Published in final edited form as:

J Biomol NMR. 2019 July ; 73(6-7): 333–346. doi:10.1007/s10858-019-00233-9.

Accuracy and Precision of Protein Structures Determined by Magic Angle Spinning NMR Spectroscopy - for Some 'With a Little Help from a Friend'

Ryan W. Russell^{1,2}, Matthew P. Fritz^{1,2}, Jodi Kraus^{1,2}, Caitlin M. Quinn^{1,2}, Tatyana Polenova^{1,2,*}, Angela M. Gronenborn^{2,3,*}

¹Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, United States;

²Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, 1051 Biomedical Science Tower 3, 3501 Fifth Ave., Pittsburgh, PA 15261, United States;

³Department of Structural Biology, University of Pittsburgh School of Medicine, 3501 Fifth Ave., Pittsburgh, PA 15261, United States

Abstract

We present a systematic investigation into the attainable accuracy and precision of protein structures determined by heteronuclear magic angle spinning solid-state NMR for a set of four proteins of varied size and secondary structure content. Structures were calculated using synthetically generated random sets of C-C distances up to 7 Å at different degrees of completeness. For single-domain proteins, 9–15 restraints per residue are sufficient to derive an accurate model structure, while maximum accuracy and precision are reached with over 15 restraints per residue. For multi-domain proteins and protein assemblies, additional information on domain orientations, quaternary structure and/or protein shape is needed. As demonstrated for the HIV-1 capsid protein assembly, this can be accomplished by integrating MAS NMR with cryoEM maps. In all cases, inclusion of TALOS-derived backbone torsion angles improves the accuracy for small number of restraints, while no further increases are noted for restraint completeness above 40%. In contrast, inclusion of TALOS-derived torsion angle restraints consistently increases the precision of the structural ensemble at all degrees of distance restraint completeness.

INTRODUCTION

The determination of three-dimensional structures of biological macromolecules by NMR relies primarily on distance restraints extracted from transfer of nuclear polarization via dipolar couplings between spin pairs. In solution, the principal experimental parameters are

*Corresponding authors: Angela M. Gronenborn, Department of Structural Biology, University of Pittsburgh School of Medicine, 3501 Fifth Ave., Pittsburgh, PA 15260, USA, Tel.: (412) 648-9959, amg100@pitt.edu; Tatyana Polenova, Department of Chemistry and Biochemistry, University of Delaware, Newark, DE, USA, Tel.: (302) 831-1968, tpolenov@udel.edu.

SUPPORTING INFORMATION AVAILABLE

Summary table of synthetic C-C restraints for different degrees of completeness; Figures with C-C distance restraints mapped onto the structures of proteins for different types of restraints; structure calculation for the HIV-1 CA capsid protein showing structure bundles and precision calculated with respect to the individual domains.

NOE-derived interproton distance (r) restraints, which scale with $1/r^6$, supplemented by torsion angle restraints extracted from J-couplings^{1,2}, ^{13}C and ^1H shifts^{3,4}, residual dipolar couplings^{5–9} as well as the use of conformational database potentials^{10,11}, paramagnetic relaxation enhancements (PRE)^{12–14}, pseudocontact shifts (PCS)¹⁵ and other complementary data. Protein structure determination by magic angle spinning solid-state NMR (MAS NMR) exploits dipolar couplings between heteronuclear spin pairs and involves the use of distance restraints, extracted from carbon-carbon or carbon-nitrogen dipolar-based correlation experiments and their proton-mediated versions, such as proton-driven spin diffusion PDS^{16,17}, dipolar-assisted rotary resonance DARR^{18–20}, combined RN-symmetry driven spin diffusion CORD²¹, CHHC and NHHC²², as well as proton-assisted recoupling approaches for homo- and heteronuclear correlation spectroscopy, like PAR²³ and insensitive nuclei cross polarization PAIN-CP²⁴. Most commonly, the experimental signal intensities of the correlation cross-peaks are measured as a function of mixing time and converted to distance ranges on the basis of peak intensities^{25,26}, similar to the protocols employed for NOE cross-peak intensity-derived distance restraints in solution NMR. In addition, accurate ^{13}C - ^{15}N or ^{13}C - ^{13}C distances can be extracted from REDOR^{27,28}, TEDOR^{29–31}, and RFDR^{32,33} experiments, by measuring the dipolar dephasing or recoupling buildup curves as a function of dephasing/mixing times, and comparison with numerically simulated curves or from universal curves³⁴. Unlike for NOEs, signal intensity scales with $1/r^3$, resulting in a less steep fall-off for longer distances (Figure 1). Similar to solution NMR structure determinations, the distance restraints are frequently supplemented by backbone ϕ and ψ torsion angle restraints from databases of chemical shifts using TALOS^{10,25,26,35}.

The strictly local nature of distance and angular restraints can limit the accuracy of NMR-derived structures, especially for non-globular architectures where the cumulative error may become significant or in cases where only a few contacts are available between structural elements, such as in multi-domain proteins and protein assemblies. In addition, in assemblies (and lattices like crystals), inter-molecular interactions may complicate assignments of cross peaks, although isotopic dilution and differential labeling strategies have proven effective in this regard (reviewed in³⁶). Therefore, for systems of that nature, additional long-range restraints, potentially available from fluorine-fluorine distances^{37–39}, and/or orthogonal information on the overall shape of the molecule, as provided by SAXS experiments for solution studies^{40–43} or cryoEM for both solution and solid state investigations^{44–46}, have to be incorporated.

Here, we performed a systematic investigation of the accuracy and precision attainable in protein structures determined from MAS NMR-derived carbon-carbon distances. To this end, we carried out model calculations for four proteins depicted in Figure 1, i) the CAP-Gly domain of dynactin, an 89-residue protein whose structures, free and bound to several target proteins, have been determined by X-ray crystallography^{47–51}, by solution NMR^{50,51}, and by MAS NMR^{52,53}; ii) the 133 residue agglutinin from *Oscillatoria agardhii* (OAA), whose structure and carbohydrate interactions were studied by solution NMR and crystallography⁵⁴, and for which both solution and MAS NMR resonance assignments are available^{55,56}; iii) the carbohydrate binding domain of galectin-3C (Galectin CBD), which comprises 138 amino acids and for which extensive structural information is available

including solution and MAS NMR chemical shifts⁵⁷ and 48 X-ray crystal structures of various resolutions in the apo and ligand-bound states^{58–65}; and iv) a tubular assembly of the 231 amino acid HIV-1 capsid protein (CA), which has been studied extensively in our laboratories by solution and MAS NMR^{39,66–72}. For all four proteins, high resolution atomic structures are available: CAP-Gly domain (high-resolution MAS NMR structure; PDB: 2MPX⁵²), OAA (1.2 Å resolution; PDB: 3OBL⁵⁴), Galectin CBD (0.86 Å resolution; PDB: 3ZSJ⁶⁵), CA (2.43 Å resolution; PDB: 4XFX⁷³).

