COMPUTATIONS OF STRUCTURES OF PROTEIN ASSEMBLIES FROM EXPERIMENTAL MAGIC ANGLE SPINNING NMR RESTRAINTS

by

Ryan W. Russell

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry and Biochemistry

Winter 2023

COMPUTATIONS OF STRUCTURES OF PROTEIN ASSEMBLIES FROM EXPERIMENTAL MAGIC ANGLE SPINNING NMR RESTRAINTS

by

Ryan W. Russell

Approved: _____

Joel Rosenthal, Ph.D.
Chair of the Department of Chemistry and Biochemistry

Approved: _____

John A. Pelesko, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____

Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Tatyana Polenova, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Zhihao Zhuang, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Juan R. Perilla, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

In-Ja Byeon, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Caitlin Quinn, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

Pursuing and accomplishing a Ph.D. has been a tremendous ride with many ups and downs. At the University of Delaware, I was fortunate to have an excellent Ph.D. advisor, Tatyana Polenova, to whom I owe a great debt of gratitude; thank you. With her guidance I was able to find a way to apply two of my favorite research topics, experimental NMR and protein structure calculations. Tatyana helped me tremendously with professional guidance, personal guidance, and made my Ph.D. studies not only extremely productive but also enjoyable and rewarding.

I would like to thank the entire Polenova Group. In particular, Matthew Fritz who has not only been a key research mentor but also an excellent friend of many years. I would like to thank Manman Lu, Sucharita Sakar, Somayeh Zeinalilathori, and Chunting Zhang for working with me on Xplor-NIH calculations and for personal guidance. My thanks extend to the entire group and former members for seamless support: Roza, Brent, Roman, Mingzhang, Jodi, Jenna, and Vera to name a few.

I am grateful for the support throughout the department including Caitlin Quinn of the NMR Facility for countless help and feedback for my work and the Perilla Laboratory for hosting me as a visiting investigator where I learned so much about coding, analysis, and computational biology. None of this work would have been possible if it were not for Angela Gronenborn. Angela, thank you so much for being the motivation and driving force behind the computational techniques and hosting

numerous productive discussions. I would also like to thank Charles Schwieters for his invaluable technical help and guidance with structure calculations; my work would not have been possible without Charles. Thank you.

I would like to thank my family and close friends dearly. To my mother Cynthia, thank you so much for the unmatched love and support. To my father Ralph for your love and support. To Stan for his guidance and advice throughout the years. To my girlfriend and partner Chelsea for her support and unmatched love, especially in the final stretch.

I am grateful for my friends and my community in Wilmington. I have made countless close friends. I would like to thank my best friends Raymond, Courtney, Daniel, and Joshua. Thank you to city councilman Nathan Field who not only became a close friend but also introduced me to so many people and organizations. I am so grateful for the community that I call home in Trolley Square Wilmington which brought me so much personal joy as I pursued my Ph.D. degree.

I am very grateful for my Chateaubriand Fellowship from the Embassy of France and the invitation from the laboratory of Guido Pintacuda in Lyon, France. Thank you for hosting me and teaching me invaluable science. Thanks to my dear friends I made there: Ribal, Kevin, Adrian, Arthur, George, and Alicia.

Finally, thank you to the Rienstra Group for sparking my interest in NMR and for working with me as an undergraduate researcher at the University of Illinois. Special thanks to Chad, Marcus, Joseph, Kristin, and Deb.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
AIDS	Acquired Immunodeficiency Syndrome
BMRB	Biological Magnetic Resonance Data Bank
BVM	Bevirimat
CA	Capsid Protein
CCF	Cross-Correlation Function
CORD	Combined R2-Driven
COVID-19	CoronaVirus Disease 2019
CP	Cross Polarization
CPMAS	Cross Polarization Magic Angle Spinning
cryo-EM	Cryo-Electron Microscopy
cryo-ET	Cryo-Electron Tomography
CSI	Chemical Shift Index
CTD	C-Terminal Domain
DANGLE	Dihedral Angles from Global Likelihood Estimates
DARR	Dipolar Assisted Rotational Resonance
EEFx	Effective Energy Function for XPLOR-NIH
EM	Electron Microscopy

EMD	Electron Microscopy Data Resource
FL	Full Length
FRET	Fluorescence resonance energy transfer
HETCOR	heteronuclear correlation
HIV-1	Human Immunodeficiency Virus Type 1
IP6	Inositol hexakisphosphate
KIF	kinesin superfamily
LSF	Least Squares Function
MA	Matrix
MAS	Magic Angle Spinning
MASNMR	Magic Angle Spinning Nuclear Magnetic Resonance
MD	Molecular Dynamics
MDSA	Molecular Dynamics Simulated Annealing
MI	Maturation Inhibitor
MT	Microtubules
NC	Nucleocapsid
N	Nucleocapsid Protein
NIH	National Institutes of Health
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NTD	N-Terminal Domain

OAA	Oscillatoria agardhii
PAIN-CP	Proton-assisted insensitive nuclei cross polarization
PAR	Proton-assisted recoupling
PCS	Pseudocontact shifts
PDB	Protein Data Bank
PDSD	Proton-Driven Spin Diffusion
PHENIX	Python-based Hierarchical Environment for Integrated Xtallography
PRE	Paramagnetic relaxation enhancements
PSF	Protein Structure File
QM	Quantum mechanical
REDOR	Rotational-Echo Double-Resonance
REPEL	REPulsive PurELy
RFDR	Radiofrequency driven dipolar recoupling
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
SA	Simulated Annealing
SARS	Severe acute respiratory syndrome
SAXS	Small-angle X-ray scattering
SP	Spacer Peptide
STRIDE	Secondary Structural Identification

TALOS	Torsion Angle Likeliness Obtained from Shift and Sequence Similarity
TEDOR	transferred echo double resonance
UCSF	University of California San Fransisco
UNAIDS	Joint United Nations Programme on HIV/AIDS
XPLOR	eXPLORation of conformational space of macromolecules

ABSTRACT

This dissertation concerns with computational aspects of protein structure determination from experimental magic angle spinning nuclear magnetic resonance (MAS NMR) data and by integrating MAS NMR experimental restraints with information obtained by other structural biology techniques, such as cryogenic electron microscopy (cryo-EM) and X-ray crystallography.

In Chapter 1, protein structure calculation approaches are introduced.

Although the framework for NMR protein structure determination has existed for quite some time, the general requirements for obtaining accurate and precise structures, particularly in the solid state, have not been established until recently. Therefore, we have performed a systematic model study to quantify accuracy and precision with varying numbers of distance restraints. The results of this work are discussed in Chapter 2.

Chapter 3 focuses on structure calculations of two crystalline systems: (1) the N-terminal domain (NTD) of SARS-CoV-2 nucleocapsid protein, and (2) the crystalline array of HIV-1 CA_{CTD}-SP1 protein bound with assembly co-factor IP6 and a maturation inhibitor Bevirimat (BVM).

Chapters 4 and 5 concern with an integrated approach to determine atomic-resolution structures of large biological assemblies, whereas MAS NMR restraints are combined with information from other experimental and computational methods. In Chapter 4, the structure of tubular assemblies of HIV-1 CA capsid protein is presented,

determined by integrating MAS NMR restraints with cryo-EM density in data-driven molecular dynamics (MD) simulations. In Chapter 5, this general approach is expanded and adapted to determine the structure of a motor domain of conventional kinesin-1, KIF5B, bound to polymerized microtubules. The studies presented in these two chapters establish the integrative structural biology framework for determination of structures of large biological systems inaccessible by any single technique in isolation.

Chapter 1

INTRODUCTION

1.1 Methods for protein structure determination

There are multiple biophysical techniques for protein structure determination, each of which has its merits and shortcomings.

X-ray crystallography

The first protein structure determined was crystalline myoglobin by X-ray diffraction in 1958.[1] At the time of this dissertation, X-ray diffraction has produced the most depositions in the protein data bank (PDB) totaling 171,900 (retrieved December 30 2022). The main caveat of X-ray crystallography is the requirement of single crystals, which are often not easily obtained or possible, depending on the system.

Cryo-electron microscopy

Another technique, cryogenic electron microscopy (cryo-EM), has significantly contributed to protein structure determination, particularly in recent years with the development/implementation of direct electron detectors in 2012 resulting in numerous structures at resolution higher than 5 Å. As the name suggests, the technique is performed at cryogenic temperatures. Additionally, in many studies, high-resolution is unachievable, with only mid-to-low-resolution (~7-12 Å) density maps only revealing an overall envelope/architecture of the system.

NMR spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is another widely used technique for atomic-resolution protein structure determination. NMR experiments can be performed both on soluble samples (solution NMR) or on insoluble proteins and protein assemblies that are prepared either as microcrystals or sedimented or lyophilized samples (solid-state NMR). This dissertation concerns with protein structures determined from experimental solid-state NMR data, using magic angle spinning (MAS) NMR experiments.

Experimental MAS NMR restraints necessary for structure determination include chemical shifts, torsion angle restraints, and distance restraints. The latter are determined from dipolar-based correlation experiments, most commonly involving ^{13}C-^{13}C correlations. These ^{13}C-^{13}C correlations correspond to interatomic distances of up to 7 Å. Collectively, with enough distance restraints, the fold and tertiary protein structure is obtainable.

Although not specifically required, secondary structure information in the form of dihedral (φ/ψ) restraints aids this process considerably.[2] One such prediction program to obtain this information from experimental NMR chemical shift assignments is TALOS-N[3], with, additional programs, such as CSI 3.0[4] and DANGLE[5] gaining in popularity. Collectively, distance restraints that connect different parts of the protein, along with the secondary structure predicted from the chemical shifts in the form of

dihedral (φ/ψ) restraints, summarized and illustrated in Figure 1.1, yield an atomic scale protein structure with a structure calculation.

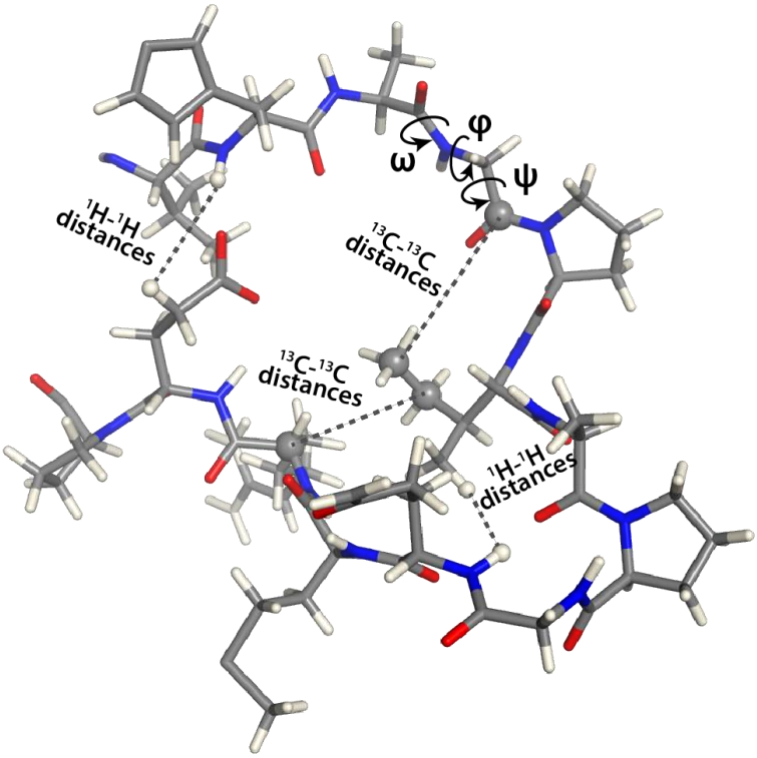

Figure 1.1: Summary and illustration of various restraints on a generic polypeptide that go into a structure calculation to yield an atomic resolution structure by NMR spectroscopy. Reprinted with permission from [2].

Solution NMR, although quite powerful, is limited to relatively small systems and requires samples to be fully soluble. On the other hand, solid-state NMR does not have any limitations with respect to long-range order, solubility, or the molecular weight of the system.[6] Static solid-state NMR experiments generate a broad pattern and to

achieve resolution approaching that of solution NMR, magic angle spinning (MAS) NMR can be employed. Under MAS conditions a solid sample is spun at an angle of 54.7° relative to the magnetic field. Consequently, the dipolar and chemical shift anisotropy interactions are averaged out over the rotor period but can be recoupled through judiciously designed pulse sequences.[7]

1.2 Protein structure determination by NMR

The most efficient and one of the most common approaches to NMR protein structure calculations to date is molecular-based simulating annealing (MDSA) calculations, which are often referred to as simulated annealing.[8] The strategy for such calculations is, while sampling conformers, to compute the potential energy of the protein structure using potential energy terms (or colloquially 'potentials') where the lowest energy structure equates to the highest probable conformation.[9] The calculation begins by initiating a high temperature bath where the conformational landscape can be carefully explored. The temperature is then slowly reduced to converge towards an overall local minimum which corresponds to the most probable conformer. As with the vast majority of protein computational approaches, the heat is simply a level of disorder applied to the system and a source of energy, as applying heat is simulated.[8]

An important consideration for practically any protein structure calculation is temperature. Broadly speaking, temperature is defined most relevantly from the zeroth law of thermodynamics.[10–12] Considering three bodies, if one said body (A) is in thermal

equilibrium with the remaining two bodies (B,C), then the remaining two bodies (B,C) are in thermal equilibrium with each other. Moreover, if two bodies (B,C) are said to be in thermal equilibrium with each other, they are said to have the same "temperature".[13]

In the context of MDSA calculations, the temperature at each timestep is set to a constant value by scaling the velocities of the canonical system (constant temperature, pressure, and amount of particles).[9] The temperature, on a molecular level, is directly related to the average kinetic energy of the system where the velocities is correlated to temperature by the Maxwell-Boltzmann distribution.[10,11]

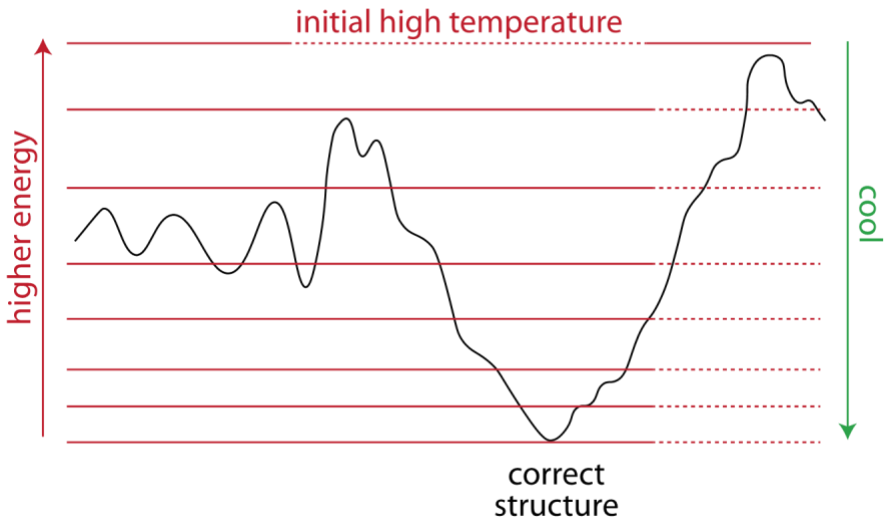

Figure 1.2: Visual representation of the energy landscapes explored during molecular-based simulating annealing (MDSA) calculations.

1.2.1 Potential energy of a protein described with potential energy terms

The potential energy of a protein can be expressed with empirical force fields with potential energy terms. The potential energy terms can be expressed as

$$U_{TOTAL} = U_{SYS} + U_{EXP} + U_{KNOW} \quad \text{Eq. 1.5}$$

where U_{SYS} describes the energy of the molecular system, U_{EXP} contains the experimental restraining energy terms derived from the NMR data, and U_{KNOW} comprises restraining terms that utilize protein knowledge to improve the accuracy and enhance the protein structure quality. Each of the three terms can be further expanded

$$U_{SYS} = U_{bonds} + U_{angle} + U_{impr} + U_{repel} \quad \text{Eq. 1.6}$$

$$U_{EXP} = U_{dist} + U_{dihed} \quad \text{Eq. 1.7}$$

$$U_{KNOW} = U_{torsionDB} + U_{HBPot} + U_{gyr} \quad \text{Eq. 1.8}$$

where U_{bonds}, U_{angle}, U_{impr} describe the bond length, bond angles, and improper angles of the protein respectively. U_{repel} is a "van der Waals-like" electrostatics term. U_{dist} and U_{dihed} contain the experimental restraints from NMR for distance and dihedral restraints, respectively. The $U_{torsionDB}$, U_{HBPot}, and U_{gyr} terms are added for the torsion angles, hydrogen bonds, and the radius of gyration for the overall size of the folded protein, respectively. [14,15]

1.2.2 Incorporation of NMR distance restraints

Confidently assigned distance restraints are incorporated in the structure calculation through a potential energy term, U_{dist}, as a hard well, as illustrated in Figure 1.3 and expressed as

$$U_{dist}(r) = \begin{cases} \left(r - d - d_{plus}\right)^2 & if\ r > d + d_{plus} \\ (r - d + d_{minus})^2 & if\ r < d - d_{minus} \\ 0 & in\ between \end{cases} \qquad \text{Eq. 1.9}$$

where d_{plus} and d_{minus} are user-provided, to set the distance restraint bounds of the potential well; r is given by

$$r = \left[\left\langle \sum_{ij} \left|\boldsymbol{q}_i - \boldsymbol{q}_j\right|^{-6} \right\rangle\right]^{-1/6} \qquad \text{Eq. 1.10}$$

where $\boldsymbol{q}_i$ and $\boldsymbol{q}_j$ are the positions of atoms i and j, the ij sum is over all atom pairs associated with the given cross-peak.[2,9]

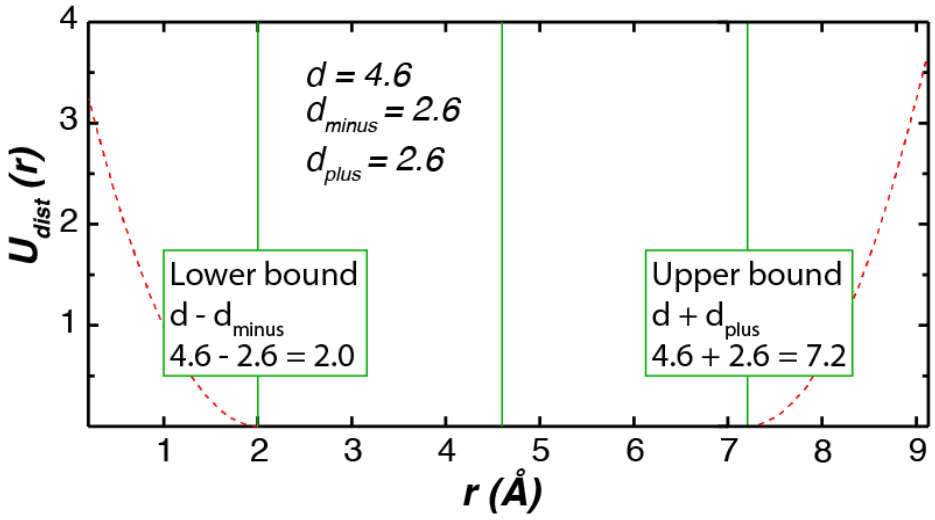

Figure 1.3: Visual depiction of a hard well potential for incorporating experimental MAS NMR cross-peaks as distance restraints as a function of $U_{dist}(r)$. The term d is the target distance; d_{minus} and d_{plus} are user defined by the user as input where the lower bound is the difference of d and d_{minus} and the upper bound is the sum of d and d_{plus}.

The distance restraint bounds are set by the user prior to the dynamics and ought to be set by the type of experiment performed. In the context of ^{13}C-^{13}C MAS NMR experiments, the distance dependence of the dipolar coupling, which gives rise to the MAS NMR signal, is given by

$$\omega_D = -\frac{1}{2\pi}\frac{\mu_o}{4\pi}\frac{\gamma^2\hbar}{r^3} \qquad \text{Eq. 1.11}$$

where r is the distance between the two nuclei, γ is the gyromagnetic ratio of the ^{13}C nuclei, $\hbar$ is the reduced Planck constant, and μ_o is the permeability of the free vacuum. Unlike the nuclear Overhauser effect (NOE) in solution NMR that scales as $1/r^6$, signal

intensity scales with $1/r^3$ in the dipolar-based MAS NMR, resulting in a less steep fall-off for longer distances.[2] This can be seen in Figure 1.4.

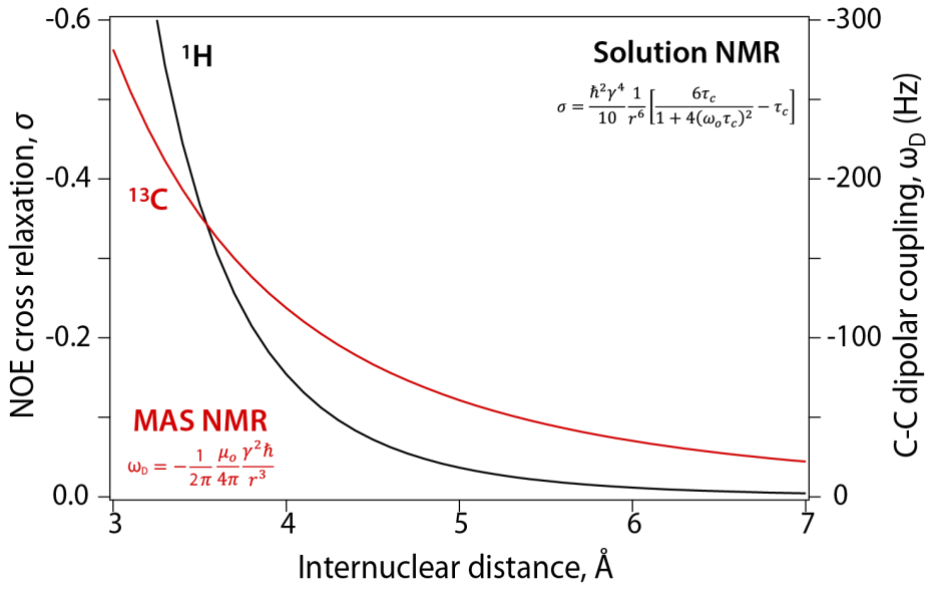

Figure 1.4: Distance dependence of the ^{1}H-^{1}H NOE and ^{13}C-^{13}C dipolar coupling for solution and MAS NMR, respectively. The NOE curve was calculated for τ_c = 7.1 ns, corresponding to a spherical protein of 14.6 kDa molecular mass at T = 37 °C. Reprinted with permission from [2].

1.2.3 Incorporation of NMR dihedral restraints

The assigned chemical shifts from experimental MAS NMR spectra can be used to accurately predict the secondary structure of a protein using an empirical database approach. Two such examples commonly employed are CSI 3.0[4] and TALOS-N[3] from the Wishart and Bax groups, respectively. In the context of structure determination, the

secondary structure prediction can be incorporated into structure calculations as φ and ψ backbone dihedral restraints where the bounds are set based on the generated confidence of the prediction. Dihedral restraints are also incorporated as a hard well potential

$$U_{dihed}(\phi) = \begin{cases} c\left(\phi_{ij} - \phi_{ij}^{u}\right)^2 & , \quad if\ \phi_{ij} > \phi_{ij}^{u} \\ 0 & , \quad if\ \phi_{ij}^{l} \leq \phi_{ij} \leq \phi_{ij}^{u} \\ c\left(\phi_{ij} - \phi_{ij}^{l}\right)^2 & , \quad if\ \phi_{ij} < \phi_{ij}^{u} \end{cases} \qquad \text{Eq. 1.12}$$

where ϕ_{ij} , ϕ_{ij}^{u} , and ϕ_{ij}^{l} is the calculated value, upper threshold, and lower threshold respectively.[16]

1.2.4 Structure elucidation with structure enhancing potential energy terms

Force fields for all atom molecular dynamics (MD) generally do not require additional terms to capture all the features of a protein because they utilize Lennard–Jones and Coulombic energy terms, which capture such interactions. However, starting coordinates are customary for such runs. In pursuit of protein coordinates the standard approach is to use an empirical force field because it is far less time-consuming and computationally demanding. This, however, does require supplementation with three additional empirical terms for hydrogen bonding, torsion angle bias, and overall ellipsoidal bias of the protein shape.

The first of such terms is the hydrogen bond potential which improves the backbone and sidechain hydrogen bond geometry by opportunistically forming them in

situations of unknown hydrogen bonding.[17] Without such inclusion the prospect of hydrogen bonds would be missed. This is incorporated mathematically by the utilization of three-dimensional potentials of mean force (*P)* created from the identity of the prospect proton donors and acceptors

$$U^{0}_{HBPot}(r,\theta,\phi) = -\ln P(r,\theta,\phi) \qquad \text{Eq. 1.13}$$

where the potential is expressed in spherical coordinates (r,θ,ϕ) and the potential of mean force was created with the adaptive kernel density estimation.[17,18]

The quality and accuracy of the protein structures is also enhanced by the addition of a statistical empirical potential energy term of probability densities of torsion angles. This term encompasses a database of torsion angle distributions to reproduce physically representative conformational features. This protocol was inspired by the X-ray crystallization protocol of the utilization of rotamer libraries for sidechain fitting in a poor resolution crevasse. Here, in the context of NMR, the potential energy term can be encompassed with the database by

$$U_{torsionDB}(\mathbb{x}) = -\beta \ln P(\mathbb{x}|a) \qquad \text{Eq. 1.14}$$

where $\mathbb{x}$ is one or more torsion angles, β is the corresponding constant, and $P(\mathbb{x}|a)$ is the probability density of $\mathbb{x}$ given a which is a variable associated from the amino acid type.[19]

The final knowledge-based potential energy term used to improve structure quality is the gyration pseudopotential. The inclusion of this term results in improved

protein packing by enforcing an overall ellipsoidal shape/boundary. Mathematically the potential energy term scales directly with the gyration volume V_g

$$U_{gyr} \propto V_g \qquad \text{Eq. 1.15}$$

which is given by

$$V_g = \frac{4}{3}\pi \sqrt{\left| \frac{1}{N} \sum_{i=1}^{N} \Delta q_i \otimes \Delta q_i \right|} \qquad \text{Eq. 1.16}$$

where N is the number if atoms and Δq_i is the difference between the position of atom i and the corresponding centroid position. The $\otimes$ indicates an outer product.[20]

1.2.5 Incorporation of cryo-EM density: batch docking

As demonstrated in the following chapters, due to the inherently local nature of NMR distance restraints, it is often necessary to incorporate cryo-EM density for the overall envelope information when there is an insufficient amount of distance restraints. To do so, the first step is to dock a folded structure calculated by NMR alone. The docking is an exhaustive exploration of the structure within the density space via translations and rotations, to identify the best fits. The quality of such fits is expressed by cross-correlation scores, which are derived from the least squares function (LSF) to identify how identical two functions are

$$LSF_{EM} = \sum_i (\rho_i^T - S\rho_i^P)^2 \qquad \text{Eq. 1.17}$$

where T is the target map from coordinates, P is the probe map that is searched, i is the position on the map, S is the scaling factor, and ρ is the voxel density. Assuming the sums of the square densities are constant they can be ignored, and the cross-correlation function (CCF) can be expressed as the negative of the approximated LSF.[21]

$$CCF_{EM} = \sum_{i} \rho_i^T \rho_i^P \qquad \text{Eq. 1.18}$$

1.2.6 Incorporation of cryo-EM density: structure calculations

The second component of cryo-EM map incorporation is during the calculation itself. A cryo-EM density map can be incorporated in structure calculations with an additional potential energy term ($U_{CrossCorr}$). This allows for simultaneous consideration of distance restraints, dihedral restraints, and cryo-EM density. The cryo-EM potential energy term ($U_{CrossCorr}$) assesses the cross-correlation between the input map and the atomic probability calculated from the protein coordinates

$$U_{CrossCorr} = \frac{1}{N} \sum_{i=1}^{n} \frac{(\, m_i^{obs} - \overline{m^{obs}}\,)(\, m_i^{calc} - \overline{m^{calc}}\,)}{\sigma^{obs}\sigma^{calc}} \qquad \text{Eq. 1.19}$$

where m_i^{obs} is the value of the input map, m_i^{calc} is the corresponding calculated value, $\overline{m^{obs}}$ is the average of the input map, σ^{obs} is the standard deviation of the input map, $\overline{m^{calc}}$ is the average of the back-calculated map, and σ^{calc} is the standard deviation of the back-calculated map.[22]

If there are sufficient NMR distance restraints (~9-15 restraints/residue), the inclusion of a cryo-EM density map is not necessary and would not noticeably impact the structure quality.[2] However, when there are an insufficient number of distance restraints, the inclusion of a cryo-EM density map in the calculation improves the structure quality.[18] To demonstrate this, control structure calculations were performed for the model system of *Oscillatoria agardhii* (OAA), Figure 1.5.

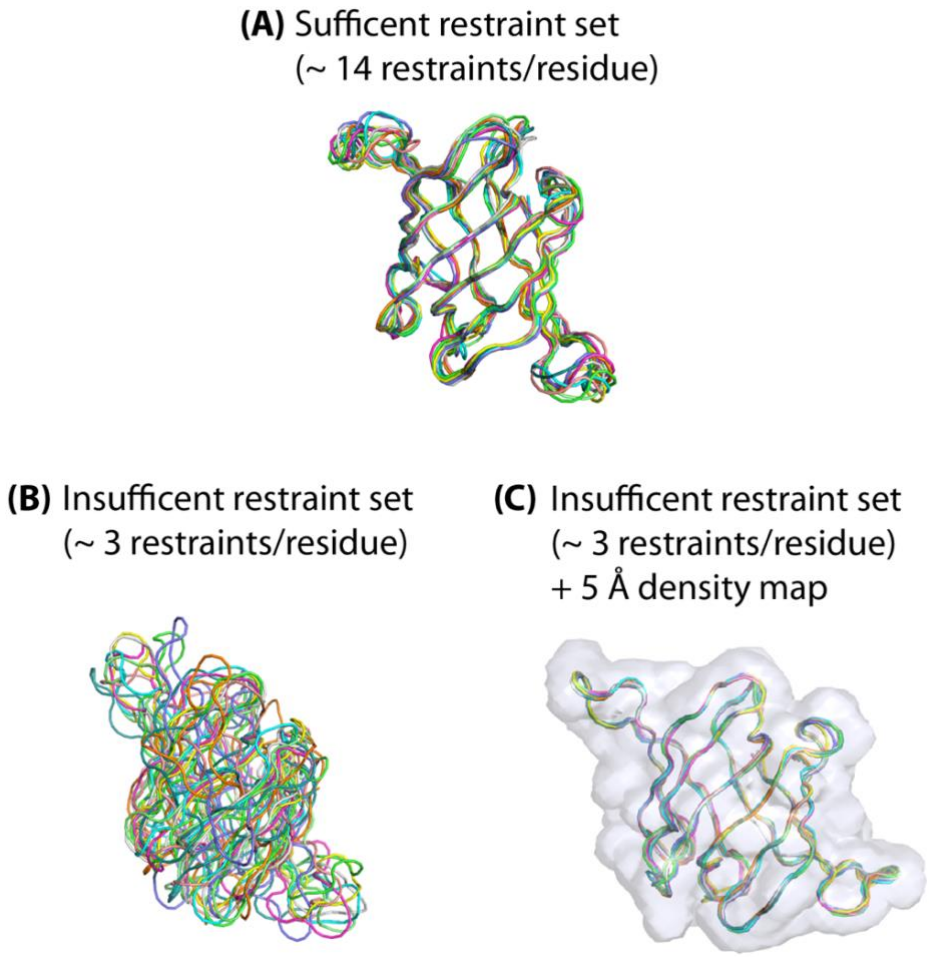

Figure 1.5: Structure calculations of *Oscillatoria agardhii* (OAA) using: (A) a sufficient distance restraint set, (B) and incomplete distance restraint set, and (C) an incomplete distance restraint set with a cryo-EM density map (5 Å resolution). The density map was prepared in UCSF Chimera[23] and calculations carried out in Xplor-NIH[22].

