SOLID-STATE NMR STUDIES OF STRUCTURE AND DYNAMICS OF HIV-1 CAPSID (CA) PROTEIN ASSEMBLIES

by

Yun Han

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

Spring 2013

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

SOLID-STATE NMR STUDIES OF STRUCTURE AND DYNAMICS OF HIV-1 CAPSID (CA) PROTEIN ASSEMBLIES

by

Yun Han

Approved: _____
Murray V. Johnston, Ph.D.
Chair of the Department of Chemistry and Biochemistry

Approved: _____
George H. Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____
James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

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: _____

Karl Booksh, 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: _____

Sharon Neal, 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: _____

Sharon Rozovsky, 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: _____

John Williams, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I am happy to have this opportunity to show my sincerest appreciation to my adviser, Professor Tatyana Polenova, for her continuous support and encouragement throughout my study at the University of Delaware. Her critical attitude and rigorous scholarship in research have a great influence on me, and I believe it will accompany me for the rest of my life. Without her, this thesis would not be completed.

I would also like to express my gratitude to the group members in Polenova lab. I had great experience working with my fellow graduate students: Sivakumar Paramasivam, Shangjin Sun, Stephanie Bolte, Olga Zapata, Si Yan, Christopher Suiter, Mingyue Li, Jenna Yehl, Manman Lu, Changmiao Guo. Thanks to the postdoc researchers and other members of the Polenova group: Dr. Guangjin Hou, Dr. Jun Yang, and Dr. Kris Ooms, Dr. Lex Vega and Ms. Vera Vladmirova for their valuable suggestions and comments on my research. The collaboration with PCHPI people: Professor Angela M. Gronenborn, Professor Jinwoo Ahn and Professor In-Ja L. Byeon etc., was very pleasantly and greatly appreciated. And the people in the high magnetic field facilities are sincerely acknowledged for their inputs during various phases of the projects: Dr. Andrew Lipton and Dr. Sarah Burton in EMSL and PNNL; Dr. William Brey, Dr. Zhehong Gan, Dr. Peter Gor'kov and Dr. Ivan Hung in NHMFL; and Dr. Robert Griffin and the Griffin group. My thesis committee members, Professor John

Williams in the Beckman Research Institute, City of Hope, Professor Sharon Rozovsky, Professor Sharon Neal and Professor Karl Booksh in our department, must also be acknowledged. And I express my sincere appreciation to Professor Cecil Dybowski, and Dr. Steve Bai for their patience in the classes and warm-hearted help after classes. All the people and friends in the Department of Chemistry and Biochemistry are appreciated. They have made my life in UD more fun and informative.

Finally, I am very grateful for my husband for his love, patience and care. The support from the family is my greatest encouragement to pursue my further career.

TABLE OF CONTENTS

LIST OF TABLES xii
LIST OF FIGURES xiii
LIST OF ABBREVIATIONS xxiv
ABSTRACT xxv

Chapter

1 INTRODUCTION 1

1.1 HIV-1 viral lifecycle 1
1.2 HIV-1 capsid (CA): structural organization 6
1.3 Gag polyprotein intermediates 8
1.4 Maturation inhibitors 11
References 13

2 SOLID-STATE NMR SPECTROSCOPY FOR STRUCTURAL AND DYNAMICS CHARACTERIZATION OF PROTEIN ASSEMBLIES 19

2.1 Introduction 19
2.2 Essential spin interactions 19
2.3 General protocol for structural characterization of proteins by SSNMR 23

2.3.1 Site-specific resonance assignments: experiments and strategies 23
2.3.2 Secondary structure analysis 27
2.3.3 Experiments for gaining distance restraints 27

2.4 Characterization of protein dynamics by SSNMR 29

2.4.1 Motional timescales and their relationship to protein function 29
2.4.2 Spin interactions and relaxation rates as a probe of protein motions 30

2.4.3 Experiments to probe nano- to micro-second timescale motions 32

2.4.3.1 ^{13}C-^{1}H and ^{15}N-^{1}H dipolar lineshapes 32
2.4.3.2 ^{15}N CSA lineshapes 33

2.4.4 Experiments to probe millisecond timescale motions 35

2.4.4.1 ^{13}C-^{15}N REDOR dipolar dephasing curves 35

References 38

3 PREPARATION AND CHARACTERIZATION OF HIV-1 CA AND CA-SP1 PROTEIN ASSEMBLIES 43

3.1 Introduction 43
3.2 Preparation of CA assemblies of conical morphology 45

3.2.1 Protein purification 45
3.2.2 Assembly 46
3.2.3 Microscopy 46

3.2.3.1 Transmission Electron Microscopy (TEM) 47
3.2.3.2 Cryo Scanning Electron Microscopy (Cryo-SEM) 47
3.2.3.3 Confocal Laser Scanning Microscopy 49

3.3 Spherical morphologies 50

3.3.1 Protein purification 51
3.3.2 Assembly 51
3.3.3 Microscopy 51

3.3.3.1 Cryo Scanning Electron Microscopy (Cryo-SEM) 52
3.3.3.2 Confocal Laser Scanning Microscopy 53

3.4 Tubular morphologies 54

3.4.1 Protein purification 54
3.4.2 Assembly 55
3.4.3 Microscopy 55

3.4.3.1 Transmission Electron Microscopy (TEM) 55
3.4.3.2 Confocal Laser Scanning Microscopy 56

References 59

4 RESONANCE ASSIGNMENTS AND SECONDARY STRUCTURE ANALYSIS OF CONICAL ASSEMBLIES OF HIV-1 CA PROTEIN 60

4.1 Introduction 60
4.2 Experiments and methods 61

4.2.1 Materials 61
4.2.2 Preparation of HIV-1 CA protein assemblies 61
4.2.3 SSNMR Spectroscopy 62
4.2.4 NMR data processing and analysis 64

4.3 The effect of morphology on solid-state NMR spectra 64
4.4 Resonance assignments of conical CA assemblies 67
4.5 Secondary structure analysis (CSI) 85
References 88

5 DYNAMICS STUDIES OF CONICAL ASSEMBLIES OF U-^{13}C, ^{15}N-TYR-LABELED HIV-1 CA 91

5.1 Introduction 91
5.2 Materials and methods 93

5.2.1 Expression and purification of U-^{13}C,^{15}N-Tyr labeled CA 93
5.2.2 Preparation of U-^{13}C, ^{15}N-Tyr labeled protein samples for SSNMR 94
5.2.3 SSNMR spectroscopy 95
5.2.4 Analysis and simulations of solid-state NMR spectra 96
5.2.5 CSA tensor convention 97

5.3 Resonance assignments of Tyr residues 97
5.4 Nano- to microsecond timescale dynamics in U-^{13}C, ^{15}N-Tyr CA 101

5.4.1 ^{1}H-^{15}N dipolar lineshapes 101
5.4.2 ^{15}N CSA lineshapes 102

5.5 Millisecond timescale dynamics in CA 103

5.5.1 Temperature dependence of heteronuclear 2D NCA spectra 104
5.5.2 CA-N spectra 105
5.5.3 ^{13}C-^{15}N REDOR dipolar dephasing curves 107

References 109

6 RESONANCE ASSIGNMENTS AND CONFORMATIONAL ANALYSIS OF TUBULAR ASSEMBLIES OF HIV-1 CA AND CA-SP1 PROTEINS 113

6.1 Introduction 113
6.2 Experiments and Methods 115

6.2.1 Sample preparation 115
6.2.2 Transmission electron microscopy and confocal laser scanning microscopy 116
6.2.3 NMR spectroscopy 116
6.2.4 NMR data processing and analysis 117

6.3 Characterization of CA and CA-SP1 tubular assemblies 118
6.4 Tubular assemblies yield ultrahigh resolution spectra 120
6.5 Resonance Assignments 121

6.5.1 2D spectra assignments 121
6.5.2 3D backbone walks 123

6.6 SP1 is unstructured in assembled CA-SP1 135

6.6.1 SP1 signals in the through-space correlation SSNMR spectra 135

6.6.1.1 Direct dipolar-based polarization transfer and spin diffusion experiments 135
6.6.1.2 Nuclear Overhauser cross relaxation based experiments 139

6.6.2 J-based correlation SSNMR spectra of HIV-1 CA and CA-SP1 assemblies 142

6.6.2.1 INADEQUATE and CP-INADEQUATE spectra of CA and CA-SP1 A92E tubular assemblies 142
6.6.2.2 INADEQUATE and CP-INADEQUATE spectra of CA WT and CA-SP1 WT tubular assemblies 153

6.7 The effect of mutation and variability in the protein sequence on the chemical shifts 159
6.8 Conclusions 161
References 162

7 CONCLUSIONS AND FUTURE OUTLOOK 166

References 168

Appendix

A SUPPORTING INFORMATION A 169
B SUPPORTING INFORMATION B 191
C PERMISSION LETTERS 200

LIST OF TABLES

Table 4.1 ^{13}C and ^{15}N Chemical Shift Assignments for the CA Protein in Conical Assemblies. 73

Table 6.1 ^{13}C and ^{15}N chemical shift assignments (ppm) for the CA protein in tubular assemblies. The CSI column is the secondary structure predicted by CSI method: H-α helix, B-β sheet, C-unstructured coil. 126

Table 6.2 Chemical shift assignments of the SP1 tail and nearby residues of CA and CA-SP1 A92E tubular assemblies. The CSI column is the secondary structure predicted by CSI method: H-α helix, B-β sheet, C-unstructured coil. 148

Table 6.3 Chemical shifts of the SP1 tail and nearby residues of CA-SP1 WT tubular assemblies. The CSI column is the secondary structure predicted by CSI method: H-α helix, B-β sheet, C-unstructured coil. 156

Table A.1 Chemical shift differences between solid-state and solution for the assigned residues in the conical assemblies of HIV-1 CA 169

Table A.2 Summary of the resonance assignments for CA conical assemblies..... 178

Table A.3 Acquisition and processing parameters for the solid-state NMR experiments probing CA conical assemblies..... 185

Table B.1 Acquisition and processing parameters for the solid-state NMR experiments probing CA tubular assemblies. 191

LIST OF FIGURES

Figure 1.1 A general view of HIV-1 viral life cycle. In the early stage of the replication (upper portion), a mature virus particle attaches itself to a host cell through glycoproteins targeting receptors on the host cells plasma membrane. Once attached, the HIV-1 virus fuses with the host cell. At some point between the fusion and integration steps the conical capsid core uncoats to release viral proteins and RNA into the cytosol where RNA is transcribed into DNA through the enzyme reverse transcriptase. This DNA is transported to the cells nucleus where it is integrated into the host cells genome with help from the enzyme integrase. Then the late stage of the replication starts (lower portion). The cellular genome expresses viral RNA, which is released to the cytosol and then translated into viral proteins. These proteins are then transported to the plasma membrane through various means, where they bind viral RNA and are released from the host cell by a budding mechanism. Sometime during or shortly after budding, the Gag polyprotein, which makes up the major structural proteins of HIV-1 is cleaved into its constituent domains in a process termed maturation which can be viewed by a condensation of an electron dense conical capsid. This maturation process results in an infectious HIV-1 virion. This figure was originally published in reference [31]. Permission for reuse in this dissertation is granted by Nature Publishing Group., Nature Reviews Microbiology, copyright 2005. 5

Figure 1.2 Structure of the mature HIV-1 capsid protein. Left: Tertiary structure of CA monomer, with the N-terminal domain (NTD) shown in dark green and C-terminal domain (CTD) colored in blue-green. The beginning of CA protein is the β-hairpin and is colored yellow. The final 11 residues are typically disordered in crystal structures an represented as dash line. Right: Fullerene model for the conical capsid. CA hexamers are colored green and pentameric declinations are colored red. This figure was originally published in reference [39]. Permission for reuse in this dissertation is granted by Elsevier Ltd., Current Opinion in Structural Biology, copyright 2008. 9

Figure 1.3 Gag cleavage cascade (a) and the TEM images of viral particles (b). (a) Gag processing cascade, illustrating the order in which the Gag precursor is cleaved by the viral protease. Each cleavage site is indicated by an

arrow, the red arrow depicts the cleavage blocked by BVM. (b) Virion morphology visualized by transmission electron microscopy. Immature particles are shown in (i) and mature particles in (ii). An aberrant viron particle produced in the presence of BVM was shown in (iii). This figure was originally published in reference [71]. Permission for reuse in this dissertation is granted by Elsevier Ltd., Drug Discovery Today, Copyright 2008. 10

Figure 1.4 The chemical structure of Bevirimat, an inhibitor of HIV-1 Gag processing [77]. 12

Figure 2.1 The magnetization transfer pathways of the dipolar-based SSNMR experiments used for resonance assignments: NCA, NCO, NCACX, NCOCX, NCACB, and DARR. The colored atoms are involved in the dipolar-based magnetization transfer. 25

Figure 2.2 Resonance assignments from 3D backbone walks. (A) 3D data sets are represented as cubes. The signals are resolved in three dimensions. NCACX and NCOCX data sets are shown in purple and blue, respectively. The extracted strips are aligned in an alternating order by same nitrogen chemical shifts for easy comparison. (B) A representation of the aligned strips for a string of residues. The nitrogen chemical shifts connect the subsequent residues in the string, and the C^{α} and C^{β} chemical shifts clearly indicate the residue type. Once the residue types have been identified, they can be matched to the protein sequence to obtain sequence-specific resonance assignments. 28

Figure 2.3 Timescale of dynamic processes in proteins (top) and the corresponding NMR techniques to probe (bottom). Figure redrawn from [43]..... 30

Figure 2.4 The two-dimensional $R18_1^7$ experiment for the recoupling of dipolar line shapes. In this experiment, C-H or N-H dipolar couplings are recorded site-specifically, separated by the ^{13}C or ^{15}N chemical shifts, respectively. 33

Figure 2.5 A two-dimensional ROCSA pulse sequence for the recoupling of $^{13}C/^{15}N$ CSA line shapes..... 34

Figure 2.6 REDOR pulse sequence [52]. Following ramped ^{1}H-^{15}N cross polarization [53], the ^{15}N signal is observed as a spin-echo. The intensity of the echo is modulated due to the ^{13}C-^{15}N dipole interactions, recoupled by applying rotor-synchronized π pulses to ^{13}C spins. Reference (S_0) experiments were performed in the absence of ^{13}C pulses..... 36

Figure 2.7 The procedure to extract results from ^{13}C-^{15}N REDOR experiments. (A) S_0 and S are the peak intensities in the absence and presence of REDOR dephasing pulses, respectively. $\Delta S = S_0 - S$. (B) In the REDOR dephasing curves, the experimental REDOR fraction $\Delta S/S_0$ (blue dots) is plotted against the dephasing time. The best-fit simulated curve (black line) yields the value of the ^{13}C-^{15}N dipolar coupling constant. 37

Figure 3.1 TEM images of CA conical assemblies. The assemblies were prepared by lyophilization and subsequent addition of 17.5% w/v PEG 20,000 solution. The resulting sample was stained with 5% ammonium molybdate solution on the coper grid. (A) The TEM image of CA protein conical assemblies taken at magnification of 50,000x; (B) The TEM image CA protein conical assemblies taken at magnification of 12,500x. 47

Figure 3.2 The cryo-SEM images of CA protein conical assemblies. The size distribution in the cryo-SEM data is much wider than in the images observed by TEM, i.e. from less than 1 micron to several microns. The shapes of assemblies observed are multi-hedral, and are not as homogeneous as observed with TEM images. (A) The cryo-SEM image of CA protein conical assemblies taken at the magnification of 5000x; (B) The cryo-SEM image of CA protein assemblies taken at the magnification of 3000x 48

Figure 3.3 The confocal images of CA conical assemblies stained with Nile Blue A dye (fluorescence mode). Since the resolution of the confocal microscope is limited in 200 nm, it is not possible to observe the shapes of the conical assemblies with confocal method. (A) The confocal image of CA protein conical assemblies; (B) The confocal image of CA protein conical assemblies after magic angle spinning 50

Figure 3.4 The cryo-SEM images of CA protein spherical assemblies. The sizes of the spheres vary from less than one micron to tens of microns. The spherical shapes are well defined and preserved. (A) The cryo-SEM image of CA protein spherical assemblies acquired at the magnification of 13,000x; (B) The cryo-SEM image of CA protein assemblies acquired at the manification of 1,000x 52

Figure 3.5 The confocal images of CA spherical assemblies reveal formation of homogeneous spherical assemblies with diameter of around 2 micron. (A) The confocal image of CA spherical assemblies in solution; (B) The confocal image of CA protein spherical assemblies after being dried under N_2 gas flow. (C) The confocal image of CA protein spherical

assemblies after magic angle spinning. The image is an overlay of the fluorescence and transmission light signals. The fluorescence channel visualizes the protein assemblies, while the transmission light channel visualizes the dried PEG 20,000 that preserves the protein. 54

Figure 3.6 The TEM images of CA protein tubular assemblies reveal tubes that are around 50 nm in diameter. The sample was prepared as described in the Assembly section. (A) The TEM image of CA protein tubular assemblies acquired at the Zeiss CEM 902 transmission electron microscope; (B) The TEM image of CA protein tubular assemblies acquired at the Zeiss Libra 120 transmission electron microscope. 56

Figure 3.7 The confocal images of CA tubular assemblies. The CA tubes are clearly visualized in the fluorescence images when stained with Nile Blue A. (A) The CA protein tubular assemblies in solution; (B) The CA protein tubular assemblies after been air dried. 57

Figure 3.8 Confocal images of (A) the CA protein tubular assemblies prepared as described in the Assembly section; (B) the same sample that was flash frozen in liquid nitrogen and then thawed at room temperature; and (C) the CA protein tubular assemblies after magic angle spinning. 58

Figure 4.1 21.1 T DARR spectrum of U-^{13}C,^{15}N HIV-1 CA conical assemblies: aliphatic region (right) and carbonyl region (left). The 1D projections in the ω_2 and ω_1 dimensions corresponding to the resolved cross-peaks for Thr residues are shown and illustrate the remarkable spectral sensitivity and resolution. The spectrum was acquired as (3072 x 216) (complex x real) matrices with 220 scans; the recycle delay was 2 s. The spectral widths were 100 kHz and 50 kHz in the ω_2 and ω_1 dimensions, respectively, and the DARR mixing time was 50 ms. Cross-peaks corresponding to the different amino acid types appear in distinct, resolved regions of the spectrum and are color coded as follows: Ala, red (C^{α}-C^{β} and C^{β}-C' correlations); Gly, orange (C^{α}-C' correlations); Ile, yellow (C^{α}-$C^{\delta 1}$, C^{β}-$C^{\gamma 1}$, C^{β}-$C^{\gamma 2}$, and $C^{\gamma 1}$-$C^{\delta 1}$ correlations); Leu, grey (C^{β}-C^{γ} correlations); Pro, purple (C^{α}-C^{δ} correlations); Ser, light green (C^{α}-C^{β} correlations); Thr, blue (C^{α}-C^{β} and C^{β}-C^{γ} correlations); Val, bright green (C^{α}-C^{β} correlations). 66

Figure 4.2 Superposition of DARR spectra recorded on tubular (red) and conical (black) assemblies of U-^{13}C,^{15}N CA at 21.1 T. The expansion of the Ser and Thr region shows more clearly about the enhanced resolution. The spectra of the tubular assemblies were acquired as (6250 x 480) (complex x real) matrices using 128 scans and a recycle delay of 2 s, a spectral

width of 78 kHz in both dimensions and the DARR mixing time of 50 ms. The acquisition parameters for the spectra of the conical assemblies are given in the caption to Figure 4.1. 67

Figure 4.3 Pulse sequences for 2D and 3D experiments employed for resonance assignments of conical CA assemblies by MAS NMR spectroscopy: a) DARR [21] b) NCA/NCA; c) NCACX/NCOCX; d) NCACB. Empty vertical bars indicate π/2 and filled vertical bars indicate π pulses. 69

Figure 4.4 14.1 T DARR spectra of conical CA assemblies acquired on a set of uniformly and specifically labeled CA: (a) U-^{13}C,^{15}N; (b) U-^{13}C,^{15}N-Ala/U-^{13}C,^{15}N-Thr, and (c) U-^{13}C,^{15}N-His/U-^{13}C,^{15}N-Tyr. In c), expansions of the aliphatic (right) and carbonyl (left) regions illustrate the one-bond correlations for His and Tyr residues. The spectra in (a) were collected as (2000 x 400) (complex x real) matrices with 200 scans, a recycle delay of 2 s, and a spectral width of 50 kHz in both dimensions. The spectra in (b) were obtained as (2000 x 512) (complex x real) matrices with 256 scans, a recycle delay of 1 s, and spectral widths of 50 kHz and 25 kHz in the ω_2 and ω_1 dimensions, respectively. The spectra in (c) were acquired as (1000 x 128) (complex x real) matrices with 640 scans, and a recycle delay of 1.5 s and spectral widths of 50 kHz and 33.3 kHz in the w_2 and w_1 dimensions, respectively. The DARR mixing time was 10 ms for all data sets. 70

Figure 4.5 Selected 2D planes of the 3D NCACX spectrum of U-^{13}C,^{15}N CA conical assemblies acquired at 14.1 T. Spectral resolution in these planes is high enough to allow assignment of spin systems. The spectrum was acquired as (1000 x 40 x 32) (complex x real x real) matrices with 256 scans and a recycle delay of 2 s. The spectral widths were 50 kHz, 4.167 kHz, and 8.333 kHz in the ω_3, ω_2 and ω_1 dimensions, respectively. SPECIFIC-CP (t_{mix} = 6.2 ms) was employed for the first NCA transfer followed by the DARR mixing (t_{mix} = 20 ms) resulting in predominantly one- and two-bond correlations. 82