Our results establish the criteria for determining accurate protein structures on the basis of the distance restraints that are typically recorded in MAS NMR experiments for ¹³C labeled proteins, following a common protocol for the preparation of isotopically labeled samples using either 1,6-¹³C-glucose or 2-¹³C-glucose as the carbon source. For compact, single domain proteins, at least 3–5 C-C restraints per residue are required to derive the global fold of the molecule, based on distances up to 7 Å, and 15 random restraints per residue are required for attaining maximum accuracy. For multi-domain proteins and protein assemblies, NMR-derived distance restraints have to be combined with additional information on the relative domain orientation, quaternary structure and/or the arrangement of the individual chains in the overall architectures. This can be accomplished using integrated approaches⁴⁴, such as combining MAS NMR with cryoEM as shown here for the HIV capsid.

EXPERIMENTAL

Structure generation

Complete sets of C-C distances of up to 7 Å between unique carbon pairs were generated from the X-ray or NMR structures of the CAP-Gly domain (PDB: 2MPX), OAA (PDB: 3OBL), Galectin CBD (PDB: 3ZSJ), and CA (PDB: 4XFX). Distance sets were generated using scripts in Python 2.7. PDB parsing and iterative distance calculations were performed using Bio-Python⁷⁴ and Bio-PDB⁷⁵ modules.

To generate sets at different degrees of completeness (1%, 3%, 5%, 7%, 10%, 20%, 40%, 60%, 80%), distance entries were randomly removed from each of the above complete sets by shuffling all distances using an arbitrarily chosen 3421 random seed in all calculations, followed by retaining the desired percentage of distances. All distance entries were converted to XPLOR bounds as follows: 1.5–6.5 Å (4.0±2.5 Å) and 2.0–7.2 Å (4.6±2.6 Å) for intra- and inter-residue restraints, respectively.

Structure calculations were performed in XPLOR-NIH version 2.45^{76–78}. Standard XPLOR-NIH terms for bond lengths, bond angles, and impropers were used to enforce correct covalent geometry. The gyration constraint term⁷⁹, an empirical hydrogen-bond database term⁸⁰ and a statistical torsion-angle potential⁸¹ were employed. In calculations that included backbone dihedral restraints, ϕ/ψ angles were predicted using TALOS-N³⁵ from the experimental solid-state ¹³C and ¹⁵N chemical shifts.

Calculations were seeded from extended strands, and 100 structures were generated by torsion angle dynamics with two successive annealing schedules and a final gradient minimization in Cartesian space. The initial annealing calculation was started at 3500 K

with a high temperature dynamics run for 800 ps or 8000 steps, whichever was completed first. The starting time step was 0.001 ps and was self-adjusted in subsequent steps based on conservation of energy. This initial calculation was followed by gradually reducing the temperature to 100 K in steps of 25 K. At each temperature, dynamics was run for 0.4 ps or 200 steps, whichever was completed first, with an initial time step of 0.002 ps. After the initial structure calculation phase, the 10 lowest energy structures were further subjected to a second phase of simulated annealing from 3000 K to 25 K in 12.5 K steps. Force constants for distance restraints were ramped from 10 to 50 kcal/mol/Å² in the initial simulated annealing and from 2 to 30 kcal/mol/Å² in the second phase. In the first annealing run, the dihedral restraint force constants were off during the high temperature dynamics at 3500 K and set to 200 kcal/mol/rad² during cooling. In the second dynamics run, the dihedral restraint force constants were set to 10 kcal/mol/rad² for high temperature dynamics at 3000 K and 200 kcal/mol/rad² during cooling. The gyration volume force constant was scaled from 0.002 to 1 in both runs. The annealed structures were minimized using a 500 step Powell energy minimization and these structures were used in the analysis.

CryoEM and MAS NMR distance-derived HIV capsid model

The lowest-energy structure of the single CA chain calculated at 40% of C-C distance restraint completeness, including backbone dihedral restraints, was used to select individual NTD (residues 1–145) and CTD (residues 148–231) domain structures as inputs for rigid docking into the cryoEM density. Six NTD and six CTD units were docked into the 8 Å cryoEM map of a CA hexamer unit from a tubular assembly, using the “*phenix.dock_in_map*” program in PHENIX 1.14⁸². This resulting structural model was subjected to real-space refinement in PHENIX 1.14, using the built-in “*phenix.real_space_refine*” routine. The refinement included a local grid search, morphing, global minimization, and simulated annealing. Annealing was started at 5000 K and run at this temperature for 0.0025 ps. This initial calculation was followed by gradually reducing the temperature to 300 K in steps of 100 K. At each temperature, dynamics was run for 50 steps, with a time step of 0.0005 ps.

Structure analysis and visualization

The 10 lowest energy structures were best-fit and RMSD values for backbone atoms (N, Cα, C) of each ensemble member with respect to the target structure (the atomic model from which synthetic distance restraints were generated) as well as pairwise RMSD values between the ten models were calculated using routines in XPLOR-NIH 2.45 distribution.

Depiction of the structural ensembles, restraint networks, and structural model of HIV-1 CA hexamer were batch rendered in PyMol⁸³ using automated in-house shell scripts. Secondary structure elements were defined using STRIDE⁸⁴.

RESULTS

In order to evaluate the effects of the number of C-C distance restraints on the accuracy and precision of the resulting simulated annealing ensembles, complete sets of distances between unique carbon pairs were generated from the X-ray or NMR models. The initial restraint set

was restricted to those carbon sites that would be ^{13}C labeled using either 1,6- ^{13}C or 2- ^{13}C glucose, since these two labeling schemes⁸⁵ are commonly employed in structure determinations by MAS NMR^{52,86}. For the CAP-Gly domain, the Galectin CBD, and OAA, this resulted in a total of 2386, 4720 and 6290 restraints, respectively, from both labeling schemes combined, using lower bounds of 1.5 Å and 2 Å and upper bounds of 6.5 and 7.2 Å for intra-residue and inter-residue distances, respectively (Table 1a and Figure 2). For the CA NTD and CTD (monomer unit), the total number of restraints was 6170 and 2560, respectively. Interestingly, there is a significant difference in the number of total distance restraints for OAA, the Galectin CBD and the CA NTD, despite the fact that all three proteins are of similar size (133 aa, 138 aa, and 145 aa). Indeed, for CA NTD the number of medium-range restraints (2166) is approximately the same as the number of long-range restraints (2021). In contrast, for OAA and the Galectin CBD the number of medium-range restraints (626 and 1124, respectively) is only a fraction of the number of long-range restraints (3264 and 2722, respectively). This is a consequence of the difference in overall secondary structure content and packing density for these three proteins, with OAA and the Galectin CBD exhibiting a tightly packed β -sheet structure (54.5% and 53.6% of all residues are in β -strands in OAA and the Galectin CBD, respectively), while the CA NTD exhibits a predominantly helical architecture with 53.1% of all residues in the α -helical ϕ, ψ space. Lowering the upper bounds to 5.5 Å reduces the overall numbers of restraints to approximately half (Table 1b). In addition, we also prepared sets of distances between all possible carbon atoms. Such sets would be potentially available using uniform ^{13}C labeling of the proteins. Naturally these sets are significantly larger, approximately doubling the numbers (increase by a factor of 2 to 2.4; Table 1c and 1d). In reality, however, medium- and long-range distance restraints cannot be efficiently extracted for uniformly- ^{13}C labeled samples because of dipolar truncation⁸⁷ as well as spectral overlap and ambiguities in assignments.