REFERENCES

(1) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *181* (4610), 662–666. https://doi.org/10.1038/181662a0.

(2) Russell, R. W.; Fritz, M. P.; Kraus, J.; Quinn, C. M.; Polenova, T.; Gronenborn, A. M. Accuracy and Precision of Protein Structures Determined by Magic Angle Spinning NMR Spectroscopy: For Some 'with a Little Help from a Friend.' *J Biomol NMR* **2019**. https://doi.org/10.1007/s10858-019-00233-9.

(3) Shen, Y.; Bax, A. Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *Journal of Biomolecular NMR* **2013**, *56* (3), 227–241. https://doi.org/10.1007/s10858-013-9741-y.

(4) Hafsa, N. E.; Arndt, D.; Wishart, D. S. CSI 3.0: A Web Server for Identifying Secondary and Super-Secondary Structure in Proteins Using NMR Chemical Shifts. *Nucleic Acids Research* **2015**, *43* (W1), W370–W377. https://doi.org/10.1093/nar/gkv494.

(5) Cheung, M.-S.; Maguire, M. L.; Stevens, T. J.; Broadhurst, R. W. DANGLE: A Bayesian Inferential Method for Predicting Protein Backbone Dihedral Angles and Secondary Structure. *Journal of Magnetic Resonance* **2010**, *202* (2), 223–233. https://doi.org/10.1016/j.jmr.2009.11.008.

(6) Quinn, C. M.; Lu, M.; Suiter, C. L.; Hou, G.; Zhang, H.; Polenova, T. Magic Angle Spinning NMR of Viruses. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2015**, *86–87*, 21–40. https://doi.org/10.1016/j.pnmrs.2015.02.003.

(7) Karlsson, T.; Popham, J. M.; Long, J. R.; Oyler, N.; Drobny, G. P. A Study of Homonuclear Dipolar Recoupling Pulse Sequences in Solid-State Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* **2003**, *125* (24), 7394–7407. https://doi.org/10.1021/ja0294360.

(8) Nilges, M.; Clore, G. M.; Gronenborn, A. M. Determination of Three-Dimensional Structures of Proteins from Interproton Distance Data by Dynamical Simulated Annealing from a Random Array of Atoms. Circumventing Problems Associated with Folding. *FEBS Lett* **1988**, *239* (1), 129–136. https://doi.org/10.1016/0014-5793(88)80559-3.

(9) Schwieters, C. D.; Clore, G. M. Internal Coordinates for Molecular Dynamics and Minimization in Structure Determination and Refinement. *Journal of Magnetic Resonance* **2001**, *152* (2), 288–302. https://doi.org/10.1006/jmre.2001.2413.

(10) Sri Harish, M.; Patra, P. K. Temperature and Its Control in Molecular Dynamics Simulations. *Molecular Simulation* **2021**, *47* (9), 701–729. https://doi.org/10.1080/08927022.2021.1907382.

(11) Wong-ekkabut, J.; Karttunen, M. The Good, the Bad and the User in Soft Matter Simulations. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2016**, *1858* (10), 2529–2538. https://doi.org/10.1016/j.bbamem.2016.02.004.

(12) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. van; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics* **1998**, *81* (8), 3684. https://doi.org/10.1063/1.448118.

(13) Planck, M.; Masius, M. *The Theory of Heat Radiation*; Philadelphia, P. Blakiston's Son & Co, [c1914].

(14) Wider, G. Structure Determination of Biological Macromolecules in Solution Using Nuclear Magnetic Resonance Spectroscopy. *BioTechniques* **2000**, *29* (6), 1278–1294. https://doi.org/10.2144/00296ra01.

(15) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. The Xplor-NIH NMR Molecular Structure Determination Package. *Journal of Magnetic Resonance* **2003**, *160* (1), 65–73.

(16) Clore, G. M.; Nilges, M.; Sukumaran, D. K.; Brünger, A. T.; Karplus, M.; Gronenborn, A. M. The Three-Dimensional Structure of A1-Purothionin in Solution: Combined Use of Nuclear Magnetic Resonance, Distance Geometry and Restrained Molecular Dynamics. *The EMBO Journal* **1986**, *5* (10), 2729–2735. https://doi.org/10.1002/j.1460-2075.1986.tb04557.x.

(17) Schwieters, C. D.; Bermejo, G. A.; Clore, G. M. A Three-Dimensional Potential of Mean Force to Improve Backbone and Sidechain Hydrogen Bond Geometry in

Xplor-NIH Protein Structure Determination. *Protein Science* **2020**, *29* (1), 100–110. https://doi.org/10.1002/pro.3745.

(18) Silverman, B. W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: Boca Raton, 1986.

(19) Bermejo, G. A.; Clore, G. M.; Schwieters, C. D. 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* **2012**, *21* (12), 1824–1836. https://doi.org/10.1002/pro.2163.

(20) Schwieters, C. D.; Clore, G. M. A Pseudopotential for Improving the Packing of Ellipsoidal Protein Structures Determined from NMR Data. *J. Phys. Chem. B* **2008**, *112* (19), 6070–6073. https://doi.org/10.1021/jp076244o.

(21) Vasishtan, D.; Topf, M. Scoring Functions for CryoEM Density Fitting. *Journal of Structural Biology* **2011**, *174* (2), 333–343. https://doi.org/10.1016/j.jsb.2011.01.012.

(22) Gong, Z.; Schwieters, C. D.; Tang, C. Conjoined Use of EM and NMR in RNA Structure Refinement. *PLOS ONE* **2015**, *10* (3), e0120445. https://doi.org/10.1371/journal.pone.0120445.

(23) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J Comput Chem* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

Chapter 2

SYSTEMATIC ASSESMENT OF THE ACCURACY AND PRECISION OF PROTEIN STRUCTRUES DETERMINED BY MAS NMR SPECTROSCOY

This chapter is a verbatim reprint of the following with permission: Russell, R. W.; Fritz, M. P.; Kraus, J.; Quinn, C. M.; Polenova, T.; Gronenborn, A. M. (**2019**) Accuracy and Precision of Protein Structures Determined by Magic Angle Spinning NMR Spectroscopy: For Some 'with a Little Help from a Friend.' *J Biomol NMR* 73(6-7):333-346. DOI: 10.1007/s10858-019-00233-9.

Author contributions are as follows: T.P. and A.M.G. conceived the study and directed the work. R.W.R wrote all scripts, performed calculations, and made figures. M.P.F., J.K., and C.M.Q. performed MAS NMR experiments on OAA and galectin-3C, analyzed experimental data assigned chemical shifts. T.P., A.M.G, and R.W.R took the lead in writing the manuscript. All authors discussed the results and contributed to manuscript preparation.

2.1 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 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 ^{1}H 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 PDSD[16,17], dipolar-assisted rotary resonance DARR[18-20], combined RN-symmetry driven spin diffusion CORD[21], CHHC and NHHC[22], as well as proton-assisted recoupling approaches for homo- and heteronuclear correlation spectroscopy, like PAR[23] and insensitive nuclei cross polarization PAIN-CP[24]. 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 curves[34]. Unlike for NOEs, signal intensity scales with $1/r^3$, resulting in a less steep fall-off for longer distances (Figure 2.1). 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 in[36]). 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 cryo-EM 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 2.1, (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 crystallography[54], 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 shifts[57] 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: 2MPX[52]), OAA (1.2 Å resolution; PDB: 3OBL[54]), Galectin CBD (0.86 Å resolution; PDB: 3ZSJ[65]), CA (2.43 Å resolution; PDB: 4XFX[73]).

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 ^{13}C labeled proteins, following a common protocol for the preparation of isotopically labeled samples using either 1,6-^{13}C-glucose or 2-^{13}C-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 approaches[44], such as combining MAS NMR with cryo-EM as shown here for the HIV capsid.

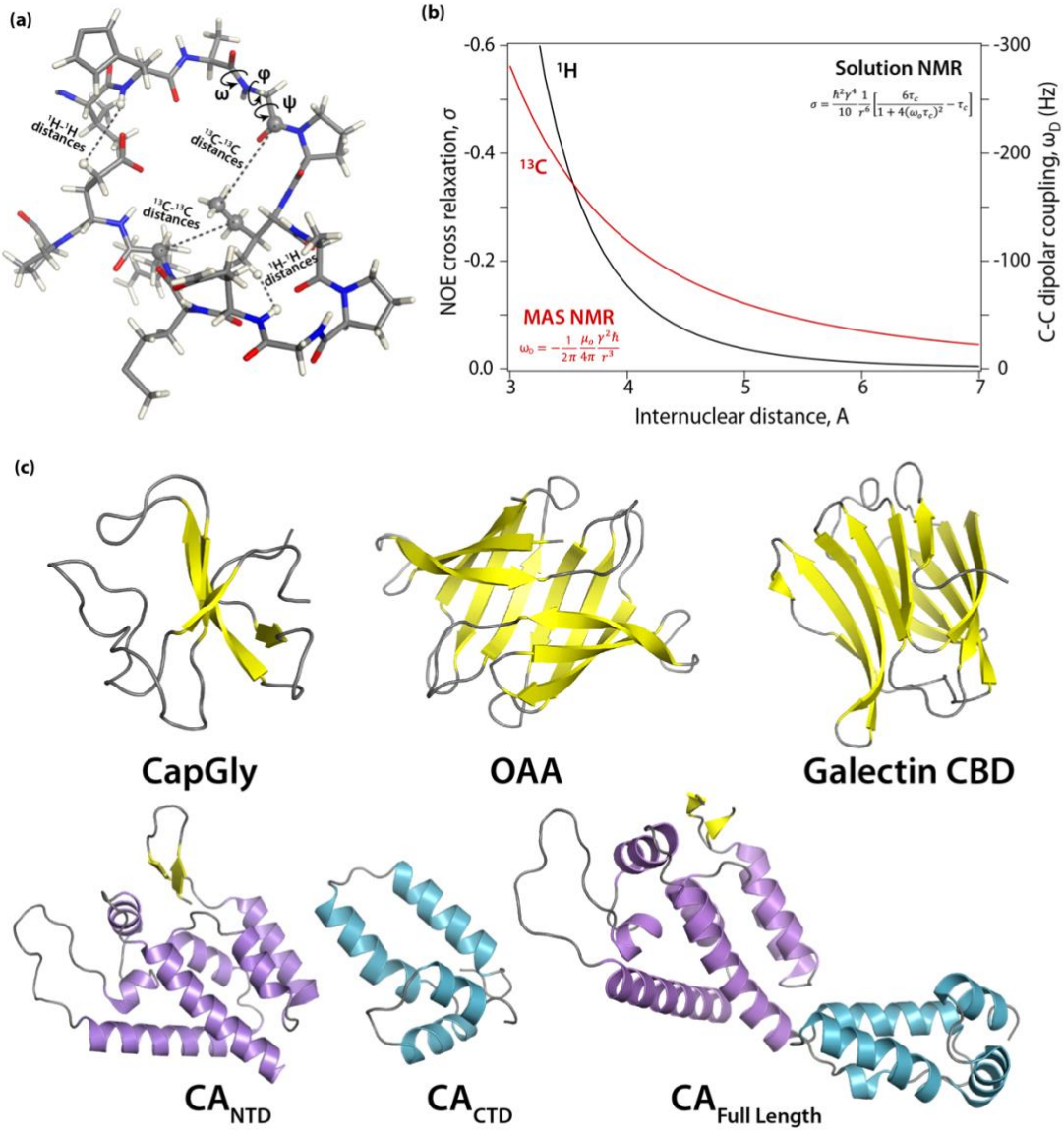

Figure 2.1: (A) Generic polypeptide chain, illustrating select backbone dihedral angles and ^{1}H-^{1}H and ^{13}C-^{13}C distances. (B) Distance dependence of

the ^{1}H-^{1}H NOE and ^{13}C-^{13}C dipolar coupling. The NOE curve was calculated for τ_c=7.1 ns, corresponding to a spherical protein of 14.6 kDa molecular mass at T = 37 ºC. (C) Ribbon representations of dynactin's CAP-Gly domain (PDBID: 2MPX), *Oscilatoria aghardii* 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).

2.2 Model calculations with synthetic distance restraints

2.2.1 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-Python[74] and Bio-PDB[75] 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.

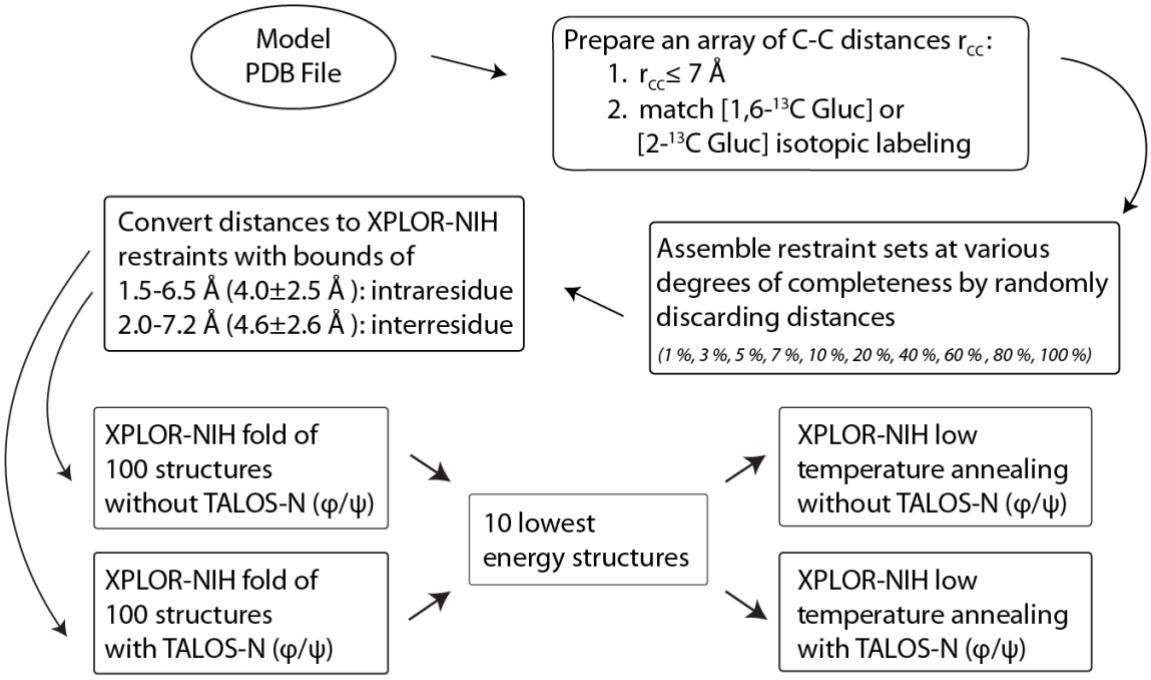

Figure 2.2: Flow diagram of computational strategy for protein structure calculations on the basis of synthetic C-C distance restraints.

Table 2.1: 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-^{13}C-glucose or 2-^{13}C-glucose (total for both labeling schemes combined)

Restraints	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|≤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-^{13}C-glucose or 2-^{13}C-glucose (total for both labeling schemes combined)

Restraints	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|≤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 ^{13}C isotopic labeling

Restraints	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|≤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 ^{13}C isotopic labeling

Restraints	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|≤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

Structure calculations were performed in XPLOR-NIH version 2.45[76-78]. Standard XPLOR-NIH terms for bond lengths, bond angles, and improper angles were used to enforce correct covalent geometry. The gyration constraint term[79], an empirical hydrogen-bond database term[80] and a statistical torsion-angle potential[81] were employed. In calculations that included backbone dihedral restraints, φ/ψ angles were

predicted using TALOS-N[35] from the experimental solid-state ^{13}C and ^{15}N 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/Å^2 in the initial simulated annealing and from 2 to 30 kcal/mol/Å^2 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/rad^2 during cooling. In the second dynamics run, the dihedral restraint force constants were set to 10 kcal/mol/rad^2 for high temperature dynamics at 3000 K and 200 kcal/mol/rad^2 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.

2.2.2 Cryo-EM 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 cryo-EM density. Six NTD and six CTD units were docked into the 8 Å cryo-EM map of a CA hexamer unit from a tubular assembly, using the "*phenix.dock_in_map*" program in PHENIX 1.14[82]. 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.

2.2.3 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 PyMol[83] using automated in-house shell scripts. Secondary structure elements were defined using STRIDE[84].

2.3 The effects of C-C distances on the accuracy and precision of protein structures

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 schemes[85] 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 2.1A and Figure 2.2). 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 2B). 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 2C and 2D). In reality, however, medium- and long-range distance restraints cannot be efficiently extracted for uniformly-^{13}C labeled samples because of dipolar truncation[87] 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 2.2. 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.

2.3.1 Dynactin's CAP-Gly domain

For the CAP-Gly domain, the accuracy (Figure 2.3A), 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 approx½tely 1/2 of the one without TALOS. This difference, however, disappears at restraint completeness above 40%. The precision (Figure 2.3B), 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

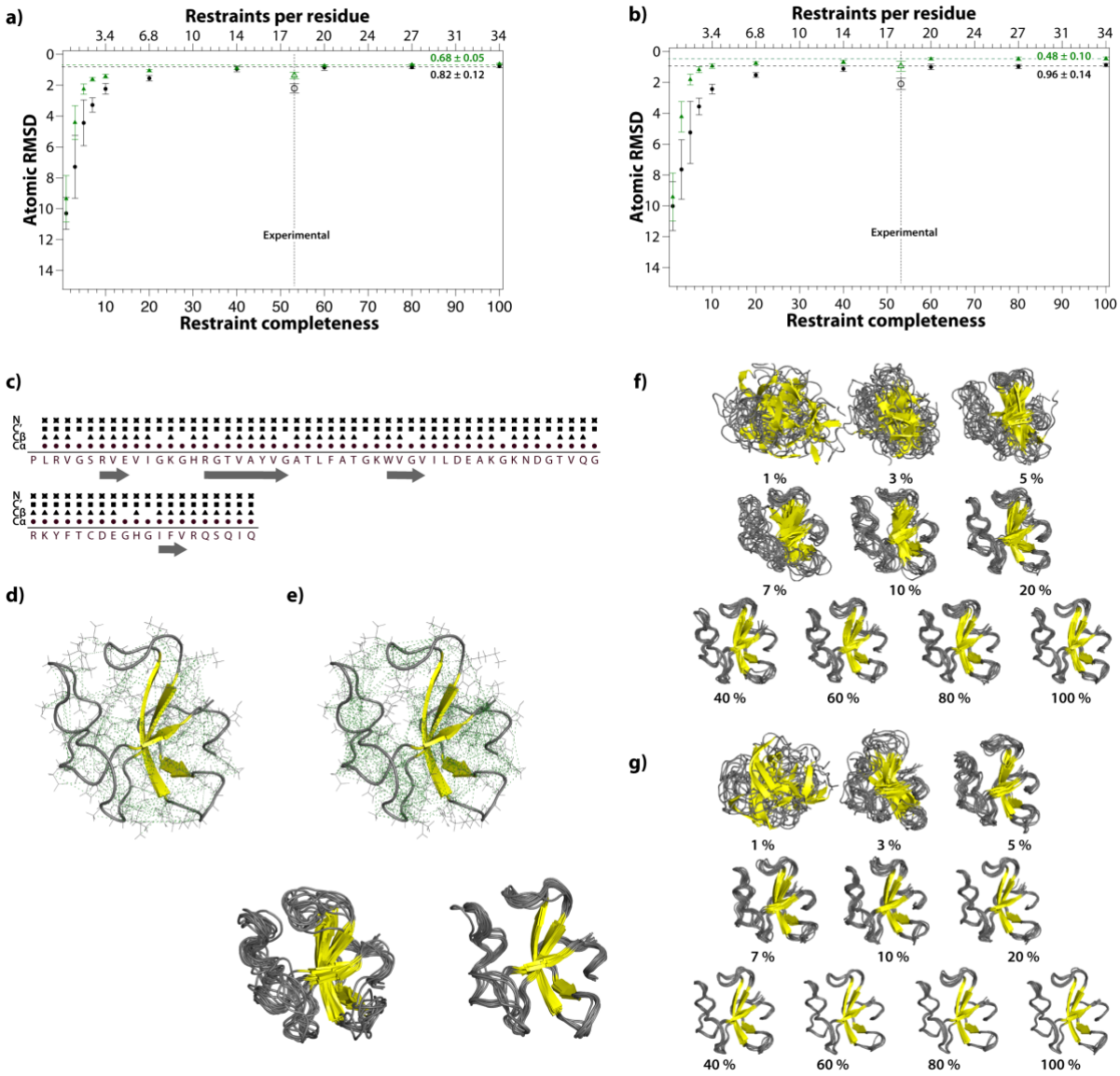

Figure 2.3: Structure calculation for dynactin's CAP-Gly domain. (A) Accuracy as defined by atomic backbone RMSD with respect to the target (input) structure and (B) Precision as defined by pairwise atomic backbone RMSD for ensemble members plotted vs. restraint completeness and number of restraints 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 ten 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 ten lowest energy structures calculated for different degrees of restraint completeness without (F) and with (G) TALOS-N derived backbone torsion angle restraints.

TALOS, respectively. In order to pictorially illustrate the increased structural precision with increasing number of restraints, the 10-member ensembles are depicted in Figures 2.3E and 2.3F. 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 2.3D)[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 manuscript[52]. 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 2.3A and 2.3B). 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 2.3A 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.

2.3.2 *Oscilatoria aghardii* agglutinin (OAA) and carbohydrate binding domain of galectin-3C

Similar qualitative behaviors are seen for OAA (Figure 2.4) and the Galectin CBD (Figure 2.5). 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 2.4B) 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.

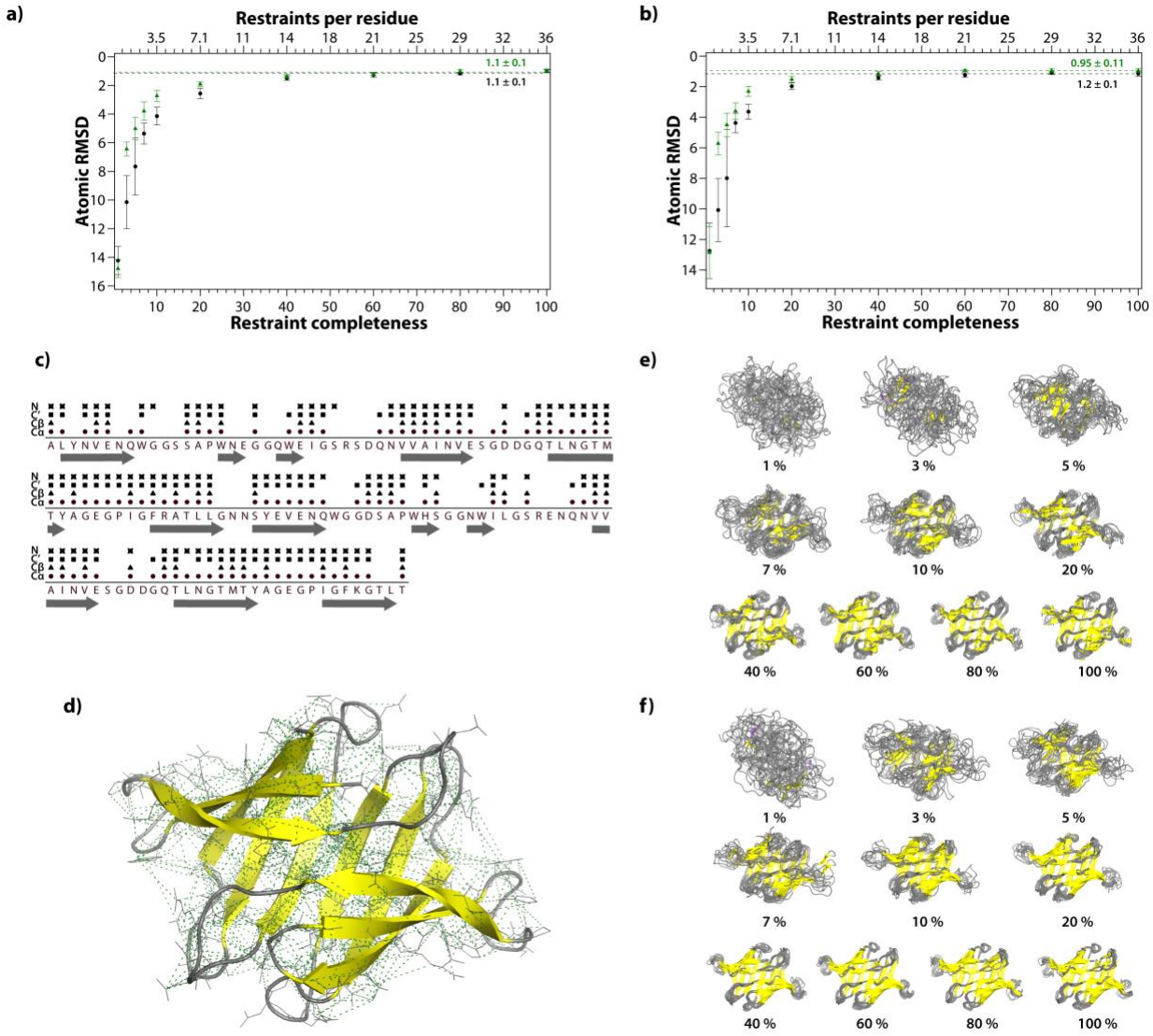

Figure 2.4: Structure calculation for *Oscilatoria aghardii* agglutinin, OAA. (A) Accuracy as defined by atomic backbone RMSD with respect to the target (input) structure and (B) Precision as defined by pairwise atomic backbone RMSD for ensemble members plotted vs. restraint completeness and number of restraints per residue. Data without (black symbols) or with (green symbols) backbone torsion angle restraints from TALOS-N using 235 experimental ^{13}C and 92 ^{15}N chemical shifts[56]. 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 OAA structure. (E,F) Best-fit superpositions of the ten lowest energy structures calculated for different degrees of restraint completeness without (E) and with (F) TALOS-N derived backbone torsion angle restraints.

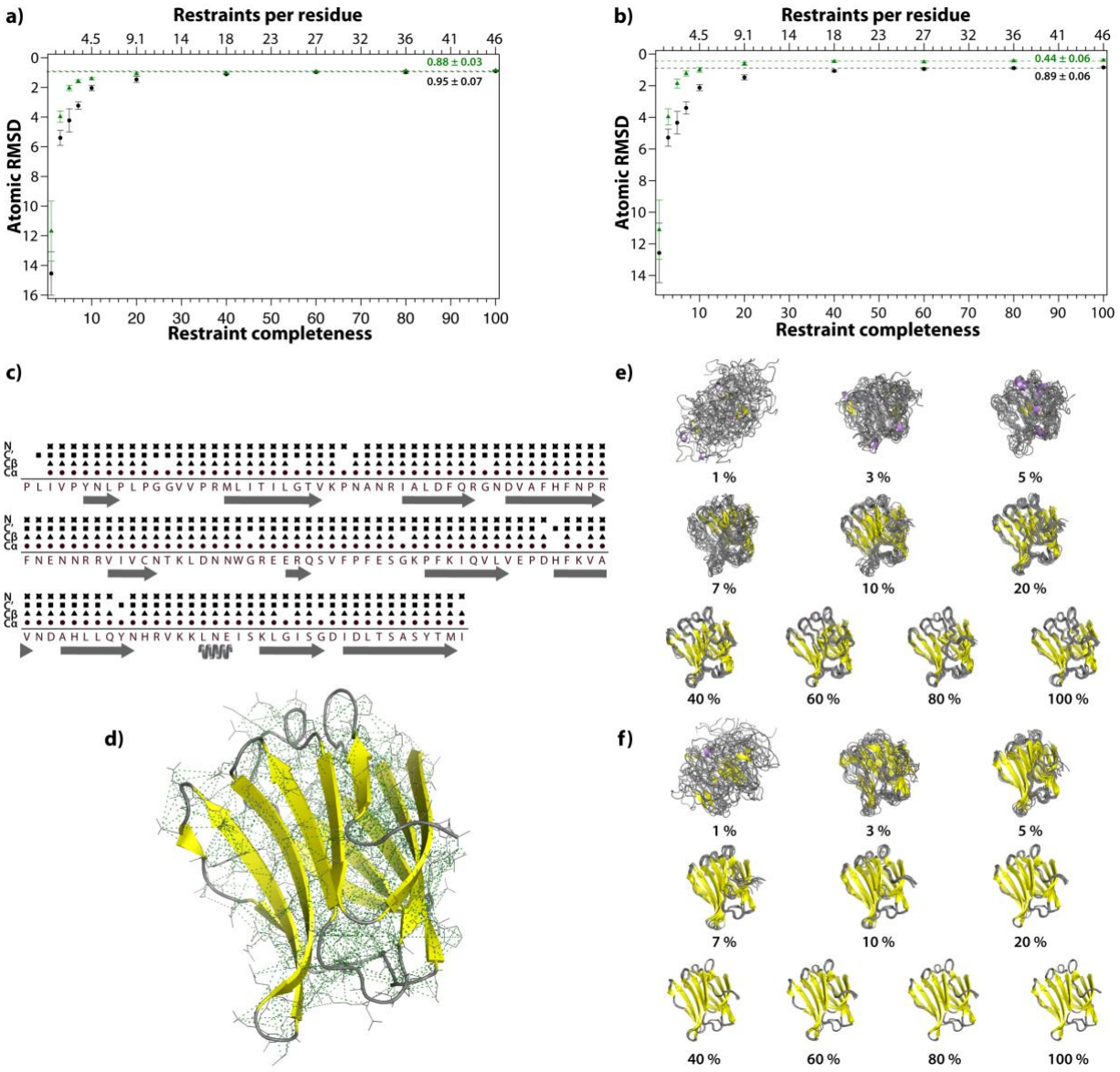

Figure 2.5: Structure calculation for the Galectin CBD. (A) Accuracy as defined by atomic backbone RMSD with respect to the target (input) structure and (B) Precision as defined by pairwise atomic backbone RMSD for ensemble members plotted vs. restraint completeness and number of restraints per residue. 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 ten lowest energy structures calculated for different degrees of restraint completeness without (E) and with (F) TALOS-N derived backbone torsion angle restraints.

