determinants of Nitroxide Motion in Spin-Labeled Proteins: Solvent-Exposed Sites in Helix B of T4 Lysozyme. *Protein Sci.* 17 (2), 228–239. doi:10.1110/ps.073174008
- Habeck, M., Nilges, M., and Rieping, W. (2008). A Unifying Probabilistic Framework for Analyzing Residual Dipolar Couplings. *J. Biomol. NMR* 40 (2), 135–144. doi:10.1007/s10858-007-9215-1
- Higman, V. A., Boyd, J., Smith, L. J., and Redfield, C. (2011). Residual Dipolar Couplings: Are Multiple Independent Alignments Always Possible? *J. Biomol. Nmr* 49 (1), 53–60. doi:10.1007/s10858-010-9457-1
- Holm, L., and Sander, C. (1993). Protein Structure Comparison by Alignment of Distance Matrices. *J. Mol. Biol.* 233 (1), 123–138. doi:10.1006/jmbi.1993.1489
- Huang, H., and Vogel, H. J. (2012). Structural Basis for the Activation of Platelet Integrin $\alpha\text{IIb}\beta\text{3}$ by Calcium- and Integrin-Binding Protein 1. *J. Am. Chem. Soc.* 134 (8), 3864–3872. doi:10.1021/ja2111306
- Hummer, G., and Köfinger, J. (2015). Bayesian Ensemble Refinement by Replica Simulations and Reweighting. *J. Chem. Phys.* 143 (24), 243150. doi:10.1063/1.4937786
- Hwang, P. M., Pan, J. S., and Sykes, B. D. (2014). Targeted Expression, Purification, and Cleavage of Fusion Proteins from Inclusion Bodies in *Escherichia Coli*. *FEBS Lett.* 588 (2), 247–252. doi:10.1016/j.febslet.2013.09.028
- Ikura, M., Clore, G., Gronenborn, A., Zhu, G., Klee, C., and Bax, A. (1992). Solution Structure of a Calmodulin-Target Peptide Complex by Multidimensional NMR. *Science* 256 (5057), 632–638. doi:10.1126/science.1585175
- Ishida, H., Nguyen, L. T., Gopal, R., Aizawa, T., and Vogel, H. J. (2016). Overexpression of Antimicrobial, Anticancer, and Transmembrane Peptides in *Escherichia Coli* through a Calmodulin-Peptide Fusion System. *J. Am. Chem. Soc.* 138 (35), 11318–11326. doi:10.1021/jacs.6b06781
- Ishida, H., and Vogel, H. J. (2010). The Solution Structure of a Plant Calmodulin and the CaM-Binding Domain of the Vacuolar Calcium-ATPase BCA1 Reveals a New Binding and Activation Mechanism. *J. Biol. Chem.* 285 (49), 38502–38510. doi:10.1074/jbc.M110.131201
- Islam, S. M., and Roux, B. (2015). Simulating the Distance Distribution between Spin-Labels Attached to Proteins. *J. Phys. Chem. B* 119 (10), 3901–3911. doi:10.1021/jp510745d
- Islam, S. M., Stein, R. A., Mchaourab, H. S., and Roux, B. (2013). Structural Refinement from Restrained-Ensemble Simulations Based on EPR/DEER Data: Application to T4 Lysozyme. *J. Phys. Chem. B* 117 (17), 4740–4754. doi:10.1021/jp311723a
- Iwahara, J., and Clore, G. M. (2006). Detecting Transient Intermediates in Macromolecular Binding by Paramagnetic NMR. *Nature* 440 (7088), 1227–1230. doi:10.1038/nature04673
- Iwahara, J., Schwieters, C. D., and Clore, G. M. (2004). Ensemble Approach for NMR Structure Refinement against ^1H Paramagnetic Relaxation Enhancement Data Arising from a Flexible Paramagnetic Group Attached to a Macromolecule. *J. Am. Chem. Soc.* 126 (18), 5879–5896. doi:10.1021/ja031580d
- Iwahara, J., Tang, C., and Marius Clore, G. (2007). Practical Aspects of ^1H Transverse Paramagnetic Relaxation Enhancement Measurements on Macromolecules. *J. Magn. Reson.* 184 (2), 185–195. doi:10.1016/j.jmr.2006.10.003
- Jaroniec, C. P., Kaufman, J. D., Stahl, S. J., Viard, M., Blumenthal, R., Wingfield, P. T., et al. (2005). Structure and Dynamics of Micelle-Associated Human Immunodeficiency Virus Gp41 Fusion Domain†. *Biochemistry* 44 (49), 16167–16180. doi:10.1021/bi051672a
- Johnson, B. A., and Blevins, R. A. (1994). NMR View: A Computer Program for the Visualization and Analysis of NMR Data. *J. Biomol. NMR* 4 (5), 603–614. doi:10.1007/BF00404272
- Kainosho, M., and Güntert, P. (2009). SAIL - Stereo-Array Isotope Labeling. *Quart. Rev. Biophys.* 42 (4), 247–300. doi:10.1017/S0033583510000016