Calculations were carried out for 10 randomly shuffled sets of distance restraints at differing degrees of completeness (1, 3, 5, 7, 10, 20, 40, 60, 80, 100%) in XPLOR-NIH. The procedure employed is outlined in the flow diagram provided in Figure 2. Using structural models for the different proteins, determined by either crystallography or MAS NMR, C-C distances up to 7 Å were extracted, using home-written scripts in Python. The lower and upper distance bounds were set to 1.5 Å to 6.5 Å and 2 Å to 7.2 Å for intra-residue and inter-residue distances, respectively. 100 structures were calculated with these restraint sets, with and without TALOS-N. Of these 100 structures, the 10 lowest energy structures were further subjected to an additional simulated annealing step down to a lower temperature of 25 K and Powell energy minimization. The final ensembles of 10 structures for all the different sets were used in the analysis.

Effect of the number of restraints on the accuracy and precision of the calculated structural models

For the CAP-Gly domain, the accuracy (Figure 3a), measured by the average atomic RMSD for members of the ensemble with respect to the starting/target structure, increases asymptotically for increasing numbers of restraints or restraint completeness, reaching final values of 0.82 ± 0.12 Å and 0.68 ± 0.05 Å without and with TALOS, respectively. For low

restraint completeness, the inclusion of TALOS increases the accuracy somewhat, whereas for restraint completeness of 40% and above, no significant differences are seen for the structures calculated with or without TALOS. In general, it appears that inclusion of TALOS helps with convergence, since the average atomic RMSD for low restraint completeness is approximately 1/2 of the one without TALOS. This difference, however, disappears at restraint completeness above 40%. The precision (Figure 3b), i.e. the average pairwise atomic RMSD for the members in the ensemble, also increases asymptotically for increasing numbers of restraints or restraint completeness, reaching final values of 0.96 ± 0.14 Å and 0.48 ± 0.10 Å without and with TALOS, respectively. In order to pictorially illustrate the increased structural precision with increasing number of restraints, the 10-member ensembles are depicted in Figures 3e and 3f. As can be appreciated, the inclusion of TALOS results in a tighter ensemble throughout all ranges of restraint completeness, possibly artificially restraining the ensemble beyond what is warranted from the paralleling accuracy. In order to illustrate the density of the distance network at the 20% completeness level, the corresponding distances were placed on the ribbon model of the CAP-Gly domain. The density of this network is lower than the one that was used in the experimental MAS NMR structure determination of dynactin's CAP-Gly domain bound to microtubules (Figure 3d)⁵². The MAS NMR structure was calculated using 1183 C-C distances below 7 Å (53% completeness; 13.4 restraints per residue), and 11 C-N distances between 4.2 – 7.1 Å. This resulted in a structural ensemble with a precision of 0.36 Å defined by the average RMSD of the 10-member ensemble with respect to the mean structure, using the nine-step calculation protocol detailed in the manuscript⁵². In order to alleviate any differences that may have arisen because of the different calculation protocols, we re-calculated the CAP-Gly structure using the experimental C-C restraints with the current, above described, two-step protocol, using the same distance bound values as in the original study (Figure 3a and 3b). The precision of the ensemble measured by pairwise backbone atomic RMSD is 2.10 ± 0.24 Å and 0.96 ± 0.22 Å without and with TALOS, respectively. These values are higher than those obtained in the model calculations based on random synthetic restraints, and the tighter experimental bundle may possibly be influenced by the non-random nature of the experimental restraints or overtightening of restraints in the previously used calculation strategy. Indeed, it is gratifying to observe that an equivalent lateral shift of the calculated values for the previously determined experimental CAP-Gly structure occurs on both the accuracy and precision plots (open symbols; Figure 3a and b). This shift positions the experimental structure at 3.4 restraints per residue or 10% restraint completeness value for the random sets. Therefore, the experimental set of restraints (18 per residue) is equivalent to ~4 truly random restraints per residue, highlighting the fact that the experimental distance restraints are non-random. As a result, the quality of the experimental structure is lower than expected for 18 random restraints per residue. Taken together, these results illustrate the consistency of our current analysis.

Similar qualitative behaviors are seen for OAA (Figure 4) and the Galectin CBD (Figure 5). The accuracy for OAA (Figure 4a) reaches final atomic RMSD values of 1.10 ± 0.10 Å for both TALOS off and on at 14 restraints per residue (40% restraint completeness) with a final precision (Figure 4b) plateauing at 1.20 ± 0.10 Å and 0.95 ± 0.11 Å without and with TALOS, respectively. For the Galectin CBD the equivalent values for accuracy are 0.95 ± 0.07 Å

(TALOS off) and 0.88 ± 0.03 Å (TALOS on) and for precision 0.89 ± 0.06 Å and 0.46 ± 0.06 Å without and with TALOS, respectively. Inclusion of sidechains results in only slightly lower accuracy and precision. RMSD values for all heavy atoms are 1.1 ± 0.05 Å (TALOS off) and 1.1 ± 0.03 Å (TALOS on) for accuracy and 1.0 ± 0.07 Å and 0.67 ± 0.07 Å without and with TALOS, respectively, for precision (Figure S5 of the Supporting Information).

As can be easily appreciated, it is possible to derive the global fold for all three proteins with as few as 3–5 restraints per residue or 10% restraint completeness, although the average backbone atomic RMSD values of ensemble members versus the target structure are only ca. 2–4 Å. In practice however, at least 9–15 restraints per residue or 30% completeness should be strived for to result in a reliable model structure. In all cases, the maximum accuracy and precision is reached above 15 restraints per residue or 40% restraint completeness.

We also carried out equivalent calculations for the HIV-1 capsid protein. This is an all-helical protein. Initially, the full-length protein was found to crystallize in a head-to-tail dimer arrangement^{88,89}, while it assembles into tubes in solution. Recently, an X-ray structure of a native wild-type hexamer unit has been solved⁷³, and here we treated the full-length protein as a single chain extracted from this hexamer.