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.

2.3.3 HIV-1 capsid protein

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 solved[73], and here we treated the full-length protein as a single chain extracted from this hexamer.

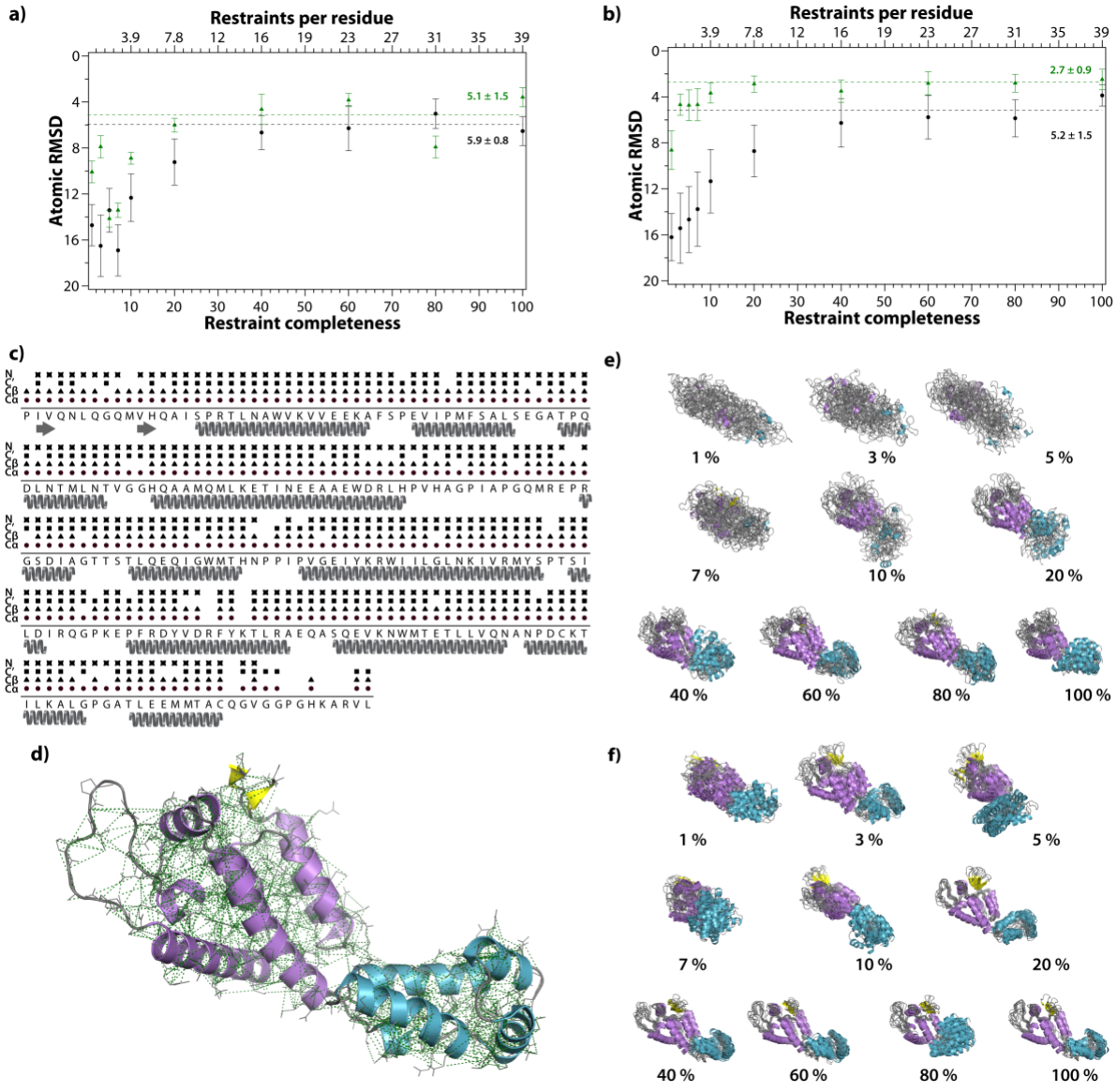

Figure 2.6: Structure calculation for the HIV-1 CA capsid protein (CA). (A) Accuracy as defined by atomic backbone RMSD with respect to the target (input) structure and (B) Precision as defined by pairwise atomic backbone RMSD for ensemble members plotted vs. restraint completeness and number of restraints 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 ten lowest energy structures calculated for different degrees of restraint completeness without (E) and with (F) TALOS-N derived backbone torsion angle restraints.

For the entire polypeptide chain of CA, accuracy and precision are much worse than for the other three systems (Figure 2.6). 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 2.7, 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.

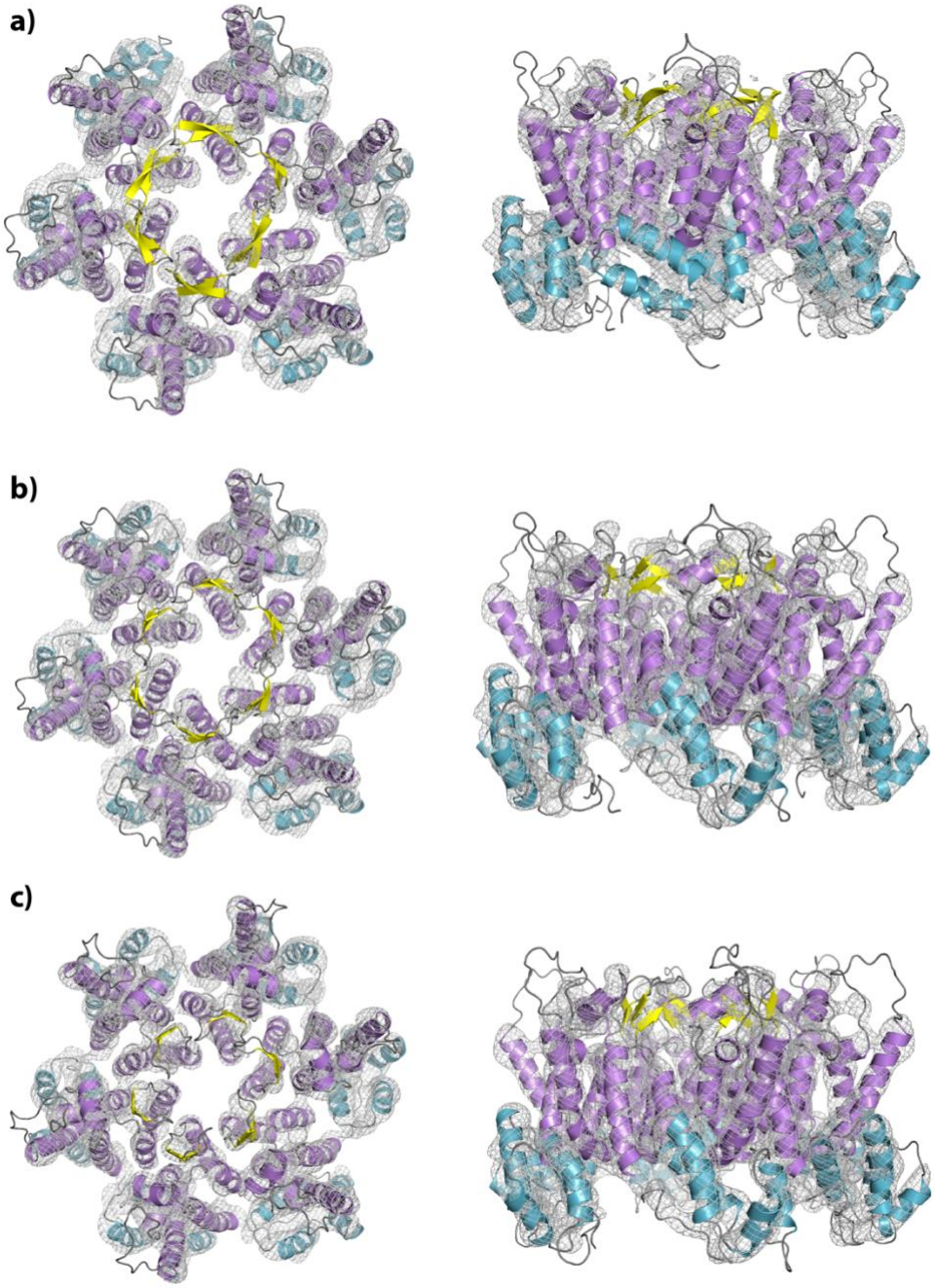

Figure 2.7: Atomic model of a hexamer unit of HIV-1 CA in tubular assemblies generated by combining MAS NMR-derived distances and cryo-EM density. The NTD (residues 1-145) and CTD (residues 148-231)

models were taken from the lowest energy structure of a single chain CA calculated at 40% restraint completeness (see Figure 2.6). (A) NTD and CTD domains were fit into the cryo-EM 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.

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 cryo-EM map of a CA hexamer unit. As illustrated in Figure 2.7A, 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 2.7B). The final hybrid NMR/cryo-EM atomic model of the CA hexamer is depicted in Figure 2.7C. 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.

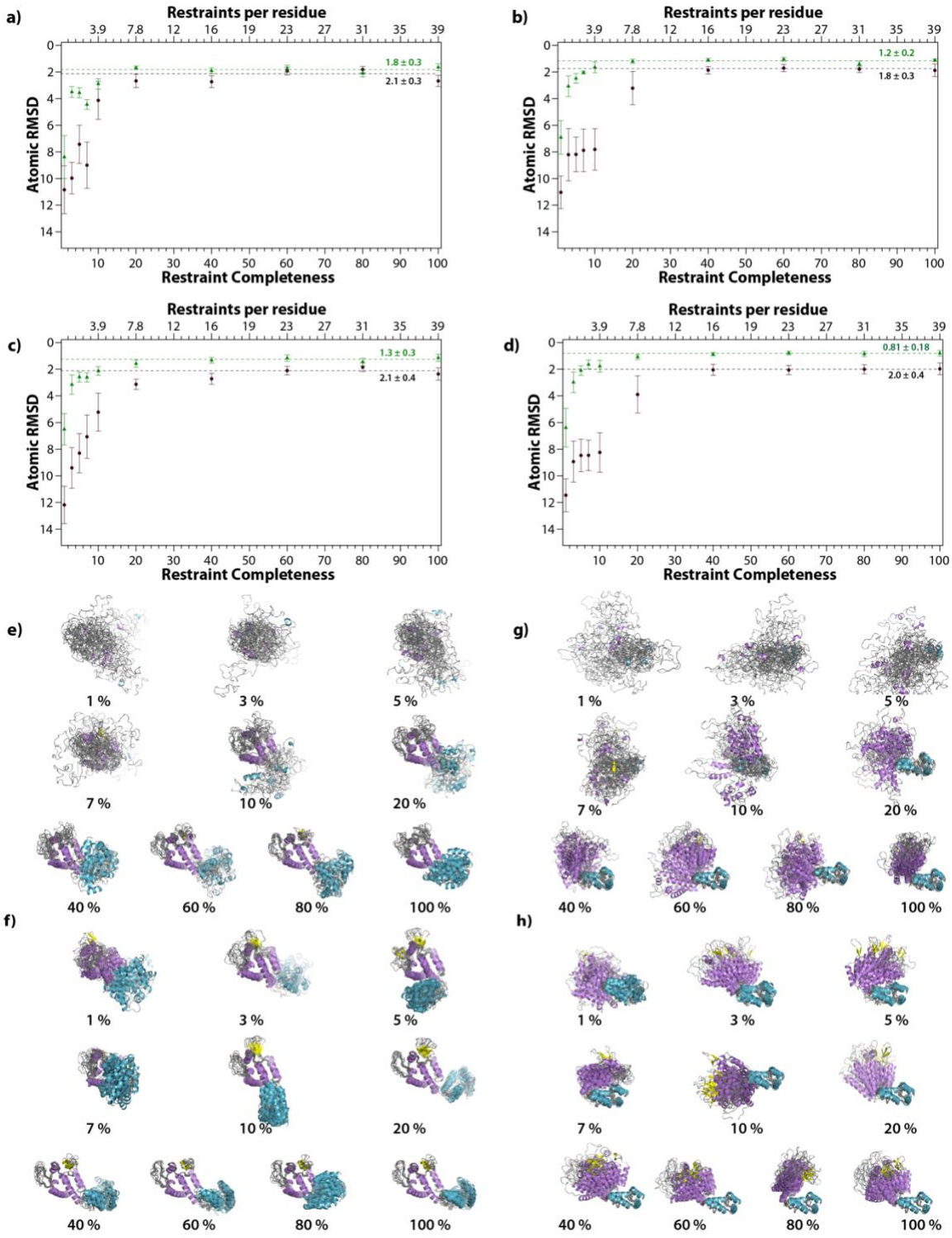
a)
Restraints per residue
3.9 7.8 12 16 19 23 27 31 35 39
Atomic RMSD
1.8 ± 0.3
2.1 ± 0.3
Restraint Completeness
b)
Restraints per residue
1.2 ± 0.2
1.8 ± 0.3
Atomic RMSD
Restraint Completeness
c)
Restraints per residue
1.3 ± 0.3
2.1 ± 0.4
Atomic RMSD
Restraint Completeness
d)
Restraints per residue
0.81 ± 0.18
2.0 ± 0.4
Atomic RMSD
Restraint Completeness
e)
1 %
3 %
5 %
7 %
10 %
20 %
40 %
60 %
80 %
100 %
f)
1 %
3 %
5 %
7 %
10 %
20 %
40 %
60 %
80 %
100 %
g)
1 %
3 %
5 %
7 %
10 %
20 %
40 %
60 %
80 %
100 %
h)
1 %
3 %
5 %
7 %
10 %
20 %
40 %
60 %
80 %
100 %

Figure 2.8: Domain Accuracy and Precision for CA. (A) Accuracy as defined by atomic backbone RMSD with respect to the target (input) structure and (C) Precision as defined by pairwise atomic backbone RMSD for ensemble members plotted vs. restraint completeness and number of restraints per residue, calculated for the NTD (residues 1-145). (B) Accuracy and (D) Precision for the CTD (residues 148-231). Data without (black symbols) or with (green symbols) backbone torsion angle restraints from TALOS-N based on experimental ^{13}C and ^{15}N chemical shifts. The dashed lines are the average values for 60, 80, and 100% restraint completeness. (E-H) Best-fit superpositions of the ten lowest energy structures calculated for different degrees of restraint completeness for the NTD without (E) or with (F) TALOS-N derived backbone torsion angle restraints. Equivalent data for the for the CTD without (G) or with (H) TALOS-N derived backbone torsion angle restraints.

2.4 Limitations of the assessment and the outlook of multi-domain proteins

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 $\leq$7 Å yield accurate structures (RMSD values with respect to the target structure of ~1 Å) with pairwise backbone atomic RMSD values of ~1 Å as well. Like for solution

NMR structures, the most critical parameter is the number of interatomic distance restraints (Table 2.1, Figures 2.3-2.9). 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 ≤7 Å 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].

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 resolution[86]. 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, ^{1}H-based distance restraints can also be supplemented in protein structure

determination, although these require extensive sample deuteration[94] 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 cryo-EM[44]. 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 in[36]) as well as extensive modeling and/or cryo-EM information. The latter has been recently employed in structural studies on several systems[44]. For the CA capsid assembly, the integration of cryo-EM data with the solid-state NMR distance restraints is clearly needed for a successful structure determination.

REFERENCES

(1) Clore, G.M. & Gronenborn, A.M. 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).

(2) Wüthrich, K. *NMR of proteins and nucleic acids,* (Wiley, Chichester, 1986).

(3) Kuszewski, J., Gronenborn, A.M. & Clore, G.M. The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Ser. B* **107**, 293-297 (1995).

(4) Kuszewski, J., Qin, J., Gronenborn, A.M. & Clore, G.M. The impact of direct refinement against $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Ser. B* **106**, 92-96 (1995).

(5) Clore, G.M. & Gronenborn, A.M. New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl. Acad. Sci. USA* **95**, 5891-5898 (1998).

(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).

(8) Tjandra, N., Grzesiek, S. & Bax, A. Magnetic field dependence of nitrogen−proton J splittings in ^{15}N-enriched human ubiquitin resulting from

relaxation interference and residual dipolar coupling. *J. Am. Chem. Soc.* **118**, 6264-6272 (1996).

(9) Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. & Bax, A. Use of dipolar $^{1}H–^{15}N$ and $^{1}H–^{13}C$ 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).

(11) Kuszewski, J., Gronenborn, A.M. & Clore, G.M. 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).

(12) Battiste, J.L. & 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).

(13) Gillespie, J.R. & 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).

(14) Girvin, M.E. & Fillingame, R.H. Hairpin folding of subunit c of F1Fo ATP synthase: ^{1}H distance measurements to nitroxide-derivatized aspartyl- *Biochemistry* **33**, 665-674 (1994).

(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).

(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, C.R., Gaponenko, V., Byrd, R.A. & Zilm, K.W. Diluting abundant spins by isotope edited radio frequency field assisted diffusion. in *J. Am. Chem. Soc.* Vol. 126 7196-7197 (2004).

(19) Takegoshi, K., Nakamura, S. & Terao, T. ^{13}C-^{1}H 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-^{1}H 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).

(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).

(23) De Paëpe, G., Lewandowski, J.R., Loquet, A., Böckmann, A. & Griffin, R.G. Proton assisted recoupling and protein structure determination. *J. Chem. Phys.* **129**, 245101 (2008).

(24) Lewandowski, J.R., De Paëpe, G. & Griffin, R.G. Proton assisted insensitive nuclei cross polarization. *J. Am. Chem. Soc.* **129**, 728-729 (2007).

(25) Castellani, F. et al. Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**, 98 (2002).

(26) Zech, S.G., Wand, A.J. & McDermott, A.E. Protein structure determination by high-resolution solid-state NMR spectroscopy: application to microcrystalline ubiquitin. *J. Am. Chem. Soc.* **127**, 8618-8626 (2005).

(27) Gullion, T. Rotational-echo, double-resonance NMR. in *Modern Magnetic Resonance* 713-718 (Springer, 2008).

(28) Michal, C.A. & Jelinski, L.W. REDOR 3D: heteronuclear distance measurements in uniformly labeled and natural abundance solids. *J. Am. Chem. Soc.* **119**, 9059-9060 (1997).

(29) Hing, A.W., Vega, S. & Schaefer, J. Transferred-echo double-resonance NMR. *J. Magn. Reson.* **96**, 205-209 (1992).

(30) Jaroniec, C.P., Filip, C. & Griffin, R.G. 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).

(31) Nieuwkoop, A.J., Wylie, B.J., Franks, W.T., Shah, G.J. & Rienstra, C.M. 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).

(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, S.L. & Watts, A. Applications of REDOR for distance measurements in biological solids. in *Annual Reports on NMR Spectroscopy*, Vol. 60 (ed. Webb, G.A.) 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).

(36) Quinn, C.M. & Polenova, T. Structural biology of supramolecular assemblies by magic-angle spinning NMR spectroscopy. *Quarterly Reviews of Biophysics* **49**, 1-44 (2016).

(37) Grage, S.L., Xu, X., Schmitt, M., Wadhwani, P. & Ulrich, A.S. ^{19}F-Labeling of peptides revealing long-range NMR distances in fluid membranes. *J. Phys. Chem. Lett.* **5**, 4256-4259 (2014).

(38) Roos, M., Wang, T., Shcherbakov, A.A. & 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).

(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).

(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).

(41) Grishaev, A., Wu, J., Trewhella, 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).

(42) Wang, I. et al. Structure, dynamics and RNA binding of the multi-domain splicing factor TIA-*Nucleic Acids Res.* **42**, 5949-5966 (2014).

(43) Debiec, K.T., Whitley, M.J., Koharudin, L.M., Chong, L.T. & Gronenborn, A.M. Integrating NMR, SAXS, and atomistic simulations: structure and dynamics of a two-domain protein. *Biophys. J.* **114**, 839-855 (2018).

(44) Cuniasse, P., Tavares, P., Orlova, E.V. & Zinn-Justin, S. Structures of biomolecular complexes by combination of NMR and cryoEM methods. *Curr. Opin. Struct. Biol.* **43**, 104-113 (2017).

(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).

(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).

(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).

(48) Honnappa, S. et al. Key interaction modes of dynamic plus TIP networks. *Mol. Cell* **23**, 663-671 (2006).

(49) Weisbrich, A. et al. Structure-function relationship of CAP-Gly domains. *Nat. Struct. Mol. Biol.* **14**, 959-967 (2007).

(50) Hayashi, I., Plevin, M.J. & Ikura, M. CLIP170 autoinhibition mimics intermolecular interactions with p150(Glued) or EB*Nat. Struct. Mol. Biol.* **14**, 980-981 (2007).

(51) Hayashi, I., Wilde, A., Mal, T.K. & Ikura, M. Structural basis for the activation of microtubule assembly by the EB1 and p150(Glued) complex. *Mol. Cell* **19**, 449-460 (2005).

(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).

(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 EB*J. Mol. Biol.* **425**, 4249-66 (2013).

(54) Koharudin, L.M. & Gronenborn, A.M. Structural basis of the anti-HIV activity of the cyanobacterial Oscillatoria Agardhii agglutinin. *Structure* **19**, 1170-81 (2011).

(55) Carneiro, M.G., Koharudin, L.M.I., Griesinger, C., Gronenborn, A.M. & Lee, D. ^{1}H, ^{13}C and ^{15}N resonance assignment of the anti-HIV lectin from Oscillatoria agardhii. *Biomol. NMR Assign.* **9**, 317-319 (2015).

(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).

(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).

(58) Sorme, P. et al. Structural and thermodynamic studies on cation-Pi 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).

(59) Atmanene, C. et al. Biophysical and structural characterization of mono/di-arylated lactosamine derivatives interaction with human galectin-*Biochem. Biophys. Res. Commun.* **489**, 281-286 (2017).

(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).

(61) Collins, P.M., Oberg, C.T., Leffler, H., Nilsson, U.J. & Blanchard, H. Taloside inhibitors of galectin-1 and galectin-*Chem. Biol. Drug Des.* **79**, 339-46 (2012).

(62) Collins, P.M., Bum-Erdene, K., Yu, X. & Blanchard, H. Galectin-3 interactions with glycosphingolipids. *J. Mol. Biol.* **426**, 1439-51 (2014).

(63) Hsieh, T.J. 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).

(64) Rajput, V.K. 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).

(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).

(66) Han, Y. et al. Solid-state NMR studies of HIV-1 capsid protein assemblies. *J. Am. Chem. Soc.* **132**, 1976-87 (2010).

(67) Byeon, I.J. 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).

(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).

(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).

(70) Liu, C. et al. Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nat. Commun.* **7**, 10714 (2016).

(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, C.M. et al. Dynamic regulation of HIV-1 capsid interaction with the restriction factor TRIM5alpha identified by magic-angle spinning NMR and molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* **115**, 11519-11524 (2018).

(73) Gres, A.T. et al. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* **349**, 99-103 (2015).

(74) Cock, P.J.A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423 (2009).

(75) Hamelryck, T. & Manderick, B. PDB file parser and structure class implemented in Python. *Bioinformatics* **19**, 2308-10 (2003).

(76) Schwieters, C.D., Kuszewski, J.J., Tjandra, N. & Clore, G.M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65-73 (2003).

(77) Schwieters, C.D., Kuszewski, J.J. & Marius Clore, G. Using Xplor–NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **48**, 47-62 (2006).

(78) Schwieters, C.D., Bermejo, G.A. & Clore, G.M. Xplor-NIH for molecular structure determination from NMR and other data sources. *Protein Sci.* **27**, 26-40 (2018).

(79) Kuszewski, J., Gronenborn, A.M. & Clore, G.M. 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).

(81) Bermejo, G.A., Clore, G.M. & Schwieters, C.D. 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).

(82) Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D* **66**, 213-221 (2010).

(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).

(85) Lundström, P. et al. Fractional ^{13}C enrichment of isolated carbons using [1-^{13}C]- or [2-^{13}C]-glucose facilitates the accurate measurement of dynamics at backbone

Cα and side-chain methyl positions in proteins. *J. Biomol. NMR* **38**, 199-212 (2007).

(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).

(87) Bayro, M.J. et al. Dipolar truncation in magic-angle spinning NMR recoupling experiments. *J. Chem. Phys.* **130**, 114506-114506 (2009).

(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).

(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).

(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).

(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, S.A. et al. Membrane-protein structure determination by solid-state NMR spectroscopy of microcrystals. *Nature methods* **9**, 1212-7 (2012).

(93) Wälti, M.A. et al. Atomic-resolution structure of a disease-relevant Aβ(1–42) amyloid fibril. *Proc. Natl. Acad. Sci. USA* **113**, E4976-E4984 (2016).

(94) Huber, M. et al. A proton-detected 4D solid-state NMR experiment for protein structure determination. *ChemPhysChem* **12**, 915-918 (2011).

(95) Andreas, L.B. et al. Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. *Proc. Natl. Acad. Sci. USA* **113**, 9187-9192 (2016).

Chapter 3

PROTEIN STRUCTURE DETERMINATION OF MICROCRYSTALLINE SYSTEMS BY MAS NMR

3.1 SARS-COV-2 nucleocapsid N-terminal domain (N^{NTD})

Content (i.e. certain figures and technical language) in § 3.1 is reprinted with permission from the published article:[1] Sucharita Sarkar, Brent Runge, Ryan W. Russell, Kumar Tekwani Movellan, Daniel Calero, Somayeh Zeinalilathori, Caitlin M. Quinn, Manman Lu, Guillermo Calero, Angela M. Gronenborn, Tatyana Polenova (**2022**) Atomic-Resolution Structure of SARS-CoV-2 Nucleocapsid Protein N-Terminal Domain. *J. Am. Chem. Soc. 144* (23), 10543-10555. DOI: 10.1021/jacs.2c03320

Author contributions are as follows: T.P. and A.M.G. conceived the project and directed the work. S.Z. designed the plasmid. B.R. prepared the protein samples for MAS NMR and X-ray diffraction. T.P. and C.M.Q. designed the MAS NMR experiments. S.S. and B.R. recorded NMR experiments and analyzed the NMR data. K.T.M. performed ^{1}H-detected experiments at the MAS frequency of 100 kHz. R.W.R. wrote scripts for structure calculations, analysis of calculation results, and structure visualizations, and with S.S. carried out structure calculations. M.L., G.C., and D.C. performed the X-ray data acquisition, analysis, and refinement of the crystal structure. All authors discussed the results. S.S, B.R., T.P., and A.M.G. took the lead in writing the manuscript.

3.1.1 Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) arose in late 2019 and is the cause of 'coronavirus disease 2019' (COVID-19). SARS-CoV-2 is a highly transmissible pathogen and has been responsible for many deaths across the globe. Although tremendous progress has been made regarding vaccine and treatment, an "ever-adapting" virus must be comprehensively studied to develop therapies, including vaccines and anti-virals, as well as a cure.

In comparison with other viruses, the overall viral architecture and lifecycle of SARS-CoV-2 is intensely studied, but atomic-resolution understanding of genome packing mechanisms are lacking. The inherent viral genome contains four structural proteins: spike (S) glycoprotein, envelope (E) protein, membrane (M) protein, and nucleocapsid (N) protein[3,4] as seen in Figure 3.1A. To date, there is no atomic resolution structure of the infectious full-length N protein, which would reveal crucial genomic organization and ribonucleoprotein formation. Such information would open the door to the development of therapies and treatment of SARS-Cov-2 infection, namely in the form of small-molecule inhibitors that could potentially suppress or disrupt viral operations.

The infectious N protein is comprised of two independently folded domains connected by a ~70 amino acid linker with intrinsically disordered regions at the N- and C-termini, shown in Figure 3.1B.[5–10] The N-terminal (N^{NTD}) and C-terminal (N^{CTD}) domains are both intricately involved with RNA genome interactions.[11–13] Herein the

atomic-resolution structure of crystalline N^{NTD} was determined by combining X-ray crystallography and solid-state magic angle spinning (MAS) NMR spectroscopy.

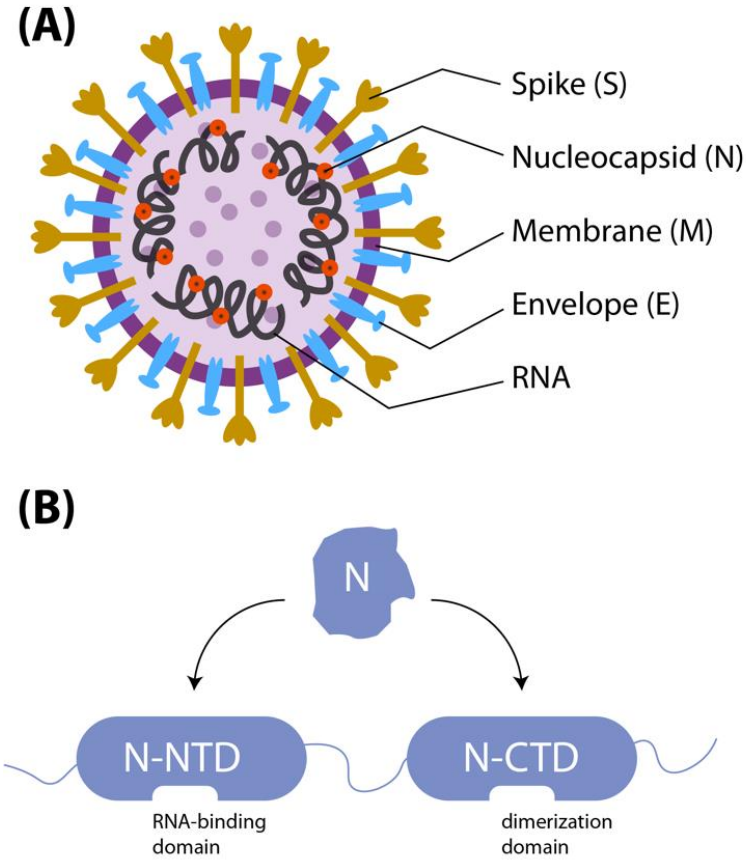

Figure 3.1: SARS-Cov-2 virus. (A) Schematic of the virus with the structural proteins and RNA labelled. (B) General organization of the nucleocapsid protein, comprising the N-terminal and C-terminal domains, the flexible linker, and the intrinsically disordered tails of the N- and C-termini.