Figure 4.6a Sequential backbone connectivity for the sequence stretch A105-L111 in conical assemblies based on 3D NCOCX, NCACX, and NCACB experiments acquired at 14.1 T. The residue names are shown on top of the spectra at their ^{15}N chemical shift plane. Negative cross-peaks resulting from two-bond N-C$^{\beta}$ correlations in the NCACB spectra are displayed in red. The NCOCX spectrum was acquired as (1000 x 36 x 40) (complex x real x real) matrices with 280 scans and a recycle delay of 2 s. The spectral widths were 50 kHz, 4.167 kHz, and 6.250 kHz in the ω_3, ω_2, and ω_1 dimensions, respectively. SPECIFIC-CP (t_{mix} = 6.2 ms)

was employed for the first NCO transfer followed by the DARR mixing (t_{mix} = 15 ms) resulting in predominantly one- and two-bond correlations. The NCACB spectrum was acquired as (1000 x 32 x 32) (complex x real x real) matrices with 400 scans and a recycle delay of 2 s. The spectral widths were 50 kHz, 4.167 kHz, and 8.333 kHz in the ω_3, ω_2, and ω_1 dimensions, respectively. SPECIFIC-CP (t_{mix} = 6.2 ms) was employed for the first NCA transfer followed by the DREAM mixing (t_{mix} = 3 ms) to establish the C^{α}-C^{β} correlations. 83

Figure 4.6b Sequential backbone connectivity for the sequence stretch L43-A47 (b) in conical assemblies based on 3D NCOCX, NCACX, and NCACB experiments acquired at 14.1 T. The residue names are shown on top of the spectra at their ^{15}N chemical shift plane. Negative cross-peaks resulting from two-bond N-C^{β} correlations in the NCACB spectra are displayed in red. The NCOCX spectrum was acquired as (1000 x 36 x 40) (complex x real x real) matrices with 280 scans and a recycle delay of 2 s. The spectral widths were 50 kHz, 4.167 kHz, and 6.250 kHz in the ω_3, ω_2, and ω_1 dimensions, respectively. SPECIFIC-CP (t_{mix} = 6.2 ms) was employed for the first NCO transfer followed by the DARR mixing (t_{mix} = 15 ms) resulting in predominantly one- and two-bond correlations. The NCACB spectrum was acquired as (1000 x 32 x 32) (complex x real x real) matrices with 400 scans and a recycle delay of 2 s. The spectral widths were 50 kHz, 4.167 kHz, and 8.333 kHz in the ω_3, ω_2, and ω_1 dimensions, respectively. SPECIFIC-CP (t_{mix} = 6.2 ms) was employed for the first NCA transfer followed by the DREAM mixing (t_{mix} = 3 ms) to establish the C^{α}-C^{β} correlations. 84

Figure 4.7 Summary of the resonance assignments of CA conical assemblies. The assigned residues are shown in red; the tentatively assigned residues are shown in blue; and the unassigned residues are shown in black. 85

Figure 4.8 Comparison between solid-state and solution secondary chemical shifts for CA conical assemblies. C^{α} shifts that reside in a-helical regions of the protein exhibit positive secondary shifts, while those located in the b-strands and other secondary structure elements are negative. 86

Figure 4.9 Comparison between C^{α} and C' secondary chemical shifts for CA conical assemblies. C^{α} and C' shifts that reside in α-helical regions of the protein exhibit positive secondary shifts, while those located in the β-strands and other secondary structure elements are negative. 87

Figure 5.1 (A) ^{15}N CPMAS spectra of full-length CA conical assemblies (top), CTD (142-231) (middle), and CTD (137-231) (bottom). (B) ^{13}C CPMAS

spectra of full-length conical CA assemblies (top), CTD (142-231) (middle), and CTD (137-231) (bottom). Resonance assignments are shown and were confirmed in 2D ^{13}C-^{13}C and ^{15}N-^{13}C correlation spectra. The solid-state NMR chemical shifts agree well with solution NMR shifts. 98

Figure 5.2 ^{13}C-^{13}C $R2_1^1$ correlation spectra of CTD (142-231) (A) and full-length conical CA assemblies (B) for the aliphatic and carbonyl regions. 1D traces extracted along the ω_2 and ω_1 frequency dimensions are shown to illustrate that weak cross peaks corresponding to the different Y145 conformers are clearly above the noise level in the CTD (142-231) spectra. No Y145 resonances were detected for full-length CA assemblies. 100

Figure 5.3 ^{1}H-^{15}N dipolar lineshapes for tyrosine residues in (A) full-length CA conical assemblies, (B) CTD (142-231), and (C) CTD (137-231). For full-length assembled CA, the 2D ^{1}H-^{15}N DIPSHIFT spectrum is shown to illustrate the high quality of the data. Experimental dipolar lineshapes and the best-fit lineshapes are shown in solid lines and dashed lines, respectively. The experimental lineshapes for Y145 are highlighted in red. For each residue, the ^{1}H-^{15}N dipolar coupling constant, w_D, extracted by numerical simulations, is listed. Note that all tyrosine residues exhibit essentially rigid-limit behavior on the timescale of the DIPSHIFT experiment. 102

Figure 5.4 ^{15}N CSA line shapes for tyrosine residues in (A) full-length CA conical assemblies, (B) CTD (142-231), and (C) CTD (137-231). For full-length CA, the 2D ^{15}N ROCSA spectrum is shown to illustrate the high quality of the data. Experimental and best-fit CSA lineshapes are shown in solid and dashed lines, respectively. The experimental lineshapes for Y145 are highlighted in red. For each residue, the ^{15}N CSA tensor parameters, δ_s and η_s, extracted from numerical simulations, are listed. Note that all tyrosines residues exhibit essentially rigid-limit parameters on the timescale of the ROCSA experiment. 103

Figure 5.5 Temperature dependence of the solid-state NCA spectra of (A) full-length CA conical assemblies and (B) CTD (142-231). The spectra were acquired at 14.1 T, and the MAS frequency was 10 kHz. The temperature at which every spectrum was collected is indicated next to the corresponding data set. Note that Y145 cross peak is not present at any temperature. 105

Figure 5.6 (A) Two-dimensional NCA spectra of full-length conical CA assemblies (top), CTD (142-231) (middle), and CTD (137-231) (bottom) at 258 K. On the right, the corresponding 1D ^{15}N CPMAS spectra are displayed. Note that the Y145 peak is present in all ^{15}N CPMAS spectra, while the corresponding Y145 cross peaks are missing in the 2D NCA spectra, suggesting that the ^{15}N-^{13}C dipolar couplings are motionally averaged. (B) CA-N (bottom) and ^{15}N CPMAS (top) spectra of full-length CA conical assemblies, illustrating that the Y145 peak intensity is severely attenuated in the CA-N data, most likely due to motional averaging of the corresponding ^{15}N-^{13}C dipolar coupling. 106

Figure 5.7 ^{15}N-^{13}C REDOR spectra (top) and dephasing curves (bottom) of (A) full-length CA conical assemblies, (B) CTD (142-231), and (C) CTD (137-231). In the REDOR spectra, S_0 and S are the peak intensities in the absence and presence of REDOR dephasing pulses, respectively. DS = S_0-S is the intensity difference. In the REDOR dephasing curves, the REDOR fraction, DS/S_0, is plotted against the mixing time. Note that for all three proteins, Y145 does not exhibit REDOR dephasing caused by motional averaging of the ^{15}N-^{13}C dipolar coupling on the timescale of the REDOR experiment (~10 ms). 108

Figure 6.1 From left to right: TEM (top) and confocal (bottom) images of CA A92E, CA-SP1 A92E, CA WT and CA-SP1 WT tubular assemblies. For the CA-SP1 WT sample, conical assemblies were observed in addition to tubes by both TEM and confocal methods (not shown)..... 119

Figure 6.2 Primary sequences of the HXB2 and NL4-3 (A92E mutant) strains used in our investigations of CA and CA-SP1 assemblies of the tubular morphology [27]. The variations between the amino acid residues in HXB2 and NL4-3 strains are shown in red, and the final fourteen amino acids of the construct belong to the SP1 spacer peptide tail, shown in blue. 120

Figure 6.3 The 21.1 T 2D NCA spectra of U-^{13}C,^{15}N HIV-1 CA (left) and CA-SP1 (right). The 1D traces in the ω_2 dimension corresponding to the resolved cross-peaks at ^{15}N shift 112.4 ppm are shown and illustrate the remarkable spectral sensitivity and resolution. The acquisition parameters for the spectra of U-^{13}C,^{15}N HIV-1 CA and CA-SP1 tubular assemblies are given in Table B.1 in the Appendix. 121

Figure 6.4 Backbone walk for the stretch of residues R100-T110 of U-^{13}C,^{15}N tubular assemblies of HIV-1 CA-SP1 A92E (NL4-3 strain) acquired at 19.9 T and 4 °C. The 1D trace at 112.4 ppm along the ^{15}N dimension of

the NCACX spectrum is shown on top. The following 2D data sets were used to establish this backbone walk: NCACX (top), NCOCX (middle right and bottom), and DARR (middle left and middle center). The DARR mixing time was 50 ms. The acquisition parameters for the spectra of U-^{13}C,^{15}N HIV-1 CA-SP1 tubular assemblies are given in Table B.1 in the Appendix..... 123

Figure 6.5 Backbone walk of the stretch of residues R100-T108 (top) and S109-G116 (buttom) of tubular assemblies of U-^{13}C,^{15}N HIV-1 CA WT (HXB2 strain). The NCACX and NCOCX spectra are shown in red and black, respectively..... 125

Figure 6.6 The superposition of the a) NCA and b) C^{α}-C' region, c) aliphatic region of DARR spectra of tubular assemblies acquired at 19.9 T and 4 °C. CA A92E is in blue and CA-SP1 A92E is in black. The DARR mixing time was 50ms. The acquisition parameters for the spectra of U-^{13}C,^{15}N HIV-1 CA-SP1 tubular assemblies are given in Table B.1 in the Appendix. 136

Figure 6.7 The superposition comparison of the a) NCA and b) C^{α}-C' region, c) aliphatic region of DARR spectra acquired at 21.1 T and 4 °C. CA WT is in blue and CA-SP1 WT is in black. The DARR mixing time was 50ms. The acquisition parameters for the spectra of U-^{13}C,^{15}N HIV-1 CA-SP1 tubular assemblies are given in Table B.1 in the Appendix. 137

Figure 6.8 A scheme of NOESY-DARR pulse sequence. The carbon nuclei are first excited with a 90-degree pulse and the correlations were established through DARR irradiation during the mixing time [35]. 139

Figure 6.9 The superposition of the NOESY-DARR spectra of the tubular assemblies of CA A92E (blue) and CA-SP1 A92E (black) acquired at 19.9 T and 4 °C. a) The C^{α}-C' region of the spectra; b) the aliphatic region of the spectra; c) expansion of the aliphatic region of Thr-Ser residue types. The DARR mixing time was 50 ms. The acquisition parameters for the spectra are given in Table B.1 in the Appendix. 141

Figure 6.10 The superposition of the DARR (blue) and NOESY-DARR (black) spectra of CA-SP1 A92E acquired at 19.9 T and 4 °C. a) The C^{α}-C' region of the spectra; b) the aliphatic region of the spectra; c) expansion of the aliphatic region of Thr-Ser residue types. The DARR mixing

times were 50 ms. The acquisition parameters for the spectra are given in Table B.1 in the Appendix..... 141

Figure 6.11 The INADEQUATE pulse sequence [43]. The carbon nuclei are first excited with a 90-degree pulse and double quantum coherences are then established through a composite of π and π/2 pulses. 143

Figure 6.12 The superposition of the INADEQUATE and NOESY-DARR spectra of CA A92E (blue) and CA-SP1 A92E (black) acquired at 19.9 T and 4 °C. a) the aliphatic region; b) and c) the C^{β}-C^{γ} region of His residues; d) and e) the C^{α}-C' region; f) and g) the expansion of Thr, Ser regions; h) and i) the superposition of NOESY-DARR spectra that correspond well with those in the INADEQUATE spectra; j) the expansion in the Thr, Ser region of NOESY-DARR spectra. The delay τ in the INADEQUATE pulse sequence was optimized before each experiment was set up and the experiment time was kept the same for both samples. The acquisition parameters for the spectra are given in Table B.1 in the Appendix. 147

Figure 6.13 A scheme of CP-INADEQUATE pulse sequence. The carbon nuclei are first polarized by cross polarization, and the double quantum coherences are established through the INADEQUATE step..... 149

Figure 6.14 The superposition of the CP-INADEQUATE (blue) and INADEQUATE spectra (black) of CA A92E tubular assemblies acquired at 19.9 T and 4 °C. a) and b) The expansion of the region that extra peaks arise; c) the aliphatic region; d) and e) the C^{β}-C^{γ} region of His residues; f) and g) the C^{α}-C' region. The delay τ in the CP-INADEQUATE pulse sequence was optimized in each experiment, and the experiment time was kept the same for both samples. The acquisition parameters for the spectra are given in Table B.1 in the Appendix..... 151

Figure 6.15 The superposition of the CP-INADEQUATE spectra of CA A92E (blue) and CA-SP1 A92E (black) acquired at 19.9 T and 4 °C. a) The aliphatic region; b) and c) the C^{α}-C' region. The signal-to-noise ratios of the spectra are the same in both samples. The acquisition parameters are given in Table B.1 in the Appendix..... 152

Figure 6.16 The comparison of the INADEQUATE spectra of CA WT (blue) and CA-SP1 WT (black) acquired at 21.1 T and 4 °C. a) The aliphatic region; b) the C^{β}-C^{γ} region of His residues; c) the C^{α}-C' region. The delay in the INADEQUATE pulse sequence was optimized before each experiment was set up and the experiment time was kept the same for

both samples. The acquisition parameters for the spectra are given in Table B.1 in the Appendix..... 155

Figure 6.17 The superposition of the CP-INADEQUATE spectra of CA WT (blue) and CA-SP1 WT (black) acquired at 21.1 T and 4 °C. a) and b) The expansion of the region that extra peaks arise; c) the aliphatic region; d) and e) the C^{α}-C' region. The delay in the CP-INADEQUATE pulse sequence was optimized in each experiment, and the experiment time was the same for both samples. The acquisition parameters for the spectra are given in Table B.1 in the Appendix. 157

Figure 6.18 The superposition of the CP-INADEQUATE (blue) and INADEQUATE (black) spectra of CA WT tubular assemblies acquired at 21.1 T and 4 °C. a) and b) The expansion of the region that extra peaks arise; c) the aliphatic region; d) the C^{β}-C^{γ} region of His residues. The signal-to-noise ratios of the spectra are the same for both samples. The acquisition parameters for the spectra are given in Table B.1 in the Appendix. 158

Figure 6.19 The superposition of the NCA (A) and NCACX (B) spectra of CA NL4-3 A92E (blue) and CA HXB-2 (black). The green and red labels represent signals from the distinct residues in the NL4-3 A92E and HXB-2 variants, respectively. (C) A view of the 3D structure of CA (HXB2 variant, PDB file 3NTE). Colored magenta are the residues that vary between the two constructs; in orange shown are the residues that are identical in the two variants but whose chemical shifts are perturbed. 160

LIST OF ABBREVIATIONS

MAS NMR	Magic Angle Spinning Nuclear Magnetic Resonance
SSNMR	Solid-State Nuclear Magnetic Resonance
HIV-1	Human immunodeficiency virus type 1
AIDS	Acquired immunodeficiency syndrome
CA	Capsid
NTD	N-terminal domain
CTD	C-terminal domain
SP1	Spacer petide 1, p2
PEG	Poly ethylene glycol
DARR	Dipolar assisted rotational resonance
NOESY	Nuclear Overhauser effect spectroscopy
INADEQUATE	Incredible natural-abundance double-quantum transfer experiment
CP-INADEQUATE	Cross polarization incredible natural-abundance double-quantum transfer experiment
TPPI	Time proportional phase incrementation
CP	Cross polarization
SPECIFIC-CP	SPECtrally Induced Filtering In Combination with Cross Polarization
CSA	Chemical shift anisotropy
ROCSA	Recoupling Of Chemical Shift Anisotropy

ABSTRACT

There are around 34 million people in the world living with HIV-1, which is a causative agent of acquired immunodeficiency syndrome (AIDS). AIDS has become the ninth leading cause for death of people ages 25-34 in the US. Even though highly active antiretroviral therapy (HAART) showed effectiveness in suppressing the virus replication and significantly prolonged the patients lives, AIDS still remains an uncured disease. To develop new therapies, atomic-level understanding of the mechanism of HIV-1 lifecycle, including the structures of the various protein assemblies, is needed.

The Gag polyprotein and its component capsid (CA) protein are essential constituents of the HIV-1 life cycle, and have recently attracted attention as targets for drug development. However, the atomic resolution structure and the dynamics of Gag and CA protein assemblies and their complexes with small-molecule inhibitors are not available because these assemblies are not amenable for characterization by traditional structural biology methods, X-ray diffraction and solution NMR spectroscopy. Solid-state NMR spectroscopy has the unique capability of providing atomic-level structural and dynamics information in large protein assemblies. The focus of this dissertation is establishing solid-state NMR spectroscopy as an atomic-level probe of structure and dynamics in HIV-1 protein assemblies. This effort required first establishing sample

conditions for formation of HIV-1 protein assemblies that give rise to high-resolution solid-state NMR spectra for subsequent structural studies. In my Ph.D. work, I have optimized protocols to prepare homogeneous HIV-1 CA protein assemblies *in vitro*, developed confocal imaging method to characterize the morphologies of the resulting assemblies. With the suitable samples in hand, I have acquired solid-state NMR spectra on assemblies of CA protein and the maturation intermediate, CA-SP1, for structural analysis. Using these solid-state NMR data, I and my colleagues have obtained novel insights into the following aspects of HIV-1 structural biology: i) conformation of CA protein in conical assemblies; ii) the role of conformational dynamics of the hinge region in the structural polymorphism of CA in conical assemblies; iii) conformation of spacer peptide SP1 in tubular CA-SP1 assemblies; iv) conformation and dynamics of CA in tubular assemblies. In this thesis, I will first discuss the preparation and characterization of HIV-1 CA and CA-SP1 protein assemblies (Chapter 2 and 3). I will then describe the resonance assignments and secondary structure analysis of conical assemblies of HIV-1 CA protein (Chapter 4). Next, in Chapter 5, I will present the dynamics studies of conical assemblies of HIV-1 CA protein assemblies. In Chapter 6, I will discuss the resonance assignments and conformational analysis of tubular assemblies of HIV-1 CA and CA-SP1 proteins.

The long-term goal of this research is developing comprehensive understanding of the mechanism of capsid assembly and disassembly, HIV-1 maturation, through the structural and dynamics analysis of the CA and Gag assemblies. The work discussed here represents the first step toward this goal and lays

out the methodological and intellectual foundations enabling the solid-state NMR analysis of HIV-1 protein assemblies.

Chapter 1

INTRODUCTION

1.1 HIV-1 viral lifecycle

According to the United Nations report, there are around 34 million people in the world living with HIV-1, which causes acquired immunodeficiency syndrome (AIDS) [1]. And AIDS has become the ninth leading cause of people ages 25-34 in the US [2]. Although highly active antiretroviral therapy (HAART) showed effectiveness in suppressing the virus replication and significantly prolonged the patients' lives, AIDS remains uncured disease till today [3,4].

HIV-1 virus is an important virus, not only because it is the causative agent of AIDS, but also because the study of AIDS leads to the discoveries of cellular mechanisms and drugs that apply to many kinds of retroviruses and the diseases caused by them.

The life cycle of HIV virus, which consists of early and late stages, lasts approximately one day under normal conditions. At the beginning of the early stage, a mature virus particle in the blood, lymph or lymphatic tissues, attaches itself to a host cell through glycoproteins targeting receptors on the host cell plasma membrane. The host cells are either resting or activated CD4+ T lymphocytes and macrophages, which

are non-dividing immune cells that also express CD4. The non-dividing cells are important targets for viral replication during the initial stages of infection because the infection of these cell populations is crucial for subsequent virus spread to lymphoid organs and T-helper lymphocytes by contributing to the establishment of virus reservoirs [5]. Once attached, the HIV-1 virus fuses with the host cell. This process is gp41-mediated, pH independent, and triggered by the interaction between gp120 and its CD4 receptor and either the CXCR4 or CCR5 co-receptor [6].

At some point between the fusion and integration steps the conical capsid core (see Section 1.2 and Figure 1.1) disassembles (the disassembly process is commonly termed uncoating) to release viral proteins and RNA into the cytosol, where RNA is reverse transcribed into full-length double-strand DNA through the enzyme reverse transcriptase [7,8]. This process begins with the elongation of a cellular $tRNA_{Lys}$, which serves as a primer, and involves the degradation of the viral RNA template by the RT RNase H activity. The reverse transcription of the viral RNA takes place in the cytoplasm within a large nucleoprotein complex termed the reverse transcription complex (RTC) containing the two copies of viral RNA and the viral proteins: RT, IN, NCp7, Vpr and some matrix molecules [9-13]. It is generally believed that the reverse transcription process is initiated in virus particles and is then completed, after virus entry, in the cystosol of the target cell [14,15].

The structural end product of the reverse transcription process, called the "preintegration complex" (PIC) is transported to the cell nucleus where it is integrated into the host cell genome with the help from the enzyme integrase [16-18]. This

enzymatic integration reaction is well understood and can be recapitulated *in vitro* with purified integrase [19]. However, the processes between the PIC nuclear entry and the integration in the nucleus, especially in the context of higher order chromatin are still a matter of debate. Chromatin-remodeling factors might play a role during these steps [20,21], and might be responsible for the HIV-1 preferential integration in active transcriptional units in human cells [22]. One of the prominent differences between oncoretroviruses and the lentiviruses is that the lentivurses such as HIV and SIV are able to import their genome through the envelope of the nucleus 4-6 h after infection without nuclear envelope disintegration during host cell mitosis [23].