For the entire polypeptide chain of CA, accuracy and precision are much worse than for the other three systems (Figure 6). Final atomic RMSD values with respect to the target structure are 5.90 ± 0.80 Å (TALOS off) and 5.10 ± 1.50 Å (TALOS on) at 39 restraints per residue (100% restraint completeness) and a final precision plateauing at 5.20 ± 1.50 Å and 2.70 ± 0.90 Å without and with TALOS, respectively. Such low accuracy reflects the two-domain nature of CA, for which very few (61) long-range inter-domain C-C distances < 7 Å were available in the synthetic data set. Indeed, as shown in Figure S6 of the Supporting Information, the accuracy calculated with respect to the individual domains, NTD (residues 1–145) and CTD (residues 148–231), is comparable and only slightly worse than for CAP-Gly, OAA and Galectin CBD. For the NTD, the RMSD values are 2.10 ± 0.30 Å and 1.80 ± 0.30 Å without and with TALOS, respectively. The corresponding values for the CTD are 1.80 ± 0.30 Å and 1.2 ± 0.30 Å. A striking difference for the calculations with and without TALOS is noted, with the inclusion of TALOS driving both accuracy and precision up when fewer than 20 restraints per residue or restraint completeness values below 60% are present.

Since only 61 NTD-CTD inter-domain restraints were available, precluding the determination of the relative domain orientation for CA based on C-C distances up to 7 Å, we used the individual NTD and CTD domain structures that were obtained at 40% restraint completeness and docked these into the 8 Å cryoEM map of a CA hexamer unit. As illustrated in Figure 7a, the initial automated global search procedure with 6 NTD and 6 CTD units reliably positioned all 12 domains into the map. The position of one CTD unit was manually adjusted to improve the fit (Figure 7b). The final hybrid NMR/cryoEM atomic model of the CA hexamer is depicted in Figure 7c. The average backbone RMSD of the six CA molecules with respect to the target structure is 1.78 Å. This value is consistent with the differences between the two structures: the target structure corresponds to a flat hexamer, as present in the crystal, while the cryo-EM map is derived from tubular assemblies with a pseudo-hexagonal lattice that possesses curvature.

DISCUSSION

The results presented here provide a benchmark for assessing the limits of accuracy and precision in protein structures determined on the basis of random sets of carbon-carbon distances extracted from commonly used solid state MAS experiments. The current work is the first systematic examination for a number of proteins of varied secondary structure topology. Provided a sufficient number of distance restraints are available, a well-defined polypeptide fold is obtained, irrespective of whether backbone torsion angle restraints derived by TALOS from the experimental ^{13}C and ^{15}N chemical shifts are used. The model calculations for the three single domain proteins, the CAP-Gly domain, the Galectin CBD and OAA all paint a uniform picture: 15–20 distances $\sim 7\text{ \AA}$ yield accurate structures (RMSD values with respect to the target structure of $\sim 1\text{ \AA}$) with pairwise backbone atomic RMSD values of $\sim 1\text{ \AA}$ as well. Like for solution NMR structures, the most critical parameter is the number of interatomic distance restraints (Table 1, Figures 3–9). It appears that fewer restraints are needed to derive similar quality structures in the case of β -sheet proteins, such as the Galectin CBD and OAA, compared to α -helical proteins, such as CA. Inclusion of backbone torsion angle restraints does not improve the accuracy when sufficient distance restraints are present, but increases the precision of the ensembles. The results reported here represent an ideal-case scenario in which a sufficient number of distances $\sim 7\text{ \AA}$ are available (corresponding to unambiguous assignments of correlations), and distance restraints are fully random. As shown for CAP-Gly here, the non-randomness of distance restraints renders both the accuracy and the precision of the calculated structures lower for the same nominal number of restraints. Furthermore, as known from the solution NMR literature, ambiguity or mistakes in resonance assignments or inclusion of incorrect distances also result in lower-quality structures^{90,91}. Finally, if only distance restraints $<5.5\text{ \AA}$ are available, even 18 restraints per residue (40% restraint completeness) are not sufficient to reach the RMSD values of $\sim 1\text{ \AA}$ with respect to the target structure and the accuracies are $1.7\pm 0.1\text{ \AA}$ and $2.2\pm 0.2\text{ \AA}$ with and without TALOS, respectively (Figure S7 of the Supporting Information). Alternatively stated, the quality of the structures with 18 restraints per residue with restraints $<5.5\text{ \AA}$ is equivalent to that with 4.5 restraints per residue with restraints $<7\text{ \AA}$.

The systematic analysis conducted in the present work was performed for sparsely ^{13}C labeled proteins using either 1,6- ^{13}C -glucose or 2- ^{13}C -glucose as the carbon source. This is a common isotopic labeling strategy to overcome dipolar truncation associated with experiments in uniformly labeled samples and to attain high spectral resolution⁸⁶. As we and others have shown, using U- ^{13}C , ^{15}N -labeled proteins and third-spin assisted recoupling experiments^{92,93} additional distance restraints can be extracted. However, these are few in number and the restraint patterns are often difficult to predict. Finally, ^1H -based distance restraints can also be supplemented in protein structure determination, although these require extensive sample deuteration⁹⁴ and/or MAS frequencies above 60 kHz ⁹⁵.

While the results on the single domain proteins are conclusive, multi-domain proteins, such as HIV-1 CA present a challenge. For such cases, additional information is needed, such as low-resolution information on the overall shape of the molecule, accessible by cryoEM⁴⁴. This requirement becomes even more acute for structure determination of supramolecular

assemblies, requiring the knowledge of intermolecular distance restraints, which potentially can be obtained using differentially labeled or isotopically diluted samples (reviewed in³⁶) as well as extensive modeling and/or cryoEM information. The latter has been recently employed in structural studies on several systems⁴⁴. For the CA capsid assembly, the integration of cryoEM data with the solid-state NMR distance restraints is clearly needed for a successful structure determination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIH Grant-P50GM082251, Technology Development Project 2) and is a contribution from the Pittsburgh Center for HIV Protein Interactions. JK is supported by the National Science Foundation Graduate Research Fellowship Program (#1247394). We acknowledge the support of the NSF CHE0959496 grant for acquisition of the 850 MHz NMR spectrometer and of the NIGMS P30GM110758 grant for the support of core instrumentation infrastructure at the University of Delaware. The cryoEM map was kindly provided by Peijun Zhang and Juan Perilla.