3.1.2 Chemical shift assignments and distance restraints of SARS-CovV-2 N-terminal domain (N^{NTD})

Using only a single, fully protonated crystalline sample of N^{NTD}, the chemical shifts and the distance restraints for the structure calculation were obtained from the following experiments: 2D CORD[14], NCACX, NCOCX at 25 ms mixing time, as well as ^{1}H-detected 2D (H)NH HETCOR, 3D (H)CANH, and (H)CONH spectra. The chemical shifts were assigned for the vast majority of the residues (128 of 136). Of the seven residues missing assignments, five were partially assigned, and two were completely unassigned, likely from disorder.

In total, the structure of the single chain of N^{NTD} was calculated form 2,968 non-redundant distance restraints and 101 ϕ/ψ torsion angle restraints from TALOS-N[15]. More specifically, there were 2,197 unambiguous ^{13}C-^{13}C, 763 ^{15}N-^{13}C, and 4 ^{1}H-^{15}N distance restraints. Of these there were 968 long-range ($|i-j|\geq 5$) restraints (Table 3.1).

Table 3.1: Summary of MAS NMR restraints of SARS-CoV-2 N^{NTD}

MAS NMR distance restraints	^{13}C-^{13}C	^{15}N-^{13}C	^{1}H-^{15}N
Unambiguous	2197	763	4
Intra-residue	807	505	0
Sequential ($\|i-j\|=1$)	119	258	4
Medium range ($1<\|i-j\|<5$)	303	0	0
Long range ($\|i-j\|\geq 5$)	968	0	0
Ambiguous	4		
Total number of restraints assigned	2968 (21.8 restraints per residue)		
MAS NMR dihedral angle restraints			
Φ	101		
Ψ	101		

3.1.3 Atomic resolution MAS NMR structure of a single N^{NTD} chain

With ~22 restraints per residue, maximum accuracy and precision was obtained for the MAS NMR structure of N^{NTD}. The average pairwise RMSD was 0.7 ± 0.2 Å and 1.2 ± 0.1 Å for the backbone atoms and all heavy atoms of the protein, respectively, for the ten lowest energy structures. This remarkable accuracy and precision are amongst the highest to date, as just one of two MAS NMR investigations to obtain more than 20 restraints per residue and whose single chain is greater that 100 amino acids long.[16]

Our MAS NMR structure exhibits the overall shape of a right hand (comprised of four-stranded β-sheet) consistent with other published coronavirus N^{NTD} structures (Figure 3.2). In the center of the structure a long β-hairpin that extends and protrudes outwards from the palm. Overall, the structure is well defined except for the except for the first eight amino acids (R2-N9) and the last residue (E136), which are likely dynamic. The β-hairpin loop (I56-K64) is also presumably dynamic from the absence of long-range restraints observed in the experimental spectra.

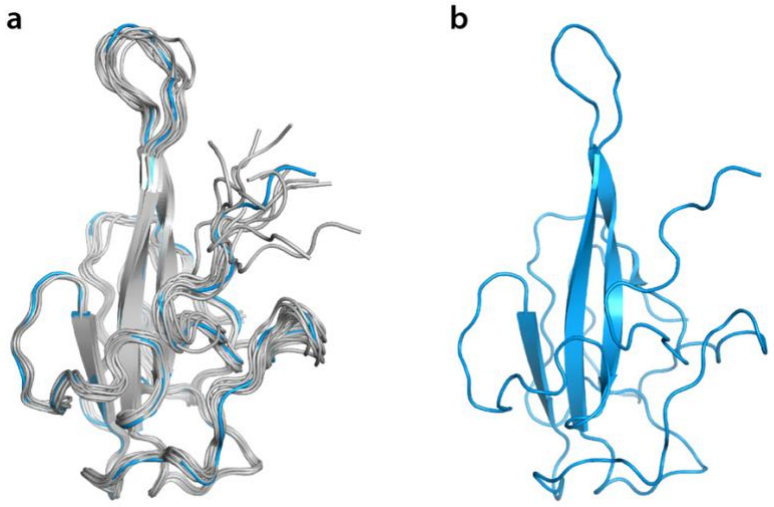

Figure 3.2: MAS NMR structure of SARS-CoV-2 N^{NTD} (A) Best-fit superposition of the ten lowest energy conformers (gray) in the MAS NMR ensemble and average structure (blue) of a single chain of SARS-CoV-2 N^{NTD}. (B) Average of ten lowest energy MAS NMR conformers (blue). Energy minimization was carried out for the average structure in Cartesian space.

Table 3.2: NMR structure statistics of SARS-CoV-2 N^{NTD}

Structure statistics from ten lowest energy subunits	
Violations (mean ± s.d.)	
Distance restraints ≥ 7.2 Å (Å)	0.144 ± 0.001
Dihedral angle restraints ≥ 5° (°)	1.528 ± 0.137
Max. distance restraint violation (Å)	1.254
Max. dihedral angle restraint violation (°)	17.267
Deviations from idealized geometry	
Bond lengths (Å)	0.008 ± 0.000
Bond angles (°)	0.774 ± 0.012
Improper angles (°)	0.516 ± 0.016
Average pairwise RMSD (Å)*	
Backbone (N, C^{α}, C')	**0.7 ± 0.2**
Heavy	**1.2 ± 0.1**

* Disordered N-terminus (residues 1-9) excluded.

3.1.4 Comparison of the MAS NMR structure and the X-ray crystal structure of N^{NTD}

As part of this study, an X-ray structure of the same sample of N^{NTD} was determined to provide complementary and validatory information. The protein crystallized in the $P2_12_12_1$ orthorhombic space group with four monomers (chains A-D) in the asymmetric unit. The average pairwise RMSD amongst the four chains is 0.5 ± 0.1 Å for the backbone atoms (excluding missing residues R2-N9, Q20-D25, R57-P68, and P124-E136). The X-ray structure revealed unique contacts about intra-tetramer interfaces, which were not available by NMR. Overall, using both techniques in concert revealed information of the packing interfaces as well as disordered regions of the RNA binding and β-hairpin.

From a validation standpoint, the MAS NMR and the X-ray structures agree well. The backbone RMSD between the X-ray structure (averaged over the four chains in the asymmetric unit) and the MAS NMR structure (averaged over the ensemble of the ten lowest energy structures) is 1.1 Å. Excluding chain D, which possesses the highest degree of disorder in the X-ray structure, the backbone RMSD becomes 0.7 Å.

Extensive details regarding the X-ray structure determination will not be discussed here as it is outside of the scope of this dissertation. However, it is worth underscoring how multiple structure determination techniques can be used synergistically.

3.1.5 Protocol for MAS NMR structure calculation of microcrystalline N^{NTD}

The MAS NMR structure of a single N^{NTD} chain was calculated in Xplor-NIH version 2.53[17–19] using ^{13}C-^{13}C, ^{15}N-^{13}C, and ^{1}H-^{15}N distance restraints, extracted from 2D CORD (100ms, 250 ms, and 500 ms mixing times), NCACX, NCOCX, and (H)NH HETCOR spectra and backbone dihedral angles predicted by TALOS-N[15] from the experimental ^{1}H, ^{13}C, and ^{15}N chemical shifts. The bounds for the distance restraints were set to 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, consistent with our previous studies[16,20].

Calculations were seeded using the primary sequence as extended strands. 1,000 structures were generated with molecular dynamics simulated annealing in the torsion angle space with two successive annealing schedules and a final gradient minimization in Cartesian space, essentially as described previously[16,20] and detailed below.

Two successive annealing schedules were used, the first in a vacuum with the REPEL module and the second with an implicit solvent refinement using the EEFx module[21]. The ten lowest energy structures were selected and served as input for the second schedule, and the ten lowest energy structures of this as input for the final ensemble (PDB: 7SD4). Standard terms for bond lengths, bond angles, and improper angles were applied to enforce correct covalent geometry.

The first annealing calculation was essentially identical to that reported previously[16,20], with initial random velocities at a 3,500 K constant temperature molecular dynamics run for the shorter of 800 ps or 8,000 steps, with the time step size

allowed to float to maintain constant energy. Subsequently, simulated annealing calculations at reduced temperatures in steps of 25 K to 100 K were carried out for the shorter of 0.4 ps or 200 steps. Force constants for distance restraints were ramped up from 10 to 50 kcal/mol•Å^2. Dihedral angle restraints were disabled for high-temperature dynamics at 3,500 K and subsequently applied with a force constant of 200 kcal/mol•rad^2. The force constant for the radius of gyration was geometrically scaled from 0.002 to 1, and a hydrogen bond term, HBPot, was used to improve hydrogen bond geometries[22]. After simulated annealing, structures were minimized using a Powell energy minimization scheme.

For the second schedule performed in the implicit solvent, all parameters were set in accordance with the EEFx example packaged with Xplor-NIH. Annealing was performed at 3,500 K for 15 ps or 15,000 steps, whichever was completed first. The starting time step was 1 fs and was self-adjusted in subsequent steps to ensure conservation of energy. Random initial velocities were assigned about a Maxwell distribution at the starting temperature of 3,500 K. Subsequently the temperatures were reduced to 25 K in steps of 12.5 K. At each temperature, 0.4 ps dynamics were run with an initial time step of 1 fs. Force constants for distance restraints were ramped up from 2 to 30 kcal/mol•Å^2. The dihedral restraint force constants were set to 10 kcal/mol•rad^2 for high-temperature dynamics at 3,000 K and 200 kcal/mol•rad^2 during cooling. After the EEFx module, structures were minimized using a Powell energy minimization scheme.

Atomic RMSD values were calculated using routines in Xplor-NIH (version 2.53)[17–19]. The visualization of structural ensembles was rendered in PyMOL[23], using in-house shell/bash scripts. Secondary structure elements were classified according to STRIDE[24] and manual inspection.

3.1.6 Conclusions and outlook

The structure presented here can provide guidance for therapeutics of SARS-CoV-2 by displaying features not previously available. In addition, the methodological protocol underscores how X-ray crystallography and MAS NMR can be used in concert to reveal functionally important regions not accessible from an individual technique alone.

3.2 MAS NMR structure of HIV-1 CA_{CTD}-SP1 and maturation inhibitors

Content (i.e. certain figures and technical language) in § 3.2 is reprinted with permission from the published article: Sucharita Sarkar, Kaneil K. Zadrozny, Roman Zadorozhnyi, Ryan W. Russell, Caitlin M. Quinn, Alex Kleinpeter, Sherimay Ablan, Hamed Meshkin, Juan R. Perilla, Eric O. Freed, Barbie K. Ganser-Pornillos, Owen Pornillos, Angela M. Gronenborn, Tatyana Polenova (**2022**) Structural Basis of HIV-1 Maturation Inhibitor Binding and Activity. *Nat. Comms.*, DOI: TBD

Author contributions are as follows: T. P., A. M. G., B. K. G.-P., and O. P. conceived the project and guided the work. S. S. and R. Z. performed NMR experiments and analyzed the experimental data. K. K. Z. prepared the samples. R. W. R. and S. S. performed the structure calculations. J. R. P. designed and guided the MD simulations. H. M. and J. R. P. parameterized the force fields for BVM. S. S., R. W. R., R. Z., H. M., and K. K. Z. prepared figures for the manuscript. R. W. R. wrote scripts for structure calculations, analysis of calculation results and structure visualizations. C. M. Q. took part in the design or analysis of NMR experiments. E. O. F. provided A1V and V7A viral sequence polymorphs and critical feedback on data analysis. A. K. and S. A. performed infectivity and BVM binding studies of BVM-resistant variants. T. P., A. M. G., B. K. G.-P., and O. P. took the lead in writing the manuscript. All authors discussed the results and contributed to the manuscript preparation.

3.2.1 HIV-1 maturation and maturation inhibitors

The maturation of the HIV-1 virus is a critical process in its lifecycle necessary for the formation of infectious virions. Despite extensive efforts, many key mechanistic aspects of maturation remain poorly understood.[25]

Maturation occurs through a proteolytic cleavage cascade of the Gag polyprotein by the viral protease and results in Gag lattice remodeling (Figure 3.3). Gag is comprised of matrix (MA), capsid (CA), nucleocapsid (NC), and p6 domains as well as spacer peptides 1 and 2 (SP1 and SP2). The final maturation step is the cleavage of a 14-residue SP1, leading to the formation of a conical capsid that harbors the NC-stabilized virial RNA.[26–29] The conical capsids are pleomorphic with varied shapes and stoichiometries: the typical stoichiometry is 216 hexamers and 12 pentamers formed by the 231-residue CA capsid protein.[30]

Understanding maturation is of great interest because it is an attractive target for anti-HIV therapies using small-molecule inhibitors. For example, infectivity is inhibited by the small molecule bevirimat (BVM) that prevents the final catalytic cleavage of SP1, by presumably binding directly to the SP1 domain.[31] Yet, direct atomic-level evidence was lacking, prior to this investigation.

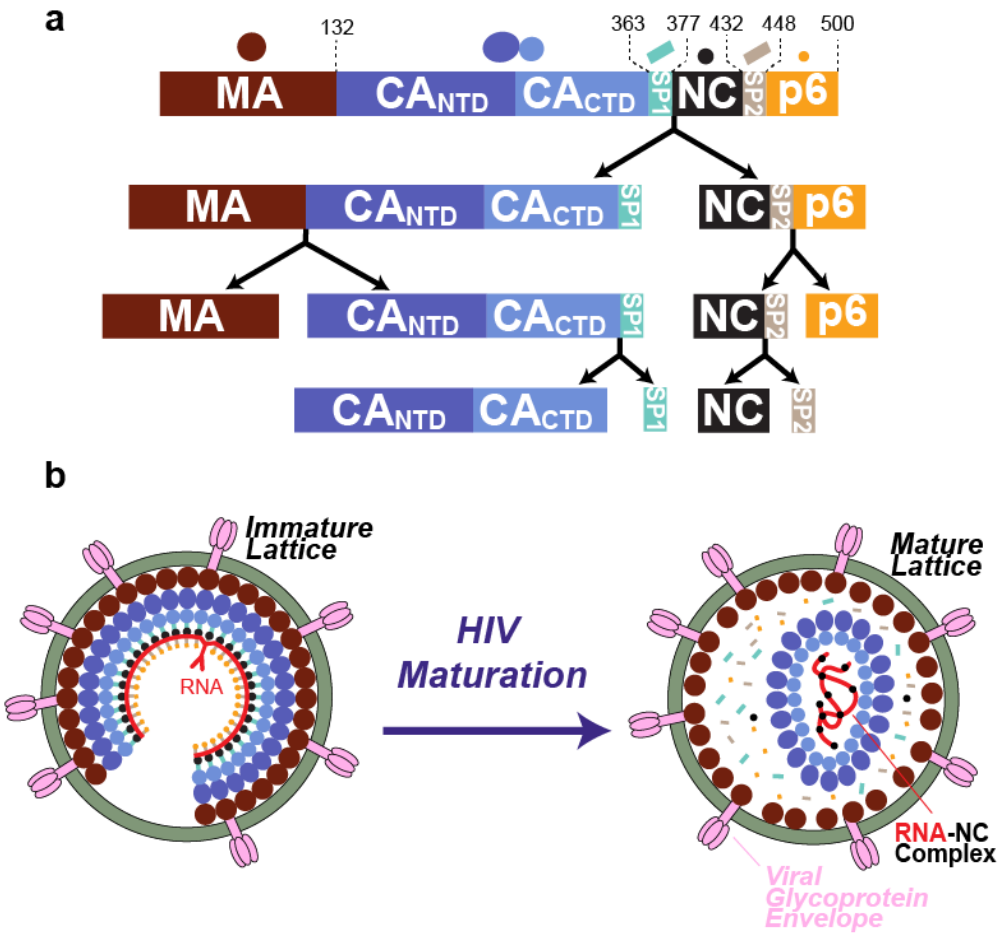

Figure 3.3: (A) Schematic representation of the proteolytic cleavage of Gag during maturation. (B) Illustration of the Gag lattice remodeling during maturation of HIV-1. Drawn using Adobe Illustrator.

The slowest step of maturation is the final cleavage of SP1 and CA as an outcome of the proteolysis site.[32–35] Within the immature HIV-1 Gag lattice (Figure 3.3A) the CA-SP1 self-associates in a six-helix bundle for stabilization of the Gag hexamer.[35,36] Moreover, for protease access of the site, there must be partial unfolding, primarily carried out by the junction of CA and SP1. It is thought that maturation

inhibitors (Mis)' 3'*O*-(3',3'-dimethylsuccinyl)-betulinic acid (bevirimat or BVM), 1-[2-(4-tert-butylphenyl)-2-(2,3-dihydro-1H-inden-2-ylamino)ethyl]-3-(trifluoromethyl)pyridin-2-one (PF-46396) and their analogs for example, do not interfere with substrate binding but rather disrupt/prevent unfolding of the six-helix bundle.[37–42]

In addition, the six-helix bundle is stabilized by Inositol hexakisphosphate (IP6) a capsid assembly cofactor. IP6 is found abundantly in cellular environments and spontaneously binds/stabilizes the six-helix bundle. Unlike BVM, which binds in the center of the pore of the six-helix bundle, as reported in previous studies, IP6 binds just above the six-helix bundle and forms slat bridges with two lysine side chains.[36,38,43]

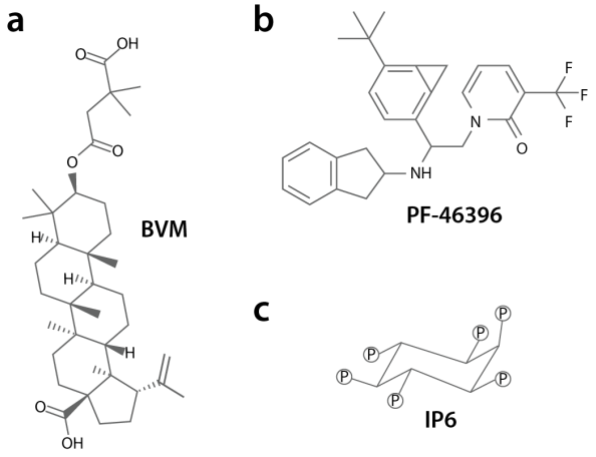

Figure 3.4: Binding molecules of the CA-SP1 six-helix bundle: (A) maturation inhibitor bevirimat, (B) maturation inhibitor PF-46396, (C) IP6 assembly co-factor.

To probe the mechanism of action of MIs at the CA-SP1 site, we determined the magic angle spinning (MAS) NMR atomic resolution structures of microcrystalline complexes of a HIV-1 Gag fragment. We solved the structure of the C-terminal domain (CA_{CTD}) and SP1 regions (CA_{CTD}-SP1) bound with BVM and/or IP6.

3.2.2 Calculation input: chemical shift assignments, dihedral restraints, and distance restraints

High-resolution MAS NMR spectra was recorded. In total, 96 % of all backbone atoms were assigned from the spectra and 8377 cross-peaks were assigned amongst various experiments. Furthermore, 3,071 non-redundant, unambiguous protein-protein distance restraints (^{13}C-^{13}C, ^{15}N-^{13}C, ^{13}C-^{1}H, and ^{15}N-^{1}H) were obtained. These comprised 641 medium-range ($1<|i-j|<4$), 610 long-range ($|i-j|\geq5$), 32 long-range inter-chain, and 20 long-range inter-hexamer restraints (Table 3.3). This equates to nearly 30 non-redundant unambiguous distance restraints per residue, which is greater than any protein MAS NMR investigation to date.

Table 3.3: Summary of MAS NMR restraints and structure statistics of CA_{CTD}-SP1/BVM/IP6

BVM-Protein restraints	$^{1}H(BVM)-^{13}C(protein)$			
Unambiguous	7			
IP6-Protein restraints (For CA_{CTD}-SP1/BVM/IP6)	$^{1}H(IP6)-^{13}C(protein)$		$^{31}P(IP6)-^{1}H(protein)$	
Unambiguous	3		3	
IP6-Protein restraints (For CA_{CTD}-SP1/BVM/IP6)	$^{1}H(IP6)-^{13}C(protein)$		$^{31}P(IP6)-^{1}H(protein)$	
Unambiguous	6		0	
Protein distance restraints	$^{13}C-^{13}C$	$^{15}N-^{13}C$	$^{15}N-^{1}H$	$^{13}C-^{1}H$
Unambiguous	2125	537	91	295
intra-residue	595	358	82	259
Sequential ($\|i-j\| = 1$)	227	132	7	26
Medium range ($1<\|i-j\|<4$)	641	29	1	3
Long range ($\|i-j\|\geq5$)	610	11	0	6
Long range ($\|i-j\|\geq5$) (inter-chain)	32	7	0	0
Long range ($\|i-j\|\geq5$) (inter-hexamer)	20	0	1	1
Ambiguous	117	4	6	2
Intra-residue	22	1	6	2
Sequential ($\|i-j\| = 1$)	13	1	0	0
Medium range ($1<\|i-j\|<4$)	34	2	0	0
Long range ($\|i-j\|\geq5$)	48	0	0	0
Total number of unambiguous restraints	3071			
Restraints/residue	30			
Percent completeness	52% (C-C only)			
Dihedral angle restraints				
ϕ	90			
ψ	90			

3.2.3 Structure of a single CA_{CTD}-SP1 chain

The single-chain structure of CA_{CTD}-SP1 was calculated using MAS NMR distance and dihedral restraints using Xplor-NIH version 2.53[17–19]. Folding calculations were seeded from primary sequence extended strands. One thousand structures were calculated using molecular dynamics simulated annealing in the torsion angle space with two successive annealing schedules and a final gradient minimization in the Cartesian space. The structure calculation began with a 3500 K constant-temperature molecular dynamics run for the shorter of 800 ps or 8,000 steps with the time step size allowed to float to maintain constant energy, within a tolerance. The initial velocities were randomized about a Maxwell distribution using a starting temperature of 3,500 K. Following this initial molecular dynamics calculation, a simulated annealing calculation was performed where the temperature was reduced to 100 K in steps of 25 K. At each temperature, dynamics were run for the shorter of 0.4 ps or 200 steps. Force constants for distance restraints were ramped up from 10 to 50 kcal mol^{-1} $Å^{-2}$. The dihedral angle restraints were disabled for high-temperature dynamics at 3500 K but enabled during simulated annealing with a force constant of 200 kcal mol^{-1} rad^{-2}. The gyration volume force constant[44] was geometrically scaled from 0.002 to 1. The torsion angle database[45] and HBPot[22] were also used. After simulated annealing, the structures were minimized using a Powell energy minimization scheme.

Subsequently, the ten lowest energy structures were selected for further refinement where 1,000 structures were refined in total. Annealing was performed at

3,000 K for 10 ps or 5,000 steps, whichever was completed first. The starting time step was 1 fs and was self-adjusted in subsequent steps to ensure conservation of energy. The initial velocities were randomized about a Maxwell distribution using the starting temperature of 3,000 K. The temperature was subsequently reduced to 25 K in steps of 12.5 K. At each temperature, the initial default time step was 1 fs, and a 0.2-ps dynamics run was performed. Force constants for distance restraints were ramped from 2 to 30 kcal mol^{-1} $Å^{-2}$. The dihedral restraint force constants were set to 10 kcal mol^{-1} rad^{-2} for high-temperature dynamics at 3,000 K and 200 kcal mol^{-1} rad^{-2} during cooling. The gyration volume force constant was scaled from 0.002 to 1. The torsion angle database and HBPot were also used. The annealed structures were minimized using a Powell energy minimization scheme.

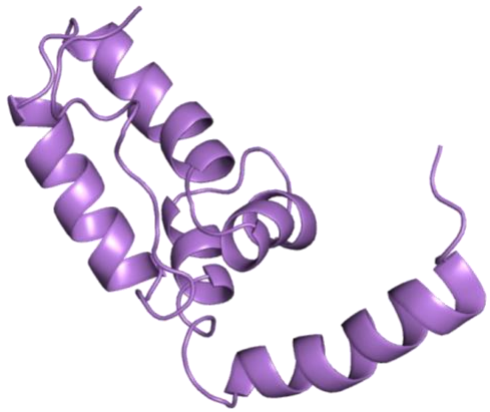

Figure 3.5: Lowest-energy structure of the single-chain calculation of CA_{CTD}-SP1 used for subsequent docking in § 3.2.4.

3.2.4 Docking of the single chain in X-ray density

The lowest energy single-chain structure calculated as described above was subjected to rigid-body docking into the envelope of the hexamer-of-hexamers (§ 1.2.5). The docking was performed using an in-house UCSF Chimera[46] script (Listing 3.1). Specifically, the 42 best positions (from 7 hexamer units) for docking of single-chain structures, were identified on the map, on the basis of lowest cross-correlation values and brief visual inspection. Prior to docking, the density was prepared using the "molmap" routine in UCSF Chimera.

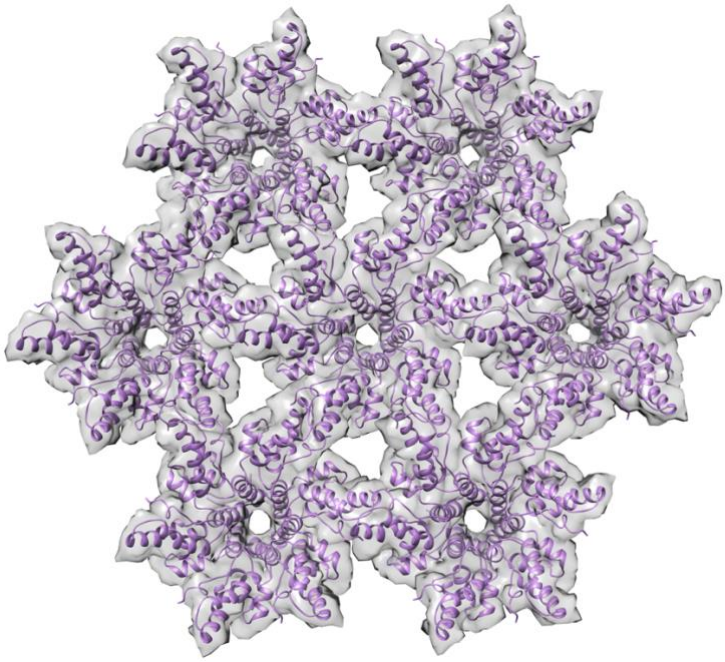

Figure 3.6: Batch docking of the experimental X-ray density map[47] of the lowest energy structure for the single-chain calculation. Shown here are the 42 positions identified on the basis of lowest cross-correlation values and brief visual inspection.

```
from chimera import runCommand as run, openModels
import re
import sys
import os

# Example of how to run this script:
# /Applications/Chimera.app/Contents/MacOS/chimera --nogui --script
dock_chimera.py

# Load in pdb and density for docking
# Must have _ after number for strucIDs, otherwise just name it manually
in_pdb = "refine1_853.pdb"
stucIDs = re.search("\\d+_", in_pdb, re.M).group(0)

run('open ./54IT_noBVM_HOH_molmap_8A_resid148to238.mrc')
run(str("open ") + str(in_pdb))
# Set step level and make density a surface
run('volume #0 step 1')
run('volume #0 level 0.0063 style surface')
# Set this number high so no slots/fits missed
strucs_in_tube = "350"

# Set indices and prepare for docking
map1_id = 1
map2_id = 0
# Set number of translations and rotations cross-correlation values
(optional)
search = 250000
# Provide resolution. Script will run fine if this is approximate, the cross-
correlation
# values will be off but the values still will reveal the fits.
res = 8

# Execute the docking
from chimera import openModels as om, selection
m1 = om.list(id = map1_id)[0]
m2 = om.list(id = map2_id)[0]
s1 = selection.ItemizedSelection([m1])
from FitMap.fitcmd import fitmap
fit_list = fitmap(s1, m2, search = search, resolution = res, listFits =
False)

# Access cross-correlation value for each docked structure
print '%d fits' % len(fit_list)
import Matrix
corrs =  []
for index, fit in enumerate(fit_list):
    if int(index) <= int(strucs_in_tube):
       print 'correlation =', fit.correlation()
    corrs.append(fit.correlation())
    if int(index) <= int(strucs_in_tube):
       fit.place_copies()
```

```
# Save pdb coordinates of fit with cross-correlation in the file-name
if not os.path.exists("Docked"):
               os.makedirs("Docked")
for index, fit in enumerate(fit_list):
               if int(index) <= int(strucs_in_tube):
                    struc_ind = float(index) + 2
                    print int(index), struc_ind
                    outname = "Docked/docked_StrucNum_" + str(int(struc_ind))
               + "_Corr_" + str(round(fit.correlation(), 4)) + ".pdb"
                    print outname
                    run(str("write ") + str(struc_ind) + str(" ") + outname)
```

Listing 3.1: UCSF Chimera script for batch docking into cryo-EM density.

3.2.5 Refinement of the seven hexamer units with BVM/IP6

After docking, a calculation was performed to identify the precise location of the IP6 and BVM ligands as well as to incorporate additional distance restraints between chains and hexamer units. The calculation was seeded from single-chain CA_{CTD}-SP1 coordinates calculated from the experimental MAS NMR restraints (see above), together with the coordinates of BVM and/or IP6 generated as described above. The placement of the molecules inside a single hexamer was estimated by visual inspection to allow the protein-ligand distance restraints to be applied properly. The coordinates were expanded from a single hexamer to a hexamer of hexamers unit containing seven hexamers (42 chains) using the "symexp" command in PyMol[23].

100 structures underwent torsion angle dynamics with an annealing schedule and a final gradient minimization in Cartesian space. The force field parameterization of the IP6 and BVM molecules were incorporated into the run via topology and parameter files, prepared specifically for Xplor-NIH. The BVM and IP6 molecules were

free to move as rigid bodies during dynamics and final minimization. Two identical runs of simulated annealing starting at 3,000 K were performed for 10 ps, with a time step of 1 fs. The initial velocities were randomized to achieve a Maxwell distribution at a starting temperature of 3,000 K. The temperature was subsequently reduced to 25 K in steps of 25 K. At each temperature step, dynamics were run for 400 fs with an initial time step of 1 fs.

Standard terms for bond lengths, bond angles, and improper angles were used to enforce proper covalent geometry. Standard potentials were used to incorporate distance and dihedral restraints.

A cross-correlation probability distribution potential often utilized for experimental cryo-EM density[20,48] enforced/conceded the overall shape and boundary of the hexamer of hexamers with the 8-Å density map (§ 1.2.6). The potential was restricted to backbone atoms (N, C, CA, and O) to ensure the density boundary would not influence sidechain conformations.

