Published in final edited form as:

J Phys Chem B. 2018 March 22; 122(11): 2931–2939. doi:10.1021/acs.jpcb.8b00853.

Chemical Shifts of the Carbohydrate Binding Domain of Galectin-3 from Magic Angle Spinning NMR and Hybrid Quantum Mechanics/Molecular Mechanics Calculations

Jodi Kraus^{1,2,#}, Rupal Gupta^{1,#}, Jenna Yehl¹, Manman Lu^{1,2}, David A. Case³, Angela M. Gronenborn^{2,4,*}, Mikael Akke^{5,*}, and Tatyana Polenova^{1,2,*}

¹Department of Chemistry and Biochemistry, University of Delaware, Newark, DE 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 Chemistry and Chemical Biology and BioMaPS Institute, Rutgers University, Piscataway, NJ 08854, United States

⁴Department of Structural Biology, University of Pittsburgh School of Medicine, 3501 Fifth Ave., Pittsburgh, PA 15261, United States

⁵Department of Biophysical Chemistry, Center for Molecular Protein Science, Lund University, P.O. Box 124, SE-22100 Lund, Sweden

Abstract

Magic angle spinning NMR spectroscopy is uniquely suited to probe the structure and dynamics of insoluble proteins and protein assemblies at atomic resolution, with NMR chemical shifts containing rich information about biomolecular structure. Access to this information, however, is problematic since accurate quantum mechanical calculation of chemical shifts in proteins remains challenging, particularly for ¹⁵NH. Here we report on isotropic chemical shift predictions for the carbohydrate recognition domain of microcrystalline galectin-3, obtained from using hybrid quantum mechanics/molecular mechanics (QM/MM) calculations, implemented using an automated fragmentation approach, and using very high resolution (0.86 Å lactose-bound and 1.25 Å apo form) X-ray crystal structures. The resolution of the X-ray crystal structure used as an input into the AF-NMR program did not affect the accuracy of the chemical shift calculations to any significant extent. Excellent agreement between experimental and computed shifts is obtained for

*Corresponding authors: Tatyana Polenova, Department of Chemistry and Biochemistry, University of Delaware, Newark, DE, USA, Tel.: (302) 831-1968; tpolenov@udel.edu; Mikael Akke, Department of Biophysical Chemistry, Center for Molecular Protein Science, Lund University, P.O. Box 124, SE-22100 Lund, Sweden, Tel.: +46 46 222 8247, mikael.akke@bpc.lu.se; 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.

#These authors have contributed equally

SUPPORTING INFORMATION AVAILABLE

MAS NMR vs. solution NMR and QM/MM chemical shifts for the lactose-binding region in apo- and lactose-bound forms of galectin-3C; cluster DFT calculations with and without lactose explicitly included in the binding site; solution NMR chemical shifts for apo- and lactose-bound forms of galectin-3C; a summary of chemical shift differences between apo- and lactose-bound galectin-3C that have been derived from solution NMR as well derived from QM/MM calculations. This information can be found on the internet at <http://pubs.acs.org>.

$^{13}\text{C}^{\alpha}$ while larger scatter is observed for $^{15}\text{N}^{\text{H}}$ chemical shifts, which are influenced to greater extent by electrostatic interactions, hydrogen bonding, and solvation.

TOC Graphic

INTRODUCTION

Isotropic NMR chemical shifts potentially provide an abundance of information on molecular structure, being dependent on numerous factors, such as the local conformational environment, hydrogen bonding networks, and electronic effects. Magic angle spinning (MAS) solid-state NMR spectroscopy is now able to obtain atomic resolution structural information on macromolecular assemblies, such as viral assemblies,¹ cytoskeletal assemblies,² and amyloid fibrils³ due to recent technological advancements.⁴ A major bottleneck for protein structure determination using MAS-NMR is the necessity to achieve site-specific resonance assignments for uniquely identifying distance restraints. In large proteins and protein assemblies, spectral overlap due to the sheer number of chemically distinct nuclei is a confounding problem, and approaches such as differential isotopic enrichment and perdeuteration are typically used to alleviate spectral degeneracy.

The ability to accurately calculate isotropic chemical shifts offers great potential for aiding resonance assignments and refining protein structures. Currently several different approaches for calculating isotropic chemical shifts exist, exhibiting varying degrees of accuracy. However, there is not yet a single, robust protocol that can be routinely used for proteins. Existing isotropic chemical shift prediction methods include sequence- and structure-based approaches. Sequence-based methods^{5, 6} exploit sequence similarity to existing structures in the protein databank (PDB) that is used in conjunction with NMR information in the biological magnetic resonance databank (BMRB). Structure-based methods use structural models to directly calculate isotropic chemical shifts for the atoms in the structure, based on already available shifts for substructures. This method is implemented in SHIFTX, SHIFTS, PROSHIFT, and CamShift.^{7–10} There are also hybrid approaches that incorporate principles from both of these methods to calculate isotropic shifts to a high level of accuracy. Among the most popular is the program SHIFTX2.¹¹ The accuracy of SHIFTX2 predictions can be partially attributed to the amount of structural and magnetic resonance data available in the various databases, which can implicitly account for factors such as the dynamic averaging of chemical shifts. However, this approach fails for

cases where the protein of interest does not possess sequence similarity to protein data available in the PDB and BMRB.