- Kay, L. E. (2016). New Views of Functionally Dynamic Proteins by Solution NMR Spectroscopy. *J. Mol. Biol.* 428 (2, Part A), 323–331. doi:10.1016/j.jmb.2015.11.028
- Keizers, P. H. J., and Ubbink, M. (2011). Paramagnetic Tagging for Protein Structure and Dynamics Analysis. *Prog. Nucl. Magn. Reson. Spectrosc.* 58 (1), 88–96. doi:10.1016/j.pnmrs.2010.08.001
- Kim, D. E., DiMaio, F., Yu-Ruei Wang, R., Song, Y., and Baker, D. (2014). One Contact for Every Twelve Residues Allows Robust and Accurate Topology-Level Protein Structure Modeling. *Proteins* 82 (0 2), 208–218. doi:10.1002/prot.24374
- Koehler, J., and Meiler, J. (2011). Expanding the Utility of NMR Restraints with Paramagnetic Compounds: Background and Practical Aspects. *Prog. Nucl. Magn. Reson. Spectrosc.* 59 (4), 360–389. doi:10.1016/j.pnmrs.2011.05.001
- Kuenze, G., Bonneau, R., Leman, J. K., and Meiler, J. (2019). Integrative Protein Modeling in RosettaNMR from Sparse Paramagnetic Restraints. *Structure* 27 (11), 1721–1734.e5. doi:10.1016/j.str.2019.08.012
- Lange, O. F., Rossi, P., Sgourakis, N. G., Song, Y., Lee, H.-W., Aramini, J. M., et al. (2012). Determination of Solution Structures of Proteins up to 40 KDa Using CS-Rosetta with Sparse NMR Data from Deuterated Samples. *Proc. Natl. Acad. Sci.* 109 (27), 10873–10878. doi:10.1073/pnas.1203013109
- Lee, A. L., Sharp, K. A., Kranz, J. K., Song, X.-J., and Wand, A. J. (2002). Temperature Dependence of the Internal Dynamics of a Calmodulin–Peptide Complex. *Biochemistry* 41 (46), 13814–13825. doi:10.1021/bi026380d
- Lipsitz, R. S., and Tjandra, N. (2004). Residual Dipolar Couplings in NMR Structure Analysis. *Annu. Rev. Biophys. Biomol. Struct.* 33, 387–413. doi:10.1146/annurev.biophys.33.110502.140306
- Liu, Z., and Vogel, H. J. (2012). Structural Basis for the Regulation of L-type Voltage-Gated Calcium Channels: Interactions between the N-Terminal Cytoplasmic Domain and Ca^{2+} -Calmodulin. *Front. Mol. Neurosci.* 5. doi:10.3389/fnmol.2012.00038
- Losonczi, J. A., Andrec, M., Fischer, M. W. F., and Prestegard, J. H. (1999). Order Matrix Analysis of Residual Dipolar Couplings Using Singular Value Decomposition. *J. Magn. Reson.* 138 (2), 334–342. doi:10.1006/jmre.1999.1754
- MacCallum, J. L., Perez, A., and Dill, K. A. (2015). Determining Protein Structures by Combining Semireliable Data with Atomistic Physical Models by Bayesian Inference. *Proc. Natl. Acad. Sci. USA* 112 (22), 6985–6990. doi:10.1073/pnas.1506788112
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theor. Comput.* 11 (8), 3696–3713. doi:10.1021/acs.jctc.5b00255
- Mal, T. K., Skrynnikov, N. R., Yap, K. L., Kay, L. E., and Ikura, M. (2002). Detecting Protein Kinase Recognition Modes of Calmodulin by Residual Dipolar Couplings in Solution NMR. *Biochemistry* 41 (43), 12899–12906. doi:10.1021/bi0264162
- Meador, W., Means, A., and Quijcho, F. (1992). Target Enzyme Recognition by Calmodulin: 2.4 Å Structure of a Calmodulin-Peptide Complex. *Science* 257 (5074), 1251–1255. doi:10.1126/science.1519061
- Onufriev, A., Bashford, D., and Case, D. A. (2004). Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* 55 (2), 383–394. doi:10.1002/prot.20033
- Otten, R., Chu, B., Krewulak, K. D., Vogel, H. J., and Mulder, F. A. A. (2010). Comprehensive and Cost-Effective NMR Spectroscopy of Methyl Groups in Large Proteins. *J. Am. Chem. Soc.* 132 (9), 2952–2960. doi:10.1021/ja907706a
- Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement of ^1H and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. *J. Magn. Reson.* 131 (2), 373–378. doi:10.1006/jmre.1998.1361
- PDB Statistics (2021). PDB Statistics: Growth of Structures from NMR Experiments Released Per Year. Available at: <https://www.rcsb.org/stats/growth/growth-nmr> (Accessed February 28, 2021).