A statistical torsion-angle potential[45] was employed, and the gyration volume term was not included to avoid conflict with the cross-correlation density potential. A hydrogen-bond database term, HBPot, was used to improve hydrogen-bond geometries[22]. Approximate non-crystallographic symmetry was imposed using Xplor-NIH's PosDiffPot term, allowing the subunits of the hexamer to differ by up to 1 Å.

Force constants for distance restraints were ramped from 2 to 30 kcal/mol•Å^2. The dihedral restraint force constants were set to 10 kcal/mol•rad^2 for high-temperature

dynamics at 3,000 K and 200 kcal/mol•rad^2 during cooling. The force constants of the cross-correlation probability distribution potential were set to 50 kcal/mol during high temperature dynamics and cooling.

After the high-temperature dynamics and cooling in dihedral space, the annealed structures were minimized using a Powell energy minimization scheme in Cartesian space. The final MAS NMR bundle comprised the five lowest energy structures of the 100 calculated ones.

RMSD values were calculated using routines in the Xplor-NIH (version 2.51). The visualizations of structural elements were batch rendered in PyMOL using in-house shell/bash scripts. Secondary structure elements were classified according to TALOS-N.[15]

3.2.6 Structure of CA_{CTD}-SP1 with IP6 and/or BVM

The structure of the CA_{CTD}-SP1 hexamer determined here takes the shape of a goblet, as reported previously[47,49–51], where the CA_{CTD} domain is the cup and the six-helix bundle of the CA-SP1 junction is the stem. Within the structure of CA_{CTD}-SP1, the SP1 domain is well defined (except M14) despite the conformational disorder.[47,50,51] The calculated structures exhibit excellent precision corroborated by the low RMSDs and structure statistics of the five lowest energy structures. (Table 3.4., Figure 3.7). Undoubtedly, this is a direct outcome of the very large number of restraints (Table 3.3), consistent with the results demonstrated in Chapter 2.

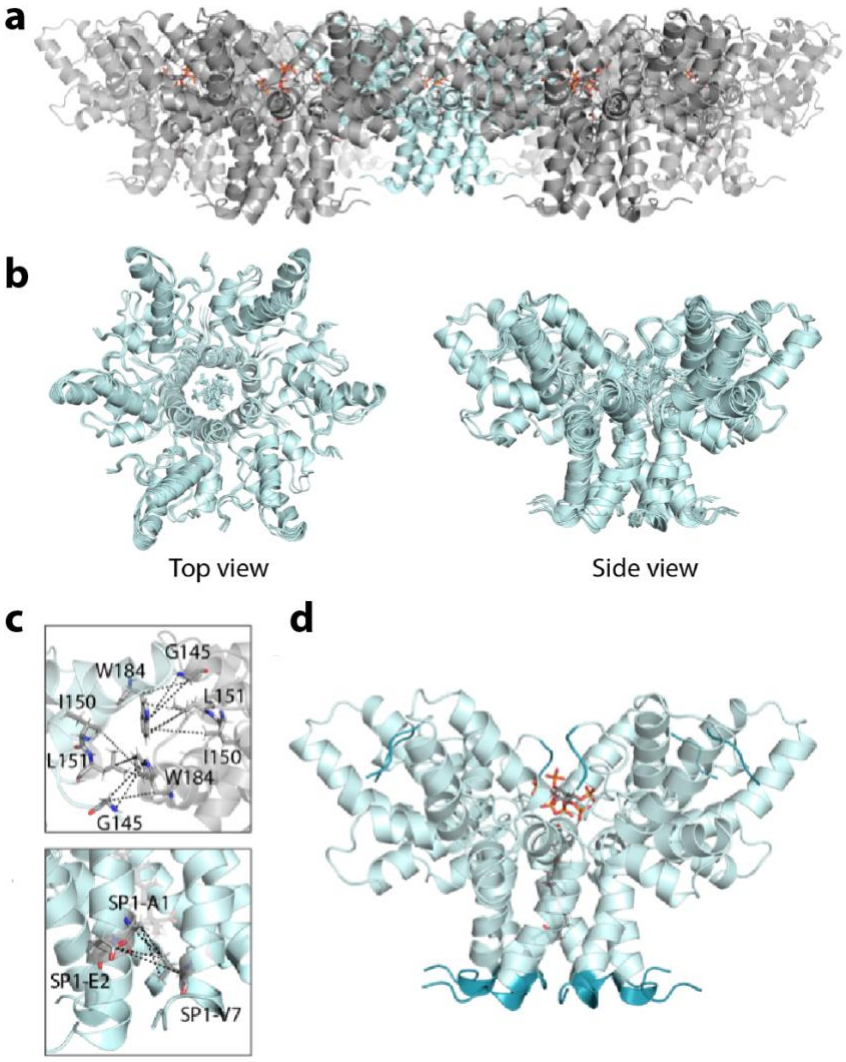

Figure 3.7: MAS NMR structure of CA_{CTD}-SP1 crystalline array. (A) Side view of hexamer of hexamers of BVM- and IP6-bound CA_{CTD}-SP1 arrays. (B) Superposition of 5 lowest energy structures of central hexamer of CA_{CTD}-SP1/BVM/IP6 crystalline arrays. (C) Expansion of inter-hexamer (top panel) and inter-chain (bottom panel) regions showing distance restraints obtained from MAS NMR correlation experiments. (D) MAS NMR structure of a single hexamer of BVM and IP6-bound CA_{CTD}-SP1 crystalline array. The residues detected by MAS NMR and

not modeled in the X-ray and cryo-EM structures[47,52] are shown in darker cyan.

Table 3.4: Structure statistics for CA_{CTD}-SP1/BVM/IP6 and CA_{CTD}-SP1/BVM

Structure statistics	
CA_{CTD}-SP1/BVM/IP6	
Violations (mean ± s.d.)	
Distance restraints ≥ 7.2 Å (Å)	0.153 ± 0.002
Dihedral angle restraints ≥ 5° (°)	2.579 ± 0.088
Max. protein-protein distance restraint violation* (Å)	1.892
Max. protein-ligand distance restraint violation* (Å)	2.787
Max. dihedral angle restraint violation* (°)	14.295
Deviations from idealized geometry	
Bond lengths (Å)	0.012 ± 0.000
Bond angles (°)	0.999 ± 0.014
Improper angles (°)	0.964 ± 0.016
Average pairwise RMSD (Å)	
Heavy	1.1 ± 0.1
Backbone (N, Cα, C)	0.9 ± 0.1
CA_{CTD}-SP1/IP6	
Violations (mean ± s.d.)	
Distance restraints ≥ 7.2 Å (Å)	0.140 ± 0.001
Dihedral angle restraints ≥ 5° (°)	2.646 ± 0.095
Max. protein-protein distance restraint violation* (Å)	2.444
Max. protein-ligand distance restraint violation* (Å)	no violations
Max. dihedral angle restraint violation* (°)	16.020
Deviations from idealized geometry	
Bond lengths (Å)	0.010 ± 0.000
Bond angles (°)	0.970 ± 0.006
Improper angles (°)	0.969 ± 0.019
Average pairwise RMSD (Å)	
Heavy	1.3 ± 0.2
Backbone (N, Cα, C)	1.1 ± 0.2

*Pairwise RMSD was calculated among 5 lowest energy central hexamers

The calculated structures revealed the important mechanisms of BVM binding as well as atomic-scale structure details including side-chain conformations. Not to exceed the scope of this dissertation, centered upon the methods of the structure

calculations, select findings will only be introduced briefly to showcase the knowledge from structure calculations.

Comparing the central hexamers of the lowest energy structure from CA_{CTD}-SP1/IP6/BVM and CA_{CTD}-SP1/IP6 reveals simultaneous binding of BVM and IP6 in the pore of the CA_{CTD}-SP1. Clearly from Figure 3.7-3.8, the orientation and precise location of BVM was determined unambiguously. Comparison of the two structures also reveals that in the presence of BVM, CA_{CTD}-SP1 undergoes pore tightening (Figure 3.8) and side-chain reorientation (Figure 3.9).

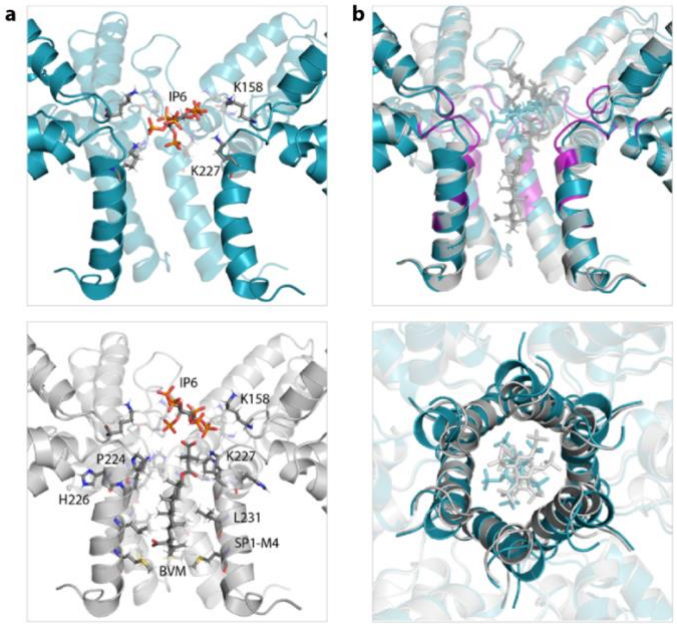

Figure 3.8: MAS NMR structure of BVM- and IP6-bound CA_{CTD}-SP1 (A) Top panel: IP6 binding mode in the hexamer of CA_{CTD}-SP1/IP6 assemblies (PDB 7R7Q, this work). Bottom panel: IP6 and BVM binding modes in the hexamer of CA_{CTD}-SP1/BVM/IP6 assemblies (PDB 7R7P, this work). Residues interacting with IP6 or BVM are shown as sticks. (B) Superposition of MAS NMR structure of CA_{CTD}-SP1/BVM/IP6 and CA_{CTD}-SP1/IP6 shown from side view (top) and top view (bottom).

BVM binding induces major structural rearrangements of the SP1 helices, resulting in the tightening of the pore and quenching the motions of the simultaneously bound IP6. Residues colored in magenta give rise to high-intensity peaks corresponding to intra- and inter-residue correlations upon BVM binding.

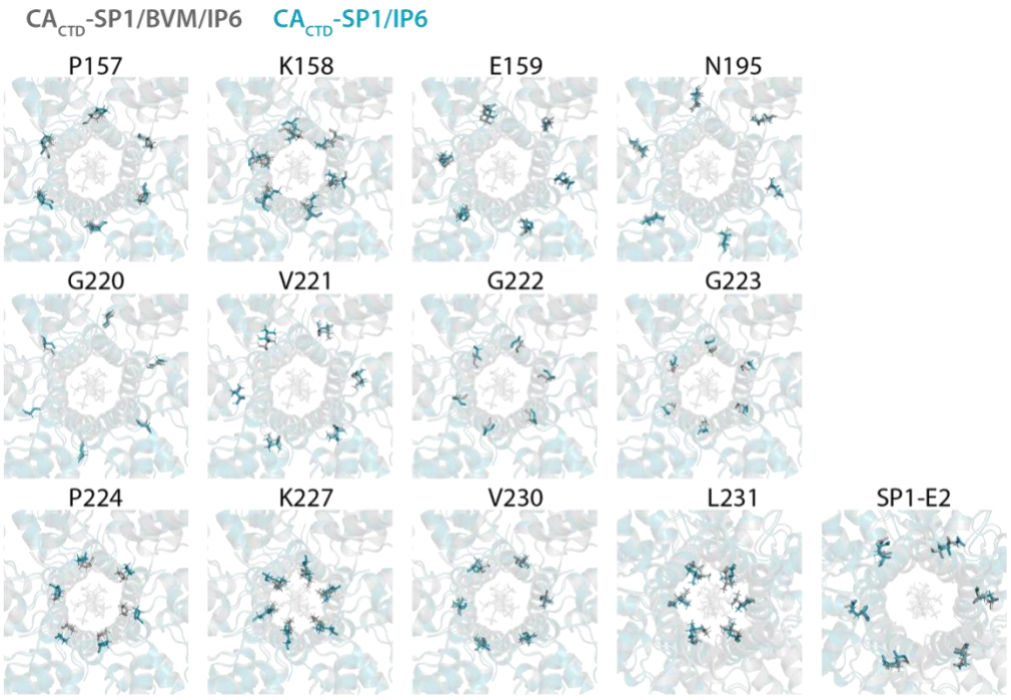

Figure 3.9: Reorientation of side chains in CA$_{CTD}$-SP1 crystalline arrays induced by BVM binding.

3.2.7 Conclusions and outlook

In conclusion, the structures of CA$_{CTD}$-SP1/IP6/BVM and CA$_{CTD}$-SP1/IP6 assemblies were determined from MAS NMR. The structures revealed the binding position of BVM and many other important features that were not established in prior studies. Our results illustrate how MAS NMR experiments can produce critical

information not available from X-ray crystallography, such as binding positions and local details, including sidechain conformational changes and the precise location of bound ligands, underscoring the power of MAS NMR structure determination.

REFERENCES

(1) Sarkar, S.; Runge, B.; Russell, R. W.; Movellan, K. T.; Calero, D.; Zeinalilathori, S.; Quinn, C. M.; Lu, M.; Calero, G.; Gronenborn, A. M.; Polenova, T. Atomic-Resolution Structure of SARS-CoV-2 Nucleocapsid Protein N-Terminal Domain. *J. Am. Chem. Soc.* **2022**, *144* (23), 10543–10555. https://doi.org/10.1021/jacs.2c03320.

(2) Sarkar, S.; Zadrozny, K. K.; Zadorozhnyi, R.; Russell, R. W.; Quinn, C. M.; Kleinpeter, A.; Ablan, S.; Xu, C.; Perilla, J. R.; Freed, E. O.; Ganser-Pornillos, B. K.; Pornillos, O.; Gronenborn, A. M.; Polenova, T. Structural Basis of HIV-1 Maturation Inhibitor Binding and Activity. bioRxiv February 22, 2022, p 2022.02.22.481470. https://doi.org/10.1101/2022.02.22.481470.

(3) Brian, D. A.; Baric, R. S. Coronavirus Genome Structure and Replication. In *Coronavirus Replication and Reverse Genetics*; Enjuanes, L., Ed.; Current Topics in Microbiology and Immunology; Springer: Berlin, Heidelberg, 2005; pp 1–30. https://doi.org/10.1007/3-540-26765-4_1.

(4) Cui, J.; Li, F.; Shi, Z.-L. Origin and Evolution of Pathogenic Coronaviruses. *Nat Rev Microbiol* **2019**, *17* (3), 181–192. https://doi.org/10.1038/s41579-018-0118-9.

(5) Chang, C.; Hou, M.-H.; Chang, C.-F.; Hsiao, C.-D.; Huang, T. The SARS Coronavirus Nucleocapsid Protein – Forms and Functions. *Antiviral Research* **2014**, *103*, 39–50. https://doi.org/10.1016/j.antiviral.2013.12.009.

(6) Cubuk, J.; Alston, J. J.; Incicco, J. J.; Singh, S.; Stuchell-Brereton, M. D.; Ward, M. D.; Zimmerman, M. I.; Vithani, N.; Griffith, D.; Wagoner, J. A.; Bowman, G. R.; Hall, K. B.; Soranno, A.; Holehouse, A. S. The SARS-CoV-2 Nucleocapsid Protein Is Dynamic, Disordered, and Phase Separates with RNA. *Nat Commun* **2021**, *12* (1), 1936. https://doi.org/10.1038/s41467-021-21953-3.

(7) Chang, C.; Chen, C.-M. M.; Chiang, M.; Hsu, Y.; Huang, T. Transient Oligomerization of the SARS-CoV N Protein – Implication for Virus Ribonucleoprotein Packaging. *PLOS ONE* **2013**, *8* (5), e65045. https://doi.org/10.1371/journal.pone.0065045.

(8) Chen, I.-J.; Yuann, J.-M. P.; Chang, Y.-M.; Lin, S.-Y.; Zhao, J.; Perlman, S.; Shen, Y.-Y.; Huang, T.-H.; Hou, M.-H. Crystal Structure-Based Exploration of the Important Role of Arg106 in the RNA-Binding Domain of Human Coronavirus OC43 Nucleocapsid Protein. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2013**, *1834* (6), 1054–1062. https://doi.org/10.1016/j.bbapap.2013.03.003.

(9) Lo, Y.-S.; Lin, S.-Y.; Wang, S.-M.; Wang, C.-T.; Chiu, Y.-L.; Huang, T.-H.; Hou, M.-H. Oligomerization of the Carboxyl Terminal Domain of the Human Coronavirus 229E Nucleocapsid Protein. *FEBS Letters* **2013**, *587* (2), 120–127. https://doi.org/10.1016/j.febslet.2012.11.016.

(10) Chang, C.; Sue, S.-C.; Yu, T.; Hsieh, C.-M.; Tsai, C.-K.; Chiang, Y.-C.; Lee, S.; Hsiao, H.; Wu, W.-J.; Chang, W.-L.; Lin, C.-H.; Huang, T. Modular

Organization of SARS Coronavirus Nucleocapsid Protein. *J Biomed Sci* **2006**, *13* (1), 59–72. https://doi.org/10.1007/s11373-005-9035-9.

(11) Schiavina, M.; Pontoriero, L.; Uversky, V. N.; Felli, I. C.; Pierattelli, R. The Highly Flexible Disordered Regions of the SARS-CoV-2 Nucleocapsid N Protein within the 1–248 Residue Construct: Sequence-Specific Resonance Assignments through NMR. *Biomol NMR Assign* **2021**, *15* (1), 219–227. https://doi.org/10.1007/s12104-021-10009-8.

(12) Yang, M.; He, S.; Chen, X.; Huang, Z.; Zhou, Z.; Zhou, Z.; Chen, Q.; Chen, S.; Kang, S. Structural Insight Into the SARS-CoV-2 Nucleocapsid Protein C-Terminal Domain Reveals a Novel Recognition Mechanism for Viral Transcriptional Regulatory Sequences. *Frontiers in Chemistry* **2021**, *8*.

(13) Peng, Y.; Du, N.; Lei, Y.; Dorje, S.; Qi, J.; Luo, T.; Gao, G. F.; Song, H. Structures of the SARS-CoV-2 Nucleocapsid and Their Perspectives for Drug Design. *The EMBO Journal* **2020**, *39* (20), e105938. https://doi.org/10.15252/embj.2020105938.

(14) 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. *Journal of Magnetic Resonance* **2013**, *232*, 18–30. https://doi.org/10.1016/j.jmr.2013.04.009.

(15) Shen, Y.; Bax, A. Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *Journal of Biomolecular NMR* **2013**, *56* (3), 227–241. https://doi.org/10.1007/s10858-013-9741-y.

(16) Russell, R. W.; Fritz, M. P.; Kraus, J.; Quinn, C. M.; Polenova, T.; Gronenborn, A. M. Accuracy and Precision of Protein Structures Determined by Magic Angle Spinning NMR Spectroscopy: For Some 'with a Little Help from a Friend.' *J Biomol NMR* **2019**. https://doi.org/10.1007/s10858-019-00233-9.

(17) Schwieters, C. D.; Bermejo, G. A.; Clore, G. M. Xplor-NIH for Molecular Structure Determination from NMR and Other Data Sources. *Protein Science* **2018**, *27* (1), 26–40. https://doi.org/10.1002/pro.3248.

(18) Schwieters, C. D.; Kuszewski, J. J.; Marius Clore, G. Using Xplor–NIH for NMR Molecular Structure Determination. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2006**, *48* (1), 47–62. https://doi.org/10.1016/j.pnmrs.2005.10.001.

(19) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Marius Clore, G. The Xplor-NIH NMR Molecular Structure Determination Package. *Journal of Magnetic Resonance* **2003**, *160* (1), 65–73. https://doi.org/10.1016/S1090-7807(02)00014-9.

(20) Lu, M.; Russell, R. W.; Bryer, A. J.; Quinn, C. M.; Hou, G.; Zhang, H.; Schwieters, C. D.; Perilla, J. R.; Gronenborn, A. M.; Polenova, T. Atomic-

Resolution Structure of HIV-1 Capsid Tubes by Magic-Angle Spinning NMR. *Nature Structural & Molecular Biology* **2020**, *27* (9), 863–869. https://doi.org/10.1038/s41594-020-0489-2.

(21) Tian, Y.; Schwieters, C. D.; Opella, S. J.; Marassi, F. M. A Practical Implicit Solvent Potential for NMR Structure Calculation. *J Magn Reson* **2014**, *243*, 54–64. https://doi.org/10.1016/j.jmr.2014.03.011.

(22) Schwieters, C. D.; Bermejo, G. A.; Clore, G. M. A Three-Dimensional Potential of Mean Force to Improve Backbone and Sidechain Hydrogen Bond Geometry in Xplor-NIH Protein Structure Determination. *Protein Science* **2020**, *29* (1), 100–110. https://doi.org/10.1002/pro.3745.

(23) *The PyMOL Molecular Graphics System, version 2.0; Schrödinger, LLC, 2000.*

(24) Heinig, M.; Frishman, D. STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins. *Nucleic Acids Research* **2004**, *32* (suppl_2), W500–W502. https://doi.org/10.1093/nar/gkh429.

(25) Freed, E. O. HIV-1 Assembly, Release and Maturation. *Nat Rev Micro* **2015**, *13* (8), 484–496. https://doi.org/10.1038/nrmicro3490.

(26) Ganser-Pornillos, B. K.; Yeager, M.; Sundquist, W. I. The Structural Biology of HIV Assembly. *Current Opinion in Structural Biology* **2008**, *18* (2), 203–217. https://doi.org/10.1016/j.sbi.2008.02.001.

(27) Freed, E. O. HIV-1 Assembly, Release and Maturation. *Nat Rev Micro* **2015**, *13* (8), 484–496. https://doi.org/10.1038/nrmicro3490.

(28) Freed, E. O. HIV-1 Gag Proteins: Diverse Functions in the Virus Life Cycle. *Virology* **1998**, *251* (1), 1–15. https://doi.org/10.1006/viro.1998.9398.

(29) *Advances in HIV-1 Assembly and Release*; Freed, E. O., Ed.; Springer New York: New York, NY, 2013.

(30) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Mature HIV-1 Capsid Structure by Cryo-Electron Microscopy and All-Atom Molecular Dynamics. *Nature* **2013**, *497* (7451), 643–646. https://doi.org/10.1038/nature12162.

(31) Keller, P. W.; Adamson, C. S.; Heymann, J. B.; Freed, E. O.; Steven, A. C. HIV-1 Maturation Inhibitor Bevirimat Stabilizes the Immature Gag Lattice. *Journal of Virology* **2011**, *85* (4), 1420–1428. https://doi.org/10.1128/JVI.01926-10.

(32) Adamson, C. S.; Ablan, S. D.; Boeras, I.; Goila-Gaur, R.; Soheilian, F.; Nagashima, K.; Li, F.; Salzwedel, K.; Sakalian, M.; Wild, C. T.; Freed, E. O. In Vitro Resistance to the Human Immunodeficiency Virus Type 1 Maturation Inhibitor PA-457 (Bevirimat). *J Virol* **2006**, *80* (22), 10957–10971. https://doi.org/10.1128/JVI.01369-06.

(33) Fontana, J.; Keller, P. W.; Urano, E.; Ablan, S. D.; Steven, A. C.; Freed, E. O. Identification of an HIV-1 Mutation in Spacer Peptide 1 That Stabilizes the Immature CA-SP1 Lattice. *J Virol* **2016**, *90* (2), 972–978. https://doi.org/10.1128/JVI.02204-15.

(34) Lee, S.-K.; Potempa, M.; Swanstrom, R. The Choreography of HIV-1 Proteolytic Processing and Virion Assembly. *J Biol Chem* **2012**, *287* (49), 40867–40874. https://doi.org/10.1074/jbc.R112.399444.

(35) Wagner, J. M.; Zadrozny, K. K.; Chrustowicz, J.; Purdy, M. D.; Yeager, M.; Ganser-Pornillos, B. K.; Pornillos, O. Crystal Structure of an HIV Assembly and Maturation Switch. *Elife* **2016**, *5*, e17063. https://doi.org/10.7554/eLife.17063.

(36) Schur, F. K. M.; Obr, M.; Hagen, W. J. H.; Wan, W.; Jakobi, A. J.; Kirkpatrick, J. M.; Sachse, C.; Kräusslich, H.-G.; Briggs, J. A. G. An Atomic Model of HIV-1 Capsid-SP1 Reveals Structures Regulating Assembly and Maturation. *Science* **2016**, *353* (6298), 506–508. https://doi.org/10.1126/science.aaf9620.

(37) Keller, P. W.; Huang, R. K.; England, M. R.; Waki, K.; Cheng, N.; Heymann, J. B.; Craven, R. C.; Freed, E. O.; Steven, A. C. A Two-Pronged Structural Analysis of Retroviral Maturation Indicates That Core Formation Proceeds by a Disassembly-Reassembly Pathway Rather than a Displacive Transition. *J Virol* **2013**, *87* (24), 13655–13664. https://doi.org/10.1128/JVI.01408-13.

(38) Purdy, M. D.; Shi, D.; Chrustowicz, J.; Hattne, J.; Gonen, T.; Yeager, M. MicroED Structures of HIV-1 Gag CTD-SP1 Reveal Binding Interactions with the Maturation Inhibitor Bevirimat. *Proc Natl Acad Sci U S A* **2018**, *115* (52), 13258–13263. https://doi.org/10.1073/pnas.1806806115.

(39) Coric, P.; Turcaud, S.; Souquet, F.; Briant, L.; Gay, B.; Royer, J.; Chazal, N.; Bouaziz, S. Synthesis and Biological Evaluation of a New Derivative of

Bevirimat That Targets the Gag CA-SP1 Cleavage Site. *Eur J Med Chem* **2013**, *62*, 453–465. https://doi.org/10.1016/j.ejmech.2013.01.013.

(40) Waki, K.; Durell, S. R.; Soheilian, F.; Nagashima, K.; Butler, S. L.; Freed, E. O. Structural and Functional Insights into the HIV-1 Maturation Inhibitor Binding Pocket. *PLoS Pathog* **2012**, *8* (11), e1002997. https://doi.org/10.1371/journal.ppat.1002997.

(41) Zhou, J.; Yuan, X.; Dismuke, D.; Forshey, B. M.; Lundquist, C.; Lee, K.-H.; Aiken, C.; Chen, C. H. Small-Molecule Inhibition of Human Immunodeficiency Virus Type 1 Replication by Specific Targeting of the Final Step of Virion Maturation. *J Virol* **2004**, *78* (2), 922–929. https://doi.org/10.1128/jvi.78.2.922-929.2004.

(42) Li, F.; Goila-Gaur, R.; Salzwedel, K.; Kilgore, N. R.; Reddick, M.; Matallana, C.; Castillo, A.; Zoumplis, D.; Martin, D. E.; Orenstein, J. M.; Allaway, G. P.; Freed, E. O.; Wild, C. T. PA-457: A Potent HIV Inhibitor That Disrupts Core Condensation by Targeting a Late Step in Gag Processing. *Proc Natl Acad Sci U S A* **2003**, *100* (23), 13555–13560. https://doi.org/10.1073/pnas.2234683100.

(43) Dick, R. A.; Zadrozny, K. K.; Xu, C.; Schur, F. K. M.; Lyddon, T. D.; Ricana, C. L.; Wagner, J. M.; Perilla, J. R.; Ganser-Pornillos, B. K.; Johnson, M. C.; Pornillos, O.; Vogt, V. M. Inositol Phosphates Are Assembly Co-Factors for HIV-1. *Nature* **2018**, *560* (7719), 509–512. https://doi.org/10.1038/s41586-018-0396-4.

(44) Schwieters, C. D.; Clore, G. M. A Pseudopotential for Improving the Packing of Ellipsoidal Protein Structures Determined from NMR Data. *J. Phys. Chem. B* **2008**, *112* (19), 6070–6073. https://doi.org/10.1021/jp076244o.

(45) Bermejo, G. A.; Clore, G. M.; Schwieters, C. D. Smooth Statistical Torsion Angle Potential Derived from a Large Conformational Database via Adaptive Kernel Density Estimation Improves the Quality of NMR Protein Structures. *Protein Science* **2012**, *21* (12), 1824–1836. https://doi.org/10.1002/pro.2163.

(46) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J Comput Chem* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(47) Wagner, J. M.; Zadrozny, K. K.; Chrustowicz, J.; Purdy, M. D.; Yeager, M.; Ganser-Pornillos, B. K.; Pornillos, O. Crystal Structure of an HIV Assembly and Maturation Switch. *eLife* **2016**, *5*, e17063. https://doi.org/10.7554/eLife.17063.

(48) Gong, Z.; Schwieters, C. D.; Tang, C. Conjoined Use of EM and NMR in RNA Structure Refinement. *PLOS ONE* **2015**, *10* (3), e0120445. https://doi.org/10.1371/journal.pone.0120445.

(49) Schur, F. K. M.; Obr, M.; Hagen, W. J. H.; Wan, W.; Jakobi, A. J.; Kirkpatrick, J. M.; Sachse, C.; Kräusslich, H.-G.; Briggs, J. A. G. An Atomic Model of HIV-1 Capsid-SP1 Reveals Structures Regulating Assembly and Maturation. *Science* **2016**, *353* (6298), 506–508. https://doi.org/10.1126/science.aaf9620.

(50) Purdy, M. D.; Shi, D.; Chrustowicz, J.; Hattne, J.; Gonen, T.; Yeager, M. MicroED Structures of HIV-1 Gag CTD-SP1 Reveal Binding Interactions with the Maturation Inhibitor Bevirimat. *Proceedings of the National Academy of Sciences* **2018**, *115* (52), 13258–13263. https://doi.org/10.1073/pnas.1806806115.

(51) Dick, R. A.; Zadrozny, K. K.; Xu, C.; Schur, F. K. M.; Lyddon, T. D.; Ricana, C. L.; Wagner, J. M.; Perilla, J. R.; Ganser-Pornillos, B. K.; Johnson, M. C.; Pornillos, O.; Vogt, V. M. Inositol Phosphates Are Assembly Co-Factors for HIV-1. *Nature* **2018**, *560* (7719), 509–512. https://doi.org/10.1038/s41586-018-0396-4.

(52) Schur, F. K. M.; Obr, M.; Hagen, W. J. H.; Wan, W.; Jakobi, A. J.; Kirkpatrick, J. M.; Sachse, C.; Kräusslich, H.-G.; Briggs, J. A. G. An Atomic Model of HIV-1 Capsid-SP1 Reveals Structures Regulating Assembly and Maturation. *Science* **2016**, *353* (6298), 506–508. https://doi.org/10.1126/science.aaf9620.