Quantum mechanical calculations are an alternative and promising strategy, which does not rely on any *a priori* knowledge of chemical shifts available in databases. Due to advances in computing power, it is now feasible to use Density Functional Theory (DFT) methods for routine calculations of systems containing up to hundreds of atoms. For a protein, however, which may contain several thousands of atoms, shielding tensor calculations for the entire system by DFT remain impractical due to prohibitively large computational resources required. Hybrid quantum mechanics/molecular mechanics (QM/MM) methods have been developed and they are able to overcome this issue. In these approaches, the protein is partitioned into fragments consisting of 150–200 atoms, accessible for standard DFT calculations. The surrounding ‘buffer’ region is treated as embedded point charges by the QM/MM framework and handled classically to take into account any influences from the remainder of the protein. This method affords an effective compromise with respect to computational power, without sacrificing accuracy. The QM/MM framework was originally developed by Merz and coworkers,¹² and has been incorporated as an automated fragmentation (AF) procedure, which was benchmarked with 20 proteins.¹³ Our group recently used AF-QM/MM to benchmark isotropic chemical shifts in the microcrystalline protein, *Oscillator agardhii* agglutinin (OAA).¹⁴ We were able to calculate $^{13}\text{C}^\alpha$ chemical shifts to high accuracy for OAA, while a larger degree of scatter was noted for $^{15}\text{N}^{\text{H}}$ chemical shifts. For OAA, influences from crystal contacts and loop dynamics were observed, affecting the accuracy of the $^{15}\text{N}^{\text{H}}$ shift calculations.

Automated fragmentation is potentially a promising strategy for the integration into iterative protein structure refinement protocols. While it has been shown previously that experimental isotropic chemical shifts can be used as input in protein structure determination^{15–17} if accurate computation of chemical shifts were available, protein structures could be refined to higher resolution. One successful example in which chemical shift tensors calculated from empirical potential energy surfaces were used for refinement is the high-resolution structure of the immunoglobulin binding domain of protein G.¹⁸

Our objective is to develop a versatile general protocol for generating highly accurate chemical shifts by QM/MM calculations and including them in protein structure refinement. Here, we examined the accuracy of AF-QM/MM chemical shift calculations for the carbohydrate-binding domain, consisting of residues P113–I250, of galectin-3 (referred to as “galectin-3C” in the remainder of the text). Galectin-3C belongs to a class of mammalian galectins, which are involved in important cellular processes, such as cell-cell adhesion, cellular signaling and recognition, and cancer pathology.^{19–21} There is ample structural information^{22–30} for galectin-3C available including 48 X-ray crystal structures of varying resolution in the apo and ligand-bound state, as well as solution NMR chemical shifts. For the computational studies reported below, the highest resolution (0.86 Å) lactose-bound structure (PDB ID: 3ZSJ) was used.³⁰

Experimental $^{13}\text{C}^\alpha$ and $^{15}\text{N}^{\text{H}}$ chemical shifts measured in a series of 2D and 3D MAS NMR experiments on a microcrystalline galectin-3C sample, crystallized under the same

conditions as the X-ray structure.³⁰ In this work, we carried out AF-QM/MM chemical shift calculations and observed a high level of agreement between experimental and predicted $^{13}\text{C}^\alpha$ chemical shifts. The agreement between experimental and predicted $^{15}\text{N}^{\text{H}}$ chemical shifts is also good, albeit with higher scatter, and we have modestly improved upon the benchmarked QM/MM predictions from previous work.¹³ We also explored the influence of the quality of the reference crystal structure on the accuracy of the chemical shift predictions. For this investigation, we chose to compare the 0.86 Å lactose-bound structure, the highest resolution structure and the same crystal form used for MAS NMR measurements, with the 1.25 Å structure of the apo form, which has near-identical backbone conformation with a C^α RMSD of 0.17 Å. Overall, our findings are encouraging, and the integrated MAS NMR – QM/MM approach reported here should be applicable to a wide variety of proteins and protein assemblies.

MATERIALS AND METHODS

Protein expression, purification, and crystallization of Galectin-3C were performed as described in work published previously.^{30, 31} The molecular weight of galectin-3C is 15.7kDa. The purity is >95% as assessed by SDS-PAGE and >98% as assessed by solution NMR. We used a total of 30 mg of galectin-3C microcrystals in a 3.2 mm thin-wall Bruker rotor for MAS NMR experiments.