- Perez, A., Gaalswyk, K., Jaroniec, C. P., and MacCallum, J. L. (2019). High Accuracy Protein Structures from Minimal Sparse Paramagnetic Solid-State NMR Restraints. *Angew. Chem. Int. Ed.* 58 (20), 6564–6568. doi:10.1002/anie.201811895
- Perez, A., MacCallum, J. L., Brini, E., Simmerling, C., and Dill, K. A. (2015). Grid-Based Backbone Correction to the ff12SB Protein Force Field for Implicit-Solvent Simulations. *J. Chem. Theor. Comput.* 11 (10), 4770–4779. doi:10.1021/acs.jctc.5b00662
- Perez, A., MacCallum, J. L., and Dill, K. A. (2015). Accelerating Molecular Simulations of Proteins Using Bayesian Inference on Weak Information.

- Proc. Natl. Acad. Sci. USA* 112 (38), 11846–11851. doi:10.1073/pnas.1515561112
- Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1997). Attenuated T2 Relaxation by Mutual Cancellation of Dipole-Dipole Coupling and Chemical Shift Anisotropy Indicates an Avenue to NMR Structures of Very Large Biological Macromolecules in Solution. *Proc. Natl. Acad. Sci.* 94 (23), 12366–12371. doi:10.1073/pnas.94.23.12366
- Pilla, K. B., Gaalswyk, K., and MacCallum, J. L. (2017). Molecular Modeling of Biomolecules by Paramagnetic NMR and Computational Hybrid Methods. *Biochim. Biophys. Acta (Bba) - Proteins Proteomics* 1865 (11, Part B), 1654–1663. doi:10.1016/j.bbapap.2017.06.016
- Prestegard, J. H., Agard, D. A., Moremen, K. W., Lavery, L. A., Morris, L. C., and Pederson, K. (2014). Sparse Labeling of Proteins: Structural Characterization from Long Range Constraints. *J. Magn. Reson.* 241, 32–40. doi:10.1016/j.jmr.2013.12.012
- Prestegard, J. H., Bougault, C. M., and Kishore, A. I. (2004). Residual Dipolar Couplings in Structure Determination of Biomolecules. *Chem. Rev.* 104 (8), 3519–3540. doi:10.1021/cr030419i
- Raman, S., Lange, O. F., Rossi, P., Tyka, M., Wang, X., Aramini, J., et al. (2010). NMR Structure Determination for Larger Proteins Using Backbone-Only Data. *Science* 327 (5968), 1014–1018. doi:10.1126/science.1183649
- Rieping, W., Habeck, M., and Nilges, M. (2005). Inferential Structure Determination. *Science* 309 (5732), 303–306. doi:10.1126/science.1110428
- Ruschak, A. M., and Kay, L. E. (2010). Methyl Groups as Probes of Supramolecular Structure, Dynamics and Function. *J. Biomol. NMR* 46 (1), 75–87. doi:10.1007/s10858-009-9376-1
- Ryabov, Y. E., and Fushman, D. (2007). A Model of Interdomain Mobility in a Multidomain Protein. *J. Am. Chem. Soc.* 129 (11), 3315–3327. doi:10.1021/ja067667r
- Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: A Hybrid Method for Predicting Protein Backbone Torsion Angles from NMR Chemical Shifts. *J. Biomol. NMR* 44 (4), 213–223. doi:10.1007/s10858-009-9333-z
- Sikic, K., Tomic, S., and Carugo, O. (2010). Systematic Comparison of Crystal and NMR Protein Structures Deposited in the Protein Data Bank. *Open Biochem. J.* 4, 83–95. doi:10.2174/1874091X01004010083
- Sullivan, D. C., Aynechi, T., Voelz, V. A., and Kuntz, I. D. (2003). Information Content of Molecular Structures. *Biophysical J.* 85 (1), 174–190. doi:10.1016/S0006-3495(03)74464-6