Chapter 4

ATOMIC RESOLUTION HIV-1 CAPSID TUBULAR ASSEMBLIES BY MAS NMR

Content (i.e. certain figures and technical language) in this chapter is reprinted with permission from the published article[1]: Manman Lu, Ryan W. Russell, Alex Bryer, Caitlin M. Quinn, Guangjin Hou, Huilan Zhang, Charles D. Schwieters, Juan R. Perilla, Angela M. Gronenborn, Tatyana Polenova (**2020**) Atomic-resolution structure of HIV-1 capsid tubes by magic-angle spinning NMR. *Nat. Mol. Struct. Biol.* **27**, 863–869. DOI: 10.1038/s41594-020-0489-2

Author contributions are as follows: T. P. and A. M. G. conceived the project and guided the work. J. R. P. designed and guided the MD simulations and structure calculations of the CA tube. M. L. prepared the samples, performed NMR experiments and analyzed the experimental data. R. W. R. and M. L. performed the structure calculations of CA hexamer unit. C. M Q. assisted in the structure calculations of the hexamer unit. A. B. conducted the MD simulations and structure calculation of the CA tube. M. L., R. W. R. and A. B. prepared figures for the manuscript. R. W. R. and A. B. wrote scripts for analysis of calculation results and visualization of the hexamer unit and tube, respectively. C. D. S. provided critical input in the NIH-Xplor based structure calculations. C. M. Q., G. H. and H. Z. took part in the design or analysis of NMR experiments. T. P. and A. M. G. took the lead in writing the manuscript. All authors discussed the results and contributed to the manuscript preparation.

4.1 Introduction

Acquired Immunodeficiency Syndrome (AIDS) is a worldwide ailment and is caused by the Human Immunodeficiency Virus (HIV). Approximately 37.7 million people [30.2 - 45.1 million] are living with HIV worldwide, with approximately 1.5 million [1.0 – 2.0 million] new cases recorded in 2020 alone, according to the UNAIDS epidemiological estimates.[2] Current treatment regimens for HIV-1 patients include antiretroviral therapy, where the patient is prescribed multiple medicines with different mechanisms of action, to prevent/suppress HIV-1 reproduction. With the usage of antiretroviral therapy, patients can live long lives. However, the current treatments are expensive, associated with numerous side effects, require to be administered for the entire life, and no cure is available to date.[3]

HIV endlessly adapts to resist new therapies and, therefore, considerable effort is being put in to trying to understand the virus on the molecular level. The genetic information of the virus, necessary for reproduction, is embedded in single-stranded RNA dimer and harbored/protected by the HIV-1 capsid (CA) protein. CA exhibits several distinct functionally important regions, such as the cyclophilin A (CypA)-binding loop and the β-hairpin. CA exhibits inherent plasticity which is a feature that is connected to numerous viral functions throughout replication. Specifically, CA is involved in uncoating[4], microtubule transport/hijacking[5,6], host-factor recruitment[7–10], genome integration[11], and ultimately nuclear import[12,13]. This, collectively, makes CA an exceptional target for therapeutic intervention.[14,15]

Structurally, CA is complex. A single chain of CA comprises the N-terminal domain (NTD) and the C-terminal domain (CTD) connected by a short flexible linker. The RNA-harboring capsid is a hexagonal surface lattice that is closed with 12 CA pentamers.[16] Indeed this totals ~1,000-1,500 copies of the capsid protein (CA), and is conical in architecture. In addition to conical assemblies, CA readily assembles into tubes *in vitro*; CA tubes recapitulate numerous structural parameters of cones and has inspired many laboratories to pursue their study.[17–19]

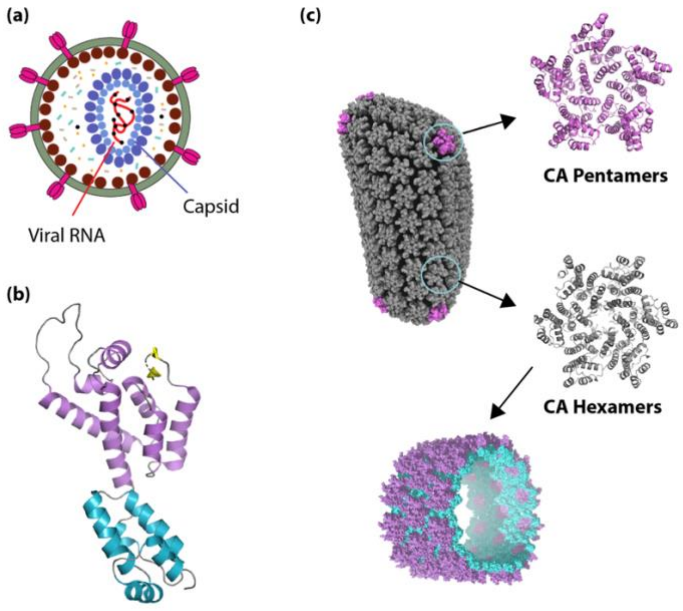

Figure 4.1: HIV-1 capsid overview. (A) Cartoon/ribbon illustration of HIV-1 virion with CA (blue/violet) harboring RNA (red). (B) Single chain of HIV-1 CA with a N-terminal domain (top) and C-terminal domain (bottom). (C) (top) All-atom model of the mature HIV-1 capsid core comprised of pentamers and hexamers, determined by an integrated

cryo-EM, cryo-ET, solution NMR, and MD approach.[19] (bottom) HIV-1 tubular assembly comprising only CA hexamers.[20]

4.2 Experimental distance and dihedral restraints by MAS NMR

High-resolution MAS NMR spectra of CA tubular assemblies were recorded. From these spectra, near-complete chemical shift assignments were performed. Using chemical shifts, cross-peaks in the through-space $R2_n^\nu$-driven (CORD) spectra of [1,6-^{13}C-glucose, U-^{15}N]-CA and [2-^{13}C-glucose, U-^{15}N]-CA, recorded with different mixing times, were converted to distance restraints. In total 1,311 distance restraints were identified, including 210 long-range ($|i-j| \geq 5$) correlations (Table 4.1). Using the experimental ^{13}C and ^{15}N chemical shift assignments, 390 backbone dihedral (ϕ/ψ) restraints were accurately derived/predicted with TALOS-N.[21]

Table 4.1: Summary of MAS NMR distance and dihedral restraints used for structure calculations of HIV-1 CA

Distance constraints	CA FL	CA NTD	CA CTD
Unambiguous			
Intra-residue	491	374	114
Inter-residue	820	620	197
Sequential ($\|i-j\| = 1$)	221	161	58
Medium range ($1 \leq \|i-j\| < 5$)	204	166	38
Long range ($\|i-j\| \geq 5$) (sidechain-sidechain)	210 (101)	166 (85)	43 (16)
Ambiguous	185	127	58
Total ^{13}C-^{13}C restraints	1311	994	311
Restraints/Residue	5.7	6.9	3.7
Dihedral Restraints	**CA FL**	**CA NTD**	**CA CTD**
ϕ	195	126	68
ψ	195	126	68

4.3 Structure of a single CA chain

The initial fold calculation for the CA single chain was seeded from an extended chain from the primary structure, distance restraints, and 390 dihedral restraints (Table 4.1). The bounds of the distance restraints were set to 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, consistent with the bounds established in Chapter 2. Ambiguous restraints exceeding five-fold ambiguity were not included in the calculations.

Structure calculations were performed in Xplor-NIH version 2.51.[22–24] Standard terms for bond lengths, bond angles and improper angles were used to enforce correct covalent geometry. A statistical torsion angle potential[25] and the gyration volume term were employed[26]. Separate gyration volume terms were applied to the NTD (residues 1–145) and CTD (residues 148–231), excluding the flexible linker. A hydrogen bond database term, HBPot, was used to improve hydrogen bond geometries[27].

In the run, 3,000 structures were calculated using molecular dynamics simulated annealing in torsion angle space with two successive annealing schedules and a final gradient minimization in Cartesian space. The structure calculation began with a 3,500 K constant-temperature molecular dynamics run for the shorter of 800 ps or 8,000 steps with the time step size allowed to float to maintain constant energy, within a tolerance. The initial velocities were randomized about a Maxwell distribution using a starting temperature of 3,500 K. Following this initial dynamics calculation, a simulated annealing calculation was performed where the temperature was reduced to 100 K in

steps of 25 K. At each temperature, dynamics were run for the shorter of 0.4 ps or 200 steps. Force constants for distance restraints were ramped up from 10 to 50 kcal mol^{-1} $Å^{-2}$. The dihedral angle restraints were disabled for high-temperature dynamics at 3,500 K but enabled during simulated annealing with a force constant of 200 kcal mol^{-1} rad^{-2}. The gyration volume force constant was geometrically scaled from 0.002 to 1. After simulated annealing, the structures were minimized using a Powell energy minimization scheme.

At the completion of the run, the lowest energy structures were identified and, as anticipated from the systematic model in Chapter 2, the orientation of the NTD and CTD relative to each other could not be correctly identified from NMR alone (Figure 4.2). This could not be identified because there were very few NTD-CTD correlations present in the MAS NMR spectra because the distances between the atoms of the NTD and CTD exceed the observable distance C-C correlations (~7 Å) in MAS NMR experiments (Figure 1.4).

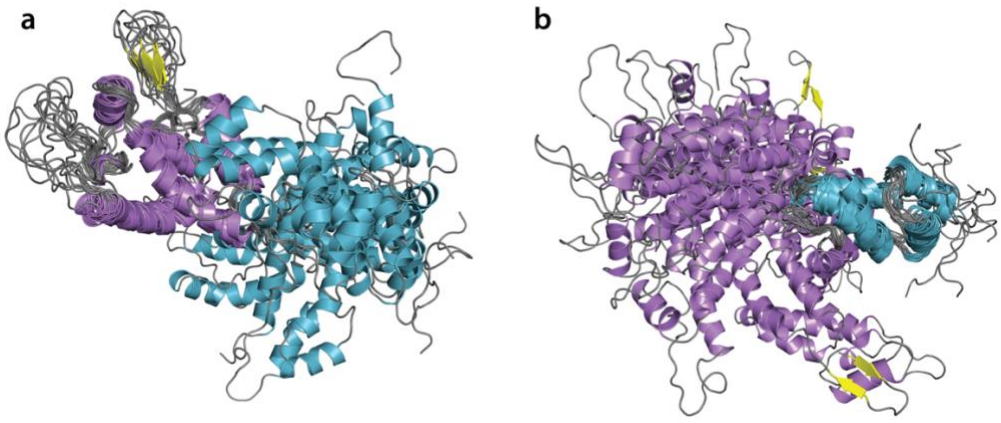

Figure 4.2: Best-fit superpositions for the NTD (A) and CTD (B), respectively, are shown. NTD helices are colored purple, the β-hairpin yellow, loops gray, and the CTD helices cyan.

Despite insufficient restraints between the two domains, the accuracy of the individual domains was adequate. This was corroborated both by visual inspection (Figure 4.2) and the tallies of the restraints: 6.9 restraints per residue (17% completeness) for the NTD (residues 1–145) and 3.7 restraints per residue (10% completeness) for the CTD (residues 148–231). The precision of the NTD and CTD ensembles measured by pairwise atomic backbone RMSDs are 2.2 ± 0.4 Å and 1.8 ± 0.5 Å, respectively, (Table 4.2) and falls within the expected values of our systematic model study (Chapter 2) of ~1-3 Å. At the completion of this stage, we then proceeded with separate treatment of the NTD and CTD in the subsequent section.

Table 4.2: Structural precision (pairwise atomic backbone RMSD) of the individual domains (NTD, CTD) of HIV-1 CA

Initial single-chain ensemble	CA NTD (1-145)	CA CTD (148-231)
Backbone (N, Cα, C')	2.2 ± 0.4 Å	1.8 ± 0.5 Å
Heavy atoms (All N, C)	2.8 ± 0.4 Å	2.7 ± 0.4 Å

4.4 Structure of the individual NTD and CTD domains of CA

The 300 lowest energy structures from the full-length CA single-chain run were selected and the coordinates of the NTD and CTD were refined in separate calculations. In the separate calculations 3,000 structures for both the NTD and CTD were refined. Simulated annealing at 3,000 K was performed for 10 ps or 5,000 steps, whichever was completed first. The starting time step was 1 fs and was self-adjusted in subsequent steps to ensure conservation of energy. The initial velocities were randomized about a Maxwell distribution using the starting temperature of 3,000 K. The temperature was subsequently reduced to 25 K in steps of 12.5 K. At each temperature, the initial default time step was 1 fs, and a 0.2-ps dynamics run was performed. Force constants for distance restraints were ramped up from 2 to 30 kcal mol^{-1} $Å^{-2}$. The dihedral restraint force constants were set to 10 kcal mol^{-1} rad^{-2} for high-temperature dynamics at 3,000 K and 200 kcal mol^{-1} rad^{-2} during cooling. The gyration volume force constant was scaled from 0.002 to 1. The annealed structures were minimized using a Powell energy minimization scheme. The 50 lowest energy structures from each run were selected for the next step and the ten lowest are shown visually in Figure 4.3.

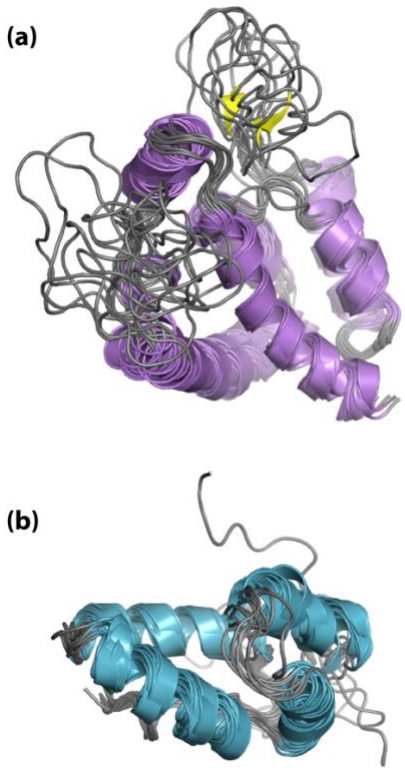

Figure 4.3: Ensemble of the ten lowest energy structures for the single-chain calculation and refinement of the NTD (A) and CTD (B) separately. The NTD helices are colored purple, the β-hairpin yellow, loops gray, and the CTD helices cyan.

4.5 Docking of the individual domains into the cryo-EM density

To incorporate the overall envelope information, and thus the orientations of the NTD and CTD relative to each other, the experimental cryo-EM density of EMD-8595 (PDB 5UPW)[20] was introduced. The first step involving the density was to dock the NTD and CTD into the cryo-EM density and to prepare the density by post-processing and down sampling to 8-Å resolution using the phenix.auto_sharpen routine in PHENIX

1.14[28]. The docking was carried out with an in-house UCSF Chimera[29] python script (Listing 4.1).

```
# Fit an atomic model in a map and save new coordinates and map to files.
from chimera import runCommand as run, openModels
import re
import sys
import os
in_pdb = "INPUT_PDB_FILE"
stucIDs = re.search("\\d+", in_pdb, re.M).group(0)
run('open ./3sympostprocess_cut_trim4_sharp_8A.mrc')
run(str("open ") + str(in_pdb))
map1_id = 1
map2_id = 0
search = 5000
res = 8
from chimera import openModels as om, selection
m1 = om.list(id = map1_id)[0]
m2 = om.list(id = map2_id)[0]
s1 = selection.ItemizedSelection([m1])
from FitMap.fitcmd import fitmap
fit_list = fitmap(s1, m2, search = search, resolution = res, listFits =
False)
print '%d fits' % len(fit_list)
import Matrix
corrs = []
for index, fit in enumerate(fit_list):
    if int(index) <= int(10):
       print 'correlation =', fit.correlation()
#     if int(index) == int(0):
    corrs.append(fit.correlation())
    if int(index) <= int(5):
       fit.place_copies()
if not os.path.exists("DockedNTDsCorrs"):
       os.makedirs("DockedNTDsCorrs")
with open(str("DockedNTDsCorrs/" + stucIDs + "_Corrs.txt"), 'w') as outfile:
       for x in corrs:
              outfile.write("%s\n" % x)

largestCorr = round(corrs[0], 4)
print largestCorr
if not os.path.exists("DockedNTDs"):
       os.makedirs("DockedNTDs")
outname = "DockedNTDs/dockedNTDs_" + str(stucIDs) + "_Corr_" +
str(largestCorr) + ".pdb"
print outname
run('combine #2#3#4#5#6#7')
run(str("write #8 ") + outname)
```

Listing 4.1: In-house UCSF Chimera python script for batch docking of protein subunits into cryo-EM density by the means of exhaustive translations and rotations corroborated by the basis of cross-correlation.

Here, the 50 lowest energy structures for both the NTD and CTD from the previous step served as input for batch docking and were subjected to the global search of 5,000 random different translations and rotations. On the basis of the lowest cross-correlation values (§ 1.2.5) and visual inspection, 14 NTDs and 14 CTDs were identified (Figure 4.4).

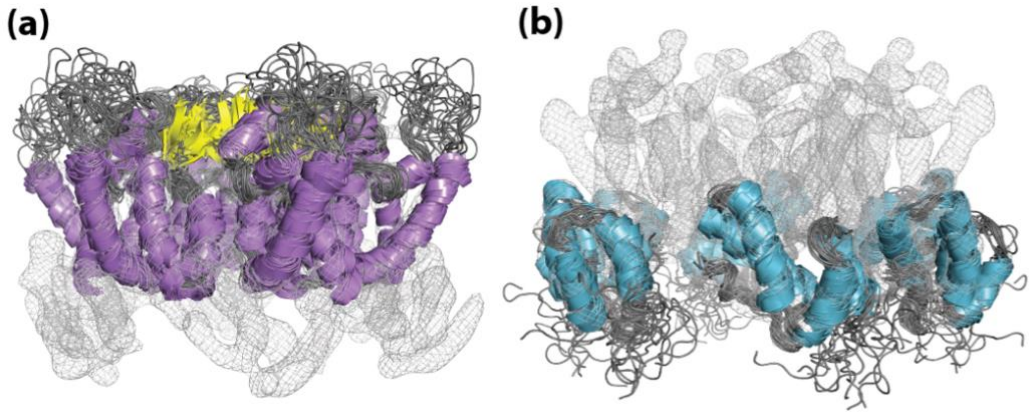

Figure 4.4: Batch docking of NTD (A) and CTD (B) domains separately into low-resolution cryo-EM map of EMD-8595 (PDB 5UPW)[20] of a hexamer unit of tubular assemblies. Shown are 14 NTDs and 14 CTDs with the lowest cross-correlation (and checked with visual inspection). The NTD helices are colored purple, the β-hairpin yellow, loops gray, and the CTD helices cyan.

4.6 Joint refinement of the CA hexamer with NMR restraints and cryo-EM density

Joint refinement calculations of NMR and cryo-EM (introduced in §§ 1.2.5-1.2.6) were performed in Xplor-NIH, using NMR distance restraints, NMR-derived dihedral restraints, and the 8-Å cryo-EM density map of the hexamer unit. Separate calculations were performed for the NTD and CTD in the hexamer (Figure 4.5). Calculations were seeded from the bundle of 14 NTD or CTD starting structures, and 100 structures for each were generated by torsion angle dynamics with an annealing schedule and a final gradient minimization in Cartesian space.

Two identical runs of simulated annealing starting at 250 K were performed for 10 ps, with a time step of 1 fs. The initial velocities were randomized to achieve a Maxwell distribution at a starting temperature of 250 K. The temperature was subsequently reduced to 25 K in steps of 25 K. At each temperature, dynamics were run for 0.4 ps with an initial time step of 1 fs.

The cryo-EM potential[30] was restricted to N, C′, Cα, O and Cβ atoms to preserve the side chain orientations defined by NMR distance and dihedral restraints. Approximate non-crystallographic symmetry was imposed using Xplor-NIH's PosDiffPot term, allowing the subunits of the hexamer to differ by up to 1 Å, and force constants for distance restraints were ramped from 2 to 30 kcal mol^{-1} $Å^{-2}$. The dihedral restraint force constants were set to 10 kcal mol^{-1} rad^{-2} for high-temperature dynamics at 3,000 K and 200 kcal mol^{-1} rad^{-2} during cooling. The EM density map was used to

generate a cross-correlation probability distribution potential, and the force constants of the cross-correlation probability distribution potential, initiated from the cryo-EM density map, was set to 50 kcal mol^{-1} during high-temperature dynamics and cooling. The gyration volume potential was turned off to avoid conflicts with the cryo-EM density map. The annealed structures were minimized using a Powell energy minimization scheme.

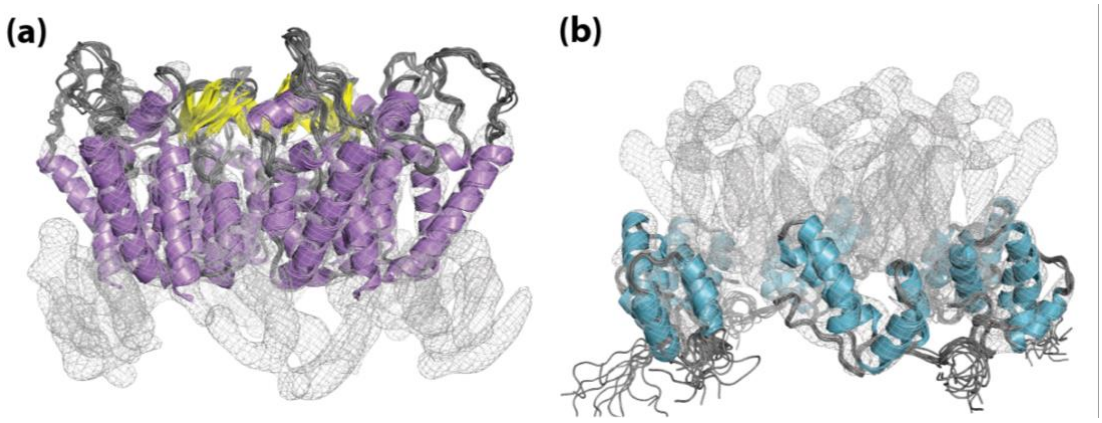

Figure 4.5: Ensemble of the ten lowest energy structures for the hexamer refinement of the NTD (A) and CTD (B) domains separately. The refinement incorporated NMR distance/dihedral restraints and the cryo-EM density simultaneously. The NTD helices are colored purple, the β-hairpin yellow, loops gray, and the CTD helices cyan.

After separate refinement of the NTD and CTD hexamer units (Figure 4.5), the ten lowest energy structures from each bundle were combined into a ten-member ensemble containing both NTD and CTD hexamer units. The protein structure file (PSF) of the hexamer was generated by loading the sequence file of the single CA chain of the

hexamer. The PSF was expanded to six chains using the Xplor-NIH function psfGen.duplicateSegment. For each structure, the starting coordinates of the NTD and CTD regions were set from the input files, while the linker region connecting the NTD and CTD was built by the protocol.addUnknownAtoms routine. Three iterations of the hexamer refinement calculation were performed. After each iteration, the ten lowest energy structures were selected and used as input for the next iteration. The final MAS-NMR bundle comprised the ten lowest energy structures from the final refinement calculation (Figure 4.6).

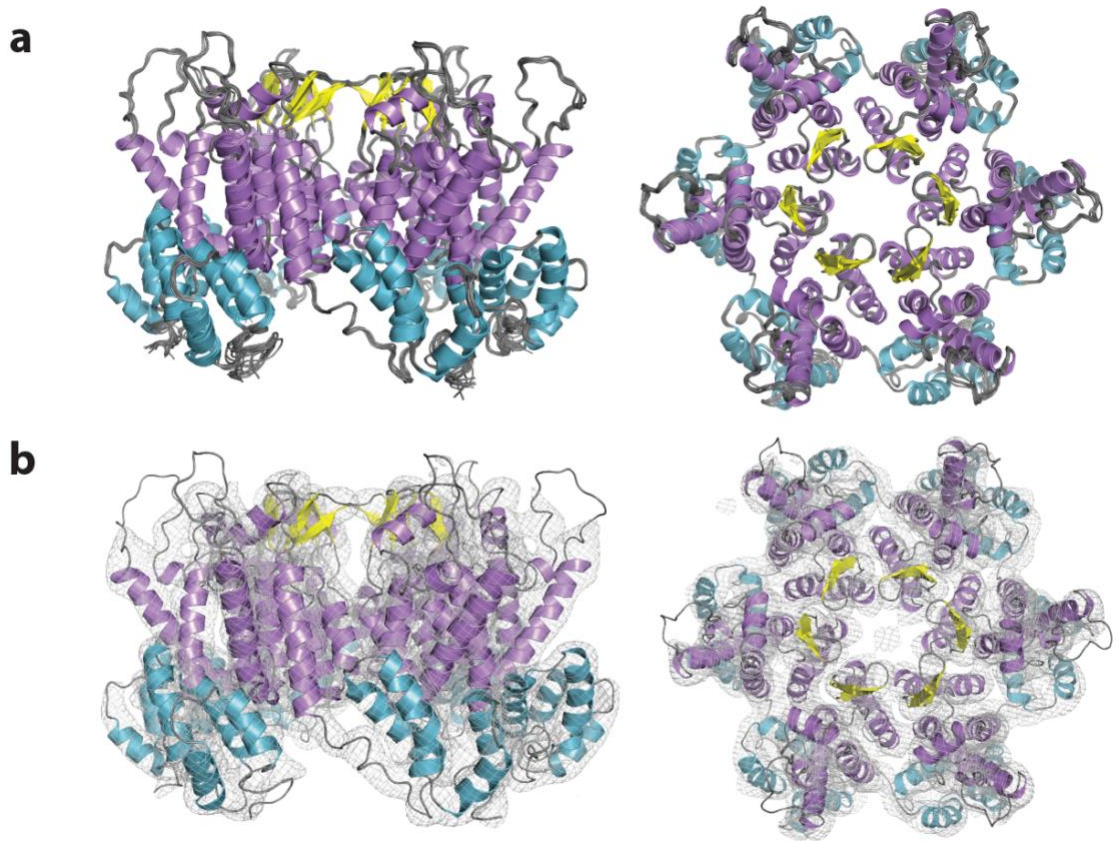

Figure 4.6: Structure of the hexameric unit in CA tubular assemblies. (A) Side and top views of the final ensemble of the ten lowest energy structures of the CA hexamer unit in the tubular CA assembly. (B) Side and top views of the superposition of the lowest energy structure of the NMR-derived CA hexamer unit and the 8-Å resolution cryo-EM density map.

4.7. Final NMR ensemble of the hexameric unit in CA tubular assemblies

Our five-step procedure is summarized in Figure 4.7. Upon completion of this procedure, we obtained a well-defined ensemble of the ten lowest energy hexamer structures. (Figure 4.6) The pairwise atomic RMSDs for the ensemble is 0.5 ± 0.1 Å and 1.2 ± 0.1 Å for backbone and heavy atoms, respectively. (All structural statistics are provided in Table 4.3). These values are remarkably low, indicating a high level of precision on the atomic scale, revealing precise sidechain details.

Table 4.3: Structure statistics of final NMR ensemble of capsid protein

Violations (mean ± s.d.)	
Distance constraints (Å)	0.049 ± 0.002
Dihedral angle constraints (°)	1.076 ± 0.101
Max. distance constraint violation (Å)	0.773
Max. dihedral angle violation (°)	12.364
Deviations from idealized geometry	
Bond lengths (Å)	0.005 ± 0.000
Bond angles (°)	0.680 ± 0.006
Improper angles (°)	0.575 ± 0.007
Average pairwise RMSD* (Å)	
Heavy	1.2 ± 0.1
Backbone	0.5 ± 0.1

* Pairwise RMSD was calculated among 10 refined structures.

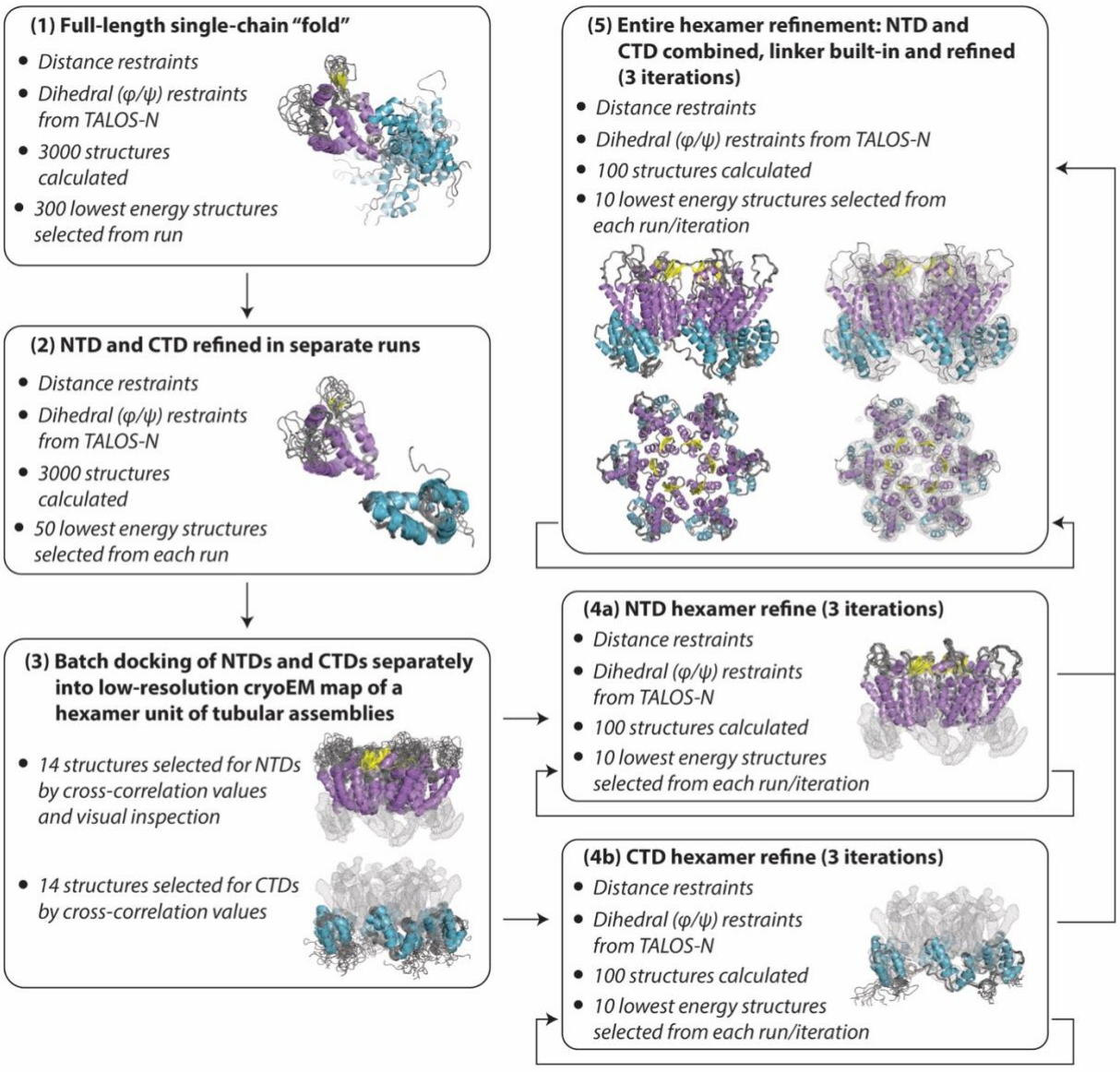

Figure 4.7: Summary of the novel five-step NMR/cryo-EM joint procedure. Ensembles of the ten lowest energy structures at each step are depicted in ribbon representation, with NTD helices colored purple, the β-hairpin yellow, loops gray, and CTD helices cyan.