REFERENCES

1. Clore GM & Gronenborn AM Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. *Crit. Rev. Biochem. Mol. Biol* 24, 479–564 (1989). [PubMed: 2676353]
2. Wüthrich K *NMR of proteins and nucleic acids*, (Wiley, Chichester, 1986).
3. Kuszewski J, Gronenborn AM & Clore GM The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Ser. B* 107, 293–297 (1995). [PubMed: 7788102]
4. Kuszewski J, Qin J, Gronenborn AM & Clore GM The impact of direct refinement against ¹³Ca and ¹³Cβ chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Ser. B* 106, 92–96 (1995). [PubMed: 7850178]
5. Clore GM & Gronenborn AM New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl. Acad. Sci. USA* 95, 5891–5898 (1998). [PubMed: 9600889]
6. Prestegard J New techniques in structural NMR—anisotropic interactions. *Nat. Struct. Mol. Biol* 5, 517 (1998).
7. Tjandra N & Bax A Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278, 1111–1114 (1997). [PubMed: 9353189]
8. Tjandra N, Grzesiek S & Bax A Magnetic field dependence of nitrogen–proton J splittings in ¹⁵N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J. Am. Chem. Soc* 118, 6264–6272 (1996).
9. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM & Bax A Use of dipolar ¹H–¹⁵N and ¹H–¹³C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Mol. Biol* 4, 732 (1997).
10. Cornilescu G, Delaglio F & Bax A Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* 13, 289–302 (1999). [PubMed: 10212987]
11. Kuszewski J, Gronenborn AM & Clore GM Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. *J. Magn. Reson* 125, 171–177 (1997). [PubMed: 9245376]
12. Battiste JL & Wagner G 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, 5355–5365 (2000). [PubMed: 10820006]

13. Gillespie JR & Shortle D Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol* 268, 170–84 (1997). [PubMed: 9149150]
14. Girvin ME & Fillingame RH Hairpin folding of subunit c of F1Fo ATP synthase: ^1H distance measurements to nitroxide-derivatized aspartyl-61. *Biochemistry* 33, 665–674 (1994). [PubMed: 8292594]
15. Gochin M & Roder H Protein structure refinement based on paramagnetic NMR shifts: Applications to wild-type and mutant forms of cytochrome c. *Protein Sci.* 4, 296–305 (1995). [PubMed: 7757018]
16. Bloembergen N On the interaction of nuclear spins in a crystalline lattice. *Physica* 15, 386–426 (1949).
17. Scholz I, Huber M, Manolikas T, Meier B & Ernst M MIRROR recoupling and its application to spin diffusion under fast magic-angle spinning. in *Chem. Phys. Lett* Vol. 460 278–283 (2008).
18. Morcombe CR, Gaponenko V, Byrd RA & Zilm KW Diluting abundant spins by isotope edited radio frequency field assisted diffusion. in *J. Am. Chem. Soc* Vol. 126 7196–7197 (2004). [PubMed: 15186155]
19. Takegoshi K, Nakamura S & Terao T ^{13}C - ^1H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett* 344, 631–637 (2001).
20. Takegoshi K, Nakamura S & Terao T ^{13}C - ^1H dipolar-driven ^{13}C - ^{13}C recoupling without ^{13}C rf irradiation in nuclear magnetic resonance of rotating solids. *J. Chem. Phys* 118, 2325–2341 (2003).
21. Hou G, Yan S, Trébosc J, Amoureux J-P & Polenova T Broadband homonuclear correlation spectroscopy driven by combined R2nv sequences under fast magic angle spinning for NMR structural analysis of organic and biological solids. *J. Magn. Reson* 232, 18–30 (2013). [PubMed: 23685715]
22. Lange A, Luca S & Baldus M Structural constraints from proton-mediated rare-spin correlation spectroscopy in rotating solids. *J. Am. Chem. Soc* 124, 9704–9705 (2002). [PubMed: 12175218]
23. De Paëpe G, Lewandowski JR, Loquet A, Böckmann A & Griffin RG Proton assisted recoupling and protein structure determination. *J. Chem. Phys* 129, 245101 (2008). [PubMed: 19123534]
24. Lewandowski JR, De Paëpe G & Griffin RG Proton assisted insensitive nuclei cross polarization. *J. Am. Chem. Soc* 129, 728–729 (2007). [PubMed: 17243786]
25. Castellani F et al. Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* 420, 98 (2002). [PubMed: 12422222]
26. Zech SG, Wand AJ & McDermott AE Protein structure determination by high-resolution solid-state NMR spectroscopy: application to microcrystalline ubiquitin. *J. Am. Chem. Soc* 127, 8618–8626 (2005). [PubMed: 15954766]
27. Gullion T Rotational-echo, double-resonance NMR. in *Modern Magnetic Resonance* 713–718 (Springer, 2008).
28. Michal CA & Jelinski LW REDOR 3D: heteronuclear distance measurements in uniformly labeled and natural abundance solids. *J. Am. Chem. Soc* 119, 9059–9060 (1997).
29. Hing AW, Vega S & Schaefer J Transferred-echo double-resonance NMR. *J. Magn. Reson* 96, 205–209 (1992).
30. Jaroniec CP, Filip C & Griffin RG 3D TEDOR NMR experiments for the simultaneous measurement of multiple carbon–nitrogen distances in uniformly ^{13}C , ^{15}N -labeled solids. *J. Am. Chem. Soc* 124, 10728–10742 (2002). [PubMed: 12207528]
31. Nieuwkoop AJ, Wylie BJ, Franks WT, Shah GJ & Rienstra CM Atomic resolution protein structure determination by three-dimensional transferred echo double resonance solid-state nuclear magnetic resonance spectroscopy. *J. Chem. Phys* 131, 095101–095101 (2009). [PubMed: 19739873]
32. Bennett A, Griffin R, Ok J & Vega S Chemical shift correlation spectroscopy in rotating solids: Radio frequency-driven dipolar recoupling and longitudinal exchange. *J. Chem. Phys* 96, 8624–8627 (1992).