MAS NMR data acquisition was performed on a 14.1 T narrow bore Bruker AVIII spectrometer equipped with a 3.2 mm HCN EFree MAS probe. Larmor frequencies were 599.8 MHz (^1H), 150.8 MHz (^{13}C), and 60.8 MHz (^{15}N). The MAS frequency was 14 kHz for all experiments, and was controlled to within ± 5 Hz by a Bruker MAS III controller. The temperature of the sample inside the MAS rotor was maintained to within 4 ± 0.1 °C using the Bruker BCU temperature controller, where KBr was used as a temperature sensor. 90° pulse lengths during our experiments were 2.9 μs (^1H), 3.7 μs (^{13}C), and 4.8 μs (^{15}N), and the contact time of ^1H - $^{15}\text{N}/^{13}\text{C}$ cross polarization (CP) was 2.0/1.0 ms. ^1H - $^{15}\text{N}/^{13}\text{C}$ CP used a 95–105% linear amplitude ramp on the ^1H channel with the center Hartmann-Hahn matched to the first spinning side band. The band-selective magnetization transfer from ^{15}N to $^{13}\text{C}^\alpha$ was performed through a 5.0 ms SPECIFIC-CP³² with a tangent amplitude ramp on the ^{15}N channel and a constant rf field on the ^{13}C channel. SPINAL-64 decoupling (48 kHz) was applied during the direct (t_3) and indirect (t_2) acquisition periods. A recycle delay of 2.0 seconds was used. In ^1H - $^{15}\text{N}/^{13}\text{C}$ RN-PARS 3D experiments³³, R12₁⁴-based symmetry sequence was used to recouple the ^1H - $^{15}\text{N}/^{13}\text{C}$ dipolar interaction during t_1 evolution, and the phase-alternated rf field irradiation (84 kHz) was applied on the $^{15}\text{N}/^{13}\text{C}$ channel. ^{13}C and ^{15}N chemical shifts were referenced with respect to the external standards adamantane and NH_4Cl .

NMR data processing was carried out in NMRPipe³⁴; the spectra were analyzed with Sparky.³⁵ In all 2D and 3D datasets, 30°, 45°, or 60° shifted sine bell apodization was followed by Lorentz-to-Gaussian transformation. The R-symmetry ^1H -X dipolar correlation data was processed using the real Fourier transform of the line shape dimension.

Fitting of MAS dipolar lineshapes was performed using SIMPSON³⁶ version 1.1.2. To produce a powder average, 320 pairs of (α , β) angles were generated according to the REPULSION³⁷ algorithm, and 16 γ angles (resulting in a total of 5,120 angle triplets) were used for all simulations. NMR parameters in the experiment matched those used during the fitting routine.

QM/MM calculations of the protein backbone atom isotropic chemical shifts were carried out in Gaussian09 software at the OLYP³⁸/tzvp³⁹ level of theory for the quantum mechanical region. We used the scripts generated in AF-NMR¹³, and referenced PDB ID: 3ZSJ as initial structure. The input coordinates are prepared for Amber minimization in AF-NMR by removing any ligands, removing crystallographic water molecules, and adding H⁺ positions that are not available in the reference structure. The structure was minimized using the Amber FF99SB molecular mechanics force field and referenced to ubiquitin (PDB ID: 1D3Z) calculated at the same level of theory (¹H = 32.0 ppm, ¹³C = 182.5, and ¹⁵N = 237.8 ppm).¹³

RESULTS AND DISCUSSION

Chemical shift assignments of microcrystalline galectin-3C

Galectin-3C is a 15.7 kDa protein that readily crystallizes, and very high-resolution crystal structures have been determined by X-ray and neutron diffraction.^{30,40} It therefore represents an ideal protein for testing the accuracy of chemical shift calculations. The primary sequence and a ribbon representation of the 0.86 Å lactose-bound galectin-3C structure are provided in Figure 1A. The sequence is comprised primarily of anti-parallel β -strands (84 residues) with loop regions (49 residues) and a short 5-residue α -helix. The MAS NMR spectra collected at 14.1 T exhibit impressive sensitivity and resolution, which allowed us to perform resonance assignments for 136/138 residues, using 2D ¹³C-¹³C correlation spectra as well as 3D dipolar based correlation experiments (Figure 1B). Backbone connectivities, extracted from 3D NCACX and NCOCX data sets for the stretch of residues from T133 through V138, are shown in Figure 1C. The only residues without assigned resonances are P113 and L114. No signals were observed in the spectra for these two N-terminal residues of the galectin-3C construct, likely due to mobility in the N-terminal region.

¹³C and ¹⁵N calculated isotropic chemical shifts of galectin-3C: comparison to MAS NMR

We used the AF-QM/MM framework to calculate chemical shifts for each residue in galectin-3C. Examples of the fragments used are shown in Figure 2. Comparisons between the ¹³C $^{\alpha}$ and ¹⁵N^H solid-state chemical shifts with solution NMR shifts, as well as with those predicted by SHIFTX2 and QM/MM are provided in Figure 3. As can be appreciated, the ¹³C $^{\alpha}$ chemical shift predictions exhibit excellent agreement with experiment. The solid-state NMR shifts and the solution NMR shifts also correlate well, except for minor discrepancies for resonances of residues K196, F198, Q220, Y221, N222, and V225. The agreement between the MAS ¹³C $^{\alpha}$ NMR isotropic shifts is high both for SHIFTX2 and QM/MM predictions, albeit the SHIFTX2 calculations are somewhat better. This is not unexpected, considering that SHIFTX2 is based on experimental database information from