At this point, the late stages of the viral replication begin and extend until virion release and maturation. The stably integrated 10 kb-long DNA molecule termed HIV-1 provirus serves as a template for the transcription of RNA by the cellular Pol II polymerase [24], which is released to the cytosol and then translated into viral proteins. The viral transactivator protein Tat binds to the stem-loop trans-acting response element (TAR) and promotes the transcriptional processivity for the synthesis of full-length HIV-1 transcripts. The viral Rev protein binds both to the cellular export factor Crm1 and to the viral RNA RRE (Rev response element) to allow the unspliced and partially spliced mRNA to exit the nucleus, which is normally forbidden [25]. The RNA released in the cytoplasm drives the expression of the viral proteins such as Env, Gag and Gag-Pol.

Viral proteins are then transported to the plasma membrane through various means where they bind viral RNA and are released from the host by a budding

mechanism [26]. During this process, the NC domain of Gag binds with the two copies of the genomic RNA through the zinc fingers, which are also involved in the tight assembly of Gag [24]. Around 1500-2000 copies of Gag are radially arranged, and the lattice formed is curved into a spherical morphology. This membrane-coated spherical particle has been proposed to be still linked to the host cell by a stem-like structure [10]. The p6 domain of Gag mediates the final steps of the HIV-1 virion release, which is a component of the cellular vesicular sorting pathway (VPS) [27].

Sometime during or shortly after budding, the Gag polyprotein, which makes up the major structural proteins of HIV-1, is cleaved into its constituent domains in a process termed maturation, which can be viewed as a condensation of an electron-dense conical capsid. This maturation process results in an infectious HIV-1 virion containing the characteristic conical inner cores [28-30].

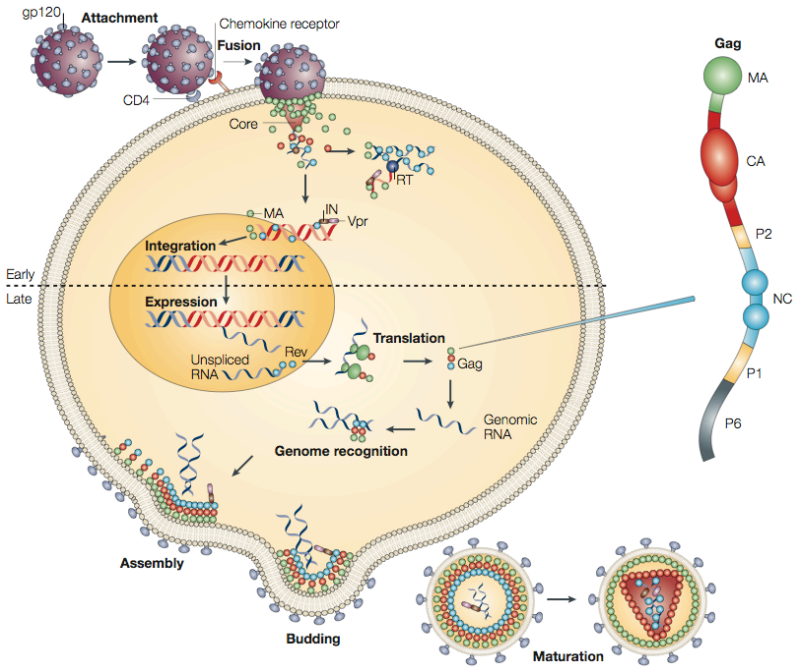


Figure 1.1 A general view of HIV-1 viral life cycle. In the early stage of the replication (upper portion), a mature virus particle attaches itself to a host cell through glycoproteins targeting receptors on the host cells plasma membrane. Once attached, the HIV-1 virus fuses with the host cell. At some point between the fusion and integration steps the conical capsid core uncoats to release viral proteins and RNA into the cytosol where RNA is transcribed into DNA through the enzyme reverse transcriptase. This DNA is transported to the cells nucleus where it is integrated into the host cells genome with help from the enzyme integrase. Then the late stage of the replication starts (lower portion). The cellular genome expresses viral RNA, which is released to the cytosol and then translated into viral proteins. These proteins are then transported to the plasma membrane through various means, where they bind viral RNA and are released from the host cell by a budding mechanism. Sometime during or shortly after budding, the Gag polyprotein, which makes up the major structural proteins of HIV-1 is cleaved into its constituent domains in a process termed maturation which can be viewed by a condensation of an electron dense conical capsid. This maturation process results in an infectious HIV-1 virion. This figure was originally published in reference [31]. Permission for reuse in this dissertation is granted by Nature Publishing Group., Nature Reviews Microbiology, copyright 2005.

1.2 HIV-1 capsid (CA): structural organization

HIV-1 CA protein is a 26.6 kDa, 231 amino acid protein [32,33]. Around 1500 copies of CA protein comprise the conical core of the mature HIV-1 virion, to protect the genetic material of the virus. During the maturation process, CA protein forms fullerene-type lattice, which is comprised of ca. 250 hexameric units as well as exactly 12 pentamers to close the cone, distinct from the unclosed Gag hexamer lattice [28,29,34-38]. The lattice spacing of mature CA is around 8 nm, whereas that for Gag assemblies is ca. 9.6 nm, according to the cryoEM studies [30,39-41]. In solution, CA protein forms predominantly oligomers with the monomer concentration being below the detectable limit [42,43]. At high salt concentration (at or higher than 1 M NaCl), HIV-1 CA protein assembles into tubes with the diameter of around 50 nm [44]. Spherical assemblies can also form and this process is pH dependent, according to Ehrlich et al [45]. The HIV-1 maturation intermediate, CA-NC protein, was also found to assemble into cones, spheres and tubes with the diameter of ca. 50 nm, in the presence of *E. coli* RNA [28,46].

The structure and function of CA protein have been extensively investigated using conventional methods, i.e. X-ray diffraction, solution NMR, cryoEM, as well as many other biochemical and cell-biology techniques. CA protein is composed of two distinct domains separated by a linker region: the N-terminal domain that contains seven α helices (NTD) and one amino-terminal β-hairpin, and the C-terminal domain (CTD) that contains four short α helices and one single-turn 3_{10}-helix [34,43,47-56] (Figure 1.2). The flexible hinge region linking the two domains has been proposed to

contribute to the polymorphism of the CA assembly [57]. In order for the CA protein to form the capsid lattice, three essential protein-protein interfaces between NTD and CTD are required: the NTD-NTD interface that forms inner rings of the hexamers; the CTD-CTD interface that connects two adjacent hexamers; and the NTD-CTD interface that is formed by insertion of Helix 4 into a groove of CTD [58]. CypA loop that spans residues V83-R100 is a functionally important region in the CA protein. It binds the cellular protein Cyclophilin A during the virus replication and the mutations in the loop result in defective virus particles [59-61]. Importantly, a mutagenesis study found that the stability of the capsid core has been optimized for proper viral maturation and disassembly [62]. This is critical for the viral infectivity because the CA protein needs to assemble into condensed and well-ordered structures, while upon entering the host cell, the same CA protein assemblies needs to perform opposing functions of uncoating and disassembly.

Recently it has been discovered that Rhesus Monkey is immune to AIDS, and the immunity is conferred by a restriction factor protein TRIM5α, which targets the CA uncoating process causing rapid and premature disassembly of CA [63,64]. Through this premature capsid uncoating mechanism, the HIV-1 virus replication is blocked, rendering the virus non-infectious [65-68]. In order to gain insight into the molecular mechanism of the TRIM5α and capsid interactions for further developments of anti-AIDS treatments, a myriad of investigations have been carried out. A cryoEM study showed that Trim5α binds to the CA assembly to damage the CA monomer structures thereby causing accelerated uncoating [69].

Despite these exciting findings and insights into the various aspects of CA organization, it has been challenging to explain how the different seemingly disparate functions are performed by the same protein. The CA assembly and uncoating processes of CA are poorly understood at atomic level due to the limitation of the conventional methods, drawing our attention and interest to this system, as will be elaborated upon in this dissertation.

1.3 Gag polyprotein intermediates

Gag polyprotein and its proteolytic cleavage products comprise the major structural elements of the virus. The polyprotein was named Gag because the gene that codes this protein is group-specific antigen (Gag). During the extracellular maturation process, Gag is proteolytically cleaved into several components: matrix protein (MA, p17), capsid protein (CA, p24), spacer peptide 1 (SP1, p2), nucleocapsid protein (NC, p7), spacer peptide 2 (SP2, p1), and p6 [70]. The Gag polyprotein cleavage process is sequentially regulated by the differential kinetics at the five cleavage sites (Figure 1.3 (a)). The first cleavage divides the Gag polyprotein into the MA-CA-SP1 (N-terminal) part and the NC-SP1-p6 (C-terminal) part. Then during the second step, MA protein is separated from the CA-SP1 part, and NC protein is separated from the SP2-p6 part. The last step of the cleavage is to liberate the SP1 peptide from the CA-SP1 part, and SP2 from the SP2-p6 part [70]. After the cleavage, the MA assemblies that are associated with viral membrane substitute the spherical Gag lattice. At this stage, the CA protein condenses into a conical capsid core to

protect the RNA/NC complex and other essential viral proteins. The immature and mature viral particles are observable by electron microscopy with distinct images of Gag and capsid core (Figure 1.3 (b)).

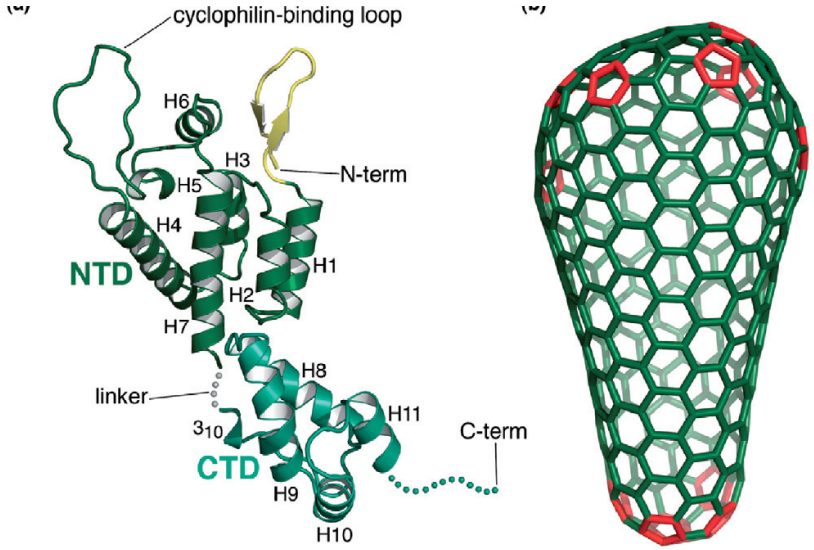


Figure 1.2 Structure of the mature HIV-1 capsid protein. Left: Tertiary structure of CA monomer, with the N-terminal domain (NTD) shown in dark green and C-terminal domain (CTD) colored in blue-green. The beginning of CA protein is the β-hairpin and is colored yellow. The final 11 residues are typically disordered in crystal structures an represented as dash line. Right: Fullerene model for the conical capsid. CA hexamers are colored green and pentameric declinations are colored red. This figure was originally published in reference [39]. Permission for reuse in this dissertation is granted by Elsevier Ltd., Current Opinion in Structural Biology, copyright 2008.

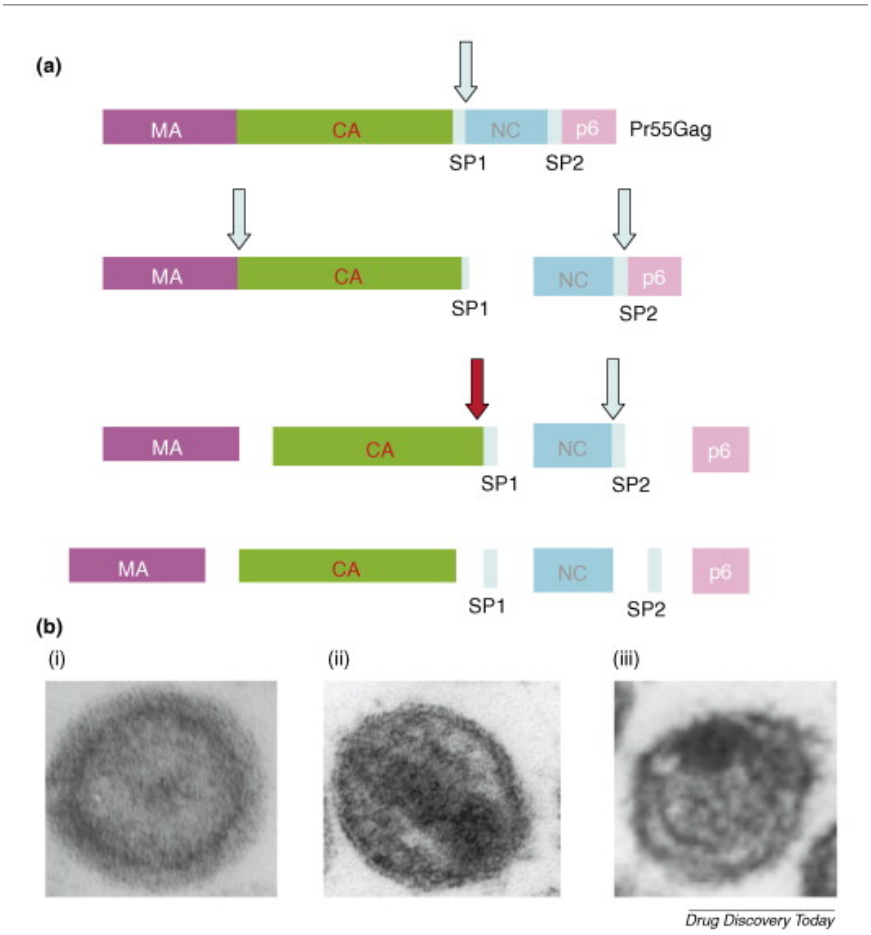


Figure 1.3 Gag cleavage cascade (a) and the TEM images of viral particles (b). (a) Gag processing cascade, illustrating the order in which the Gag precursor is cleaved by the viral protease. Each cleavage site is indicated by an arrow, the red arrow depicts the cleavage blocked by BVM. (b) Virion morphology visualized by transmission electron microscopy. Immature particles are shown in (i) and mature particles in (ii). An aberrant viron particle produced in the presence of BVM was shown in (iii). This figure was originally published in reference [71]. Permission for reuse in this dissertation is granted by Elsevier Ltd., Drug Discovery Today, Copyright 2008.

1.4 Maturation inhibitors

For the treatment of AIDS, HAART (highly active antiretroviral therapy) regimens have been developed and clinically proved efficient. The drugs used in HAART contain over 20 retroviral drugs, which inhibit the various viral enzymes: reverse transcriptase (RT), protease (PR), and integrase (IN) [72-75]. With the current therapies, it is not possible to completely cure AIDS due to the drug resistance, and scientists are constantly working on identification of new drug targets and on elucidation of their mechanisms [76].

The newly discovered drug Bevirimat (BVM) (3-O-(3, 3′-dimethylsuccinyl) betulinic acid), which was originally isolated from the leaves of the Chinese herb, *Syzygium claviflorum* (Figure 1.4). It has demonstrated potent *in vitro* and *in vivo* activities, with an *in vitro* 90% inhibitory concentration of 22.1 ng/ml (37.8 nM), and is shown to retain activity against viruses that are resistant to other classes of antiretrovirals (Figure 1.3 (a)).

Bevirimat targets a late step in the Gag processing cascade, namely, the release of the capsid protein (CA/p24) from the capsid precursor (CA-SP1/p25). This disruption to the Gag polyprotein processing results in defective core condensation and the release of noninfectious virus particles, blocking the spread of the infection to new cells (Figure 1.3 (b)). Therefore, BVM was termed maturation inhibitor, which is a novel class among the current drugs. The molecular mechanism of BVM interaction with CA-SP1 is not well understood due to the lack of atomic-level structural

information on the assembled CA-SP1, and in particular because the conformation of SP1 is unknown.

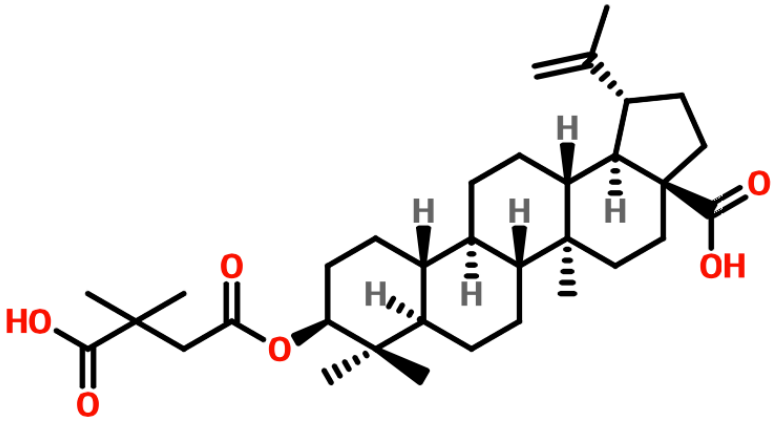


Figure 1.4 The chemical structure of Bevirimat, an inhibitor of HIV-1 Gag processing [77].

References

[1] Unaids. UNAIDS report on the global AIDS epidemic, 2012.

[2] D. L. Hoyert, Xu, Jiaquan. National vital statistics reports, 2012.

[3] R. H. Haubrich; S. J. Little; J. S. Currier; D. N. Forthal; C. A. Kemper; G. N. Beall; D. Johnson; M. P. Dube; J. Y. Hwang; J. A. Mccutchan; C. C. T. Grp. The value of patient-reported adherence to antiretroviral therapy in predicting virologic and immunologic response. *Aids,* 13(9): 1099-1107, 1999.

[4] D. L. Paterson; S. Swindells; J. Mohr; M. Brester; E. N. Vergis; C. Squier; M. M. Wagener; N. Singh. Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. *Annals of Internal Medicine,* 133(1): 21-30, 2000.

[5] O. J. Cohen; A. S. Fauci. Current strategies in the treatment of HIV infection. *Adv. Intern. Med.,* 46207-246, 2001.

[6] A. C. Saphire; M. D. Bobardt; P. A. Gallay. Human immunodeficiency virus type 1 hijacks host cyclophilin A for its attachment to target cells. *Immunol. Res.,* 21(2-3): 211-217, 2000.

[7] W. A. Banks; E. O. Freed; K. M. Wolf; S. M. Robinson; M. Franko; V. B. Kumar. Transport of human immunodeficiency virus type 1 pseudoviruses across the blood-brain barrier: role of envelope proteins and adsorptive endocytosis. *J. Virol.,* 75(10): 4681-4691, 2001.

[8] E. Le Rouzic; S. Benichou. The Vpr protein from HIV-1: distinct roles along the viral life cycle. *Retrovirology,* 2(1): 11-25, 2005.

[9] C. M. Farnet; W. A. Haseltine. Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *J. Virol.,* 65(4): 1910-1915, 1991.

[10] A. Fassati; S. P. Goff. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. *J. Virol.,* 75(8): 3626-3635, 2001.

[11] M. D. Miller; C. M. Farnet; F. D. Bushman. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.,* 71(7): 5382-5390, 1997.

[12] M. V. Nermut; A. Fassati. Structural analyses of purified human immunodeficiency virus type 1 intracellular reverse transcription complexes. *J. Virol.,* 77(15): 8196-8206, 2003.

[13] M. I. Bukrinsky; S. Haggerty; M. P. Dempsey; N. Sharova; A. Adzhubel; L. Spitz; P. Lewis; D. Goldfarb; M. Emerman; M. Stevenson. A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature,* 365(6447): 666-669, 1993.

[14] S. P. Goff. Intracellular trafficking of retroviral genomes during the early phase of infection: viral exploitation of cellular pathways. *J. Gene. Med.,* 3(6): 517-528, 2001.

[15] H. Zhang; G. Dornadula; J. Orenstein; R. J. Pomerantz. Morphologic changes in human immunodeficiency virus type 1 virions secondary to intravirion reverse transcription: evidence indicating that reverse transcription may not take place within the intact viral core. *J. Hum. Virol.,* 3(3): 165-172, 2000.

[16] R. A. Fouchier; M. H. Malim. Nuclear import of human immunodeficiency virus type-1 preintegration complexes. *Adv. Virus Res.,* 52275-299, 1999.

[17] B. R. Cullen. Journey to the center of the cell. *Cell,* 105(6): 697-700, 2001.

[18] M. Bukrinsky; A. Adzhubei. Viral protein R of HIV-1. *Rev. Med. Virol.,* 9(1): 39-49, 1999.

[19] D. Esposito; R. Craigie. HIV integrase structure and function. *Adv. Virus. Res.,* 52319-333, 1999.

[20] C. M. Farnet; F. D. Bushman. HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes *in vitro*. *Cell,* 88(4): 483-492, 1997.

[21] G. V. Kalpana; S. Marmon; W. Wang; G. R. Crabtree; S. P. Goff. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science,* 266(5193): 2002-2006, 1994.

[22] A. R. Schroder; P. Shinn; H. Chen; C. Berry; J. R. Ecker; F. Bushman. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell,* 110(4): 521-529, 2002.

[23] U. F. Greber; A. Fassati. Nuclear import of viral DNA genomes. *Traffic,* 4(3): 136-143, 2003.

[24] E. O. Freed. HIV-1 replication. *Somat Cell Mol. Genet.,* 26(1-6): 13-33, 2001.

[25] W. C. Greene; B. M. Peterlin. Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. *Nat. Med.,* 8(7): 673-680, 2002.

[26] A. G. Bukrinskaya. HIV-1 assembly and maturation. *Arch. Virol.,* 149(6): 1067-1082, 2004.

[27] J. E. Garrus; U. K. Von Schwedler; O. W. Pornillos; S. G. Morham; K. H. Zavitz; H. E. Wang; D. A. Wettstein; K. M. Stray; M. Cote; R. L. Rich; D. G. Myszka; W. I. Sundquist. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell,* 107(1): 55-65, 2001.

[28] B. K. Ganser; S. Li; V. Y. Klishko; J. T. Finch; W. I. Sundquist. Assembly and analysis of conical models for the HIV-1 core. *Science,* 283(5398): 80-83, 1999.