- Sullivan, D. C., and Kuntz, I. D. (2004). Distributions in Protein Conformation Space: Implications for Structure Prediction and Entropy. *Biophysical J.* 87 (1), 113–120. doi:10.1529/biophysj.104.041723
- Tolman, J. R. (2002). A Novel Approach to the Retrieval of Structural and Dynamic Information from Residual Dipolar Couplings Using Several Oriented Media in Biomolecular NMR Spectroscopy. *J. Am. Chem. Soc.* 124 (40), 12020–12030. doi:10.1021/ja0261123
- Tugarinov, V., Hwang, P. M., and Kay, L. E. (2004). Nuclear Magnetic Resonance Spectroscopy of High-Molecular-Weight Proteins. *Annu. Rev. Biochem.* 73, 107–146. doi:10.1146/annurev.biochem.73.011303.074004
- Tugarinov, V., Kanelis, V., and Kay, L. E. (2006). Isotope Labeling Strategies for the Study of High-Molecular-Weight Proteins by Solution NMR Spectroscopy. *Nat. Protoc.* 1 (2), 749–754. doi:10.1038/nprot.2006.101
- Vallat, B., Webb, B., Westbrook, J. D., Sali, A., and Berman, H. M. (2018). Development of a Prototype System for Archiving Integrative/Hybrid Structure Models of Biological Macromolecules. *Structure* 26 (6), 894–904.e2. doi:10.1016/j.str.2018.03.011
- van Dijk, A. D. J., Fushman, D., and Bonvin, A. M. J. J. (2005). Various Strategies of Using Residual Dipolar Couplings in NMR-Driven Protein Docking: Application to Lys48-Linked Di-ubiquitin and Validation against 15N-Relaxation Data. *Proteins* 60 (3), 367–381. doi:10.1002/prot.20476
- Vlasie, M. D., Comuzzi, C., van den Nieuwendijk, A. M. C. H., Prudêncio, M., Overhand, M., and Ubbink, M. (2007). Long-Range-Distance NMR Effects in a Protein Labeled with a Lanthanide-DOTA Chelate. *Chem. Eur. J.* 13 (6), 1715–1723. doi:10.1002/chem.200600916
- Vlasie, M. D., Fernández-Busnadiego, R., Prudêncio, M., and Ubbink, M. (2008). Conformation of Pseudoazurin in the 152 KDa Electron Transfer Complex with Nitrite Reductase Determined by Paramagnetic NMR. *J. Mol. Biol.* 375 (5), 1405–1415. doi:10.1016/j.jmb.2007.11.056
- Ward, A. B., Sali, A., and Wilson, I. A. (2013). Integrative Structural Biology. *Science* 339 (6122), 913–915. doi:10.1126/science.1228565
- Yao, L., Ying, J., and Bax, A. (2009). Improved Accuracy of 15N-1H Scalar and Residual Dipolar Couplings from Gradient-Enhanced IPAP-HSQC Experiments on Protonated Proteins. *J. Biomol. NMR* 43 (3), 161–170. doi:10.1007/s10858-009-9299-x
- Yuan, T., Gomes, A. V., Barnes, J. A., Hunter, H. N., and Vogel, H. J. (2004). Spectroscopic Characterization of the Calmodulin-Binding and Autoinhibitory Domains of Calcium/Calmodulin-dependent Protein Kinase I. *Arch. Biochem. Biophys.* 421 (2), 192–206. doi:10.1016/j.abb.2003.11.012
- Zhang, M., Yuan, T., Aramini, J. M., and Vogel, H. J. (1995). Interaction of Calmodulin with its Binding Domain of Rat Cerebellar Nitric Oxide Synthase. *J. Biol. Chem.* 270 (36), 20901–20907. doi:10.1074/jbc.270.36.20901

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Gaalswyk, Liu, Vogel and MacCallum. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.