the Biological Magnetic Resonance Bank (BMRB), therefore implicitly accounting for dynamic averaging of chemical shifts. We previously observed that integrating MD simulations with QM/MM chemical shift calculations results in greatly improved accuracy for chemical shift predictions of dynamic regions.⁴¹ However, galectin-3C is a rigid system with significant motions only occurring in the ligand-binding pocket (approximately residues 160–200). For these residues, the agreement between experimental and QM/MM calculated chemical shifts is generally high. Therefore, it appears that the differences between QM/MM and experimental shifts for the above six residues are not solely due to dynamic averaging. The accuracy of QM/MM calculations is generally determined by QM fragment size, DFT functional and basis set, long-range electrostatics, and strongly coupled H-bonding networks. The latter two contributions have significant effect on the accuracy of the $^{15}\text{N}^{\text{H}}$ shift predictions. The influence of these factors on the accuracy of QM/MM calculations for galectin is not further discussed here, but will be the subject of future investigations.

When analyzing the $^{15}\text{N}^{\text{H}}$ chemical shift predictions, the agreement is poorer than $^{13}\text{C}^{\alpha}$ with a slope of 0.72 and larger scatter ($r^2 = 0.60$). To accurately predict $^{15}\text{N}^{\text{H}}$ chemical shifts, several additional factors aside from torsion angles and local geometry must be considered. Some of these are hydrogen bonding networks, electrostatics, and solvent effects, see above. To this end it is worth noting that galectin-3C is a predominantly β -sheet protein with extensive hydrogen bonding between the strands. Overall, the current $^{15}\text{N}^{\text{H}}$ chemical shift predictions from QM/MM are encouraging since a modest improvement (from a linear correlation of $R^2 = 0.62$ to 0.68) over the predictions from the benchmarking study of 20 proteins¹³ is noted.

Reference crystal structure quality dependence on the accuracy of QM/MM chemical shift predictions

The influence of the reference X-ray crystal structure on the accuracy of the QM/MM predictions was evaluated, placing emphasis on the $^{13}\text{C}^{\alpha}$ outliers and the $^{15}\text{N}^{\text{H}}$ predictions. For Galectin-3C 48 crystal structures of varying resolution and with a variety of ligands are available in the PDB. We selected the low temperature 0.86 Å lactose-bound structure (PDBID 3ZSJ) and the room temperature 1.25 Å structure of the apo form (PDBID 3ZSM). Although these two protein structures are different with respect to the presence of a ligand, both are part of the same data collection series,³⁰ with the crystals obtained using the same conditions. However, our experimental MAS NMR chemical shifts of lactose-bound galectin-3C were measured at 277 K, a temperature between that of the low temperature and room temperature X-ray structures. This was necessitated by the fact that the protein sample needs to remain stable (without freezing) for long times for sufficient signal acquisition as well as hardware limitations when considering cooling an MAS rotor to 100 K. In the future, it may be instructive to record chemical shifts at 293 K to evaluate whether a higher temperature causes any significant chemical shift changes or notable enhancements in resolution.

For both the $^{13}\text{C}^{\alpha}$ and $^{15}\text{N}^{\text{H}}$ isotropic chemical shifts, no significant differences were observed when used for the QM/MM calculations, despite the difference with respect to

ligand bound (Table 1 and Figure 4), in accord with the very small C^α backbone RMSD of 0.17 Å between the two structures. Furthermore, the C^α and N^H backbone RMSD between the structures following Amber minimization (and prior to fragmentation) was even smaller at 0.15 Å. We analyzed the predictions specifically for the ligand-binding region (where the two structures differ) and, again, did not find a significant difference (Table S1 in the Supporting Information). The current version of the AF-NMR program prepares the input coordinates by first removing any ligands, followed by AMBER minimization. Since removal of the lactose coordinates prior to minimization and fragmentation may influence the degree of agreement between the two reference structures, cluster DFT calculations were carried out for the binding site residues with and without lactose. For ¹³C^α chemical shifts, we observe some interesting differences. Notably, N174 and E184 exhibit chemical shift differences of >1.4 ppm when comparing the calculated chemical shift from AF-NMR (no lactose) and cluster DFT (includes lactose). Additionally, R144, H158, and N160 exhibit chemical shift differences between 0.5 and 0.6 ppm. These results are summarized Figure S1 of the Supporting Information. More extensive analysis will be performed in the future to understand how explicit inclusion of bound ligands affects the chemical shift calculations using AF-NMR.