[29] S. Li; C. P. Hill; W. I. Sundquist; J. T. Finch. Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature,* 407(6802): 409-413, 2000.

[30] J. A. Briggs; T. Wilk; R. Welker; H. G. Krausslich; S. D. Fuller. Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J.,* 22(7): 1707-1715, 2003.

[31] V. D'souza; M. F. Summers. How retroviruses select their genomes. *Nat. Rev. Microbiol.,* 3(8): 643-655, 2005.

[32] H. R. Gelderblom; E. H. S. Hausmann; M. Ozel; G. Pauli; M. A. Koch. Fine-structure of human-immunodeficiency-virus (HIV) and immunolocalization of structural proteins. *Virology,* 156(1): 171-176, 1987.

[33] M. V. Nermut; D. J. Hockley; J. B. M. Jowett; I. M. Jones; M. Garreau; D. Thomas. Fullerene-like organization of HIV Gag-protein shell in virus-like particles produced by recombinant baculovirus. *Virology,* 198(1): 288-296, 1994.

[34] Z. Jin; L. Jin; D. L. Peterson; C. L. Lawson. Model for lentivirus capsid core assembly based on crystal dimers of EIAV p26. *J. Mol. Biol.,* 286(1): 83-93, 1999.

[35] J. B. Heymann; C. Butan; D. C. Winkler; R. C. Craven; A. C. Steven. Irregular and semi-regular polyhedral models for Rous sarcoma virus cores. *Comput. Math. Methods. Med.,* 9(3-4): 197-210, 2008.

[36] I. J. Byeon; X. Meng; J. Jung; G. Zhao; R. Yang; J. Ahn; J. Shi; J. Concel; C. Aiken; P. Zhang; A. M. Gronenborn. Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell,* 139(4): 780-790, 2009.

[37] O. Pornillos; B. K. Ganser-Pornillos; B. N. Kelly; Y. Hua; F. G. Whitby; C. D. Stout; W. I. Sundquist; C. P. Hill; M. Yeager. X-ray structures of the hexameric building block of the HIV capsid. *Cell,* 137(7): 1282-1292, 2009.

[38] O. Pornillos; B. K. Ganser-Pornillos; M. Yeager. Atomic-level modelling of the HIV capsid. *Nature,* 469(7330): 424-427, 2011.

[39] B. K. Ganser-Pornillos; M. Yeager; W. I. Sundquist. The structural biology of HIV assembly. *Curr. Opin. Struct. Biol.,* 18(2): 203-217, 2008.

[40] L. A. Carlson; A. De Marco; H. Oberwinkler; A. Habermann; J. A. Briggs; H. G. Krausslich; K. Grunewald. Cryo electron tomography of native HIV-1 budding sites. *PLoS Pathog.,* 6(11): e1001173, 2010.

[41] J. A. Briggs; H. G. Krausslich. The molecular architecture of HIV. *J. Mol. Biol.,* 410(4): 491-500, 2011.

[42] B. Chen; R. Tycko. Simulated self-assembly of the HIV-1 capsid: protein shape and native contacts are sufficient for two-dimensional lattice formation. *Biophys. J.,* 100(12): 3035-3044, 2011.

[43] T. R. Gamble; F. F. Vajdos; S. Yoo; D. K. Worthylake; M. Houseweart; W. I. Sundquist; C. P. Hill. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell,* 87(7): 1285-1294, 1996.

[44] I. Gross; H. Hohenberg; H. G. Krausslich. *In vitro* assembly properties of purified bacterially expressed capsid proteins of human immunodeficiency virus. *Eur. J. Biochem.,* 249(2): 592-600, 1997.

[45] L. S. Ehrlich; T. Liu; S. Scarlata; B. Chu; C. A. Carter. HIV-1 capsid protein forms spherical (immature-like) and tubular (mature-like) particles *in vitro*: structure switching by pH-induced conformational changes. *Biophys. J.,* 81(1): 586-594, 2001.

[46] S. Campbell; V. M. Vogt. Self-assembly *in vitro* of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. *J. Virol.,* 69(10): 6487-6497, 1995.

[47] C. Berthet-Colominas; S. Monaco; A. Novelli; G. Sibai; F. Mallet; S. Cusack. 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(5): 1124-1136, 1999.

[48] R. Campos-Olivas; J. L. Newman; M. F. Summers. Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses. *J. Mol. Biol.,* 296(2): 633-649, 2000.

[49] R. K. Gitti; B. M. Lee; J. Walker; M. F. Summers; S. Yoo; W. I. Sundquist. Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science,* 273(5272): 231-235, 1996.

[50] S. Khorasanizadeh; R. Campos-Olivas; M. F. Summers. Solution structure of the capsid protein from the human T-cell leukemia virus type-I. *J. Mol. Biol.,* 291(2): 491-505, 1999.

[51] S. Kadam; G. S. Mcalpine; M. L. Phelan; R. E. Kingston; K. A. Jones; B. M. Emerson. Functional selectivity of recombinant mammalian SWI/SNF subunits. *Genes Dev.,* 14(19): 2441-2451, 2000.

[52] C. Momany; L. C. Kovari; A. J. Prongay; W. Keller; R. K. Gitti; B. M. Lee; A. E. Gorbalenya; L. Tong; J. Mcclure; L. S. Ehrlich; M. F. Summers; C. Carter; M. G. Rossmann. Crystal structure of dimeric HIV-1 capsid protein. *Nat. Struct. Biol.,* 3(9): 763-770, 1996.

[53] G. B. Mortuza; L. F. Haire; A. Stevens; S. J. Smerdon; J. P. Stoye; I. A. Taylor. High-resolution structure of a retroviral capsid hexameric amino-terminal domain. *Nature,* 431(7007): 481-485, 2004.

[54] C. Tang; Y. Ndassa; M. F. Summers. Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. *Nat. Struct. Biol.,* 9(7): 537-543, 2002.

[55] D. K. Worthylake; H. Wang; S. Yoo; W. I. Sundquist; C. P. Hill. Structures of the HIV-1 capsid protein dimerization domain at 2.6 A resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.,* 55(Pt 1): 85-92, 1999.

[56] T. R. Gamble; S. Yoo; F. F. Vajdos; U. K. Von Schwedler; D. K. Worthylake; H. Wang; J. P. Mccutcheon; W. I. Sundquist; C. P. Hill. Structure of the

carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science,* 278(5339): 849-853, 1997.

[57] I. J. Byeon; G. Hou; Y. Han; C. L. Suiter; J. Ahn; J. Jung; C. H. Byeon; A. M. Gronenborn; T. Polenova. 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(14): 6455-6466, 2012.

[58] B. K. Ganser-Pornillos; A. Cheng; M. Yeager. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell,* 131(1): 70-79, 2007.

[59] A. C. S. Saphire; M. D. Bobardt; P. A. Gallay. Cyclophilin a plays distinct roles in human immunodeficiency virus type 1 entry and postentry events, as revealed by spinoculation. *J. Virol.,* 76(9): 4671-4677, 2002.

[60] D. Braaten; E. K. Franke; J. Luban. Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription. *J. Virol.,* 70(6): 3551-3560, 1996.

[61] L. Dietrich; L. S. Ehrlich; T. J. Lagrassa; D. Ebbets-Reed; C. Carter. Structural consequences of cyclophilin a binding on maturational refolding in human immunodeficiency virus type 1 capsid protein. *J. Virol.,* 75(10): 4721-4733, 2001.

[62] B. M. Forshey; U. Von Schwedler; W. I. Sundquist; C. Aiken. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.,* 76(11): 5667-5677, 2002.

[63] M. Stremlau; C. M. Owens; M. J. Perron; M. Kiessling; P. Autissier; J. Sodroski. The cytoplasmic body component TRIM5 alpha restricts HIV-1 infection in Old World monkeys. *Nature,* 427(6977): 848-853, 2004.

[64] D. M. Sayah; E. Sokolskaja; L. Berthoux; J. Luban. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature,* 430(6999): 569-573, 2004.

[65] M. Stremlau; M. Perron; M. Lee; Y. Li; B. Song; H. Javanbakht; F. Diaz-Griffero; D. J. Anderson; W. I. Sundquist; J. Sodroski. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. *Proc. Natl. Acad. Sci. U. S. A.,* 103(14): 5514-5519, 2006.

[66] M. J. Perron; M. Stremlau; M. Lee; H. Javanbakht; B. Song; J. Sodroski. The human TRIM5alpha restriction factor mediates accelerated uncoating of the N-tropic murine leukemia virus capsid. *J. Virol.,* 81(5): 2138-2148, 2007.

[67] J. Shi; C. Aiken. Saturation of TRIM5 alpha-mediated restriction of HIV-1 infection depends on the stability of the incoming viral capsid. *Virology,* 350(2): 493-500, 2006.

[68] S. Sebastian; J. Luban. TRIM5alpha selectively binds a restriction-sensitive retroviral capsid. *Retrovirology,* 240, 2005.

[69] G. Zhao; D. Ke; T. Vu; J. Ahn; V. B. Shah; R. Yang; C. Aiken; L. M. Charlton; A. M. Gronenborn; P. Zhang. Rhesus TRIM5alpha disrupts the HIV-1 capsid at the inter-hexamer interfaces. *PLoS Pathog.,* 7(3): e1002009, 2011.

[70] C. S. Adamson; K. Salzwedel; E. O. Freed. Virus maturation as a new HIV-1 therapeutic target. *Expert Opin. Ther. Targets,* 13(8): 895-908, 2009.

[71] C. S. Adamson; E. O. Freed. Recent progress in antiretrovirals--lessons from resistance. *Drug Discov. Today,* 13(9-10): 424-432, 2008.

[72] T. Cihlar; A. S. Ray. Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Res.,* 85(1): 39-58, 2010.

[73] A. M. Wensing; N. M. Van Maarseveen; M. Nijhuis. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res.,* 85(1): 59-74, 2010.

[74] M. P. De Bethune. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989-2009). *Antiviral Res.,* 85(1): 75-90, 2010.

[75] D. J. Mccoll; X. Chen. Strand transfer inhibitors of HIV-1 integrase: bringing IN a new era of antiretroviral therapy. *Antiviral Res.,* 85(1): 101-118, 2010.

[76] C. S. Adamson; E. O. Freed. Novel approaches to inhibiting HIV-1 replication. *Antiviral Res.,* 85(1): 119-141, 2010.

[77] E. De Clercq. The design of drugs for HIV and HCV. *Nat. Rev. Drug Discov.,* 6(12): 1001-1018, 2007.

Chapter 2

SOLID-STATE NMR SPECTROSCOPY FOR STRUCTURAL AND DYNAMICS CHARACTERIZATION OF PROTEIN ASSEMBLIES

2.1 Introduction

Protein functions are closely related with protein structures and dynamics. Solid-state NMR is able to probe protein structures at atomic level and dynamics from picoseconds to seconds and longer. This chapter will mainly discuss the solid-state NMR principles and methods that were applied or developed during the investigation of structure and dynamics of HIV-1 CA and CA-SP1 assemblies. In principle, solid-state NMR techniques detect various interactions between nuclei with different radio frequency waves and generate corresponding spectra. From the spectra acquired, structural and dynamics information can be extracted and analyzed.

2.2 Essential spin interactions

In the static magnetic fields (B_0), the nuclear magnetic moments are aligned preferentially with the applied field and net magnetization arises from the nuclei in a sample. This is the Zeeman interaction, which is represented by the following Hamiltonian

 , (2.1)

where $\hat{\mu}$ is the nuclear magnetic moment operator and B_0 is the external static magnetic field applied. For the nucleus of spin $\hat{I}$,

$$\hat{\mu} = \gamma \hbar z \hat{I}, \tag{2.2}$$

where γ is the magnetogyric ratio and $\hbar$ is the reduced Planck constant. Therefore, the Hamiltonian for the Zeeman interaction becomes

$$\hat{H}_Z = -\gamma \hbar \hat{I}_Z B_0. \tag{2.3}$$

When a radio frequency (rf) wave, which is perpendicular to the B_0 field, is applied to the sample, it interacts with the nuclear spin, and the precession of the magnetization vector will be significantly affected. The rf interaction Hamiltonian is noted as $\hat{H}_{RF}$. The operator to describe the interaction of the applied B_1 field and the nucleus of spin $\hat{I}$ is,

$$\hat{H}_{RF} = -\hat{\mu} \cdot B_1 = -\gamma \hat{I} \cdot B_1. \tag{2.4}$$

In addition to the external applied magnetic fields, there are internal sources of magnetic field that may affect the nuclear spin systems. For the nucleus of spin $\hat{I}$, the most important interactions include chemical shielding $\hat{H}_{CS}(t)$, dipole coupling $\hat{H}_D(t)$, J-coupling $\hat{H}_J$ and quadrupole coupling $\hat{H}_Q$. The total Hamiltonian acting on a nuclear spin in the presence of the above interaction can be represented as:

$$\hat{H}(t) = \hat{H}_Z + \hat{H}_{RF} + \hat{H}_D(t) + \hat{H}_{CS}(t) + \hat{H}_J + \hat{H}_Q, \tag{2.5}$$

In solid-state NMR, the through-space dipolar coupling has a profound effect on the NMR spectra. The Hamiltonian representing the dipolar interaction is:

$$\hat{H}_D = -2\hat{I} \cdot D \cdot \hat{S}, \tag{2.6}$$

where $\hat{I}$ and $\hat{S}$ are the two spin operators of the spin pair involved in the dipolar interactions, D is the dipolar-coupling tensor which describes the strength and the orientation of the dipolar interaction that spin S imposes on spin I. The value of D relates to several factors: the gyromagnetic ratios of the interacting nuclei, the cube of the internuclear distance r, and the orientational term $(3\cos^2\theta - 1)$,

$$D \propto \frac{\gamma_I \gamma_S (3\cos^2\theta - 1)}{r^3}. \tag{2.7}$$

Therefore, for nuclei possessing high gyromagnetic ratio such as protons, the dipolar coupling will be significant. And because of the r^3 dependence, dipolar coupling is difficult to observe at large internuclear distances. For example, for ^{13}C, dipolar couplings corresponding to distances over 7 Å are too small to be measured.

To obtain high-resolution solid-state NMR spectra, orientational dependence of the nuclear frequencies has to be eliminated, and in one approach (the so-called magic angle spinning) samples are typically spun at the angle of 54.74°, so that the orientational dependence of the nuclear spin interactions is averaged to zero over the full rotation cycle.

The magnetic shielding term $\hat{H}_{CS}(t)$ is another important local interaction. In the external field B_0, the electrons that are close to the nucleus produce induced currents, which create distinct local environment at the nucleus, which is the source of

the chemical shift. The chemical shift is the key NMR observable. The interaction Hamiltonian is expressed as

$$\hat{H}_{CS} = \gamma \hat{I} \bullet \sigma \bullet B_0, \tag{2.8}$$

where σ is the shielding tensor that describes the extent and direction of the shielding effect. The symmetric part of the shielding tensor can be broken down into isotropic and anisotropic terms as follows:

$$\sigma = \sigma_{iso} + \sigma_{aniso}, \tag{2.9}$$

and the anisotropic portion has the following orientational dependence:

$$\sigma_{aniso} \propto 3\cos^2\theta - 1, \tag{2.10}$$

θ is the angle between the external field and the orientation of the ellipsoid axes.

The chemical shift is defined as the difference in resonance frequencies between the nucleus of interest and a reference nucleus:

$$\delta = 10^6 \frac{\sigma_{ref} - \sigma}{1 - \sigma_{ref}} \approx 10^6 (\sigma_{ref} - \sigma), \tag{2.11}$$

In solution NMR and magic angle spinning (MAS) solid-state NMR experiments, because the dipolar interaction and the σ_{aniso} terms are averaged out by the motion,

$$\delta = \delta_{iso}, \tag{2.12}$$

The J-coupling $\hat{H}_J$ is the scalar interaction between two chemically bonded nuclear spins [1], and it therefore provides information on the connectivity of

molecules. The Hamiltonian of the J-coupling interaction between spins I and S (the corresponding atoms typically belong to the same molecule) is:

$$\hat{H}_J = 2\pi\hat{I} \cdot J_{IS} \cdot \hat{S}, \tag{2.13}$$

where J_{IS} is the J-coupling tensor. In solid-state NMR, the magnitude of this interaction is small compared to other terms and J-coupling usually results in additional line broadening. With the advent of high-field instruments and fast-MAS probes, the spectral resolution in solids is becoming higher and higher, and techniques have been developed to detect and resolve the J-coupling signals in isotopically labeled protein samples [2-4].

The quadrupole coupling $\hat{H}_Q$ is another internal spin interaction. It exists in the nuclei that are not spin $\frac{1}{2}$ and will not be further discussed in this work.

2.3 General protocol for structural characterization of proteins by SSNMR

SSNMR is an emerging method in the field of protein structure characterization, especially well suited for studies of insoluble, non-crystalline or microcrystalline samples [5-7]. Below, a brief overview is presented of the main steps for protein structure determination using MAS NMR approaches.

2.3.1 Site-specific resonance assignments: experiments and strategies

The starting step of structure elucidation with SSNMR methods is the site-specific resonance assignments that serve to attribute the observed NMR signals to the

specific atoms belonging to a given residue in the protein [8-10]. The chemical shifts themselves are a rich source of information, even in the absence of a 3D structure. From the assigned chemical shift, information such as the protein secondary structure, local electrostatic environment (e. g., hydrogen bonding interactions), can be extracted. Furthermore, with the resonance assignments in hand, one can carry out residue-specific dynamics studies [11,12]. Therefore, even though resonance assignments are always time-consuming and have to be accomplished manually at the present time, they are the first and fundamental step that cannot be bypassed.

To perform resonance assignments, a series of multi-dimensional SSNMR spectra first need to be acquired, such as NCACX, NCOCX, NCACB, DARR and NCA. Even though these are the most commonly used spectra, the methods in SSNMR never stopped evolving, and many additional experiments are currently available or under development by us and by others that facilitate this procedure [13,14]. Some of the spectra, such as NCACX, correlate the resonances of the amide nitrogen, alpha carbon, beta carbon and other carbon atoms within one amino acid in the protein sequence. Other spectra, such as NCOCX, correlate the amide nitrogen and the carbonyl carbon of the previous residue, followed by correlations to other carbon atoms that are close to the carbonyl carbon in space. All of the experiments mentioned above are based on the dipolar interactions between the nuclei, and because dipolar interaction is strongly dependent on distances, the experimental conditions can be easily tuned to give rise to exclusively one-bond correlations (Figure 2.1). Some of the highest-quality 2D spectra yielded nearly full assignment in several proteins [15-

19], including three examples from our laboratory on intact and reassembled thioredoxin [19-21]. For large protein systems, the cross peaks in SSNMR spectra generally highly overlap precluding complete resonance assignments from 2D spectra alone [22,23]. Multi-dimensional spectra resolve the cross peaks in more dimensions, and usually 3D spectra are required for characterization of large proteins [24].

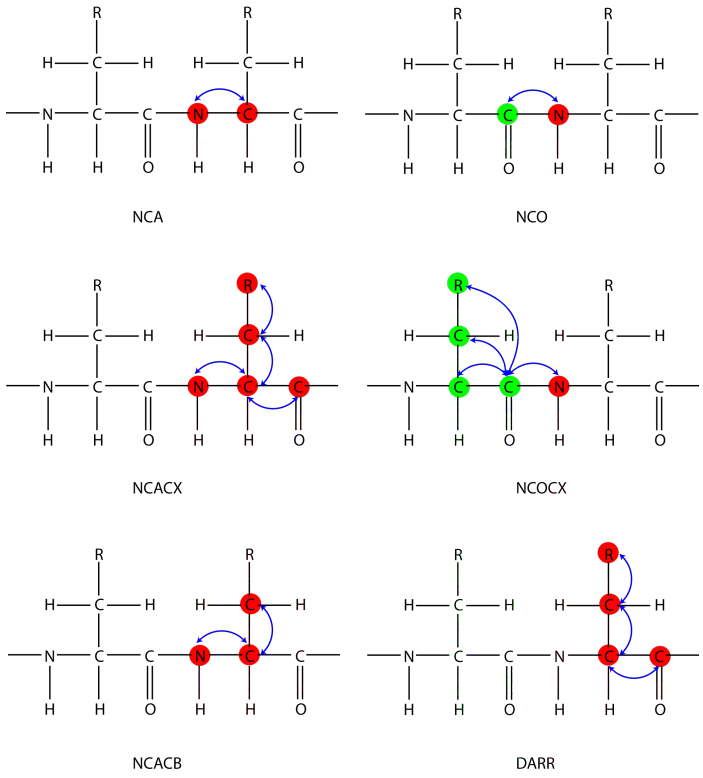


Figure 2.1 The magnetization transfer pathways of the dipolar-based SSNMR experiments used for resonance assignments: NCA, NCO, NCACX, NCOCX, NCACB, and DARR. The colored atoms are involved in the dipolar-based magnetization transfer.

After a series of spectra for resonance assignments are acquired and processed, the data are subjected to further analysis. For 2D spectra, different types of amino

acids give rise to characteristic sets of carbon and nitrogen chemical shifts. Biological Magnetic Resonance Bank (BMRB) contains a compilation of chemical shifts for a large number of assigned proteins, and the analysis of these shifts may facilitate the initial step of resonance assignments of an unknown protein [25]. For example, glycine typically has C^{α} shift at around 45 ppm, and threonine has C^{α} and C^{β} shifts around 70 ppm. These shifts have pronounced dependence on the secondary structure. On the basis of the connectivities between the cross peaks and their chemical shifts, groups of cross peaks are attributed to certain amino acid types. Usually, one can obtain an amino acid count on the basis of several 2D data sets [26].