4.8 Novel structural details of functionally important regions

Our integrated MAS-NMR/cryo-EM structure reveals several unique structural details previously inaccessible from other studies, such as conformations of the β-hairpin and the CypA binding loop, which is known to be dynamic[17]. These conformations are afforded from the extensive experimental MAS NMR distance restraints (Figure 4.8A/B). The β-hairpin is unambiguously defined and in an open conformation as indicated by restraints within the residues of the loop (P1,H12) and with other NTD residues (A47, T48, D51, L111), as shown in Figure 4.8B. That the conformation of the dynamic CypA loop could be determined is thanks to the fact that the motions of the loop residues occurring on nano- to microsecond timescales do not interfere with the signals in correlation spectra. Moreover, at cryogenic temperatures individual conformers are frozen out, and this is likely the cause for the poorly defined density of these residues in the cryo-EM density.[31]

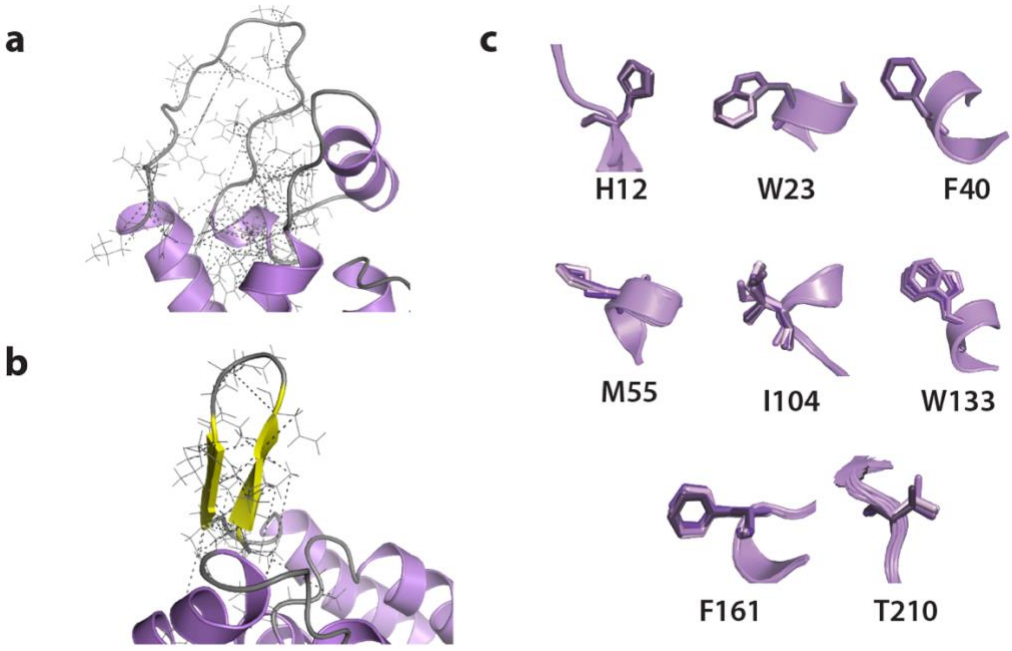

Figure 4.8: Structural details of the hexameric unit in CA tubular assemblies. (A, B) Details of the MAS-NMR-derived distance restraint network for the CypA loop and the β-hairpin, respeIvely. (C) Selected side chain conformations in the final ten-conformer ensemble.

As expected, there are several key differences between the MAS NMR structure and the structure of hydrated crystals determined by X-ray diffraction (PDB 4XFX)[32]. In particular, CA crystallizes in flat hexamers and hence the structure has strict six-fold symmetry. Our MAS NMR structure, on the other hand, exhibits no six-fold symmetry as the tubes are not flat, and hence more similar to the *in virio* conical capsids[19] thus underscoring the biological relevance of the study.

4.9 Conclusions and future outlook

Our joint NMR/cryo-EM structure is not only biologically relevant but is computationally also the first-of-its-kind that incorporates NMR-level detail and low-resolution cryo-EM density for a large assembly where protein folding was performed solely from experimental MAS NMR distance and dihedral restraints. The motivation for this approach was the lack of distance restraints between the individual CA domains. This protocol devised in our study ensures that no symmetry was imposed and that the critical atomic-scale NMR details are retained including, but not limited to, side chain conformations and restraints.

More broadly, integration of results obtained by two experimental techniques is a powerful approach to overcome single method-inherent limitations. As illustrated in our study, determination of the atomic-resolution structure of a single CA chain, based on MAS-NMR restraints combined with a low-resolution cryo-EM map for defining the overall shape of a hexameric unit, provided the means for computationally deriving an integrated all-atom structure of the hexamer building block as well as the *in-vitro*-assembled tube. An even more effective result can be obtained when combined with large-scale all-atom data-guided MD simulations, as performed for the CA tubular assembly by the Perilla group using the coordinates from the structural ensemble of the CA hexamer determined in our study.

REFERENCES

(1) Lu, M.; Russell, R. W.; Bryer, A. J.; Quinn, C. M.; Hou, G.; Zhang, H.; Schwieters, C. D.; Perilla, J. R.; Gronenborn, A. M.; Polenova, T. Atomic-Resolution Structure of HIV-1 Capsid Tubes by Magic-Angle Spinning NMR. *Nature Structural & Molecular Biology* **2020**, *27* (9), 863–869. https://doi.org/10.1038/s41594-020-0489-2.

(2) *UNAIDS Data 2021*; Geneva: Joint United Nations Programme on HIV/AIDS, 2021.

(3) Kemnic, T. R.; Gulick, P. G. HIV Antiretroviral Therapy. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2022.

(4) Ambrose, Z.; Aiken, C. HIV-1 Uncoating: Connection to Nuclear Entry and Regulation by Host Proteins. *Virology* **2014**, *454–455*, 371–379. https://doi.org/10.1016/j.virol.2014.02.004.

(5) Malikov, V.; da Silva, E. S.; Jovasevic, V.; Bennett, G.; de Souza Aranha Vieira, D. A.; Schulte, B.; Diaz-Griffero, F.; Walsh, D.; Naghavi, M. H. HIV-1 Capsids Bind and Exploit the Kinesin-1 Adaptor FEZ1 for Inward Movement to the Nucleus. *Nat Commun* **2015**, *6*, 6660. https://doi.org/10.1038/ncomms7660.

(6) Lukic, Z.; Dharan, A.; Fricke, T.; Diaz-Griffero, F.; Campbell, E. M. HIV-1 Uncoating Is Facilitated by Dynein and Kinesin 1. *J Virol* **2014**, *88* (23), 13613–13625. https://doi.org/10.1128/JVI.02219-14.

(7) Stremlau, M.; Owens, C. M.; Perron, M. J.; Kiessling, M.; Autissier, P.; Sodroski, J. The Cytoplasmic Body Component TRIM5alpha Restricts HIV-1 Infection in Old World Monkeys. *Nature* **2004**, *427* (6977), 848–853. https://doi.org/10.1038/nature02343.

(8) Liu, Z.; Pan, Q.; Ding, S.; Qian, J.; Xu, F.; Zhou, J.; Cen, S.; Guo, F.; Liang, C. The Interferon-Inducible MxB Protein Inhibits HIV-1 Infection. *Cell Host Microbe* **2013**, *14* (4), 398–410. https://doi.org/10.1016/j.chom.2013.08.015.

(9) Goujon, C.; Moncorgé, O.; Bauby, H.; Doyle, T.; Ward, C. C.; Schaller, T.; Hué, S.; Barclay, W. S.; Schulz, R.; Malim, M. H. Human MX2 Is an Interferon-Induced Post-Entry Inhibitor of HIV-1 Infection. *Nature* **2013**, *502* (7472), 559–562. https://doi.org/10.1038/nature12542.

(10) Luban, J.; Bossolt, K. L.; Franke, E. K.; Kalpana, G. V.; Goff, S. P. Human Immunodeficiency Virus Type 1 Gag Protein Binds to Cyclophilins A and B. *Cell* **1993**, *73* (6), 1067–1078. https://doi.org/10.1016/0092-8674(93)90637-6.

(11) Ocwieja, K. E.; Brady, T. L.; Ronen, K.; Huegel, A.; Roth, S. L.; Schaller, T.; James, L. C.; Towers, G. J.; Young, J. A. T.; Chanda, S. K.; König, R.; Malani, N.; Berry, C. C.; Bushman, F. D. HIV Integration Targeting: A Pathway Involving Transportin-3 and the Nuclear Pore Protein RanBP2. *PLoS Pathog* **2011**, *7* (3), e1001313. https://doi.org/10.1371/journal.ppat.1001313.

(12) König, R.; Zhou, Y.; Elleder, D.; Diamond, T. L.; Bonamy, G. M. C.; Irelan, J. T.; Chiang, C.-Y.; Tu, B. P.; De Jesus, P. D.; Lilley, C. E.; Seidel, S.; Opaluch,

A. M.; Caldwell, J. S.; Weitzman, M. D.; Kuhen, K. L.; Bandyopadhyay, S.; Ideker, T.; Orth, A. P.; Miraglia, L. J.; Bushman, F. D.; Young, J. A.; Chanda, S. K. Global Analysis of Host-Pathogen Interactions That Regulate Early-Stage HIV-1 Replication. *Cell* **2008**, *135* (1), 49–60. https://doi.org/10.1016/j.cell.2008.07.032.

(13) Brass, A. L.; Dykxhoorn, D. M.; Benita, Y.; Yan, N.; Engelman, A.; Xavier, R. J.; Lieberman, J.; Elledge, S. J. Identification of Host Proteins Required for HIV Infection through a Functional Genomic Screen. *Science* **2008**, *319* (5865), 921–926. https://doi.org/10.1126/science.1152725.

(14) Campbell, E. M.; Hope, T. J. HIV-1 Capsid: The Multifaceted Key Player in HIV-1 Infection. *Nat Rev Microbiol* **2015**, *13* (8), 471–483. https://doi.org/10.1038/nrmicro3503.

(15) Novikova, M.; Zhang, Y.; Freed, E. O.; Peng, K. Multiple Roles of HIV-1 Capsid during the Virus Replication Cycle. *Virol. Sin.* **2019**, *34* (2), 119–134. https://doi.org/10.1007/s12250-019-00095-3.

(16) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Mature HIV-1 Capsid Structure by Cryo-Electron Microscopy and All-Atom Molecular Dynamics. *Nature* **2013**, *497* (7451), 643–646. https://doi.org/10.1038/nature12162.

(17) Lu, M.; Hou, G.; Zhang, H.; Suiter, C. L.; Ahn, J.; Byeon, I.-J. L.; Perilla, J. R.; Langmead, C. J.; Hung, I.; Gor'kov, P. L.; Gan, Z.; Brey, W.; Aiken, C.; Zhang,

P.; Schulten, K.; Gronenborn, A. M.; Polenova, T. Dynamic Allostery Governs Cyclophilin A-HIV Capsid Interplay. *Proc Natl Acad Sci U S A* **2015**, *112* (47), 14617–14622. https://doi.org/10.1073/pnas.1516920112.

(18) Han, Y.; Hou, G.; Suiter, C. L.; Ahn, J.; Byeon, I.-J. L.; Lipton, A. S.; Burton, S.; Hung, I.; Gor'kov, P. L.; Gan, Z.; Brey, W.; Rice, D.; Gronenborn, A. M.; Polenova, T. 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* **2013**, *135* (47), 17793–17803. https://doi.org/10.1021/ja406907h.

(19) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Mature HIV-1 Capsid Structure by Cryo-Electron Microscopy and All-Atom Molecular Dynamics. *Nature* **2013**, *497* (7451), 643–646. https://doi.org/10.1038/nature12162.

(20) Perilla, J. R.; Zhao, G.; Lu, M.; Ning, J.; Hou, G.; Byeon, I.-J. L.; Gronenborn, A. M.; Polenova, T.; Zhang, P. CryoEM Structure Refinement by Integrating NMR Chemical Shifts with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2017**, *121* (15), 3853–3863. https://doi.org/10.1021/acs.jpcb.6b13105.

(21) Shen, Y.; Bax, A. Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *Journal of Biomolecular NMR* **2013**, *56* (3), 227–241. https://doi.org/10.1007/s10858-013-9741-y.

(22) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. The Xplor-NIH NMR Molecular Structure Determination Package. *Journal of Magnetic Resonance* **2003**, *160* (1), 65–73.

(23) Schwieters, C.; Kuszewski, J.; Mariusclore, G. Using Xplor–NIH for NMR Molecular Structure Determination. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2006**, *48* (1), 47–62. https://doi.org/10.1016/j.pnmrs.2005.10.001.

(24) Schwieters, C. D.; Bermejo, G. A.; Clore, G. M. Xplor-NIH for Molecular Structure Determination from NMR and Other Data Sources. *Protein Science* **2018**, *27* (1), 26–40. https://doi.org/10.1002/pro.3248.

(25) Bermejo, G. A.; Clore, G. M.; Schwieters, C. D. Smooth Statistical Torsion Angle Potential Derived from a Large Conformational Database via Adaptive Kernel Density Estimation Improves the Quality of NMR Protein Structures. *Protein Science* **2012**, *21* (12), 1824–1836. https://doi.org/10.1002/pro.2163.

(26) Schwieters, C. D.; Clore, G. M. A Pseudopotential for Improving the Packing of Ellipsoidal Protein Structures Determined from NMR Data. *J. Phys. Chem. B* **2008**, *112* (19), 6070–6073. https://doi.org/10.1021/jp076244o.

(27) Schwieters, C. D.; Bermejo, G. A.; Clore, G. M. A Three-Dimensional Potential of Mean Force to Improve Backbone and Sidechain Hydrogen Bond Geometry in Xplor-NIH Protein Structure Determination. *Protein Science* **2020**, *29* (1), 100–110. https://doi.org/10.1002/pro.3745.

(28) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr* **2010**, *66* (2), 213–221. https://doi.org/10.1107/S0907444909052925.

(29) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J Comput Chem* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(30) Gong, Z.; Schwieters, C. D.; Tang, C. Conjoined Use of EM and NMR in RNA Structure Refinement. *PLOS ONE* **2015**, *10* (3), e0120445. https://doi.org/10.1371/journal.pone.0120445.

(31) Gupta, R.; Zhang, H.; Lu, M.; Hou, G.; Caporini, M.; Rosay, M.; Maas, W.; Struppe, J.; Ahn, J.; Byeon, I.-J. L.; Oschkinat, H.; Jaudzems, K.; Barbet-Massin, E.; Emsley, L.; Pintacuda, G.; Lesage, A.; Gronenborn, A. M.; Polenova, T. Dynamic Nuclear Polarization Magic-Angle Spinning Nuclear Magnetic Resonance Combined with Molecular Dynamics Simulations Permits

Detection of Order and Disorder in Viral Assemblies. *J. Phys. Chem. B* **2019**, *123* (24), 5048–5058. https://doi.org/10.1021/acs.jpcb.9b02293.

(32) Gres, A. T.; Kirby, K. A.; KewalRamani, V. N.; Tanner, J. J.; Pornillos, O.; Sarafianos, S. G. X-Ray Crystal Structures of Native HIV-1 Capsid Protein Reveal Conformational Variability. *Science* **2015**, *349* (6243), 99–103. https://doi.org/10.1126/science.aaa5936.

Chapter 5

MAS NMR STRUCTURE OF THE KINESIN-1 MOTOR DOMAIN BOUND TO POLYMERIZED MICROTUBULES

Content (i.e. certain figures and technical language) in this section is reprinted with permission from the published article:[1] Chunting Zhang, Changmiao Guo, Ryan W. Russell, Caitlin M. Quinn, Mingyue Li, John C. Williams, Angela M. Gronenborn, and Tatyana Polenova (**2022**) Magic-Angle-Spinning NMR Structure of the Kinesin-1 Motor Domain Assembled with Microtubules Reveals the Elusive Neck Linker Orientation. *Nat. Comms.* DOI: 10.1038/s41467-022-34026-w

Author contributions are as follows: T.P. designed the experiments and directed the project. C.Z. and C.G. prepared the fully protonated KIF5B/MT samples and performed MAS NMR experiments. T.P., C.M.Q., and A.M.G. acquired MAS NMR spectra with the CPMAS CryoProbe. C.Z. performed data analysis. R.W.R. and C.Z. performed the structure calculation and refinements and prepared figures and tables. M.L. prepared the sample of deuterated KIF5B/MT samples and performed the proton-detected MAS NMR experiments. J.C.W. designed the KIF5B constructs and provided sample preparation protocol. All authors discussed the results. C.Z., C.G., T.P., A. M. G., and R.W.R. took the lead in writing the manuscript.

5.1 Introduction

Microtubules (MTs) are cytoskeleton filaments built of α/β tubulin heterodimers. Eukaryotic MTs play critical biological roles, including intracellular transport of cargos and organelles as well as cellular mitigation and migration.[1,2] MTs are also responsible for cell structure/shape and have diameters of up to 23-27 nm and lengths of up to 50 μm.[3] Kinesins are protein motors that move along microtubules to transport various cargos (Figure 5.1). Kinesin-1 (also referred to as conventional kinesin) is the founding member of the kinesin superfamily. Kinesin-1 is critically involved during cell mitigation and migration, notably in the formation of mitotic spindles.[4,5] Mutations of kinesins are responsible for numerous diseases.[6]

Herein the structure of the ubiquitously expressed kinesin-1 isoform KIF5B, bound with polymerized MTs, was determined, by integrating MAS NMR restraints with medium-resolution cryo-EM density. To date, there have been several models of KIF5B published in different states such as free KIF5B[7], KIF5B bound with a tubulin dimer[8], and KIF5B bound with polymerized MTs.[9–12] However, a structure of KIF5B with a well-defined neck linker in complex with polymerized MTs had not been determined prior to this study. The structure reveals the position and orientation of the functionally important neck linker and how ADP induces structural and dynamic changes that ensue in the neck linker. These results demonstrate that the neck linker is in the undocked conformation and oriented in the direction opposite to the KIF5B

movement. Chemical shift perturbations and intensity changes indicate that a significant portion of ADP-KIF5B is in the neck linker docked state. This study also highlights the unique capability of MAS NMR to provide atomic-level information on dynamic regions of biological assemblies.

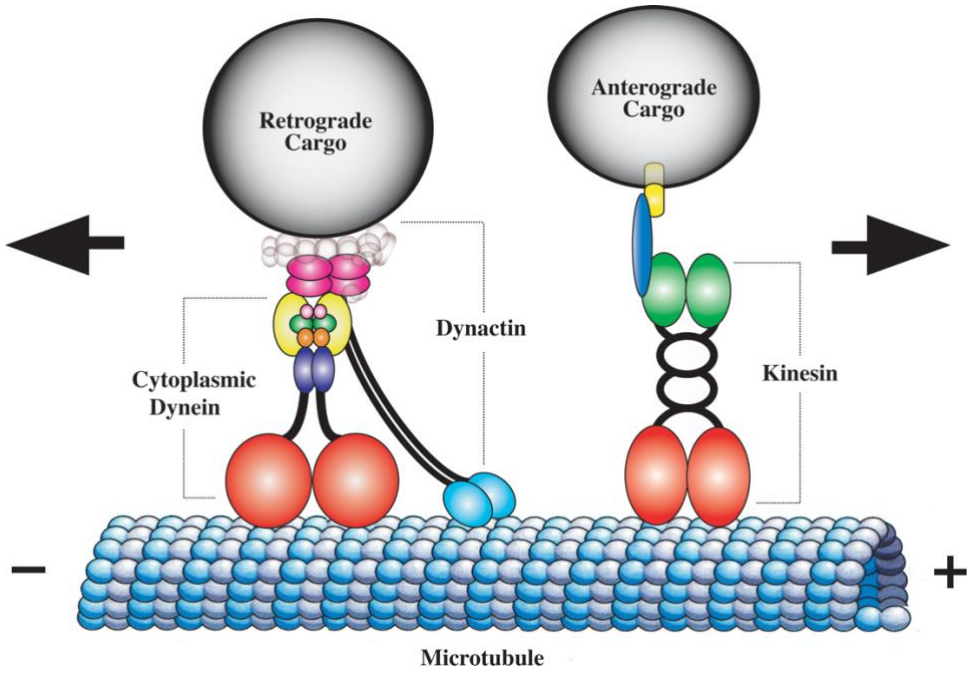

Figure 5.1: Schematic diagram of the microtubule motor proteins cytoplasmic dynein and kinesin. Cytoplasmic dynein transports cargo in the retrograde direction toward the minus ends of microtubules whereas kinesin transports cargo in the anterograde direction toward the plus ends. This figure was originally published in refrence[13]. Permission for reuse in this dissertation granted by the Creative Commons Attribution License (Appendix A).

5.2 Experimental distance and dihedral restraints by MAS NMR

Utilizing experimental chemical shift assignments determined from multiple MAS NMR datasets acquired by our team,[14] signals in $R2_n^\nu$-driven (CORD) spectra of [1,6-^{13}C-glucose, U-^{15}N]- KIF5B/MT and [2-^{13}C-glucose, U-^{15}N]- KIF5B/MT were identified as distance restraints. In total, there are 1,339 non-redundant distance restraints as summarized in Table 5.1. Ambiguous restraints exceeding five-fold ambiguity were not considered. The bounds of the distance restraints were set to 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, consistent with our previous study[15]. The chemical shifts were used to accurately predict the secondary structure elements as well as 494 dihedral (ϕ/ψ) restraints with the TALOS-N program.[16]

Table 5.1: Summary of MAS NMR distance and dihedral restraints used for structure calculations of KIF5B bound with polymerized MTs.

Distance constraints	
Unambiguous	1146
Intra-residue	937
Inter-residue	209
Sequential ($\|i\text{-}j\| = 1$)	55
Medium range ($1 \leq \|i\text{-}j\| < 5$)	43
Long range ($\|i\text{-}j\| \geq 5$) (sidechain-sidechain)	111 (56)
Ambiguous	193
Total ^{13}C-^{13}C restraints	1339
Restraints/Residue	3.85
Summary of dihedral angle restraints	
ϕ	247
ψ	247

5.3 Structure of a single subunit of KIF5B bound to MTs

Using the distance and dihedral restraints reported in § 5.2, structure calculations for the single unit were seeded from the coordinates of the cryo-EM structure (PDB 3J8X)[11] and 100 structures were annealed using Xplor-NIH 2.53[17–19]. Missing residues were built with the protocol.addUnknownAtoms routine. Two rounds of annealing were performed in the run at 3,000 K for 10 ps or 10,000 steps, whichever was completed first. The starting time step was 1 fs and was self-adjusted in subsequent steps to ensure conservation of energy. The initial velocities were randomized about a Maxwell distribution using the starting temperature of 3,000 K. Subsequently the temperatures were reduced to 25 K in steps of 25 K. At each temperature, the initial time step was set to the default value of 1 fs, and a 0.4-ps dynamics run was performed. Force constants for distance restraints were ramped up from 2 to 30 kcal mol^{-1} $Å^{-2}$. The dihedral restraint force constants were set to 10 kcal mol^{-1} rad^{-2} for high-temperature dynamics at 3,000 K and 200 kcal mol^{-1} rad^{-2} during cooling.

A global envelope in the form of synthetic density with an 8 Å resolution was prepared with UCSF Chimera[20] to preserve the overall shape of the system using the coordinates of the cryo-EM structure (PDB 3J8X)[11]. The resulting map was implemented into the Xplor-NIH run with the cryo-EM potential to generate a cross-correlation probability distribution potential[21]. The potential was restricted to backbone (N, C′, Cα, O) atoms so as not to distribute the sidechain orientations defined by NMR

distance and dihedral restraints. The cryo-EM potential was only applied to residues that are present in this work and the cryo-EM structure (PDB 3J8X)[11] (residues 3-6, 8-167, 169-173, 175-320) and not those that are sequence mismatches from the starting coordinates. The force constant of the cross-correlation probability distribution potential was set to 50 kcal mol^{-1} during high-temperature dynamics and cooling. The gyration volume force constant was turned off to avoid conflicts with the cross-correlation potential. The annealed structures were minimized using a Powell energy minimization scheme in Cartesian space. Standard terms for bond lengths, bond angles and improper angles were used to enforce correct covalent geometry. A statistical torsion angle potential[22] and the gyration volume term were employed[23]. A hydrogen bond database term, HBPot, was used to improve hydrogen bond geometries[24]. The lowest energy structure (Figure 5.2) from the run was selected for the next step.

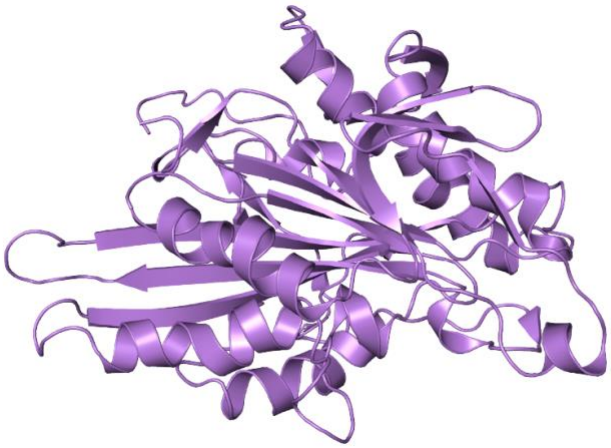

Figure 5.2: Lowest energy structure for the single-chain calculation of KIF5B/MT.

5.4 Docking into cryo-EM density

The lowest energy structure from the run described in the last subsection (Figure 5.2) was subjected to rigid-body docking about the experimental cryo-EM density map (EMD-6187, PDB 3J8X, 6 Å resolution)[11] using an in-house UCSF Chimera[20] Python script (Listing 5.1). The protocol bears similarities to previous work from our laboratory[15] with an important adaptation: instead of employing docking to identify the best fitting structure amongst many candidates in the cryo-EM density, the script identifies the best docking positions of a single structure. Here, 22 positions were identified throughout the map on the basis of lowest cross-correlation values and brief visual inspection.

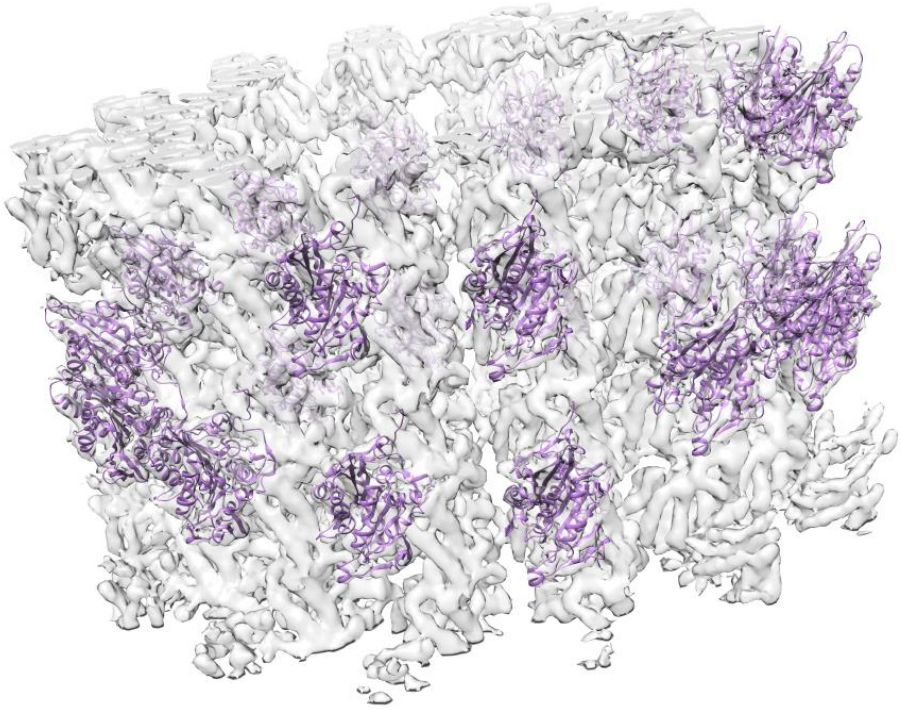

Figure 5.3: Batch docking of the lowest energy structure for the single-chain calculation about the experimental cryo-EM density map[12] (EMD-6187, PDB 3J8X, 6 Å resolution). Shown here are the 22 positions identified on the basis of lowest cross-correlation values and brief visual inspection.

```
from chimera import runCommand as run, openModels
import re
import sys
import os

# Example of how to run this script:
# /Applications/Chimera.app/Contents/MacOS/chimera --nogui --script
dock_chimera.py

# Load in pdb and density for docking
# Must have _ after number for strucIDs, otherwise just name it manually
in_pdb = "refine1_853.pdb"
stucIDs = re.search("\\d+_", in_pdb, re.M).group(0)

run('open ./54IT_noBVM_HOH_molmap_8A_resid148to238.mrc')
run(str("open ") + str(in_pdb))
# Set step level and make density a surface
run('volume #0 step 1')
run('volume #0 level 0.0063 style surface')
# Set this number high so no slots/fits missed
strucs_in_tube = "350"

# Set indices and prepare for docking
map1_id = 1
map2_id = 0
# Set number of translations and rotations cross-correlation values
(optional)
search = 250000
# Provide resolution. Script will run fine if this is approximate, the cross-
correlation
# values will be off but the values still will reveal the fits.
res = 6

# Execute the docking
from chimera import openModels as om, selection
m1 = om.list(id = map1_id)[0]
m2 = om.list(id = map2_id)[0]
s1 = selection.ItemizedSelection([m1])
from FitMap.fitcmd import fitmap
fit_list = fitmap(s1, m2, search = search, resolution = res, listFits =
False)

# Access cross-corrleation value for each docked structure
print '%d fits' % len(fit_list)
import Matrix
corrs = []
for index, fit in enumerate(fit_list):
    if int(index) <= int(strucs_in_tube):
       print 'correlation =', fit.correlation()
    corrs.append(fit.correlation())
    if int(index) <= int(strucs_in_tube):
       fit.place_copies()

# Save pdb coordinates of fit with cross-correlation in the file-name
if not os.path.exists("Docked"):
       os.makedirs("Docked")
```

```
for index, fit in enumerate(fit_list):
      if int(index) <= int(strucs_in_tube):
            struc_ind = float(index) + 2
            print int(index), struc_ind
            outname = "Docked/docked_StrucNum_" + str(int(struc_ind)) +
"_Corr_" + str(round(fit.correlation(), 4)) + ".pdb"
            print outname
            run(str("write ") + str(struc_ind) + str(" ") + outname)
```

Listing 5.1: UCSF Chimera script for batch docking.