We also compared solution NMR ¹³C^α and ¹⁵N^H chemical shifts from the apo and lactose-bound galectin-3C (Figures S2–S5 of the Supporting Information). Both sets of shifts exhibit generally tight linear correlations between apo and lactose-bound states (see Figure S2). Nevertheless, as shown in Figure S3 and S4, there are multiple perturbations (>0.02 ppm for ¹H^N, >0.2 ppm for ¹⁵N^H, >0.05 ppm for weighted ¹H-¹⁵N, and >0.1 ppm for ¹³C^α) between the lactose-bound and apo- forms. Not surprisingly, most lactose-bound residues exhibit significant perturbations with a weighted ¹H-¹⁵N value >0.05 ppm. This has been reported previously,³¹ and is summarized in Figure S3 of the Supporting Information. Finally, we used our two sets of QM/MM calculations from the lactose-bound and apo- structures to determine if QM/MM can predict chemical shift perturbations. This is summarized in Figure S5 of the Supporting Information. We observe that the QM/MM calculations do not accurately recapitulate the perturbations in the binding site observed in the solution NMR data. This further suggests that the removal of lactose prior to fragmentation is the likely source of agreement between the two sets of calculated chemical shifts.

It is worth noting that we originally expected that the ¹⁵N^H predictions would depend to a greater extent on the resolution as these are affected by subtle changes in the electronic shielding. We only observed very minor changes when the room temperature structure was used, with a slope going from 0.89 to 0.87, and no detectable change in the scatter. This is perhaps not surprising as the final input structures following Amber minimization in AF-NMR are nearly identical for backbone atoms having a backbone RMSD of 0.15 Å. It is important to note that the resolution in both X-ray structures is remarkably high (higher than that for 99.8% of proteins in the PDB), so this result is overall gratifying.

Influence of nano- to microsecond dynamics on the accuracy of QM/MM chemical shift predictions

We also examined whether dynamics may play a role for the differences observed between experimental and predicted chemical shifts, as seen previously for OAA and HIV-1 CA proteins.^{14, 41} Analyses of ^{15}N - ^1H and ^{13}C - ^1H dipolar lineshapes from MAS NMR, solution NMR order parameters, obtained from ^{15}N relaxation data (R_1 , R_2 , and NOEs), and crystallographic B-factors³¹ were carried out. ^{15}N - ^1H and ^{13}C - ^1H dipolar lineshapes from MAS NMR report on conformational heterogeneity and are indirect reporters of the presence of motions.⁴² Comparing the dipolar order parameters and crystallographic B-factors (Figure 5) reveals, for residues 160 through 200, reduced dipolar order parameters as well as increased crystallographic B-factors. This is the region of the protein containing the lactose binding site. Previous solution NMR results substantiate these observations,³¹ with order parameters derived from solution and solid-state NMR experiments exhibiting excellent agreement (an average difference of only 0.04). This is shown in Fig. 5C, with the most pronounced deviation being at residue I115, near the C-terminus. For the other parts of galectin-3 C, a fairly rigid backbone conformation is observed, with all the dipolar order parameters > 0.75 (except the terminal P113), and closest to 1.0. In addition, all of the crystallographic B-factors are less than 20, indicating a rigid system.

Discussion

Our long-term goal for the work presented here is twofold. First, we aim to develop comprehensive methodology to accurately predict chemical shifts in proteins. Second, we aim to incorporate chemical shift calculations into iterative protein structure refinement protocols. The results of our current study using microcrystalline galectin-3C demonstrate that QM/MM calculations of $^{13}\text{C}^\alpha$ chemical shifts have reached the required level of accuracy, necessary for incorporation into the iterative structure calculation routines. In contrast, accurate calculations of $^{15}\text{N}^\text{H}$ chemical shifts still remain a challenge. We previously conducted an initial investigation for another microcrystalline protein, the β -barrel protein OAA, on the effects of N-H bond length, inclusion of crystallographic water molecules, and the kind of functional/basis set in the DFT calculation. Surprisingly, all of these were found to exert only relatively minor influences, suggesting that dynamics plays a larger role, as we have previously noted for the HIV-1 capsid protein.⁴¹ Even though galectin-3C does not exhibit large-amplitude motions on the nano- to microsecond timescales, dynamic contributions may still come into play and need to be considered for accurate $^{15}\text{N}^\text{H}$ chemical shift predictions. We will therefore pursue such MD/QM/MM calculations of galectin-3C's chemical shifts in the future. It will also be interesting to perform measurements of chemical shift tensors at cryogenic temperatures (110 K), to match the temperature of the low-temperature X-ray structure. While such conditions are currently not easily accessible due to hardware (probe) limitations these experiments will be pursued in the future.

CONCLUSIONS AND FUTURE OUTLOOK

Using hybrid QM/MM calculations, excellent agreement between predicted and experimental $^{13}\text{C}^\alpha$ isotropic chemical shifts for the microcrystalline galectin-3 protein was

found. $^{13}\text{C}^{\alpha}$ chemical shifts mostly depend on local geometry, which is accurately described at the QM/MM level. At the same time, considerable degree of scatter is still observed for $^{15}\text{N}^{\text{H}}$ shifts, most likely due to the intricate dependence of these shifts on H-bonding networks, electrostatics, and solvent environment. Nevertheless, the agreement between $^{15}\text{N}^{\text{H}}$ experiment and calculation has been improved relative to previous benchmarking studies.¹³ Incorporating chemical shift based restraints from quantum chemical based calculations into iterative structure refinement protocols is a promising approach for gaining very high resolution of NMR-derived structures, which will potentially enable the determination of structures of currently intractable non-crystalline macromolecular assemblies by MAS NMR.