As mentioned above, 3D experiments are required for assigning larger and more challenging proteins. On the basis of 3D experiments, several protein systems have been completely (or almost completely) assigned by us and by others, including DLC8 [27], CAP-Gly [28], β1 immunoglobulin binding domain of protein G (GB1) [29] and HET-s prion protein [30]. A common way to visualize 3D spectra is strips. With the same nitrogen chemical shift, the signals are resolved in two carbon directions, resulting in a small number of peaks in one strip plot. For analysis, strips with the same nitrogen (or carbon) shifts are aligned. The amino acid type can be readily inferred from the chemical shifts of the alpha and beta carbon atoms. Some amino acid types, such as Ala, Thr and Gly are easy to identify. Once amino acid types have been identified, sequence-specific assignments are made. In Figure 2.2, the approach for resonance assignments from the 3D data is illustrated.

2.3.2 Secondary structure analysis

CSI (Chemical Shift Index) was originally developed for $^{1}H^{\alpha}$ chemical shifts [31]. In 1994, Sykes et al. extended CSI to include $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and carbonyl ^{13}C chemical shifts, which enables the secondary structure analysis on the basis of the solid-state NMR chemical shifts [32,33]. Although CSI is an empirical method to determine the protein secondary structure, the accuracy of prediction can reach as high as 92% with consensus results of C^{α}, C^{β} and C'. In 2001, CSI was improved by introducing sequence-dependent correction factors [34]. The random coil and unfolded protein secondary structures are further improved with the amendment.

2.3.3 Experiments for gaining distance restraints

NMR protein structures are calculated on the basis of numerous internuclear distance restraints that define the overall fold and the detailed architectures. The quality of the structure depends critically on the availability of a sufficient number of long-range distance restraints defining spatial proximities between residues, which are distant in the primary sequence. There is a wide range of homonuclear and heteronuclear experiments in MAS NMR spectroscopy that are utilized to gain distance information. For uniformly labeled protein samples, the conventional spin diffusion experiments (such as DARR [35] and PDSD [36]), and the newly developed

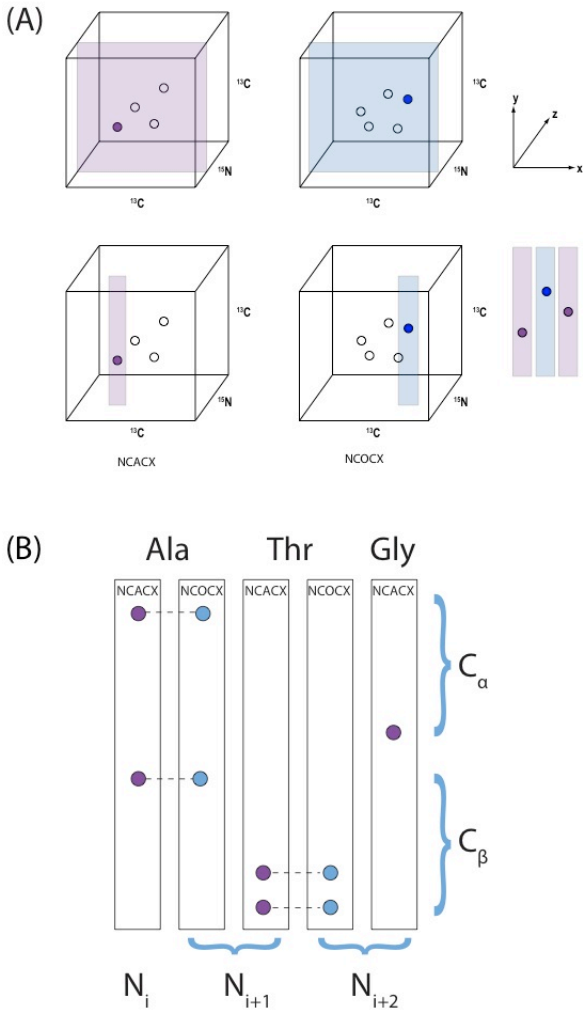


Figure 2.2 Resonance assignments from 3D backbone walks. (A) 3D data sets are represented as cubes. The signals are resolved in three dimensions. NCACX and NCOCX data sets are shown in purple and blue, respectively. The extracted strips are aligned in an alternating order by same nitrogen chemical shifts for easy comparison. (B) A representation of the aligned strips for a string of residues. The nitrogen chemical shifts connect the subsequent residues in the string, and the C^{α} and C^{β} chemical shifts clearly indicate the residue type. Once the residue types have been identified, they can be matched to the protein sequence to obtain sequence-specific resonance assignments.

rotor-synchronized $R2_n^v$ symmetry sequences (RDSD) have proven useful for long-range distance measurement [37]. For the selectively labeled samples, the individual ^{13}C-^{13}C and ^{15}N-^{13}C distances can be measured using rotational resonance (R^2) [38] or its variants [39]. Recently, the development of paramagnetic relaxation enhancement (PRE's) [40,41] and paramagnetic pseudocontact shift (PCS's) [42] techniques enabled measurements of long-range distance up to 20-30 Å.

2.4 Characterization of protein dynamics by SSNMR

Dynamics plays an important role in biological functions of proteins and other macromolecules. SSNMR is a powerful method to investigate protein dynamics over a wide range of time scales from picoseconds to seconds and longer.

2.4.1 Motional timescales and their relationship to protein function

Protein functions, such as protein folding, target recognition, protein-protein interaction and complex formation rely on molecular dynamics. The motional timescales span a wide range. For example, in the time scale of picoseconds to nanoseconds, methyl group rotation, local flexibility and bond vibrations occur. On the time scales of nano- to microseconds, H-transfer and Brownian tumbling take place. In the regime spanning micro- to milliseconds, domain motions in proteins occur, some of which are directly related to catalysis and/or binding of substrates and inhibitors. Slower motions spanning milliseconds to seconds and slower may be associated with global conformational rearrangements important in protein folding as

well as with catalytically relevant motions. The timescale of protein motions and the corresponding NMR techniques are shown in Figure 2.3.

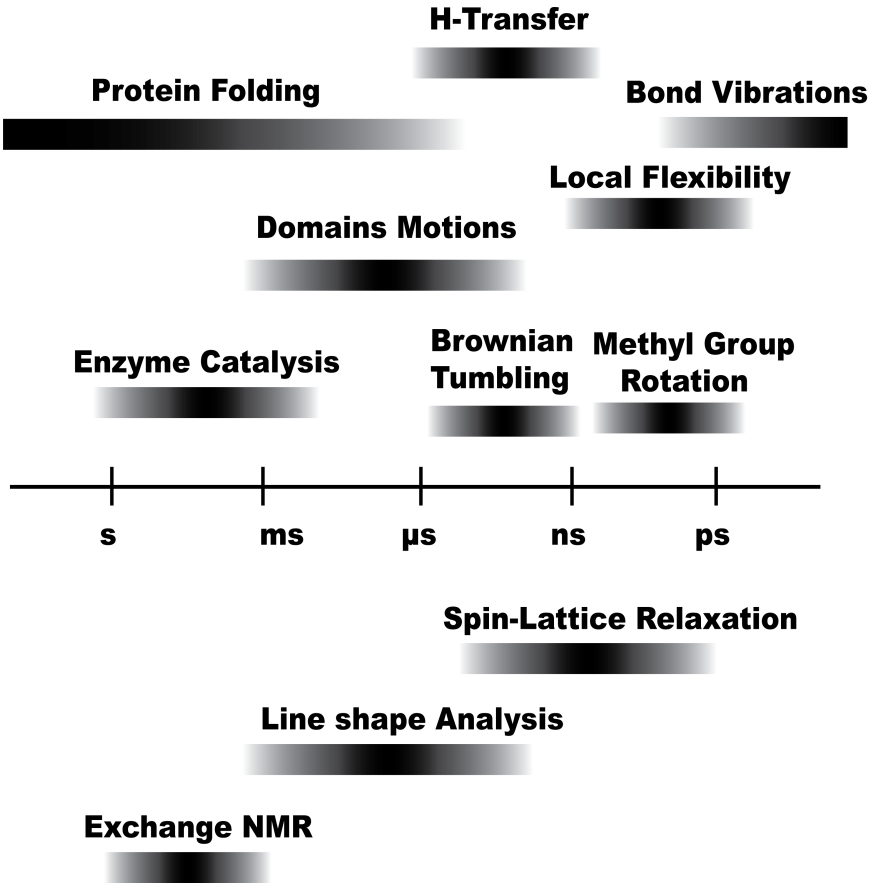


Figure 2.3 Timescale of dynamic processes in proteins (top) and the corresponding NMR techniques to probe (bottom). Figure redrawn from [43].

2.4.2 Spin interactions and relaxation rates as a probe of protein motions

Most relaxation methods for probing dynamics have essentially the same origin: molecular motion renders internuclear magnetic interaction time-dependent. At the time scale of picoseconds to microseconds, spin-lattice relaxation (T_1) measurements are useful means to probe the molecular dynamics. The molecular

motions modulate the internuclear dipolar coupling during spin-lattice relaxation time, T_1 [44].

On the time scales of nanoseconds to microseconds, line shapes (^1H-^{15}N/^1H-^{13}C dipolar [45] and chemical shift anisotropy [46]) are sensitive to motions, and can be utilized to probe the motions occurring on these time scales. In MAS NMR, pulse sequences are designed to recouple the dipolar and chemical shift interactions, which are otherwise averaged out by the rotation, to record the corresponding line shapes in a residue-specific manner. In the presence of motions occurring on the timescales of the same order as dipolar and CSA interactions, both types of line shapes experience narrowing, from the extent of which one can determine the amplitude and the symmetry of the motion.

On slower time scales, i.e. seconds to milliseconds, measurements of smaller-magnitude interactions (such as ^{13}C-^{15}N dipolar) and exchange NMR methods [47] are employed to probe slow conformational changes, such as ligand binding and release, topological interconversion of secondary structure and cis-trans isomerization.

One of the most prominent advantages of NMR spectroscopy is that it is able to detect chemical exchange even when the system is at equilibrium – the magnetization is perturbed in only one nuclear spin state to study rates without perturbing the chemical system at large.

2.4.3 Experiments to probe nano- to micro-second timescale motions

Measurement of heteronuclear (^{1}H–^{13}C and ^{1}H–^{15}N) dipolar couplings and CSA provide information both about the structure and mobility of the proteins on nano- to microsecond time scales.

2.4.3.1 ^{13}C-^{1}H and ^{15}N-^{1}H dipolar lineshapes

For the heteronuclear recoupling of the dipolar lineshapes, various sequences have been designed to work for different interactions. One class of such sequences is R-type sequences, which are utilized for recoupling the heteronuclear spin-spin interactions, while at the same time suppressing the effect of homonuclear interactions. The R-type sequences are usually denoted as RN_n^v, in which R means the basic construction block of rotor-synchronized pulses; N is an even number and the duration of R is equal to

$$\tau_R = n\tau_r / N, \qquad [2.14]$$

where τ_r is the rotational period; the phase alternation of RN_n^v is in the form of

$$RN_n^v = (R)_\phi (R')_{-\phi} (R)_\phi (R')_{-\phi} \cdots (R')_{-\phi}, \qquad [2.15]$$

where

$$\phi = \frac{\pi v}{N}, \qquad [2.16]$$

and R' is identical to R except that the phase of the π pulses are reversed; the n is the number of rotor periods in a complete RN_n^v sequence [45,48].

For example, at the MAS frequency of 10 kHz, the $R18_1^7$ block works efficiently for recoupling the dipolar lineshapes, where 9 repetitions of the pulse pairs occupy exactly one rotational period and require an rf field corresponding to a nutation frequency of nine times the spinning frequency. A 2D $R18_1^7$-based experiment for recording the ^{1}H-^{15}N/^{1}H-^{13}C dipolar lineshapes is shown in Figure 2.4.

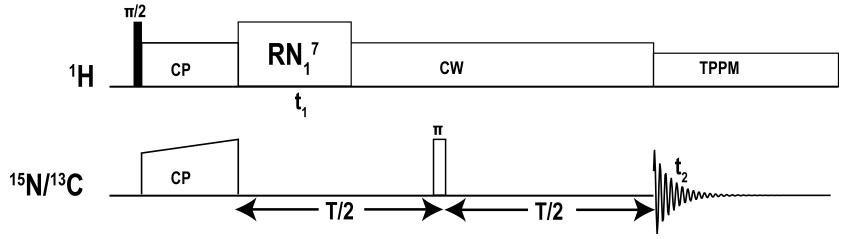


Figure 2.4 The two-dimensional $R18_1^7$ experiment for the recoupling of dipolar line shapes. In this experiment, C-H or N-H dipolar couplings are recorded site-specifically, separated by the ^{13}C or ^{15}N chemical shifts, respectively.

2.4.3.2 ^{15}N CSA lineshapes

ROCSA (recoupling of chemical shift anisotropy) is a well-established approach to recouple the CSA interaction under MAS spinning frequencies up to 20 kHz. In the ROCSA pulse sequence, the CN_n^v symmetry blocks are incorporated in the ^{15}N CSA recoupling experiments during the t_1 period for ^{15}N CSA evolution. In the CN_n^v symmetry pulse sequences established by Levitt and coworkers, each C element spans a total of n rotor periods τ_r. The basic C-type building blocks are 2π pulses [48]. ROCSA experiment utilizes the simplest symmetry Cn_n^1, in which C is

an rf sequence of span $\tau_C = n\tau_r / N$ with phase 0 or π that produces a net rotation that is an integer multiple of 2π. The rf phase shift $2\pi qv / N$ is applied to the qth repetition of C, where q is an integer running from 0 to $N-1$ [49].

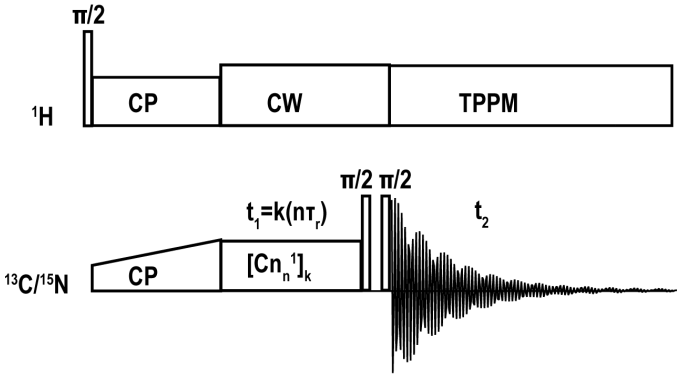


Figure 2.5 A two-dimensional ROCSA pulse sequence for the recoupling of $^{13}C/^{15}N$ CSA line shapes.

The advantages of this method include: (i) the effect of homonuclear dipole-dipole interactions are considerably suppressed, (ii) the line shape acquired from this method is identical to the corresponding static powder pattern, (iii) the required rf field power level is around four times the spinning frequency and accessible to common hardware, provided spinning frequencies not exceeding 20 kHz [49]. The disadvantages are quite severe and include: (i) extreme sensitivity to experimental settings; (ii) limited recoupling bandwidth and inability to recouple broad powder patterns, such as those corresponding to carbonyl and aromatic sidechain carbons; (iii) high demands on rf power making it impossible to utilize ROCSA at MAS frequencies higher than 20 kHz.

To overcome the above limitations of the ROCSA sequence, our laboratory has recently developed the RN_n^v based experiments for recoupling of the heteronuclear (nuclei other than ^{1}H) and ^{1}H CSA tensors [50,51].

2.4.4 Experiments to probe millisecond timescale motions

The presence of millisecond time scale motions can be inferred from ^{13}C-^{15}N dipolar coupling, which is averaged partially or completely in the presence of such motions. As the result, millisecond time scale motions result in peak broadening or disappearance in the NCA and CAN spectra, which serve as the diagnostic signature of the presence of such motions. Explicit measurements of ^{13}C-^{15}N dipolar couplings is also a useful approach to probe the presence and the amplitude of these motions, and these can be performed using a REDOR experiment, as discussed below.

2.4.4.1 ^{13}C-^{15}N REDOR dipolar dephasing curves

^{13}C-^{15}N rotational echo double resonance (REDOR) spectroscopy is an experiment that was established by Gullion and Schaefer in the late 1980's to characterize heteronuclear dipolar interactions in isolated pairs of nuclei [52]. This method allows the recoupling of dipolar interactions between different nuclei (e.g. ^{15}N and ^{13}C) by periodically inverting the sign of the dipolar interaction with pulse sequences. In the presence of motions occurring on the timescale of the same order as the magnitude of the ^{13}C-^{15}N dipolar interaction (milliseconds), the dipolar coupling is

attenuated or averaged out completely, and therefore, the extent of dipolar dephasing is a sensitive reporter of the millisecond timescale motions.

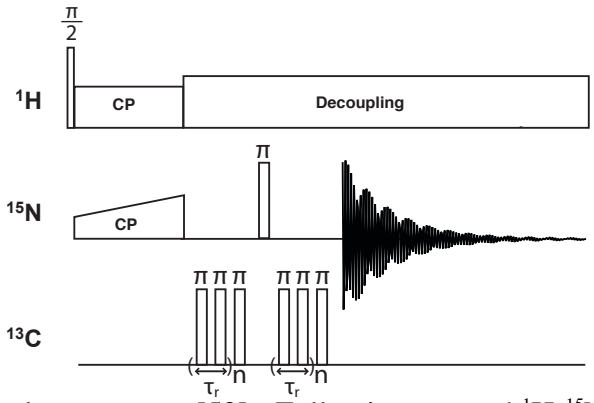


Figure 2.6 REDOR pulse sequence [52]. Following ramped ^{1}H-^{15}N cross polarization [53], the ^{15}N signal is observed as a spin-echo. The intensity of the echo is modulated due to the ^{13}C-^{15}N dipole interactions, recoupled by applying rotor-synchronized π pulses to ^{13}C spins. Reference (S_0) experiments were performed in the absence of ^{13}C pulses.

In the REDOR spectra, S_0 and S are the peak intensities in the absence and presence of the dephasing π pulses, respectively. $\Delta S = S_0 - S$ is the intensity difference. In the analysis of the REDOR dephasing curves, the REDOR fraction ΔS over S_0 is plotted against the mixing time. Numerical simulations of REDOR dephasing curves using the actual experimental conditions yield the value of the dipolar coupling constant. In the presence of motion, the resulting value is attenuated compared to the rigid-limit dipolar coupling.

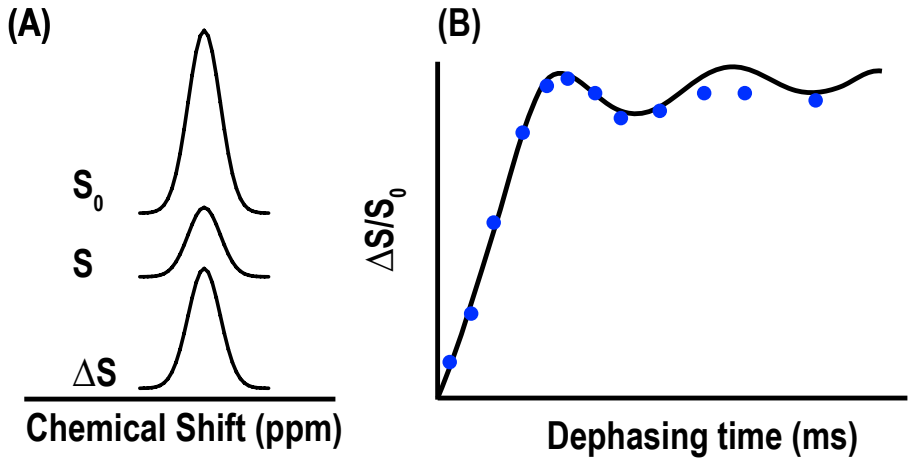


Figure 2.7 The procedure to extract results from ^{13}C-^{15}N REDOR experiments. (A) S_0 and S are the peak intensities in the absence and presence of REDOR dephasing pulses, respectively. $\Delta S = S_0 - S$. (B) In the REDOR dephasing curves, the experimental REDOR fraction $\Delta S/S_0$ (blue dots) is plotted against the dephasing time. The best-fit simulated curve (black line) yields the value of the ^{13}C-^{15}N dipolar coupling constant.

References

[1] E. Hahn; D. Maxwell. Spin Echo Measurements of Nuclear Spin Coupling in Molecules. *Phys. Rev.,* 88(5): 1070-1084, 1952.

[2] W. C. Lai; N. Mclean; A. Gansmuller; M. A. Verhoeven; G. C. Antonioli; M. Carravetta; L. Duma; P. H. M. Bovee-Geurts; O. G. Johannessen; H. J. M. De Groot; J. Lugtenburg; L. Emsley; S. P. Brown; R. C. D. Brown; W. J. Degrip; M. H. Levitt. Accurate measurements of C-13-C-13 J-couplings in the rhodopsin chromophore by double-quantum solid-state NMR spectroscopy. *J. Am. Chem. Soc.,* 128(12): 3878-3879, 2006.

[3] A. B. Siemer; C. Ritter; M. Ernst; R. Riek; B. H. Meier. High-resolution solid-state NMR spectroscopy of the prion protein HET-s in its amyloid conformation. *Angew. Chem. Int. Ed. Engl.,* 44(16): 2441-2444, 2005.

[4] L. Chen; J. M. Kaiser; T. Polenova; J. Yang; C. M. Rienstra; L. J. Mueller. Backbone assignments in solid-state proteins using J-based 3D Heteronuclear correlation spectroscopy. *J. Am. Chem. Soc.,* 129(35): 10650-10651, 2007.

[5] F. Castellani; B. Van Rossum; A. Diehl; M. Schubert; K. Rehbein; H. Oschkinat. Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature,* 420(6911): 98-102, 2002.

[6] A. C. Zeri; M. F. Mesleh; A. A. Nevzorov; S. J. Opella. Structure of the coat protein in fd filamentous bacteriophage particles determined by solid-state NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.,* 100(11): 6458-6463, 2003.

[7] R. Tycko. Molecular structure of amyloid fibrils: insights from solid-state NMR. *Q. Rev. Biophys.,* 39(1): 1-55, 2006.

[8] M. Baldus. Correlation experiments for assignment and structure elucidation of immobilized polypeptides under magic angle spinning. *Prog. Nucl. Mag. Res. Sp.,* 41(1-2): 1-47, 2002.