33. Ishii Y ^{13}C – ^{13}C dipolar recoupling under very fast magic angle spinning in solid-state nuclear magnetic resonance: Applications to distance measurements, spectral assignments, and high-throughput secondary-structure determination. *J. Chem. Phys.* 114(2001).
34. Grage SL & Watts A Applications of REDOR for distance measurements in biological solids in *Annual Reports on NMR Spectroscopy*, Vol. 60 (ed. Webb GA) 191–228 (Academic Press, 2006).
35. Shen Y & Bax A Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* 56, 227–241 (2013). [PubMed: 23728592]
36. Quinn CM & Polenova T Structural biology of supramolecular assemblies by magic-angle spinning NMR spectroscopy. *Quarterly Reviews of Biophysics* 49, 1–44 (2016).
37. Grage SL, Xu X, Schmitt M, Wadhwani P & Ulrich AS ^{19}F -Labeling of peptides revealing long-range NMR distances in fluid membranes. *J. Phys. Chem. Lett* 5, 4256–4259 (2014). [PubMed: 26273971]
38. Roos M, Wang T, Shcherbakov AA & Hong M Fast magic-angle-spinning ^{19}F spin exchange NMR for determining nanometer ^{19}F – ^{19}F distances in proteins and pharmaceutical compounds. *J. Phys. Chem. B* 122, 2900–2911 (2018). [PubMed: 29486126]
39. Wang M et al. Fast magic-angle spinning ^{19}F NMR spectroscopy of HIV-1 capsid protein assemblies. *Angew. Chem. Int. Ed. Engl* 57, 16375–16379 (2018). [PubMed: 30225969]
40. Wang J et al. Determination of multicomponent protein structures in solution using global orientation and shape restraints. *J. Am. Chem. Soc* 131, 10507–10515 (2009). [PubMed: 19722627]
41. Grishaev A, Wu J, Trewella J & Bax A Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J. Am. Chem. Soc* 127, 16621–16628 (2005). [PubMed: 16305251]
42. Wang I et al. Structure, dynamics and RNA binding of the multi-domain splicing factor TIA-1. *Nucleic Acids Res.* 42, 5949–5966 (2014). [PubMed: 24682828]
43. Debiec KT, Whitley MJ, Koharudin LM, Chong LT & Gronenborn AM Integrating NMR, SAXS, and atomistic simulations: structure and dynamics of a two-domain protein. *Biophys. J* 114, 839–855 (2018). [PubMed: 29490245]
44. Cuniassé P, Tavares P, Orlova EV & Zinn-Justin S Structures of biomolecular complexes by combination of NMR and cryoEM methods. *Curr. Opin. Struct. Biol* 43, 104–113 (2017). [PubMed: 28056362]
45. Demers J-P et al. High-resolution structure of the *Shigella* type-III secretion needle by solid-state NMR and cryo-electron microscopy. *Nat. Commun* 5, 4976 (2014). [PubMed: 25264107]
46. Sborgi L et al. Structure and assembly of the mouse ASC inflammasome by combined NMR spectroscopy and cryo-electron microscopy. *Proc. Natl. Acad. Sci. USA* 112, 13237–13242 (2015). [PubMed: 26464513]
47. Bjelic S et al. Interaction of mammalian end binding proteins with CAP-Gly domains of CLIP-170 and p150(Glued). *J. Struct. Biol* 177, 160–167 (2012). [PubMed: 22119847]
48. Honnappa S et al. Key interaction modes of dynamic plus TIP networks. *Mol. Cell* 23, 663–671 (2006). [PubMed: 16949363]
49. Weisbrich A et al. Structure-function relationship of CAP-Gly domains. *Nat. Struct. Mol. Biol* 14, 959–967 (2007). [PubMed: 17828277]
50. Hayashi I, Plevin MJ & Ikura M CLIP170 autoinhibition mimics intermolecular interactions with p150(Glued) or EB1. *Nat. Struct. Mol. Biol* 14, 980–981 (2007). [PubMed: 17828275]
51. Hayashi I, Wilde A, Mal TK & Ikura M Structural basis for the activation of microtubule assembly by the EB1 and p150(Glued) complex. *Mol. Cell* 19, 449–460 (2005). [PubMed: 16109370]
52. Yan S et al. Atomic-resolution structure of the CAP-Gly domain of dynactin on polymeric microtubules determined by magic angle spinning NMR spectroscopy. *Proc. Natl. Acad. Sci. USA* 112, 14611–6 (2015). [PubMed: 26604305]
53. Yan S et al. Three-dimensional structure of CAP-Gly domain of mammalian dynactin determined by magic angle spinning NMR spectroscopy: conformational plasticity and interactions with end-binding protein EB1. *J. Mol. Biol* 425, 4249–66 (2013). [PubMed: 23648839]
54. Koharudin LM & Gronenborn AM Structural basis of the anti-HIV activity of the cyanobacterial *Oscillatoria Agardhii* agglutinin. *Structure* 19, 1170–81 (2011). [PubMed: 21827952]

55. Carneiro MG, Koharudin LMI, Griesinger C, Gronenborn AM & Lee D 1H, ¹³C and ¹⁵N resonance assignment of the anti-HIV lectin from *Oscillatoria agardhii*. *Biomol. NMR Assign* 9, 317–319 (2015). [PubMed: 25680849]
56. Fritz M et al. Toward closing the gap: quantum mechanical calculations and experimentally measured chemical shifts of a microcrystalline lectin. *J. Phys. Chem. B* 121, 3574–3585 (2017). [PubMed: 28001418]
57. Kraus J et al. Chemical shifts of the carbohydrate binding domain of galectin-3 from magic angle spinning NMR and hybrid quantum mechanics/molecular mechanics calculations. *J. Phys. Chem. B* 122, 2931–2939 (2018). [PubMed: 29498857]
58. Sorme P et al. Structural and thermodynamic studies on cation- π interactions in lectin-ligand complexes: high-affinity galectin-3 inhibitors through fine-tuning of an arginine-arene interaction. *J. Am. Chem. Soc* 127, 1737–43 (2005). [PubMed: 15701008]
59. Atmanene C et al. Biophysical and structural characterization of mono/di-arylated lactosamine derivatives interaction with human galectin-3. *Biochem. Biophys. Res. Commun* 489, 281–286 (2017). [PubMed: 28554839]
60. Bum-Erdene K et al. Investigation into the feasibility of thioditaloside as a novel scaffold for galectin-3-specific inhibitors. *Chembiochem* 14, 1331–42 (2013). [PubMed: 23864426]
61. Collins PM, Oberg CT, Leffler H, Nilsson UJ & Blanchard H Taloside inhibitors of galectin-1 and galectin-3. *Chem. Biol. Drug Des* 79, 339–46 (2012). [PubMed: 22136701]
62. Collins PM, Bum-Erdene K, Yu X & Blanchard H Galectin-3 interactions with glycosphingolipids. *J. Mol. Biol* 426, 1439–51 (2014). [PubMed: 24326249]
63. Hsieh TJ et al. Dual thio-digalactoside-binding modes of human galectins as the structural basis for the design of potent and selective inhibitors. *Sci. Rep* 6, 29457 (2016). [PubMed: 27416897]
64. Rajput VK et al. A selective galactose-coumarin-derived galectin-3 inhibitor demonstrates involvement of galectin-3-glycan interactions in a pulmonary fibrosis model. *J. Med. Chem* 59, 8141–7 (2016). [PubMed: 27500311]
65. Saraboji K et al. The carbohydrate-binding site in galectin-3 is preorganized to recognize a sugarlike framework of oxygens: ultra-high-resolution structures and water dynamics. *Biochemistry* 51, 296–306 (2012). [PubMed: 22111949]
66. Han Y et al. Solid-state NMR studies of HIV-1 capsid protein assemblies. *J. Am. Chem. Soc* 132, 1976–87 (2010). [PubMed: 20092249]
67. Byeon IJ et al. Motions on the millisecond time scale and multiple conformations of HIV-1 capsid protein: implications for structural polymorphism of CA assemblies. *J. Am. Chem. Soc* 134, 6455–66 (2012). [PubMed: 22428579]
68. Han Y et al. Magic angle spinning NMR reveals sequence-dependent structural plasticity, dynamics, and the spacer peptide 1 conformation in HIV-1 capsid protein assemblies. *J. Am. Chem. Soc* 135, 17793–803 (2013). [PubMed: 24164646]
69. Gupta R et al. Dynamic nuclear polarization enhanced MAS NMR spectroscopy for structural analysis of HIV-1 protein assemblies. *J. Phys. Chem. B* 120, 329–39 (2016). [PubMed: 26709853]
70. Liu C et al. Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nat. Commun* 7, 10714 (2016). [PubMed: 26940118]
71. Zhang H et al. HIV-1 capsid function is regulated by dynamics: quantitative atomic-resolution insights by integrating magic-angle-spinning NMR, QM/MM, and MD. *J. Am. Chem. Soc* (2016).
72. Quinn CM et al. Dynamic regulation of HIV-1 capsid interaction with the restriction factor TRIM5 α identified by magic-angle spinning NMR and molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* 115, 11519–11524 (2018). [PubMed: 30333189]
73. Gres AT et al. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* 349, 99–103 (2015). [PubMed: 26044298]
74. Cock PJA et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423 (2009). [PubMed: 19304878]
75. Hamelryck T & Manderick B PDB file parser and structure class implemented in Python. *Bioinformatics* 19, 2308–10 (2003). [PubMed: 14630660]
76. Schwieters CD, Kuszewski JJ, Tjandra N & Clore GM The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson* 160, 65–73 (2003). [PubMed: 12565051]