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 P30 GM110758-01 grant for the support of core instrumentation infrastructure at the University of Delaware. MA was supported by the Swedish Research Council (2014-5815) and the Knut and Alice Wallenberg Foundation (20013.0022).

References

1. Morag O, Sgourakis NG, Baker D, Goldbourn A. The NMR-Rosetta Capsid Model of M13 Bacteriophage Reveals a Quadrupled Hydrophobic Packing Epitope. *Proc. Natl. Acad. Sci. U. S. A.* 2015; 112:971–976. [PubMed: 25587134]
2. Yan S, Guo CM, Hou GJ, Zhang HL, Lu XY, Williams JC, Polenova T. Atomic-Resolution Structure of the CAP-Gly Domain of Dynactin on Polymeric Microtubules Determined by Magic Angle Spinning NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* 2015; 112:14611–14616. [PubMed: 26604305]
3. Schutz AK, Vagt T, Huber M, Ovchinnikova OY, Cadalbert R, Wall J, Guntert P, Bockmann A, Glockshuber R, Meier BH. Atomic-Resolution Three-Dimensional Structure of Amyloid Beta Fibrils Bearing the Osaka Mutation. *Angew. Chem., Int. Ed.* 2015; 54:331–335.
4. Separovic, F., Naito, A. *Advances in Biological Solid-State NMR: Proteins and Membrane-Active Peptides.* The Royal Society of Chemistry; Cambridge, U.K: 2014.
5. Shen Y, Bax A. Protein Backbone Chemical Shifts Predicted from Searching a Database for Torsion Angle and Sequence Homology. *J. Biomol. NMR.* 2007; 38:289–302. [PubMed: 17610132]
6. Sanz-Hernandez M, De Simone A. The PROSECCO Server for Chemical Shift Predictions in Ordered and Disordered Proteins. *J. of Biomol. NMR.* 2017; 69:147–156. [PubMed: 29119515]
7. Neal S, Nip AM, Zhang HY, Wishart DS. Rapid and Accurate Calculation of Protein H-1, C-13 and N-15 Chemical Shifts. *J. Biomol. NMR.* 2003; 26:215–240. [PubMed: 12766419]
8. Xu XP, Case DA. Automated Prediction of $^{15}\text{N}^{\text{H}}$, $^{13}\text{C}^{\alpha}$, $^{13}\text{C}^{\beta}$ and $^{13}\text{C}'$ Chemical Shifts in Proteins Using a Density Functional Database. *J. Biomol. NMR.* 2001; 21:321–333. [PubMed: 11824752]
9. Meiler J. PROSHIFT: Protein Chemical Shift Prediction Using Artificial Neural Networks. *J. of Biomol. NMR.* 2003; 26:25–37. [PubMed: 12766400]
10. Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M. Fast and Accurate Predictions of Protein NMR Chemical Shifts from Interatomic Distances. *J. Am. Chem. Soc.* 2009; 131:13894–13895. [PubMed: 19739624]