5.5 Joint refinement with cryo-EM density

After the docking a joint refinement was performed in Xplor-NIH using the 22 molecules and the experimental cryo-EM density (EMD-6187, PDB 3J8X, 6 Å resolution) following the same protocol, parameters, and force constants as in the earlier step (§ 5.3). Each of the 22 molecules was assigned to a unique segment identifier (A-V). The protein structure file (PSF) of a single unit was loaded from the sequence file and expanded to all the 22 subunits with the psfGen.duplicateSegment function in Xplor-NIH. Coordinates were loaded from 22 files with the initCoords protocol in Xplor-NIH. The experimental distance and dihedral restraints from the first Xplor-NIH structure calculation were applied to the 22 subunits with a loop in the Python infrastructure of Xplor-NIH.

Approximate non-crystallographic symmetry was imposed using the PosDiffPot term in Xplor-NIH, allowing the 22 subunits to differ by up to 1 Å. The experimental cryo-EM map (EMD-6187, PDB 3J8X, 6 Å resolution) was incorporated into the Xplor-NIH run with the cryo-EM potential to generate a cross-correlation probability

distribution potential[21]. The calculation force constants were set to the same values as § 5.3. As with § 5.3, while the annealed structures were minimized using a Powell energy minimization scheme in Cartesian space. The lowest energy structure comprising 22 subunits from the run was selected for the next step.

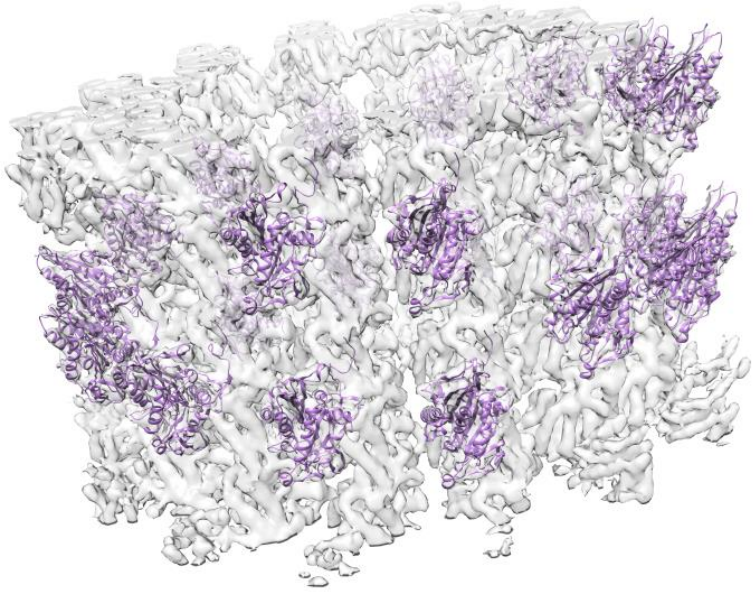

Figure 5.4: Resulting assembly of a simultaneous joint refinement of NMR distance/dihedral restraints and the cryo-EM density.

5.6 Neck-linker refinement

A final iteration of the Xplor-NIH structure calculation was performed to incorporate 17 additional restraints for the neck-linker region. For each of the 22 subunits we performed an individual run where 1,000 structures were calculated. We followed the identical protocol, parameters, and force constants as in § 5.3, except for one modification: to preserve the joint refinement with cryo-EM density only neck-linker and terminal residues (321-349) were permitted to move freely during dynamics, the remaining residues (1-320) were set to a rigid body. As with §§ 5.3,5.5, the annealed structures underwent a Powell energy minimization in Cartesian space. The lowest energy structure from each of the runs, corresponding to each of the 22 subunits, was selected for the final ensemble. Lastly, each member of the final ensemble was returned to their initial fitting in the density after joint refinement by aligning those residues that were not in the neck-linker or terminus (1-320) followed by a local density fitting in UCSF Chimera[20].

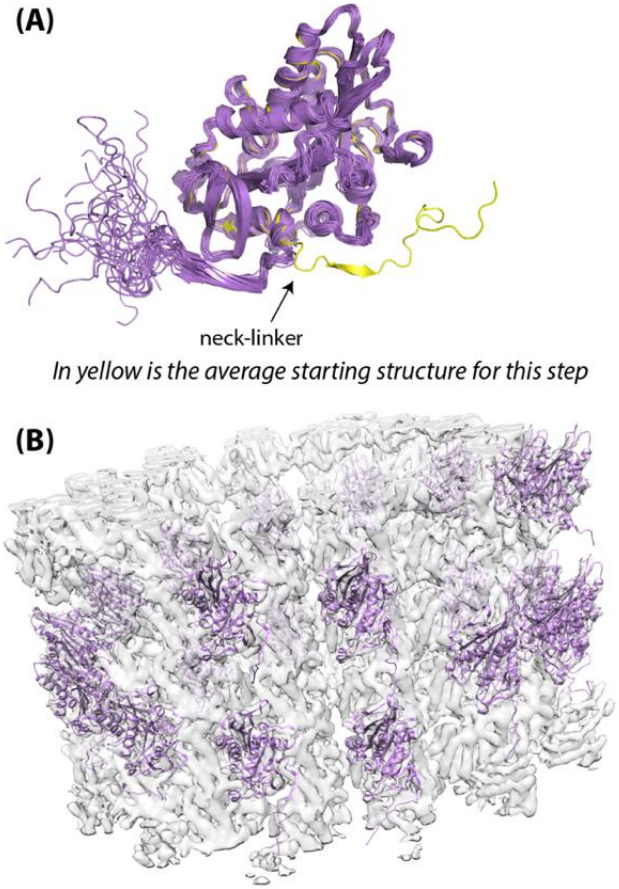

Figure 5.5: Resulting and final assembly from the neck-linker refinement of KIF5B/MT. (A) Ensemble of 22 units aligned to starting structure for the step. (B) Ensemble of the 22 units about the cryo-EM density.

5.7 Final Ensemble of KIF5B bound to polymerized MTs

The protocol for the structure determination of KIF5B bound to MTs is summarized as a flowchart, Figure 5.6. Upon completion of this procedure, we obtained a well-defined ensemble of 22 structures on the basis of lowest energy. The pairwise atomic RMSDs for the ensemble are 0.89 ± 0.09 Å and 1.24 ± 0.09 Å for backbone and

heavy atoms, respectively. (All structural statistics are provided in Table 5.2). These values are very low, especially considering the very large size of the complex (349 aa, 39.3 kDa) and the 22-member ensemble.

Table 5.2: Summary of MAS NMR distance and dihedral restraints used for structure calculations of KIF5B bound with polymerized MTs

Structure statistics from 22 lowest energy subunits	
Violations (mean ± s.d.)	
Distance restraints ≥ 7.2 Å (Å)	0.168 ± 0.001
Dihedral angle restraints ≥ 5° (°)	1.421 ± 0.077
Max. distance restraint violation* (Å)	0.855
Max. dihedral angle restraint violation (°)	13.230
Deviations from idealized geometry	
Bond lengths (Å)	0.005 ± 0.010
Bond angles (°)	0.667 ± 0.388
Improper angles (°)	0.756 ± 1.152
Average pairwise RMSD* (Å)	
Heavy	1.24 ± 0.09
Backbone (N, Cα, C)	0.89 ± 0.09

* Pairwise RMSD was calculated among 22-member ensemble.

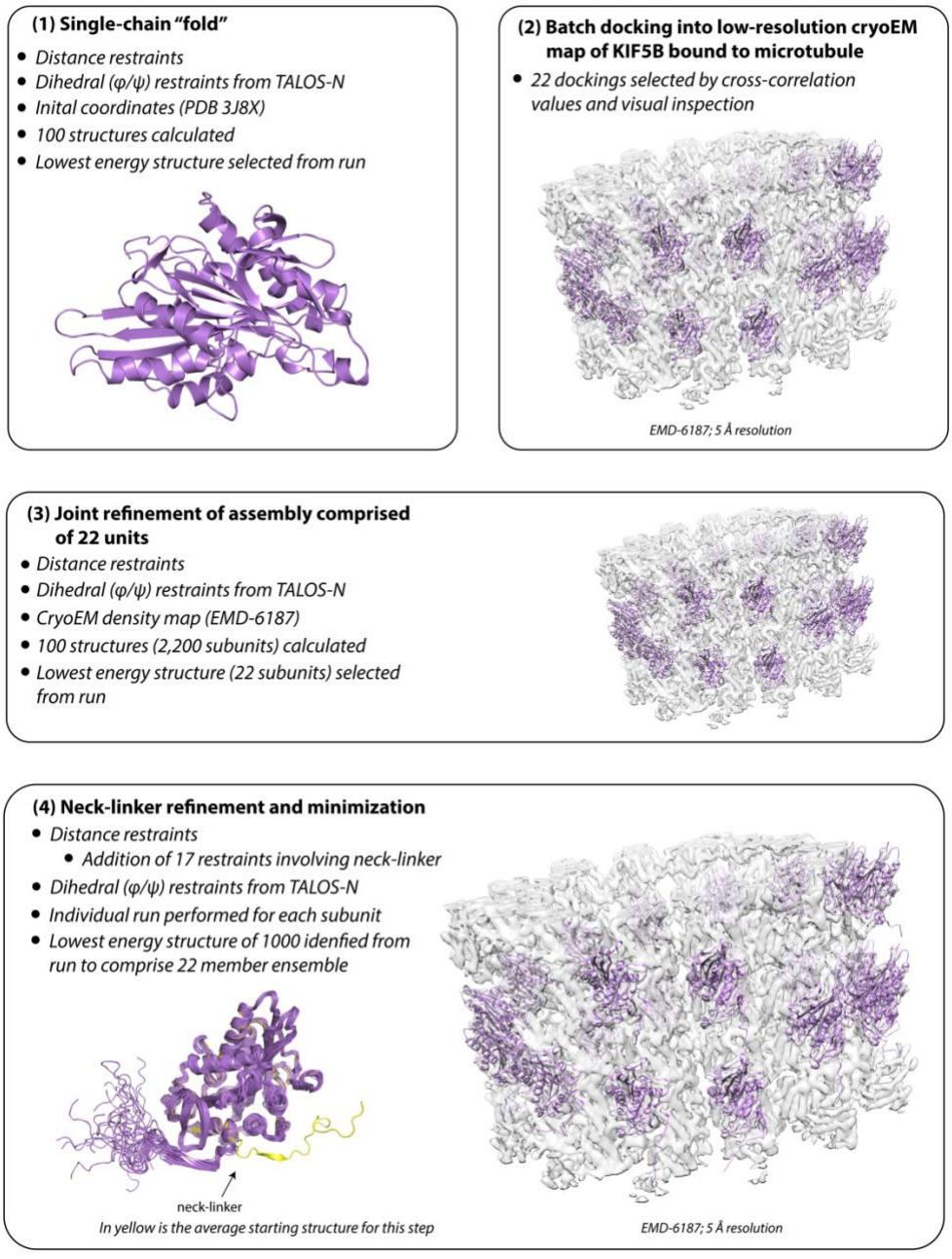

Figure 5.6: Flowchart of the structure determination protocol for KIF5B bound to MTs.

5.8 Discussion

MTs and their biological roles are vast, and only briefly discussed in § 5.1 to not exceed the scope of the dissertation. Herein we have determined the structure of a kinesin motor protein (KIF5B) bound to polymerized MTs using MAS NMR. Indeed cryo-EM structures were published prior to this work but were lacking many structural features due to insufficient resolution and disorder.[11,12] The structure presented herein revealed the previously uncharacterized functionally critical regions, including the orientation of neck-linker (Figure 5.7A), presumably unresolved in cryo-EM due to its unstructured nature. This "undocked" neck-linker conformation, evident in all 22 subunits and corroborated by experimental NMR distance restraints (Figure 5.7B), agrees with the hypothesized conformation in low-resolution FRET and MD simulation studies.[25,26]

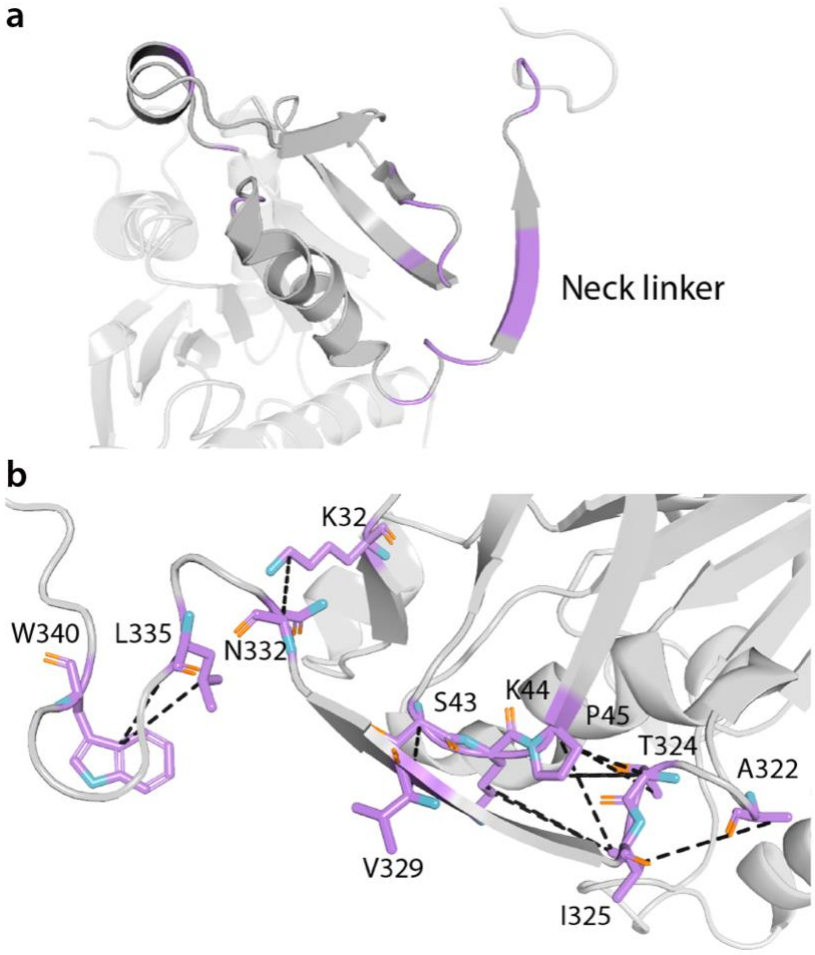

Figure 5.7: Neck-linker conformation (A) revealed by the MAS NMR calculation corroborated by long-range correlations (B) of KIF5B bound to MTs.

In addition, the tubulin dimer interface with heightened details (Figure 5.8A) is revealed, including sidechain conformations. Another important region of interest is the nucleotide-binding region, which was revealed to be in an “open” state (Figure 5.8B).

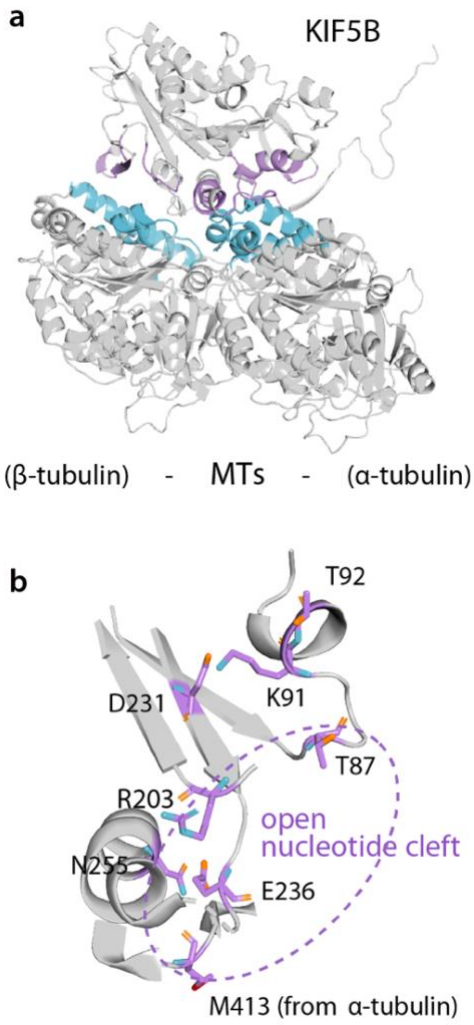

Figure 5.8: Tubulin dimer interface (A) and the "open" nucleotide-binding region (B) of KIF5B bound to MTs.

5.9 Conclusions

Overall, the structure of nucleotide-free KIF5B bound to MTs represents an important advance, enabled by exciting recent developments in MAS NMR technology and progress in integrated methodological approaches combining experimental data

from several techniques and computation. Specifically, by integrating MAS NMR restraints and the cryo-EM density map, several critical structural features were revealed for understanding the molecular mechanism of KIF5B's processivity on MTs. Looking forward, emerging MAS NMR technologies, such as ultrahigh magnetic fields (28.2 T) and ultrafast MAS probes capable of spinning frequencies of 100-150 kHz, delivering dramatic sensitivity and resolution enhancements[27–30], will open doors for structural and dynamics studies of very large microtubule-based protein assemblies, with atomic level detail.

REFERENCES

(1) Vale, R. D. The Molecular Motor Toolbox for Intracellular Transport. *Cell* **2003**, *112* (4), 467–480. https://doi.org/10.1016/s0092-8674(03)00111-9.

(2) Garcin, C.; Straube, A. Microtubules in Cell Migration. *Essays Biochem* **2019**, *63* (5), 509–520. https://doi.org/10.1042/EBC20190016.

(3) Ledbetter, M. C.; Porter, K. R. A "Microtubule" In Plant Cell Fine Structure. *Journal of Cell Biology* **1963**, *19* (1), 239–250. https://doi.org/10.1083/jcb.19.1.239.

(4) Hirokawa, N.; Noda, Y.; Tanaka, Y.; Niwa, S. Kinesin Superfamily Motor Proteins and Intracellular Transport. *Nat Rev Mol Cell Biol* **2009**, *10* (10), 682–696. https://doi.org/10.1038/nrm2774.

(5) Wordeman, L. How Kinesin Motor Proteins Drive Mitotic Spindle Function: Lessons from Molecular Assays. *Semin Cell Dev Biol* **2010**, *21* (3), 260–268. https://doi.org/10.1016/j.semcdb.2010.01.018.

(6) Huszar, D.; Theoclitou, M.-E.; Skolnik, J.; Herbst, R. Kinesin Motor Proteins as Targets for Cancer Therapy. *Cancer Metastasis Rev* **2009**, *28* (1–2), 197–208. https://doi.org/10.1007/s10555-009-9185-8.

(7) Cao, L.; Cantos-Fernandes, S.; Gigant, B. The Structural Switch of Nucleotide-Free Kinesin. *Sci Rep* **2017**, *7* (1), 42558. https://doi.org/10.1038/srep42558.

(8) Gigant, B.; Wang, W.; Dreier, B.; Jiang, Q.; Pecqueur, L.; Plückthun, A.; Wang, C.; Knossow, M. Structure of a Kinesin–Tubulin Complex and Implications for Kinesin Motility. *Nat Struct Mol Biol* **2013**, *20* (8), 1001–1007. https://doi.org/10.1038/nsmb.2624.

(9) Jon Kull, F.; Sablin, E. P.; Lau, R.; Fletterick, R. J.; Vale, R. D. Crystal Structure of the Kinesin Motor Domain Reveals a Structural Similarity to Myosin. *Nature* **1996**, *380* (6574), 550–555. https://doi.org/10.1038/380550a0.

(10) Cao, L.; Wang, W.; Jiang, Q.; Wang, C.; Knossow, M.; Gigant, B. The Structure of Apo-Kinesin Bound to Tubulin Links the Nucleotide Cycle to Movement. *Nat Commun* **2014**, *5* (1), 5364. https://doi.org/10.1038/ncomms6364.

(11) Shang, Z.; Zhou, K.; Xu, C.; Csencsits, R.; Cochran, J. C.; Sindelar, C. V. High-Resolution Structures of Kinesin on Microtubules Provide a Basis for Nucleotide-Gated Force-Generation. *eLife* **2014**, *3*, e04686. https://doi.org/10.7554/eLife.04686.

(12) Debs, G. E.; Cha, M.; Liu, X.; Huehn, A. R.; Sindelar, C. V. Dynamic and Asymmetric Fluctuations in the Microtubule Wall Captured by High-Resolution Cryoelectron Microscopy. *Proceedings of the National Academy of Sciences* **2020**, *117* (29), 16976–16984. https://doi.org/10.1073/pnas.2001546117.

(13) Duncan, J. E.; Goldstein, L. S. B. The Genetics of Axonal Transport and Axonal Transport Disorders. *PLOS Genetics* **2006**, *2* (9), e124. https://doi.org/10.1371/journal.pgen.0020124.

(14) Zhang, C.; Guo, C.; Russell, R. W.; Quinn, C. M.; Li, M.; Williams, J. C.; Gronenborn, A. M.; Polenova, T. Magic-Angle-Spinning NMR Structure of the Kinesin-1 Motor Domain Assembled with Microtubules Reveals the Elusive Neck Linker Orientation. *Nat Commun* **2022**, *13* (1), 6795. https://doi.org/10.1038/s41467-022-34026-w.

(15) Lu, M.; Russell, R. W.; Bryer, A. J.; Quinn, C. M.; Hou, G.; Zhang, H.; Schwieters, C. D.; Perilla, J. R.; Gronenborn, A. M.; Polenova, T. Atomic-Resolution Structure of HIV-1 Capsid Tubes by Magic-Angle Spinning NMR. *Nature Structural & Molecular Biology* **2020**, *27* (9), 863–869. https://doi.org/10.1038/s41594-020-0489-2.

(16) Shen, Y.; Bax, A. Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *Journal of Biomolecular NMR* **2013**, *56* (3), 227–241. https://doi.org/10.1007/s10858-013-9741-y.

(17) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. The Xplor-NIH NMR Molecular Structure Determination Package. *Journal of Magnetic Resonance* **2003**, *160* (1), 65–73.

(18) Schwieters, C. D.; Bermejo, G. A.; Clore, G. M. Xplor-NIH for Molecular Structure Determination from NMR and Other Data Sources. *Protein Science* **2018**, *27* (1), 26–40. https://doi.org/10.1002/pro.3248.

(19) Schwieters, C.; Kuszewski, J.; Mariusclore, G. Using Xplor–NIH for NMR Molecular Structure Determination. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2006**, *48* (1), 47–62. https://doi.org/10.1016/j.pnmrs.2005.10.001.

(20) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J Comput Chem* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(21) Gong, Z.; Schwieters, C. D.; Tang, C. Conjoined Use of EM and NMR in RNA Structure Refinement. *PLOS ONE* **2015**, *10* (3), e0120445. https://doi.org/10.1371/journal.pone.0120445.

(22) Bermejo, G. A.; Clore, G. M.; Schwieters, C. D. Smooth Statistical Torsion Angle Potential Derived from a Large Conformational Database via Adaptive Kernel Density Estimation Improves the Quality of NMR Protein Structures. *Protein Science* **2012**, *21* (12), 1824–1836. https://doi.org/10.1002/pro.2163.

(23) Schwieters, C. D.; Clore, G. M. A Pseudopotential for Improving the Packing of Ellipsoidal Protein Structures Determined from NMR Data. *J. Phys. Chem. B* **2008**, *112* (19), 6070–6073. https://doi.org/10.1021/jp076244o.

(24) Schwieters, C. D.; Bermejo, G. A.; Clore, G. M. A Three-Dimensional Potential of Mean Force to Improve Backbone and Sidechain Hydrogen Bond Geometry in Xplor-NIH Protein Structure Determination. *Protein Science* **2020**, *29* (1), 100–110. https://doi.org/10.1002/pro.3745.

(25) Tomishige, M.; Stuurman, N.; Vale, R. D. Single-Molecule Observations of Neck Linker Conformational Changes in the Kinesin Motor Protein. *Nat Struct Mol Biol* **2006**, *13* (10), 887–894. https://doi.org/10.1038/nsmb1151.

(26) Budaitis, B. G.; Jariwala, S.; Reinemann, D. N.; Schimert, K. I.; Scarabelli, G.; Grant, B. J.; Sept, D.; Lang, M. J.; Verhey, K. J. Neck Linker Docking Is Critical for Kinesin-1 Force Generation in Cells but at a Cost to Motor Speed and Processivity. *eLife* **2019**, *8*, e44146. https://doi.org/10.7554/eLife.44146.

(27) Callon, M.; Malär, A. A.; Pfister, S.; Římal, V.; Weber, M. E.; Wiegand, T.; Zehnder, J.; Chávez, M.; Cadalbert, R.; Deb, R.; Däpp, A.; Fogeron, M.-L.; Hunkeler, A.; Lecoq, L.; Torosyan, A.; Zyla, D.; Glockshuber, R.; Jonas, S.; Nassal, M.; Ernst, M.; Böckmann, A.; Meier, B. H. Biomolecular Solid-State NMR Spectroscopy at 1200 MHz: The Gain in Resolution. *J Biomol NMR* **2021**, *75* (6), 255–272. https://doi.org/10.1007/s10858-021-00373-x.

(28) Nimerovsky, E.; Movellan, K. T.; Zhang, X. C.; Forster, M. C.; Najbauer, E.; Xue, K.; Dervişoğlu, R.; Giller, K.; Griesinger, C.; Becker, S.; Andreas, L. B. Proton Detected Solid-State NMR of Membrane Proteins at 28 Tesla (1.2 GHz)

and 100 KHz Magic-Angle Spinning. *Biomolecules* **2021**, *11* (5), 752. https://doi.org/10.3390/biom11050752.

(29) Struppe, J.; Quinn, C. M.; Lu, M.; Wang, M.; Hou, G.; Lu, X.; Kraus, J.; Andreas, L. B.; Stanek, J.; Lalli, D.; Lesage, A.; Pintacuda, G.; Maas, W.; Gronenborn, A. M.; Polenova, T. Expanding the Horizons for Structural Analysis of Fully Protonated Protein Assemblies by NMR Spectroscopy at MAS Frequencies above 100 KHz. *Solid State Nuclear Magnetic Resonance* **2017**, *87*, 117–125. https://doi.org/10.1016/j.ssnmr.2017.07.001.

(30) Schledorn, M.; Malär, A. A.; Torosyan, A.; Penzel, S.; Klose, D.; Oss, A.; Org, M.-L.; Wang, S.; Lecoq, L.; Cadalbert, R.; Samoson, A.; Böckmann, A.; Meier, B. H. Protein NMR Spectroscopy at 150 KHz Magic-Angle Spinning Continues To Improve Resolution and Mass Sensitivity. *ChemBioChem* **2020**, *21* (17), 2540–2548. https://doi.org/10.1002/cbic.202000341.

Appendix A

COPYRIGHT PERMISSIONS

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jun 14, 2022

This Agreement between University of Delaware -- Ryan Russell ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5327820444352
License date	Jun 14, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Journal of Biomolecular NMR
Licensed Content Title	Accuracy and precision of protein structures determined by magic angle spinning NMR spectroscopy: for some ‘with a little help from a friend’
Licensed Content Author	Ryan W. Russell et al
Licensed Content Date	Mar 7, 2019
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	yes
Title	Computations of atomic-resolution structures of protein assemblies from experimental magic angle spinning NMR restraints
Institution name	University of Delaware
Expected presentation date	Sep 2022

Requestor Location	University of Delaware 041 Brown Laboratory NEWARK, DE 19716 United States Attn: University of Delaware
Total	0.00 USD

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months

Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:*** Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8. 2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Atomic-resolution structure of HIV-1 capsid tubes by magic-angle spinning NMR

Author: Manman Lu et al
Publication: Nature Structural & Molecular Biology
Publisher: Springer Nature
Date: Sep 8, 2020

Author Request

PLOS GENETICS

The Genetics of Axonal Transport and Axonal Transport Disorders

Jason E Duncan, Lawrence S. B Goldstein

Published: September 29, 2006 • https://doi.org/10.1371/journal.pgen.0020124

Abstract

Neurons are specialized cells with a complex architecture that includes elaborate dendritic branches and a long, narrow axon that extends from the cell body to the synaptic terminal. The organized transport of essential biological materials throughout the neuron is required to support its growth, function, and viability. In this review, we focus on insights that have emerged from the genetic analysis of long-distance axonal transport between the cell body and the synaptic terminal. We also discuss recent genetic evidence that supports the hypothesis that disruptions in axonal transport may cause or dramatically contribute to neurodegenerative diseases.

Citation: Duncan JE, Goldstein LSB (2006) The Genetics of Axonal Transport and Axonal Transport Disorders. PLoS Genet 2(9): e124. https://doi.org/10.1371/journal.pgen.0020124

Editor: Elizabeth M. C. Fisher, University College London, United Kingdom

Published: September 29, 2006

Funding: LSBG is an investigator of the Howard Hughes Medical Institute.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: AchE, acetylcholinesterase; AD, Alzheimer disease; ALS, amyotrophic lateral sclerosis; APP, amyloid precursor protein; ChAT, choline acetyltransferase; CMT, Charcot-Marie-Tooth disease; CSP, cysteine-string protein; Dhc, cytoplasmic dynein heavy chain; Dlc, cytoplasmic dynein light chain; *Dync1h1* , dynein heavy chain gene; GAP-43, growth associated protein 43; GSK 3β, glycogen synthase kinase 3 β; HAP1, Huntingtin-associated protein 1; HD, Huntington disease; HSP, Hereditary Spastic Paraplegia; HSP(SPG 10), Hereditary Spastic Paraplegia Type 10; Htt, huntingtin protein; JIP, JNK interacting protein; JNK, cJun NH_2-terminal kinase; Khc, kinesin heavy chain; KIFs, kinesin superfamily members; Klc, kinesin light chain; PS1, presenilin-1; Snb-GFP, synaptobrevin-GFP; SOD1, Cu/Zn superoxide dismutase; Syt, synaptotagmin; UNC-104, UNCoordinated-104