[9] M. Sattler; J. Schleucher; C. Griesinger. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Mag. Res. Sp.,* 34(2): 93-158, 1999.

[10] K. Wèuthrich. NMR of proteins and nucleic acids; Wiley: New York, 1986.

[11] N. Popovych; S. J. Sun; R. H. Ebright; C. G. Kalodimos. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.,* 13(9): 831-838, 2006.

[12] J. Yang; M. L. Tasayco; T. Polenova. Dynamics of reassembled thioredoxin studied by magic angle spinning NMR: snapshots from different time scales. *J. Am. Chem. Soc.,* 131(38): 13690-13702, 2009.

[13] A. Lesage. Recent advances in solid-state NMR spectroscopy of spin I=1/2 nuclei. *Phys. Chem. Chem. Phys.,* 11(32): 6876-6891, 2009.

[14] A. Bockmann; A. Lange; A. Galinier; S. Luca; N. Giraud; M. Juy; H. Heise; R. Montserret; F. Penin; M. Baldus. Solid state NMR sequential resonance assignments and conformational analysis of the 2 x 10.4 kDa dimeric form of the Bacillus subtilis protein Crh. *J. Biomol. NMR,* 27(4): 323-339, 2003.

[15] W. T. Franks; B. J. Wylie; H. L. F. Schmidt; A. J. Nieuwkoop; R. M. Mayrhofer; G. J. Shah; D. T. Graesser; C. M. Rienstra. Dipole tensor-based atomic-resolution structure determination of a nanocrystalline protein by solid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.,* 105(12): 4621-4626, 2008.

[16] M. Tang; G. Comellas; L. J. Mueller; C. M. Rienstra. High resolution C-13-detected solid-state NMR spectroscopy of a deuterated protein. *J. Biomol. NMR,* 48(2): 103-111, 2010.

[17] J. Yang; S. Paramasivan; D. Marulanda; M. Cataidi; M. L. Tasayco; T. Polenova. Magic angle spinning NMR spectroscopy of thioredoxin reassemblies. *Magn. Reson. Chem.,* 45(S1): S73-S83, 2007.

[18] D. Marulanda; M. L. Tasayco; M. Cataldi; V. Arriaran; T. Polenova. Resonance assignments and secondary structure analysis of *E.coli* thioredoxin by magic angle spinning solid-state NMR spectroscopy. *J. Phys. Chem. B,* 109(38): 18135-18145, 2005.

[19] D. Marulanda; M. L. Tasayco; A. Mcdermott; M. Cataldi; V. Arriaran; T. Polenova. Magic angle spinning solid-state NMR spectroscopy for structural studies of protein interfaces. Resonance assignments of differentially enriched *Escherichia coli* thioredoxin reassembled by fragment complementation. *J. Am. Chem. Soc.,* 126(50): 16608-16620, 2004.

[20] D. Marulanda; M. L. Tasayco; M. Cataldi; V. Arriaran; T. Polenova. Resonance assignments and secondary structure analysis of *E. coli* thioredoxin by magic angle spinning solid-state NMR spectroscopy. *J. Phys. Chem. B,* 109(38): 18135-18145, 2005.

[21] J. Yang; S. Paramasivam; D. Marulanda; M. Cataldi; M. L. Tasayco; T. Polenova. Magic angle spinning NMR spectroscopy of thioredoxin reassemblies. *Magn. Reson. Chem.,* 45 Suppl 1S73-83, 2007.

[22] A. Mcdermott. Structure and dynamics of membrane proteins by magic angle spinning solid-state NMR. *Annu. Rev. Biophys.,* 38385-403, 2009.

[23] R. Tycko. Solid-state NMR studies of amyloid fibril structure. *Annu. Rev. Phys. Chem., Vol 62,* 62279-299, 2011.

[24] W. T. Franks; H. S. Atreya; T. Szyperski; C. M. Rienstra. GFT projection NMR spectroscopy for proteins in the solid state. *J. Biomol. NMR,* 48(4): 213-223, 2010.

[25] D. S. Wishart; M. S. Watson; R. F. Boyko; B. D. Sykes. Automated H-1 and C-13 chemical shift prediction using the BioMagResBank. *J. Biomol. NMR,* 10(4): 329-336, 1997.

[26] A. Mcdermott; T. Polenova; A. Bockmann; K. W. Zilm; E. K. Paulsen; R. W. Martin; G. T. Montelione. Partial NMR assignments for uniformly (C-13, N-15)-enriched BPTI in the solid state. *J. Biomol. NMR,* 16(3): 209-219, 2000.

[27] S. Sun; A. H. Butterworth; S. Paramasivam; S. Yan; C. M. Lightcap; J. C. Williams; T. Polenova. Resonance assignments and secondary structure analysis of dynein light chain 8 by magic angle spinning NMR spectroscopy. *Can. J. Chem.,* 89(7): 909-918, 2011.

[28] S. Sun; A. Siglin; J. C. Williams; T. Polenova. Solid-state and solution NMR studies of the CAP-Gly domain of mammalian dynactin and its interaction with microtubules. *J. Am. Chem. Soc.,* 131(29): 10113-10126, 2009.

[29] H. N. B. Moseley; L. J. Sperling; C. M. Rienstra. Automated protein resonance assignments of magic angle spinning solid-state NMR spectra of beta 1 immunoglobulin binding domain of protein G (GB1). *J. Biomol. NMR,* 48(3): 123-128, 2010.

[30] A. Schuetz; C. Wasmer; B. Habenstein; R. Verel; J. Greenwald; R. Riek; A. Bockmann; B. H. Meier. Protocols for the sequential solid-state NMR spectroscopic assignment of a uniformly labeled 25 kDa protein: HET-s(1-227). *ChemBioChem,* 11(11): 1543-1551, 2010.

[31] D. S. Wishart; B. D. Sykes; F. M. Richards. The Chemical-Shift Index - a sast and simple method for the assignment of protein secondary structure through NMR-spectroscopy. *Biochemistry,* 31(6): 1647-1651, 1992.

[32] D. S. Wishart; B. D. Sykes. The C-13 Chemical-Shift Index - a simple method for the identification of protein secondary structure using C-13 chemical-shift data. *J. Biomol. NMR,* 4(2): 171-180, 1994.

[33] D. S. Wishart; B. D. Sykes. Chemical-shifts as a tool for structure determination. *Method Enzymol.,* 239363-392, 1994.

[34] S. Schwarzinger; G. J. A. Kroon; T. R. Foss; J. Chung; P. E. Wright; H. J. Dyson. Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.,* 123(13): 2970-2978, 2001.

[35] A. Bockmann. 3D protein structures by solid-state NMR spectroscopy: ready for high resolution. *Angew. Chem. Int. Ed. Engl.,* 47(33): 6110-6113, 2008.

[36] T. Manolikas; T. Herrmann; B. H. Meier. Protein structure determination from ^{13}C spin-diffusion solid-state NMR spectroscopy. *J. Am. Chem. Soc.,* 130(12): 3959-3966, 2008.

[37] G. Hou; S. Yan; S. Sun; Y. Han; I. J. Byeon; J. Ahn; J. Concel; A. Samoson; A. M. Gronenborn; T. Polenova. Spin diffusion driven by R-symmetry sequences: applications to homonuclear correlation spectroscopy in MAS NMR of biological and organic solids. *J. Am. Chem. Soc.,* 133(11): 3943-3953, 2011.

[38] D. P. Raleigh; M. H. Levitt; R. G. Griffin. Rotational resonance in solid-state NMR. *Chem. Phys. Lett.,* 146(1-2): 71-76, 1988.

[39] P. R. Costa; B. Q. Sun; R. G. Griffin. Rotational resonance tickling: Accurate internuclear distance measurement in solids. *J. Am. Chem. Soc.,* 119(44): 10821-10830, 1997.

[40] C. P. Jaroniec. Solid-state nuclear magnetic resonance structural studies of proteins using paramagnetic probes. *Solid State Nucl. Magn. Reson.,* 43-441-13, 2012.

[41] P. S. Nadaud; J. J. Helmus; N. Hofer; C. P. Jaroniec. Long-range structural restraints in spin-labeled proteins probed by solid-state nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.,* 129(24): 7502-7503, 2007.

[42] M. J. Knight; I. C. Felli; R. Pierattelli; I. Bertini; L. Emsley; T. Herrmann; G. Pintacuda. Rapid measurement of pseudocontact shifts in metalloproteins by proton-detected solid-state NMR spectroscopy. *J. Am. Chem. Soc.,* 134(36): 14730-14733, 2012.

[43] K. Henzler-Wildman; D. Kern. Dynamic personalities of proteins. *Nature,* 450(7172): 964-972, 2007.

[44] M. Hologne; K. Faelber; A. Diehl; B. Reif. Characterization of dynamics of perdeuterated proteins by MAS solid-state NMR. *J. Am. Chem. Soc.,* 127(32): 11208-11209, 2005.

[45] X. Zhao; J. L. Sudmeier; W. W. Bachovchin; M. H. Levitt. Measurement of NH bond lengths by fast magic-angle spinning solid-state NMR spectroscopy: a new method for the quantification of hydrogen bonds. *J. Am. Chem. Soc.,* 123(44): 11097-11098, 2001.

[46] J. C. Chan; R. Tycko. Solid-state NMR spectroscopy method for determination of the backbone torsion angle psi in peptides with isolated uniformly labeled residues. *J. Am. Chem. Soc.,* 125(39): 11828-11829, 2003.

[47] A. Krushelnitsky; D. Reichert; G. Hempel; V. Fedotov; H. Schneider; L. Yagodina; A. Schulga. Superslow backbone protein dynamics as studied by 1D solid-state MAS exchange NMR spectroscopy. *J. Magn. Reson.,* 138(2): 244-255, 1999.

[48] M. Carravetta; M. Eden; X. Zhao; A. Brinkmann; M. H. Levitt. Symmetry principles for the design of radiofrequency pulse sequences in the nuclear magnetic resonance of rotating solids. *Chem. Phys. Lett.,* 321(3-4): 205-215, 2000.

[49] J. C. C. Chan; R. Tycko. Recoupling of chemical shift anisotropies in solid-state NMR under high-speed magic-angle spinning and in uniformly C-13-labeled systems. *J. Chem. Phys.,* 118(18): 8378-8389, 2003.

[50] G. Hou; I. J. Byeon; J. Ahn; A. M. Gronenborn; T. Polenova. Recoupling of chemical shift anisotropy by R-symmetry sequences in magic angle spinning NMR spectroscopy. *J. Chem. Phys.,* 137(13): 134201, 2012.

[51] G. Hou; S. Paramasivam; S. Yan; T. Polenova; A. J. Vega. Multidimensional magic angle spinning NMR spectroscopy for site-resolved measurement of

proton chemical shift anisotropy in biological solids. *J. Am. Chem. Soc.*, 135(4): 1358-1368, 2013.

[52] T. Gullion; J. Schaefer. Rotational-echo double-resonance NMR. 1989. *J. Magn. Reson.*, 213(2): 413-417, 2011.

[53] G. Metz; X. L. Wu; S. O. Smith. Ramped-amplitude cross-polarization in magic-angle-spinning NMR. *J. Magn. Reson. Ser. A,* 110(2): 219-227, 1994.

Chapter 3

PREPARATION AND CHARACTERIZATION OF HIV-1 CA AND CA-SP1 PROTEIN ASSEMBLIES

3.1 Introduction

HIV-1 CA protein assemblies exhibit structural polymorphism. While *in vivo* the mature virions contain capsids of predominantly conical morphology, the immature viral particles contain spherical Gag assemblies, which are the precursors of the conical capsid in the mature virions [1,2]. *In vitro*, depending on the assembly conditions, CA assemblies can adopt conical, spherical, donut-like, or tubular morphologies [3-6]. The characterizations of those morphologies are important not only because those morphologies are closely related to different stages of the viral life cycles, but also because that morphologically homogeneous samples are essential for the quality of solid-state NMR spectra.

TEM as a conventional method has provided high-resolution images of HIV-1 CA protein assemblies of various morphologies. The tubular and conical assemblies of CA protein have been successfully characterized by this technique, either with heavy metal solution staining or by cryoEM.

Laser scanning confocal microscopy is able to detect protein assemblies in solution with the typical resolution of optical microscopes. By using various fluorescence staining techniques, protein assemblies are distinguished from the solvent as they emit fluorescence signals. In 2009, Barklis and coworkers devised an FM-based assay and characterized the formation of HIV-1 CA tubular assemblies [7]. They utilized a primary antibody to recognize the protein assemblies and a fluorophore-conjugated secondary antibody used for the fluorescence emission. In this approach, the fluorescence light explicitly images the protein assemblies. Because confocal microscopy is based on optical light, the resolution could not exceed 0.2 μm, which is around the size of HIV-1 CA conical assemblies.

Cryo-SEM is a complementary method to characterize the HIV-1 CA assemblies. In cryo-SEM experiments, the sample is frozen in a gold cup at high pressure and low temperature, to best preserve the original protein assemblies' structures [8]. Then the cover of the gold cup is fractured with a precooled knife, after which the fractured surface of the sample reveals the details of the assembly structure as in solution. The sample preparation procedures are strongly dependent on humidity and environment temperatures, which makes it difficult to prevent formation of ice crystals without disturbing the protein assembly structures. The resolution of cryo-SEM is lower than that of TEM but higher than of optical microscopes. One advantage of cryo-SEM method for observing the HIV-1 CA protein spherical assemblies is that those the spheres do not get flattened while imaging. Furthermore,

it is possible to perform a dissection through a sphere with the special knife. In this way, inner structures can be observed inside the section of the sphere.

3.2 Preparation of CA assemblies of conical morphology

HIV-1 CA forms conical cores *in vivo* after the maturation process, and the conical assemblies have been successfully assembled *in vitro*. CA can be assembled into conical morphologies using PEG solutions, as detailed below.

3.2.1 Protein purification

Expression and purification of the HIV-1 capsid protein were performed by Dr. Jinwoo Ahn and his lab members (University of Pittsburgh) and have been reported previously [9].

The cDNA encoding Gag polyprotein, $pr55^{gag}$, was obtained from the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH, [10] and the sequence coding for CA was amplified and subcloned into pET21 (EMD chemicals, Inc. San Diego, CA) using NdeI and XhoI sites. U-^{13}C,^{15}N isotopically enriched proteins were expressed in *Escherichia coli*, Rosetta 2 (DE3), cultured in modified minimal medium using $^{15}NH_4Cl$ and U-$^{13}C_6$-glucose as sole nitrogen and carbon sources, after induction with 0.4 mM IPTG at 23 °C for 16 h. Four additional HIV-1 CA samples were prepared containing isotopic labels in a single amino acid residue type: U-^{13}C,^{15}N-Ala, U-^{13}C,^{15}N-His, U-^{13}C,^{15}N-Thr, and U-^{13}C,^{15}N-Tyr. Proteins were expressed in modified minimal media adding each specific U-^{13}C,^{15}N-

amino acid and purified via ionic exchange and gel filtration column chromatography to greater than 99% homogeneity.

3.2.2 Assembly

To prepare the CA assemblies of conical morphology, the sample was dialyzed first to reduce the salt content. The dialysis buffer contains 25 mM sodium phosphate, 1 mM DTT and 0.02% NaN_3 (pH 5.5). The Amicon cell was used to perform the dialysis and to concentrate the protein solution to the final volume of 1 ml (the final concentration of the protein is typically 400-600 μM). Dialysis and buffer exchange is performed three times to reduce the salt concentration to the final value of 25 mM.

Next, the protein solution prepared from the above dialysis/concentration step is lyophilized. If lyophilization is not desired, protein solution is concentrated to 32 mg/ml or higher. Then 17.5% w/v PEG-20,000 solution or 30% w/v PEG-4000 solution is added to the lyophilized or concentrated protein solutions at a volume ratio of 1:1. The slurry is then incubated at 37 °C for one hour and subjected to centrifugation to remove the supernatant at 14,000 g at 4 °C. The resulting pellet is used in the various imaging experiments, to corroborate the conical morphology of the CA sample.

3.2.3 Microscopy

To characterize the morphology of the CA assemblies, TEM, confocal, and cryo-SEM methods were utilized, as detailed below.

3.2.3.1 Transmission Electron Microscopy (TEM)

CA assemblies were analyzed using a Zeiss CEM 902 transmission electron microscope operating at 80 kV. Samples were deposited onto 400 mesh, Formval/ carbon-coated copper grids, washed with two drops of water, then stained with ammonium molybdate (5% w/v) to enhance contrast, and dried for 40 min to prevent moisture from entering the vacuum system.

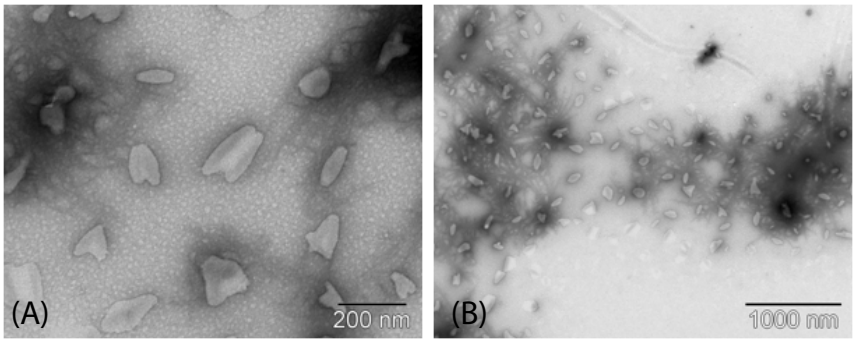


Figure 3.1 TEM images of CA conical assemblies. The assemblies were prepared by lyophilization and subsequent addition of 17.5% w/v PEG 20,000 solution. The resulting sample was stained with 5% ammonium molybdate solution on the coper grid. (A) The TEM image of CA protein conical assemblies taken at magnification of 50,000x; (B) The TEM image CA protein conical assemblies taken at magnification of 12,500x.

3.2.3.2 Cryo Scanning Electron Microscopy (Cryo-SEM)

Slurries containing conical assemblies (~1 mm^3) were applied to gold carrier plates and frozen in a Leica EM PACT high-pressure freezer. The cooling rate should be no less than 6000 K/s to ensure the complete immobilization of the sample fine structures. Subsequently, the samples were fractured with a precooled knife and the

fractured surface was coated with 10 nm of gold. The frozen samples were then transferred to the cryostage and warmed to -90 °C for 5-7 min to remove surface water. The images were taken on a Hitachi S-4700 FESEM (Tokyo, Japan) at -125 °C and 1.0 kV with a working distance of approximately 4-5 mm.

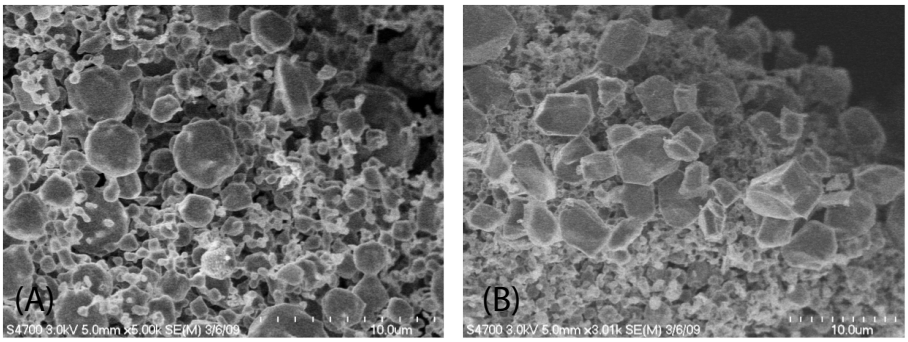


Figure 3.2 The cryo-SEM images of CA protein conical assemblies. The size distribution in the cryo-SEM data is much wider than in the images observed by TEM, i.e. from less than 1 micron to several microns. The shapes of assemblies observed are multi-hedral, and are not as homogeneous as observed with TEM images. (A) The cryo-SEM image of CA protein conical assemblies taken at the magnification of 5000x; (B) The cryo-SEM image of CA protein assemblies taken at the magnification of 3000x.

The shapes observed in the cryo-SEM images are irregular multi-hedral objects. The size distribution is from less than one micron to several microns. Because of the inherently limited resolution of SEM method, smaller objects that showed up in TEM images were not observed with cryo-SEM. However, it is clear that the protein slurry under investigation contains assemblies that are neither spheres nor tubes in shape. This method corroborated that when incubated with the PEG-20,000 solution, the CA protein formed conical assemblies at larger length scales. We

also note that the sample preparation process involves the warming step at -90 °C, which may have removed the surface water as well as the fine structures of the protein assemblies.

3.2.3.3 Confocal Laser Scanning Microscopy

Confocal images were acquired on a Zeiss LSM 510 NLO laser scanning microscope (Carl Zeiss, Germany) using a Zeiss 40× (NA 1.3) oil immersion objective lens. All data were collected using the 584 nm laser line of a 25 mW argon laser (LASOS, Germany), and 633 nm laser line of a 35 mW HeNe laser (LASOS, Germany) with a rhodamine reflector and the transmitted-light channel when appropriate. Protein solutions were stained with the Nile Blue A dye (10 ppm) according to the standard protocol, and all images were acquired in fluorescence, transmittance, and reflection modes.

The confocal microscope resolution is limited to 200 nm because of the diffraction limit of visible light. The size of the CA conical assemblies is smaller than 200 nm according to the TEM results. Therefore, it is not possible to observe the specific shapes of the CA conical assemblies with confocal microscope. The Nile Blue A dye emits fluorescence in the hydrophobic environment such as CA protein assemblies. The fluorescent dots appeared in the confocal microscope images are the protein assemblies distinguished from the hydrophilic PEG 20,000 and the water solvent. The comparison of the confocal images of CA samples before and after magic angle spinning showed essentially the same result, telling us we can study the intact CA conical assemblies under solid-state NMR experimental conditions.