11. Han B, Liu YF, Ginzinger SW, Wishart DS. SHIFTX2: Significantly Improved Protein Chemical Shift Prediction. *J. Biomol. NMR.* 2011; 50:43–57. [PubMed: 21448735]
12. He X, Wang B, Merz KM. Protein NMR Chemical Shift Calculations Based on the Automated Fragmentation QM/MM Approach. *J. Phys. Chem. B.* 2009; 113:10380–10388. [PubMed: 19575540]
13. Swails J, Zhu T, He X, Case DA. AFNMR: Automated Fragmentation Quantum Mechanical Calculation of NMR Chemical Shifts for Biomolecules. *J. Biomol. NMR.* 2015; 63:125–139. [PubMed: 26232926]
14. Fritz M, Quinn CM, Wang MZ, Hou GJ, Lu XG, Koharudin LMI, Polenova T, Gronenborn AM. Toward Closing the Gap: Quantum Mechanical Calculations and Experimentally Measured Chemical Shifts of a Microcrystalline Lectin. *J. Phys. Chem. B.* 2017; 121:3574–3585. [PubMed: 28001418]
15. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. Protein Structure Determination from NMR Chemical Shifts. *Proc. Natl. Acad. Sci. U. S. A.* 2007; 104:9615–9620. [PubMed: 17535901]
16. Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu GH, Eletsky A, Wu YB, Singarapu KK, Lemak A, et al. Consistent Blind Protein Structure Generation from NMR Chemical Shift Data. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:4685–4690. [PubMed: 18326625]
17. Shen Y, Vernon R, Baker D, Bax A. De Novo Protein Structure Generation from Incomplete Chemical Shift Assignments. *J. Biomol. NMR.* 2009; 43:63–78. [PubMed: 19034676]
18. Wylie BJ, Sperling LJ, Nieuwkoop AJ, Franks WT, Oldfield E, Rienstra CM. Ultrahigh Resolution Protein Structures Using NMR Chemical Shift Tensors. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:16974–16979. [PubMed: 21969532]
19. Dennis JW, Lau KS, Demetriou M, Nabi IR. Adaptive Regulation at the Cell Surface by N-Glycosylation. *Traffic.* 2009; 10:1569–1578. [PubMed: 19761541]
20. Lis H, Sharon N. Lectins: Carbohydrate-Specific Proteins that Mediate Cellular Recognition. *Chem. Rev.* 1998; 98:637–674. [PubMed: 11848911]
21. Liu FT, Rabinovich GA. Galectins as Modulators of Tumour Progression. *Nat. Rev. Cancer.* 2005; 5:29–41. [PubMed: 15630413]
22. Sorme P, Arnoux P, Kahl-Knutsson B, Leffler H, Rini JM, Nilsson UJ. 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.* 2005; 127:1737–1743. [PubMed: 15701008]
23. Atmanene C, Ronin C, Teletchea S, Gautier FM, Djedaini-Pilard F, Ciesielski F, Vivat V, Grandjean C. Biophysical and Structural Characterization of Mono/Di-Arylated Lactosamine Derivatives Interaction with Human Galectin-3. *Biochem. Biophys. Res. Commun.* 2017; 489:281–286. [PubMed: 28554839]
24. Bum-Erdene K, Gagarinov IA, Collins PM, Winger M, Pearson AG, Wilson JC, Leffler H, Nilsson UJ, Grice ID, Blanchard H. Investigation into the Feasibility of Thioditaloside as a Novel Scaffold for Galectin-3-Specific Inhibitors. *ChemBioChem.* 2013; 14:1331–1342. [PubMed: 23864426]
25. Collins PM, Oberg CT, Leffler H, Nilsson UJ, Blanchard H. Taloside Inhibitors of Galectin-1 and Galectin-3. *Chem. Biol. Drug Des.* 2012; 79:339–346. [PubMed: 22136701]
26. Collins PM, Bum-Erdene K, Yu X, Blanchard H. Galectin-3 Interactions with Glycosphingolipids. *J. Mol. Biol.* 2014; 426:1439–1451. [PubMed: 24326249]
27. Bian CF, Zhang Y, Sun H, Li DF, Wang DC. Structural Basis for Distinct Binding Properties of the Human Galectins to Thomsen-Friedenreich Antigen. *PLoS One.* 2011; 6:e25007. [PubMed: 21949831]
28. Hsieh TJ, Lin HY, Tu Z, Lin TC, Wu SC, Tseng YY, Liu FT, Hsu STD, Lin CH. Dual Thio-Digalactoside-Binding Modes of Human Galectins as the Structural Basis for the Design of Potent and Selective Inhibitors. *Sci. Rep.* 2016; 6:29457. [PubMed: 27416897]
29. Rajput VK, MacKinnon A, Mandal S, Collins P, Blanchard H, Leffler H, Sethi T, Schambye H, Mukhopadhyay B, Nilsson UJ. A Selective Galactose-Coumarin-Derived Galectin-3 Inhibitor Demonstrates Involvement of Galectin-3-Glycan Interactions in a Pulmonary Fibrosis Model. *J. Med. Chem.* 2016; 59:8141–8147. [PubMed: 27500311]

30. Saraboji K, Hakansson M, Genheden S, Diehl C, Qvist J, Weininger U, Nilsson UJ, Leffler H, Ryde U, Akke M, 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*. 2012; 51:296–306. [PubMed: 22111949]
31. Diehl C, Engstrom O, Delaine T, Hakansson M, Genheden S, Modig K, Leffler H, Ryde U, Nilsson UJ, Akke M. Protein Flexibility and Conformational Entropy in Ligand Design Targeting the Carbohydrate Recognition Domain of Galectin-3. *J. Am. Chem. Soc.* 2010; 132:14577–14589. [PubMed: 20873837]
32. Baldus M, Petkova AT, Herzfeld J, Griffin RG. Cross Polarization in the Tilted Frame: Assignment and Spectral Simplification in Heteronuclear Spin Systems. *Mol. Phys.* 1998; 95:1197–1207.
33. Hou GJ, Lu XY, Vega AJ, Polenova T. Accurate Measurement of Heteronuclear Dipolar Couplings by Phase-Alternating R-Symmetry (PARS) Sequences in Magic Angle Spinning NMR Spectroscopy. *J. Chem. Phys.* 2014; 141:104202. [PubMed: 25217909]
34. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe - A Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR.* 1995; 6:277–293. [PubMed: 8520220]
35. Stevens TJ, Fogh RH, Boucher W, Higman VA, Eisenmenger F, Bardiaux B, van Rossum BJ, Oschkinat H, Laue ED. A Software Framework for Analysing Solid-State MAS NMR Data. *J. Biomol. NMR.* 2011; 51:437–447. [PubMed: 21953355]
36. Bak M, Rasmussen JT, Nielsen NC. SIMPSON: A General Simulation Program for Solid-State NMR Spectroscopy. *J. Magn. Reson.* 2000; 147:296–330. [PubMed: 11097821]
37. Bak M, Nielsen NC. REPULSION, a Novel Approach to Efficient Powder Averaging in Solid-State NMR. *J. Magn. Reson.* 1997; 125:132–139. [PubMed: 9245368]
38. Cohen AJ, Handy NC. Dynamic Correlation. *Mol. Phys.* 2001; 99:607–615.
39. Schafer A, Huber C, Ahlrichs R. Fully Optimized Contracted Gaussian-Basis Sets of Triple Zeta Valence Quality for Atoms Li to Kr. *J. Chem. Phys.* 1994; 100:5829–5835.
40. Manzoni F, Saraboji K, Sprenger J, Kumar R, Noresson AL, Nilsson UJ, Leffler H, Fisher Z, Schrader TE, Ostermann A, et al. perdeuteration, Crystallization, Data Collection and Comparison of Five Neutron Diffraction Data Sets of Complexes of Human Galectin-3C. *Acta Crystallogr., Sect. D: Struct. Biol.* 2016; 72:1194–1202. [PubMed: 27841752]
41. Zhang HL, Hou GJ, Lu MM, Ahn J, Byeon IJL, Langmead CJ, Perilla JR, Hung I, Gor'kov PL, Gan ZH, 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; 138:14066–14075.
42. Yang J, Tasayco ML, Polenova T. Magic Angle Spinning NMR Experiments for Structural Studies of Differentially Enriched Protein Interfaces and Protein Assemblies. *J. Am. Chem. Soc.* 2008; 130:5798–5807. [PubMed: 18393505]