77. Schwieters CD, Kuszewski JJ & Marius Clore G Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc* 48, 47–62 (2006).
78. Schwieters CD, Bermejo GA & Clore GM Xplor-NIH for molecular structure determination from NMR and other data sources. *Protein Sci* 27, 26–40 (2018). [PubMed: 28766807]
79. Kuszewski J, Gronenborn AM & Clore GM Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J. Am. Chem. Soc* 121, 2337–2338 (1999).
80. Grishaev A & Bax A An empirical backbone–backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation. *J. Am. Chem. Soc* 126, 7281–7292 (2004). [PubMed: 15186165]
81. Bermejo GA, Clore GM & Schwieters CD Smooth statistical torsion angle potential derived from a large conformational database via adaptive kernel density estimation improves the quality of NMR protein structures. *Protein Sci* 21, 1824–1836 (2012). [PubMed: 23011872]
82. Adams PD et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D* 66, 213–221 (2010). [PubMed: 20124702]
83. Schrödinger L The PyMOL molecular graphics system. 2.0 edn.
84. Heinig M & Frishman D STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res* 32, W500–W502 (2004). [PubMed: 15215436]
85. Lundström P et al. Fractional ^{13}C enrichment of isolated carbons using $[1-^{13}\text{C}]$ - or $[2-^{13}\text{C}]$ -glucose facilitates the accurate measurement of dynamics at backbone $\text{C}\alpha$ and side-chain methyl positions in proteins. *J. Biomol. NMR* 38, 199–212 (2007). [PubMed: 17554498]
86. Loquet A, Lv G, Giller K, Becker S & Lange A ^{13}C spin dilution for simplified and complete solid-state NMR resonance assignment of insoluble biological assemblies. *J. Am. Chem. Soc* 133, 4722–5 (2011). [PubMed: 21401039]
87. Bayro MJ et al. Dipolar truncation in magic-angle spinning NMR recoupling experiments. *J. Chem. Phys* 130, 114506–114506 (2009). [PubMed: 19317544]
88. Berthet-Colominas C et al. Head-to-tail dimers and interdomain flexibility revealed by the crystal structure of HIV-1 capsid protein (p24) complexed with a monoclonal antibody Fab. *EMBO J* 18, 1124–1136 (1999). [PubMed: 10064580]
89. Du S et al. Structure of the HIV-1 full-length capsid protein in a conformationally trapped unassembled state induced by small-molecule binding. *J. Mol. Biol* 406, 371–386 (2011). [PubMed: 21146540]
90. Nilges M Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol* 245, 645–60 (1995). [PubMed: 7844833]
91. Spronk C, B. Nabuurs, S, Krieger E, Vriend G & Vuister G Validation of protein structures derived by NMR spectroscopy, 315–337 (2004).
92. Shahid SA et al. Membrane-protein structure determination by solid-state NMR spectroscopy of microcrystals. *Nature methods* 9, 1212–7 (2012). [PubMed: 23142870]
93. Wälti MA et al. Atomic-resolution structure of a disease-relevant $\text{A}\beta(1-42)$ amyloid fibril. *Proc. Natl. Acad. Sci. USA* 113, E4976–E4984 (2016). [PubMed: 27469165]
94. Huber M et al. A proton-detected 4D solid-state NMR experiment for protein structure determination. *ChemPhysChem* 12, 915–918 (2011). [PubMed: 21442705]
95. Andreas LB et al. Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. *Proc. Natl. Acad. Sci. USA* 113, 9187–9192 (2016). [PubMed: 27489348]

Figure 1.

a) Generic polypeptide chain, illustrating select backbone dihedral angles and ^1H - ^1H and ^{13}C - ^{13}C distances. b) Distance dependence of the ^1H - ^1H NOE and ^{13}C - ^{13}C dipolar coupling. The NOE curve was calculated for $\tau_c=7.1$ ns, corresponding to a spherical protein of 14.6 kDa molecular mass at $T = 37^\circ\text{C}$. c) Ribbon representations of dynactin's CAP-Gly domain (PDBID: 2MPX), *Oscillatoria agardii* agglutinin, OAA (PDBID: 3OB2), the carbohydrate binding domain (CBD) of galectin-3C (PDBID: 3ZSJ), and full-length chain of HIV-1 capsid protein (CA) in the assembled state (PDBID: 4XFX).

Figure 2.

Flow diagram of computational strategy for protein structure calculations on the basis of synthetic C-C distance restraints.

Figure 3.