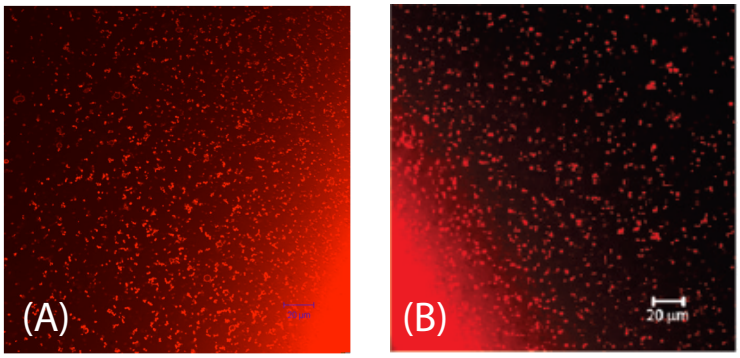


Figure 3.3 The confocal images of CA conical assemblies stained with Nile Blue A dye (fluorescence mode). Since the resolution of the confocal microscope is limited in 200 nm, it is not possible to observe the shapes of the conical assemblies with confocal method. (A) The confocal image of CA protein conical assemblies; (B) The confocal image of CA protein conical assemblies after magic angle spinning.

3.3 Spherical morphologies

Prior to the maturation process, HIV-1 CA protein, which exists as part of the Gag poly protein, is densely packed into spherical assemblies *in vivo*. The large spherical assemblies get flattened on the TEM grids and form finely ordered 2D crystals that are ideal subject for cryo-EM studies [11]. In this way, the overall image of CA protein spherical assemblies can be indirectly deduced as the upper and lower layer of the 2D crystals possess mirror symmetry. According to the cryo EM models, the large spherical assemblies of CA are composed of hexamers and evenly distributed pentamers. Using cryo-SEM and confocal microscopy, we directly visualized those large spherical assemblies for the first time.

3.3.1 Protein purification

This part of the work was carried out in the laboratory of our collaborator, Dr. Jinwoo Ahn (University of Pittsburgh). Expression and purification of the HIV-1 capsid protein solution was performed by Dr. Jinwoo Ahn and his lab members, according to the same protocols described for CA conical assemblies.

3.3.2 Assembly

For preparation of the CA assemblies, the protein was first dialyzed against 25 mM sodium phosphate buffer, containing 1 mM DTT, and 0.02% NaN3 (pH 5.5).

CA assemblies of spherical morphology were prepared by mixing the 32 mg/mL CA solution with 7.5% w/v PEG-20 000 (1:1 volume ratio). The resulting mixture was incubated on ice for 30 min and then diluted 4-fold. To concentrate the protein assemblies for solid-state NMR experiments, the solution containing the spherical assemblies was dried with N_2 gas to remove any excess water.

3.3.3 Microscopy

To characterize the morphology of the conical assemblies, we had to develop a new protocol as there were no prior studies. We established novel confocal and cryo-SEM methods to characterize the shape of the CA spherical assemblies. The results from both microscopy methods are consistent. Confocal microscopy showed that the spheres were preserved upon being air-dried, making our solid-state NMR studies of this assembly morphology feasible.

3.3.3.1 Cryo Scanning Electron Microscopy (Cryo-SEM)

The procedure to acquire cryo-SEM images of CA protein spherical assemblies is the same as described for the CA conical assemblies.

According to the cryo-SEM images, the CA spherical assemblies vary a lot in sizes, from less than one micron to tens of microns. The spherical shapes were well preserved with the high pressure freezing steps during sample preparation. We attempted to observe the hexameric and pentameric units that comprise the spheres. However, the fine structures could not be resolved in the cryo-SEM images. We speculate that this is due to the limited resolution of the SEM and/or due to the warming steps for surface water removal.

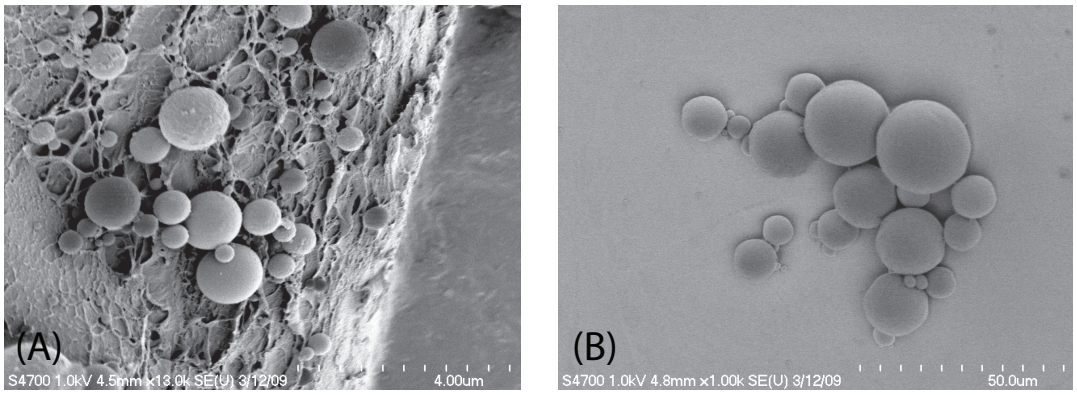


Figure 3.4 The cryo-SEM images of CA protein spherical assemblies. The sizes of the spheres vary from less than one micron to tens of microns. The spherical shapes are well defined and preserved. (A) The cryo-SEM image of CA protein spherical assemblies acquired at the magnification of 13,000x; (B) The cryo-SEM image of CA protein assemblies acquired at the manification of 1,000x.

3.3.3.2 Confocal Laser Scanning Microscopy

Confocal images were acquired on a Zeiss LSM 510 NLO laser scanning microscope (Carl Zeiss, Germany) using a Zeiss 40× (NA 1.3) oil immersion objective lens. All data were collected using the 584 nm laser line of a 25 mW argon laser (LASOS, Germany) with a rhodamine reflector and the transmitted-light channel when appropriate. Protein solutions were stained with the Nile Blue A dye (10 ppm) according to the standard protocol, and all images were acquired in fluorescence, transmittance, and reflection modes.

To prepare the spherical assemblies for the solid-state NMR experiments, we need to concentrate the protein to accommodate as much material as possible into the rotors, which are around 10 μl in volume. We have spent considerable efforts to establish the conditions for concentrating the samples, such as ultracentrifugation at 40,000 g for more than one hour and super filtration. However, the spherical assemblies could not be pelleted using these methods. Fortunately and accidentally, we discovered during the confocal imaging experiment that the sample slurry got air-dried, and the PEG 20,000 reagent formed a substrate to preserve the protein assemblies inside. The fluorescence signals clearly delineated the intact spherical shapes inside the PEG 20,000 solid. Later on, we tested the stability of these CA spherical assemblies against magic angle spinning in the solid-state NMR instruments. The confocal image of the sample containing spherical assemblies subjected to magic angle spinning shows that the spherical shapes are retained, even though the size distribution is less homogeneous than prior to the magic angle spinning. Therefore,

we proceeded with this method for the solid-state NMR sample preparation of spherical CA assemblies.

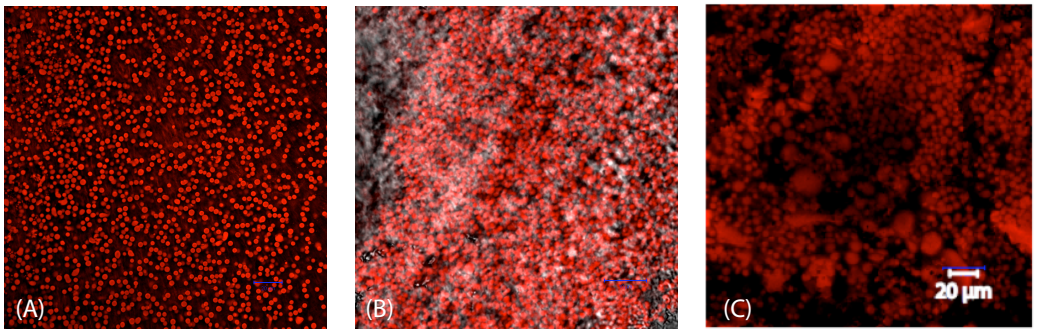


Figure 3.5 The confocal images of CA spherical assemblies reveal formation of homogeneous spherical assemblies with diameter of around 2 micron. (A) The confocal image of CA spherical assemblies in solution; (B) The confocal image of CA protein spherical assemblies after being dried under N_2 gas flow. (C) The confocal image of CA protein spherical assemblies after magic angle spinning. The image is an overlay of the fluorescence and transmission light signals. The fluorescence channel visualizes the protein assemblies, while the transmission light channel visualizes the dried PEG 20,000 that preserves the protein.

3.4 Tubular morphologies

HIV-1 CA tubular assemblies are the most investigated morphology. CA readily assembles into tubes under high salt concentration (above 500 mM). The TEM results show that the CA tubular assemblies are built of homogeneous hexameric lattice, which resembles the hexamers in mature virion capsid core.

3.4.1 Protein purification

This part of the work was carried out in the laboratory of our collaborator, Dr. Jinwoo Ahn (University of Pittsburgh). Expression and purification of the HIV-1

capsid protein solution was performed by Dr. Jinwoo Ahn and his lab members, following the same procedures as described for CA conical assemblies.

3.4.2 Assembly

CA assemblies of tubular morphology were prepared as follows. The CA protein solution was first concentrated to 32 mg/ml; during this process, the buffer was changed to 25 mM NaPi, pH 5.5; then aqueous solution of 1 M NaCl was added to the protein solution. In the last step, the protein solution was incubated at 37 °C for 1 h, and stored at 4 °C. It is important to note that tubular assemblies should never be frozen as this results in their irreversible decomposition.

3.4.3 Microscopy

To characterize the morphology of the tubular assemblies, TEM and confocal microscopies were utilized. The results from different microscopies are consistent.

3.4.3.1 Transmission Electron Microscopy (TEM)

Some of the CA tubular assemblies were analyzed using a Zeiss CEM 902 transmission electron microscope operating at 80 kV. Samples were stained with ammonium molybdate (5% w/v), deposited onto 400 mesh, formval/ carbon-coated copper grids, and dried for 40 min.

Some of the CA tubular assemblies were analyzed using a Zeiss Libra 120 transmission electron microscope operating at 120 kV. Samples were stained with

uranyl acetate (5% w/v), deposited onto 400 mesh, formval/carbon-coated copper grids, and dried for 40 min in the air. The copper grid was pretreated with Pelco easiGlow Discharge Unit to deposit a charge, so that the tubular assemblies are uniformly spread on the grid surface and adhere to it.

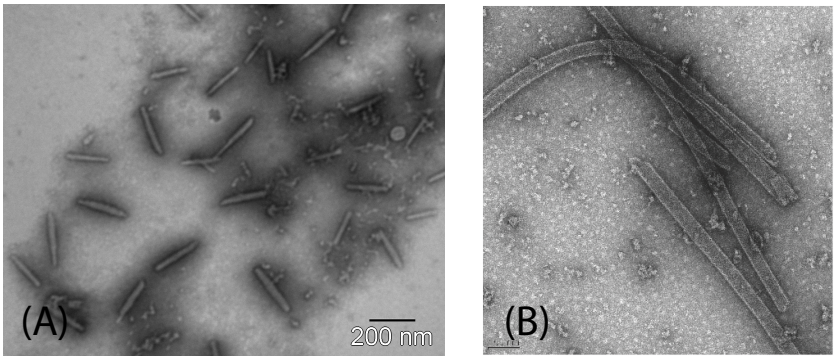


Figure 3.6 The TEM images of CA protein tubular assemblies reveal tubes that are around 50 nm in diameter. The sample was prepared as described in the Assembly section. (A) The TEM image of CA protein tubular assemblies acquired at the Zeiss CEM 902 transmission electron microscope; (B) The TEM image of CA protein tubular assemblies acquired at the Zeiss Libra 120 transmission electron microscope.

3.4.3.2 Confocal Laser Scanning Microscopy

Confocal images were acquired on a Zeiss LSM 510 NLO laser scanning microscope (Carl Zeiss, Germany) using a Zeiss 40× (NA 1.3) oil immersion objective lens. All data were collected using the 584 nm laser line of a 25 mW argon laser (LASOS, Germany), and 633 nm laser line of a 35 mW HeNe laser (LASOS, Germany) with a rhodamine reflector and the transmitted-light channel when appropriate. Protein solutions were stained with the Nile Blue A dye (10 ppm) solution after been applied on the sample stage with 1:1 volume ratio. We note that

the dye could not be incorporated directly into the sample solutions because of the high salt concentration, making Nile Blue A precipitate out of solution.

According to the results, the CA protein forms homogeneous tubes. The lengths of the tubes observed by confocal microscopy are generally more than 20 micron for the CA WT sample. The tubes remain intact upon air-drying on the confocal sample stage. These results are illustrated in Figure 3.7.

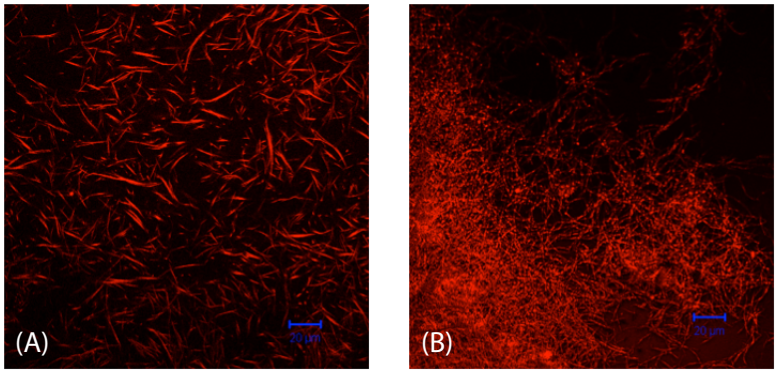


Figure 3.7 The confocal images of CA tubular assemblies. The CA tubes are clearly visualized in the fluorescence images when stained with Nile Blue A. (A) The CA protein tubular assemblies in solution; (B) The CA protein tubular assemblies after been air dried.

Each solid-state NMR experiment in HIV-1 CA assemblies usually lasts for 3-10 days, and multiple data sets are acquired for their structural and dynamics characterization. It is therefore necessary to test the sample stability under the various experimental conditions, including during the freeze-thaw cycles. These tests revealed that the CA protein tubular assemblies decompose irreversibly upon freezing and subsequent thawing, necessitating their storage at 4 °C and the experiment temperatures to be at or above 4 °C. Figure 3.8 (B) illustrates the decomposed CA

tubes after a single freeze-thaw cycle in comparison with the control sample (A). Importantly, the CA tubular assemblies are intact upon magic angle spinning as shown in Figure 3.8 (C).

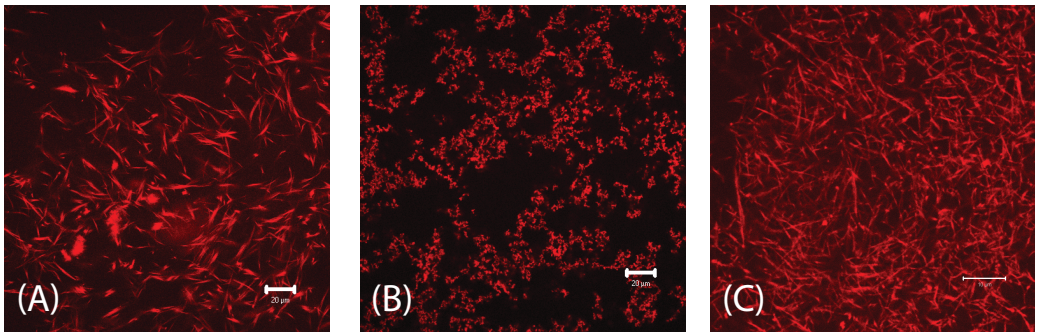


Figure 3.8 Confocal images of (A) the CA protein tubular assemblies prepared as described in the Assembly section; (B) the same sample that was flash frozen in liquid nitrogen and then thawed at room temperature; and (C) the CA protein tubular assemblies after magic angle spinning.

References

[1] U. K. Von Schwedler; K. M. Stray; J. E. Garrus; W. I. Sundquist. Functional surfaces of the human immunodeficiency virus type 1 capsid protein. *J. Virol.,* 77(9): 5439-5450, 2003.

[2] D. Ako-Adjei; M. C. Johnson; V. M. Vogt. The retroviral capsid domain dictates virion size, morphology, and coassembly of Gag into virus-like particles. *J. Virol.,* 79(21): 13463-13472, 2005.

[3] S. Campbell; V. M. Vogt. Self-assembly *in vitro* of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. *J. Virol.,* 69(10): 6487-6497, 1995.

[4] I. Gross; H. Hohenberg; H. G. Krausslich. *In vitro* assembly properties of purified bacterially expressed capsid proteins of human immunodeficiency virus. *Eur. J. Biochem.,* 249(2): 592-600, 1997.

[5] E. Barklis; J. Mcdermott; S. Wilkens; S. Fuller; D. Thompson. Organization of HIV-1 capsid proteins on a lipid monolayer. *J. Biol. Chem.,* 273(13): 7177-7180, 1998.

[6] L. S. Ehrlich; T. Liu; S. Scarlata; B. Chu; C. A. Carter. HIV-1 capsid protein forms spherical (immature-like) and tubular (mature-like) particles *in vitro*: structure switching by pH-induced conformational changes. *Biophys. J.,* 81(1): 586-594, 2001.

[7] E. Barklis; A. Alfadhli; C. Mcquaw; S. Yalamuri; A. Still; R. L. Barklis; B. Kukull; C. S. Lopez. Characterization of the *in vitro* HIV-1 capsid assembly pathway. *J. Mol. Biol.,* 387(2): 376-389, 2009.

[8] P. Walther; M. Muller. Biological ultrastructure as revealed by high resolution cryo-SEM of block faces after cryo-sectioning. *J. Microsc.,* 196279-287, 1999.

[9] Y. Han; J. Ahn; J. Concel; I. J. L. Byeon; A. M. Gronenborn; J. Yang; T. Polenova. Solid-state NMR studies of HIV-1 capsid protein assemblies. *J. Am. Chem. Soc.,* 132(6): 1976-1987, 2010.

[10] S. Ericksonviitanen; J. Manfredi; P. Viitanen; D. E. Tribe; R. Tritch; C. A. Hutchison; D. D. Loeb; R. Swanstrom. Cleavage of HIV-1 Gag polyprotein synthesized *in vitro* - sequential cleavage by the viral protease. *AIDS Res. Hum. Retroviruses,* 5(6): 577-591, 1989.

[11] B. K. Ganser-Pornillos; A. Cheng; M. Yeager. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell,* 131(1): 70-79, 2007.

Chapter 4

RESONANCE ASSIGNMENTS AND SECONDARY STRUCTURE ANALYSIS OF CONICAL ASSEMBLIES OF HIV-1 CA PROTEIN

4.1 Introduction

HIV-1 capsid (CA) has been recognized as a crucial protein, since it is the target of Rhesus Monkey TRIM5α to render this species immune to AIDS [1]. The vaccine and drug development against AIDS have been impeded by lack of atomic resolution structure of HIV-1 CA protein in the assembled state. X-ray provided detailed and accurate information on cross-linked CA mutant crystals, but is not able to directly work on the non-crystalline CA assemblies [2]. Solution NMR was used to study the separate domains of CA protein, but with the intrinsic limit on the protein sizes, the signals from full-length CA solutions were not resolved [3]. Cryo EM techniques are fast developing and able to reach molecular resolution. However, atomic resolution is big challenge that has not been overcome yet for CA assemblies [4]. Solid-state NMR is able to provide atomic level structural information on the CA protein in assembled state.

In this chapter, we present a solid-state NMR study of conical and spherical HIV-1 CA protein assemblies. We report partial site-specific resonance assignments of CA conical assemblies by a combination of two- and three-dimensional correlation experiments, as our first step to detailed structural and dynamics analysis of CA assemblies. Chemical shift analyses indicate that no major differences in the secondary

structure of the individual CA proteins in the spherical and conical assemblies exist and that the structural polymorphism most likely is due to the differences in the intermolecular packing. Most importantly, our results illustrate that the CA protein assemblies are amenable to structural analyses by MAS NMR spectroscopy.

4.2 Experiments and methods

4.2.1 Materials

All chemicals were purchased from Fisher Scientific or Sigma Aldrich and were used without further purification. 400 mesh copper grids coated with Formval and stabilized with evaporated carbon films were purchased from Electron Microscopy Sciences (Hatfield, PA).

4.2.2 Preparation of HIV-1 CA protein assemblies

The expression, purification and assembly of U-^{13}C,^{15}N isotopically enriched CA protein assemblies for high-resolution structural analysis by MAS NMR follow the protocols described in the previous chapter. Differentially mixed labeled CA samples were prepared using 17.5% w/v solutions of PEG-20,000 in 90% D_2O/10% H_2O containing 10 mM Cu-EDTA.

For the CA conical assemblies, fifteen milligrams of the protein slurry were packed into a 3.2 mm Varian rotor and sealed using an upper spacer and a top spinner.

For the CA spherical assemblies, twelve milligrams of the N_2-dried sample were packed into a 3.2 mm Varian rotor and sealed using an upper spacer and a top spinner.

For CA tubular assemblies, eighteen milligrams of the CA precipitate were packed into three 3.2 mm Bruker rotors, sealed using an upper spacer and a top spinner, and used in subsequent solid-state NMR experiments.