Figure 1.

A: Amino acid sequence (top) and ribbon representation (bottom) of the X-ray structure of lactose-bound galectin-3C (PDB ID: 3ZSJ). **B:** MAS NMR spectra of galectin-3C microcrystals: CORD (top) and NCACX (bottom). Selected assignments are labeled with residue name and number. **C:** Backbone walk for residues T133-V138 using 3D NCACX and NCOCX data sets. The spectra were acquired at 14.1 T; the MAS frequency was 14 kHz for all experiments.

Figure 2.

Galectin-3C QM/MM fragments for residues H158 to E165 (starting at top center fragment and moving clockwise) generated by AF-NMR. Each fragment contains the central residue, along with a corresponding buffer region treated at the DFT level of theory. The remainder of the protein is treated at the MM level, as a series of embedded point charges surrounding the central and buffer regions.

Figure 3.

Top: Correlation plots of $^{13}\text{C}^\alpha$ and $^{15}\text{N}^{\text{H}}$ MAS NMR versus solution NMR chemical shifts, calculated shifts from QM/MM, and shifts predicted from the SHIFTX2 program.⁸ For calculations, the reference X-ray structure determined at 100 K was used (PDB ID: 3ZSJ). Bottom: Difference between QM/MM calculated chemical shifts and MAS NMR chemical shifts, plotted versus the residue number. The secondary structure is shown on the top of the plot. There is no correlation between agreement and secondary structure type.

Figure 4.

Top: Correlation plots of $^{13}\text{C}^\alpha$ and $^{15}\text{N}^{\text{H}}$ MAS NMR and solution NMR chemical shifts, calculated shifts from QM/MM, and shifts predicted from the SHIFTX2 program.⁸ For calculations, the reference X-ray structure determined at 293 K was used (PDB ID: 3ZSM). Bottom: Difference between QM/MM calculated chemical shifts and MAS NMR chemical shifts, plotted versus the residue number. The secondary structure is shown on the top of the plot and loop regions are indicated by shaded areas. There is no correlation between agreement and secondary structure type.

Figure 5.

Nano- to micro-second dynamics in galectin-3C. A: Experimental ^1H - ^{13}C and ^1H - ^{15}N dipolar lineshapes acquired using MAS NMR experiments. B: ^1H - ^{13}C and ^1H - ^{15}N MAS NMR dipolar order parameters (grey), ^1H - ^{15}N solution NMR dipolar order parameters (teal), and crystallographic B-factors, plotted as a function of secondary structure type. C: Differences in dipolar order parameters derived from MAS NMR and solution NMR, with an average difference of 0.04. The most reduced order parameters and highest crystallographic B-factors occur from residues 160 through 200 and represent the ligand binding region of the protein. However, the order parameters are never reduced below a value of 0.75 (with the exception of P113), which is indicative of an overall rigid structure devoid of significant dynamics. D: Comparison of MAS NMR dipolar order parameters (grey) and ^1H - ^{15}N solution NMR order parameters (teal) with the X-ray crystallographic B-factors. Higher NMR order parameters generally correlate with lower crystallographic B-factors.

Table 1

Comparison of QM/MM chemical shift predictions using low temperature (100 K) and room temperature (293 K) reference X-ray crystal structures ^{*}.

	T=100 K	T=293 K
¹³ C ^α MAS vs. QM/MM	$m = 0.84$	$m = 0.85$
	$R^2 = 0.83$	$R^2 = 0.84$
¹⁵ N ^H MAS vs. QM/MM	$m = 0.89$	$m = 0.87$
	$R^2 = 0.68$	$R^2 = 0.65$

^{*} Comparison was performed by linear regression, where m is the slope and R^2 is a measure of the goodness of the fit or degree of correlation between MAS NMR and QM/MM chemical shifts.