Structure calculation for dynactin's CAP-Gly domain. a) Accuracy as defined by atomic backbone RMSD with respect to the target (input) structure and b) Precision as defined by pairwise atomic backbone RMSD for ensemble members plotted vs. restraint completeness and number of restraints per residue. Data without (black symbols) or with (green symbols) backbone torsion angle restraints from TALOS-N using 200 experimental ^{13}C and 71 ^{15}N chemical shifts (BMRB 25005). The RMSD values for calculations using experimental C-C distance restraints are shown with open symbols. The horizontal dashed lines are the average values of those at 60, 80, and 100% restraint completeness. c) Experimental ^{13}C and ^{15}N chemical shifts used in the structure calculation listed along the amino acid sequence. Secondary structure elements are depicted below the sequence. d) Superposition C-C distances at 20% restraint completeness onto the CAP Gly structure. e) Top: Superposition of the experimental set of C-C distances onto the CAP-Gly structure. Bottom: Best-fit superpositions of the 10 lowest-energy structures calculated on the basis of the experimental distance restraints without (left) and with (right) TALOS-N derived backbone torsion angle restraints. f,g) Best-fit superpositions of the 10 lowest-energy structures calculated for different degrees of restraint completeness without (f) and with (g) TALOS-N derived backbone torsion angle restraints.

Figure 4.

Structure calculation for *Oscillatoria agardii* agglutinin, OAA. a) Accuracy as defined by atomic backbone RMSD with respect to the target (input) structure and b) Precision as defined by pairwise atomic backbone RMSD for ensemble members plotted vs. restraint completeness and number of restraints per residue. Data without (black symbols) or with (green symbols) backbone torsion angle restraints from TALOS-N using 235 experimental ¹³C and 92 ¹⁵N chemical shifts⁵⁶. The dashed lines are the average values of those at 60, 80, and 100% restraint completeness. c) Experimental ¹³C and ¹⁵N chemical shifts used in the structure calculation listed along the amino acid sequence. Secondary structure elements are depicted below the sequence. d) Superposition of C-C distances at 20% restraint completeness onto the OAA structure. e, f) Best-fit superpositions of the 10 lowest-energy structures calculated for different degrees of restraint completeness without (e) and with (f) TALOS-N derived backbone torsion angle restraints.

Figure 5.

Structure calculation for the Galectin CBD. a) Accuracy as defined by atomic backbone RMSD with respect to the target (input) structure and b) Precision as defined by pairwise atomic backbone RMSD for ensemble members plotted vs. restraint completeness and number of restraints per residue. Data without (black symbols) or with (green symbols) backbone torsion angle restraints from TALOS-N using 385 experimental ^{13}C and 132 ^{15}N chemical shifts. The dashed lines are the average values of those at 60, 80, and 100% restraint completeness. c) Experimental ^{13}C and ^{15}N chemical shifts used in the structure calculation listed along the amino acid sequence. Secondary structure elements are depicted below the sequence. d) Superposition of C-C distances at 20% restraint completeness onto the galectin's CBD structure. e, f) Best-fit superpositions of the 10 lowest-energy structures calculated for different degrees of restraint completeness without (e) and with (f) TALOS-N derived backbone torsion angle restraints.

Figure 6.

Structure calculation for the HIV-1 CA capsid protein (CA). a) Accuracy as defined by atomic backbone RMSD with respect to the target (input) structure and b) Precision as defined by pairwise atomic backbone RMSD for ensemble members plotted vs. restraint completeness and number of restraints per residue. Data without (black symbols) or with (green symbols) backbone torsion angle restraints from TALOS-N using 618 experimental ^{13}C and 205 ^{15}N chemical shifts. The dashed lines are the average values of those at 60, 80, and 100% restraint completeness. c) Experimental ^{13}C and ^{15}N chemical shifts used in the structure calculation listed along the amino acid sequence. Secondary structure elements are depicted below the sequence. d) Superposition of the C-C distances at 20% restraint completeness onto the CA structure. e,f) Best-fit superpositions of the 10 lowest-energy structures calculated for different degrees of restraint completeness without (e) and with (f) TALOS-N derived backbone torsion angle restraints.

Figure 7.

Atomic model of a hexamer unit of HIV-1 CA in tubular assemblies generated by combining MAS NMR-derived distances and cryoEM density. The NTD (residues 1–145) and CTD (residues 148–231) models were taken from the lowest-energy structure of a single chain CA calculated at 40% restraint completeness (see Figure 6). a) NTD and CTD domains were fit into the cryoEM map of CA hexamer by automated rigid-body docking. b) The position of one CTD domain was manually adjusted to improve the fit. c) CA hexamer after real-space refinement. NTDs and CTDs are shown in purple and cyan, respectively; the β -hairpin is colored yellow.

Table 1.

Number of C-C distances for the different proteins under study

a. C-C distances up to 7 Å for isotopic labeling using 1,6-¹³C-glucose or 2-¹³C-glucose (total for both labeling schemes combined)						
Restrains	CAP-Gly	OAA	Galectin CBD	CA	CA NTD (1–145)	CA CTD (148–231)
Intraresidue ($ i-j =0$)	309	523	683	1030	693	322
Sequential ($ i-j =1$)	505	849	1219	1935	1290	593
Medium-Range ($1< i-j \leq 4$)	476	626	1124	3222	2166	1066
Long-Range ($ i-j >4$)	1096	2722	3264	2795	2021	579
Total Restraints	2386	4720	6290	8982	6170	2560
b. C-C distances up to 5.5 Å for isotopic labeling using 1,6-¹³C-glucose or 2-¹³C-glucose (total for both labeling schemes combined)						
Restrains	CAP-Gly	OAA	Galectin CBD	CA	CA NTD (1–145)	CA CTD (148–231)
Intraresidue ($ i-j =0$)	303	529	674	1015	681	319
Sequential ($ i-j =1$)	272	485	684	1114	741	347
Medium-Range ($1< i-j \leq 4$)	133	168	338	1122	794	357
Long-Range ($ i-j >4$)	410	1076	1248	1101	812	222
Total Restraints	1118	2258	2944	4352	3028	1245
c. C-C distances up to 7 Å for uniform ¹³C isotopic labeling						
Restrains	CAP-Gly	OAA	Galectin CBD	CA	CA NTD (1–145)	CA CTD (148–231)
Intraresidue ($ i-j =0$)	775	1368	1653	2395	1616	747
Sequential ($ i-j =1$)	1242	2053	2635	3992	2666	1235
Medium-Range ($1< i-j \leq 4$)	1111	1700	2392	6738	4572	2178
Long-Range ($ i-j >4$)	2533	6079	7082	5591	3906	1212
Total Restraints	5661	11200	13762	18716	12760	5372
d. C-C distances up to 5.5 Å for uniform ¹³C isotopic labeling						
Restrains	CAP-Gly	OAA	Galectin CBD	CA	CA NTD (1–145)	CA CTD (148–231)
Intraresidue ($ i-j =0$)	757	1368	1617	2340	1573	733
Sequential ($ i-j =1$)	800	2053	1689	2555	1717	776
Medium-Range ($1< i-j \leq 4$)	350	1700	786	2800	1919	906
Long-Range ($ i-j >4$)	899	6079	2723	2107	1496	454
Total Restraints	2806	11200	6815	9802	6705	2869