4.2.3 SSNMR Spectroscopy

Solid-state NMR spectra were acquired at 14.1 T (600 MHz) on a narrow bore Varian InfinityPlus instrument outfitted with a 3.2 mm triple resonance T3 probe; Larmor frequencies are 599.5 MHz for ^{1}H, 150.7 MHz for ^{13}C, and 60.7 MHz for ^{15}N. Dipolar-based two- and three-dimensional DARR, NCA, NCO, NCACB, NCACX, and NCOCX spectra were collected at the MAS frequency of 10.000±0.001 kHz regulated by a Varian MAS controller. The dipolar NCOCX and NCACX experiments utilized the SPECIFIC-CP heteronuclear mixing [5] followed by the DARR mixing sequence [6]; in the NCACB sequence, homonuclear DREAM mixing [7] with a linear ramp was used. The temperature was calibrated at different MAS frequencies using a $PbNO_3$ temperature sensor [8], and the actual temperature at the sample was maintained to within ±0.5 °C throughout the experiments using the Varian temperature controller. The DARR spectra were acquired at temperatures of -5.0 °C, 5.0 °C for spherical and conical assemblies respectively. The heteronuclear 2D and 3D spectra were acquired at 5 °C. ^{1}H, ^{13}C and ^{15}N chemical shifts were referenced with respect to DSS, adamantane and ammonium chloride as external referencing standards [9]. The typical pulse lengths in the double resonance mode were 2.6 μs (^{1}H), and 3.1 μs (^{13}C), and in the triple resonance mode 2.6 μs (^{1}H), 5 μs (^{13}C), and 5 μs (^{15}N). ^{1}H- ^{13}C(^{15}N) cross polarization was achieved with a linear amplitude ramp; the ^{1}H radio frequency field was 50 kHz with the center of the ramp Hartmann-Hahn matched to the first spinning sideband. The ^{1}H-^{13}C and ^{1}H-^{15}N

contact times were 900 μs and 1.1 ms for conical and spherical assemblies respectively. In the dipolar-based experiments, the DARR mixing time was 10 ms, the SPECIFIC-CP mixing time was 6.2 ms, and the DREAM mixing time was 3.0 ms. TPPM decoupling [10] was used, with the decoupling field strengths ranging between 80 and 100 kHz in the different experiments. Recycle delays in all experiments were 2 s. Time-proportional phase incrementation (TPPI) [11] was used for frequency discrimination in the indirect dimensions.

DARR spectra of the conical assemblies were also acquired on a Varian Inova 21.1 T spectrometer outfitted with a 3.2 mm BioMAS HXY probe. The Larmor frequencies were 900.575 MHz and 226.47 MHz for ^{1}H for ^{13}C, respectively. Spectra were collected at a MAS frequency of 14.000±0.001 kHz regulated by a Varian MAS controller, and the temperature was maintained at 15±0.5 °C throughout the experiments using a Varian temperature controller. The pulse lengths in the DARR experiments were 3.7 μs (^{1}H), and 6.3 μs (^{13}C), and ^{1}H-^{13}C cross polarization was performed with a tangent amplitude ramp (80-100%), having the ^{1}H radio frequency field of 68 kHz matched to the second spinning sideband. The ^{1}H-^{13}C contact time was 1 ms. TPPM decoupling (71 kHz) was applied during the evolution and acquisition periods. The ^{1}H field strength during DARR was 14 kHz, and the DARR mixing time was 50 ms. Under these conditions predominantly one-bond correlations are observed in the spectra.

Additional acquisition and processing parameters for data sets collected at 14.1 and 21.1 T are provided in the figure captions and in Table A.3 of the Appendix.

4.2.4 NMR data processing and analysis

All spectra were processed with NMRpipe [12] and analyzed using Sparky [13]. For 2D data sets, 90°, 60°, or 30° shifted sine bell apodization followed by a Lorentzian-to-Gaussian transformation was applied in both dimensions. Forward linear prediction to twice the number of the original data points was used in the indirect dimension in some data sets followed by zero filling to twice the total number of points. The 3D data sets were processed by using 30° or 60° shifted sine bells in all dimensions followed by a Lorentzian-to-Gaussian transformation (for resolution enhancement) and by using 90° degree shifted sinebells in all dimensions followed by a Lorentzian-to-Gaussian transformation (for sensitivity enhancement). Detailed information about processing parameters is provided in Table A.3 of the Appendix.

4.3 The effect of morphology on solid-state NMR spectra

The 2D ^{13}C-^{13}C DARR spectrum of CA conical assemblies recorded at 21.1 T is shown in Figure 4.1. The spectrum has remarkably high quality, and attested to the feasibility of detailed structural characterization of CA protein assemblies of conical morphology. From this spectrum, a number of important conclusions can be extracted even without the complete residue-specific resonance assignments. First, the linewidths of individual cross peaks lie in the range of 0.25-0.5 ppm, indicating that the sample is conformationally homogeneous. Second, despite the size of the protein, numerous well-resolved peaks are present. The overall number of the cross peaks indicates that most correlations are present in the spectrum. From this spectrum alone extracted the amino acid spin topology and counted most of the amino acids expected to be present in the resolved region of the spectrum: Ala (12/20), Gly (9/17), Ile (12/15), Pro (9/18), Thr

(13/16), Ser (7/9), Val (9/15). Qualitative analysis of the C^{α} and C^{β} chemical shifts of the Ala and Thr residues additionally reveals that the conformation of the protein is predominantly α-helical, consistent with the prior solution NMR, X-ray diffraction, and cryo-EM studies [14-16].

Figure 4.2 shows the superposition of the 21.1 T DARR spectra of spherical and conical CA assemblies. The spectral resolution is inferior for the spherical particles, and the linewidths are in the range of 0.4-0.8 ppm. Interestingly, the sensitivity appeared to be considerably lower than in the conical assemblies, with the signal-to-noise ratio in the one-dimensional CPMAS spectra only ca. 55% that of the conical samples. This noticeable deterioration in spectral quality may be due to dehydration of the spherical assemblies, which is necessary to prevent their conversion to the amorphous precipitate while keep high concentration of the materials.

As can be appreciated from the data presented in Figure 4.2, no significant differences in ^{13}C chemical shifts can be detected for the spherical and the conical assemblies, suggesting that the secondary and tertiary structure elements in the CA assemblies are the same, irrespective of their microscopic morphology.

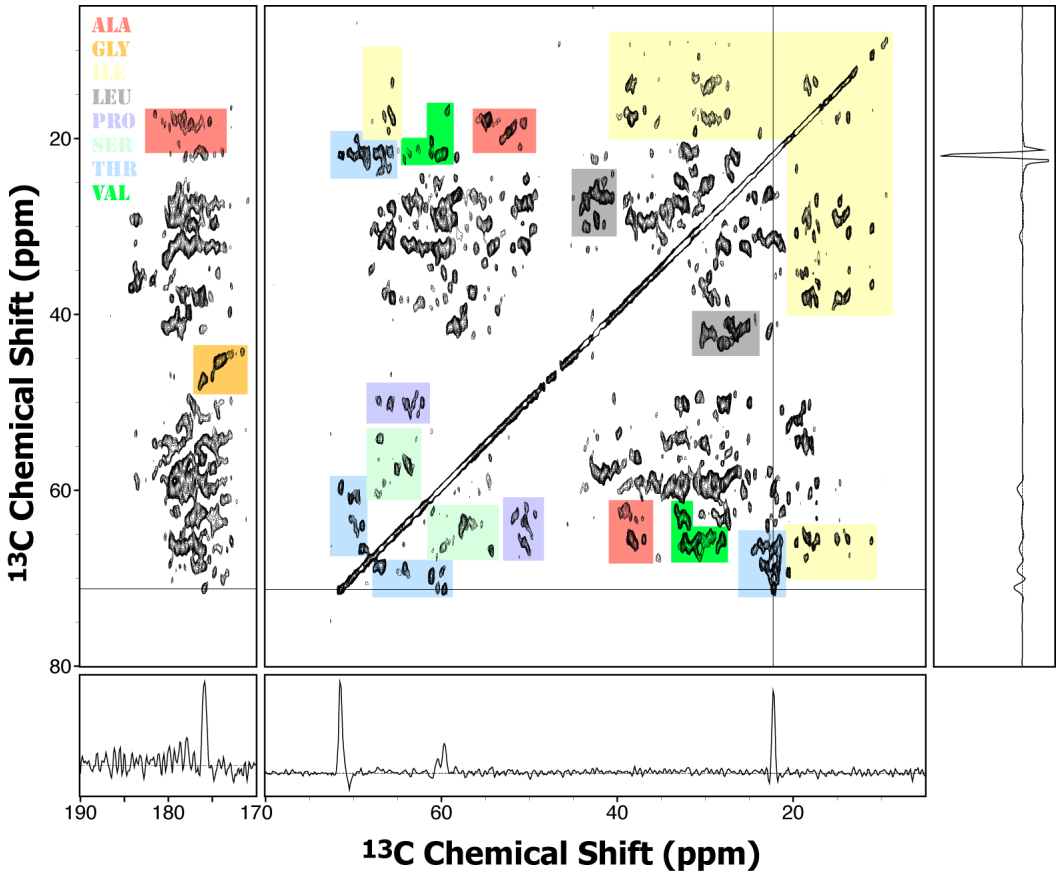


Figure 4.1 21.1 T DARR spectrum of U-^{13}C,^{15}N HIV-1 CA conical assemblies: aliphatic region (right) and carbonyl region (left). The 1D projections in the ω_2 and ω_1 dimensions corresponding to the resolved cross-peaks for Thr residues are shown and illustrate the remarkable spectral sensitivity and resolution. The spectrum was acquired as (3072 x 216) (complex x real) matrices with 220 scans; the recycle delay was 2 s. The spectral widths were 100 kHz and 50 kHz in the ω_2 and ω_1 dimensions, respectively, and the DARR mixing time was 50 ms. Cross-peaks corresponding to the different amino acid types appear in distinct, resolved regions of the spectrum and are color coded as follows: Ala, red (C^{α}-C^{β} and C^{β}-C' correlations); Gly, orange (C^{α}-C' correlations); Ile, yellow (C^{α}-$C^{\delta 1}$, C^{β}-$C^{\gamma 1}$, C^{β}-$C^{\gamma 2}$, and $C^{\gamma 1}$-$C^{\delta 1}$ correlations); Leu, grey (C^{β}-C^{γ} correlations); Pro, purple (C^{α}-C^{δ} correlations); Ser, light green (C^{α}-C^{β} correlations); Thr, blue (C^{α}-C^{β} and C^{β}-C^{γ} correlations); Val, bright green (C^{α}-C^{β} correlations).

Figure 4.2 shows the superposition of the 21.1 T DARR spectra of tubular and conical CA assemblies. The spectral resolution is superior for the tubular assemblies, and the linewidths are less than 0.3 ppm. In the Thr region, 9 out of 16 Thr peaks were resolved in the CA conical assemblies. While in the tubular assemblies, 14 out of 16 peaks were resolved.

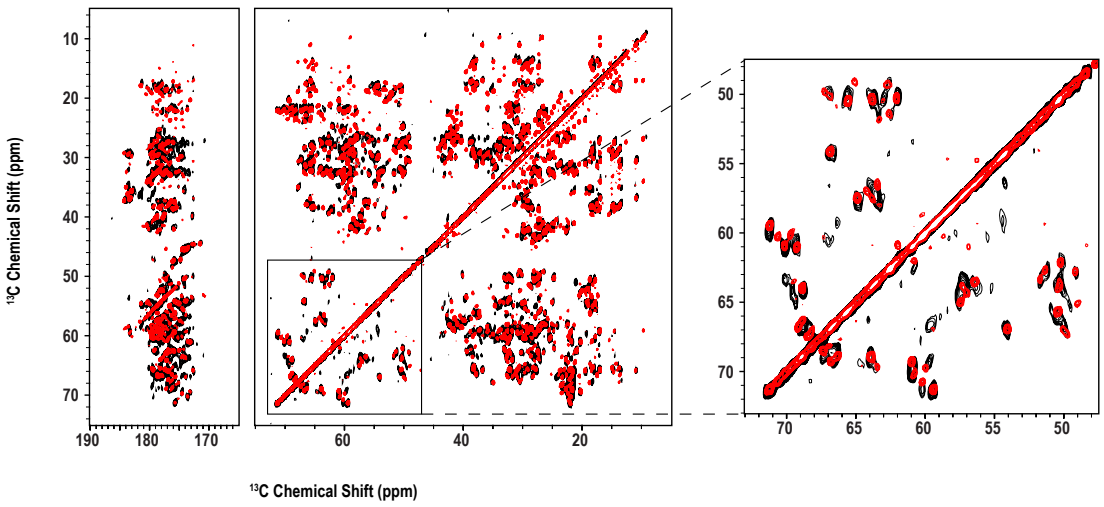


Figure 4.2 Superposition of DARR spectra recorded on tubular (red) and conical (black) assemblies of U-^{13}C,^{15}N CA at 21.1 T. The expansion of the Ser and Thr region shows more clearly about the enhanced resolution. The spectra of the tubular assemblies were acquired as (6250 x 480) (complex x real) matrices using 128 scans and a recycle delay of 2 s, a spectral width of 78 kHz in both dimensions and the DARR mixing time of 50 ms. The acquisition parameters for the spectra of the conical assemblies are given in the caption to Figure 4.1.

The spectra reveal that the ^{13}C chemical shifts for the tubular and the conical assemblies are consistent, suggesting that the secondary and tertiary elements in the CA assemblies are the same. It is not surprising that tubular CA assemblies exhibit higher spectral resolution: they are comprised of hexameric lattice while in conical assemblies, pentameric declinations are also present, causing conformational heterogeneity and line broadening.

4.4 Resonance assignments of conical CA assemblies

For resonance assignments of conical CA assemblies, we have recorded the following three-dimensional homo- and heteronuclear correlation MAS NMR spectra at 14.1 T: NCACB, NCACX, NCOCX and CCC (DARR-DARR). The pulse sequences for

these experiments are illustrated in Figure 4.3. Combining these three-dimensional data sets with the DARR spectra acquired at 21.1 T, we have confidently assigned 147 residues, about 63% of the entire protein, with an additional 21 residues tentatively assigned.

As illustrated in Figure 4.1, intra-residue side chain correlations were established for seven amino acid residue types (Ala, Gly, Ile, Pro, Ser, Thr, Val), since these appear in the unique regions of the spectra [17-20]. In addition, due to the unique position of Leu C^{β}-C^{γ} cross peaks in the spectrum, 12 of the expected 18 correlations corresponding to the Leu residues were identified. The two differentially mixed labeled samples, U-^{13}C,^{15}N-Ala/U-^{13}C,^{15}N-Thr and U-^{13}C,^{15}N-His/U-^{13}C,^{15}N-Tyr, were employed for unambiguous identification of the Ala and Thr (Figure 4.4b), and the His and Tyr (Figure 4.4c) cross peaks in the DARR spectra. For example, the DARR spectrum of the U-^{13}C,^{15}N-His/U-^{13}C,^{15}N-Tyr sample allowed the identification of the aliphatic and carbonyl cross peaks corresponding to 4 out of 6 His residues, and 3 out of 4 Tyr residues. In total, the 2D DARR spectrum yielded identification for the following amino acids: Ala (12/20), Gly (9/17), Ile (12/15), Pro (9/18), Thr (13/16), Ser (7/9), Val (9/15).

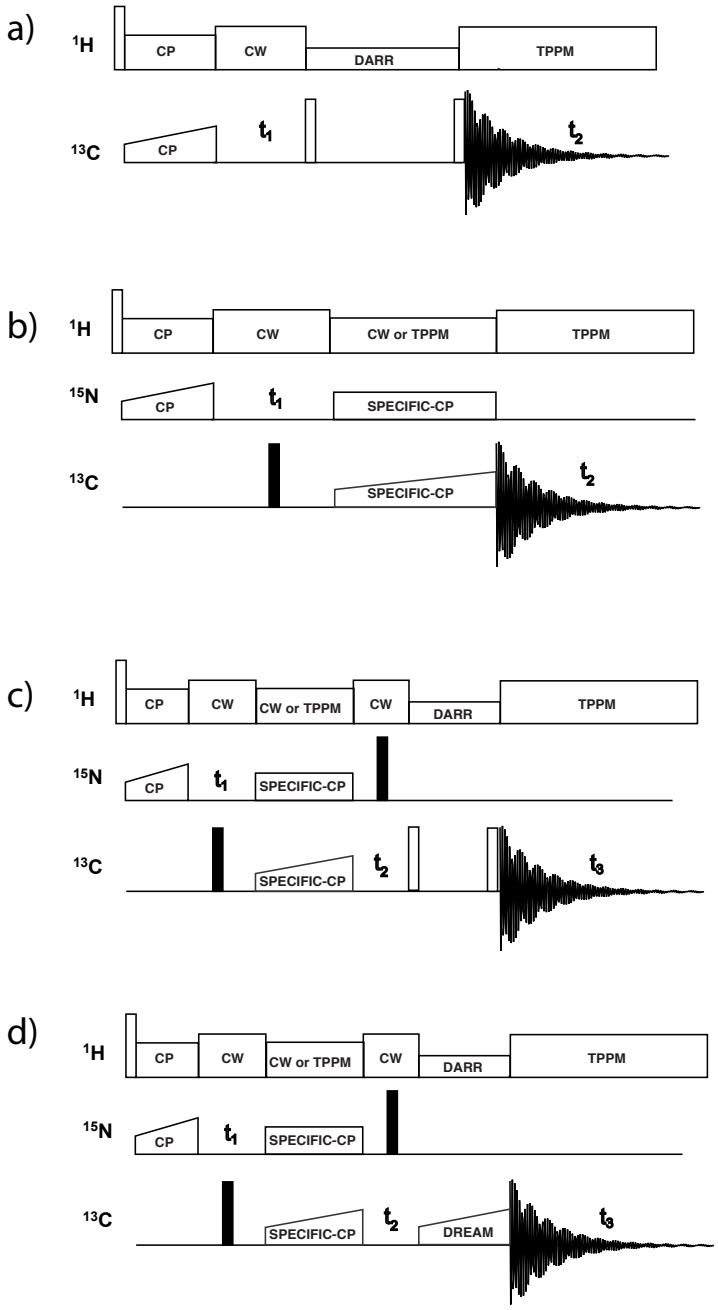


Figure 4.3 Pulse sequences for 2D and 3D experiments employed for resonance assignments of conical CA assemblies by MAS NMR spectroscopy: a) DARR [21] b) NCA/NCA; c) NCACX/NCOCX; d) NCACB. Empty vertical bars indicate π/2 and filled vertical bars indicate π pulses.

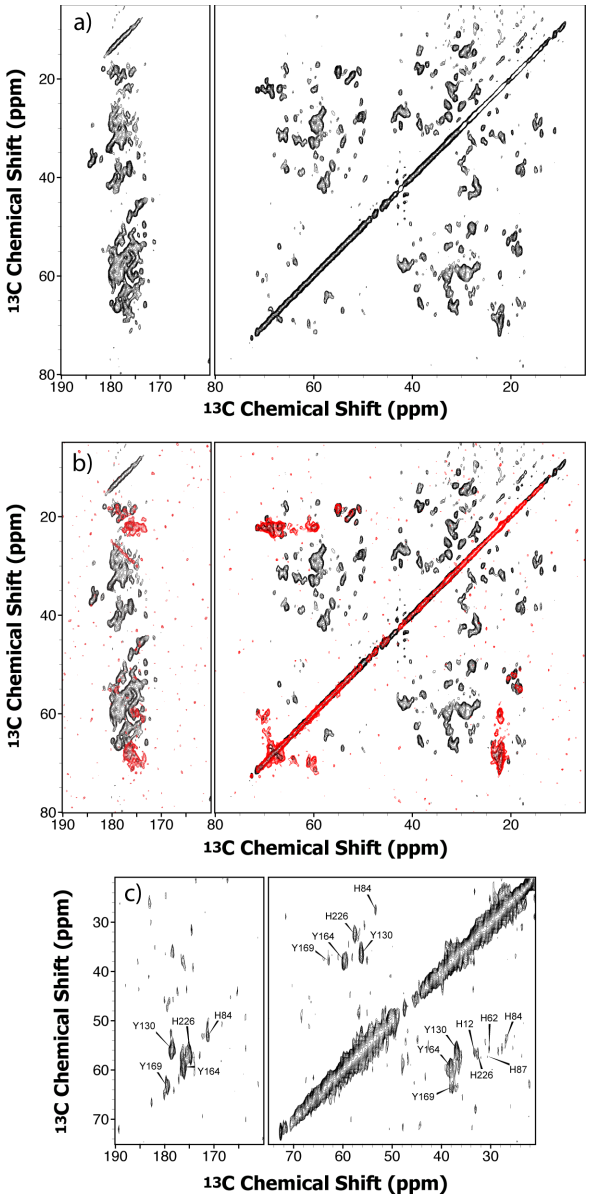


Figure 4.4 14.1 T DARR spectra of conical CA assemblies acquired on a set of uniformly and specifically labeled CA: (a) U-^{13}C,^{15}N; (b) U-^{13}C,^{15}N-Ala/U-^{13}C,^{15}N-Thr, and (c) U-^{13}C,^{15}N-His/U-^{13}C,^{15}N-Tyr. In c), expansions of the aliphatic (right) and carbonyl (left) regions illustrate the one-bond correlations for His and Tyr residues. The spectra in (a) were collected as (2000 x 400) (complex x real) matrices with 200 scans, a recycle delay of 2 s, and a spectral width of 50 kHz in both dimensions. The spectra in (b) were obtained as (2000 x 512) (complex x real) matrices with 256 scans, a recycle delay of 1 s, and spectral widths of 50 kHz and 25 kHz in the ω_2 and ω_1 dimensions, respectively. The spectra in (c) were acquired as (1000 x 128) (complex x real) matrices with 640 scans, and a recycle delay of 1.5 s and spectral widths of 50 kHz and 33.3 kHz in the w_2 and w_1 dimensions, respectively. The DARR mixing time was 10 ms for all data sets.

The above spin system assignments were confirmed in NCACX and NCACB 3D experiments, which remove the chemical shift degeneracy in the ^{13}C dimensions for a number of the above amino acid types. As illustrated in Figure 4.5, many of the 2D ^{15}N planes in the 3D NCACX spectrum exhibit excellent resolution, permitting multiple amino acid residue types to be assigned. In addition to the above listed spin systems, cross peaks corresponding to Cys (2/2), and to Asp (7/7) residues were identified in distinct, well-resolved regions of the spectra.

In total, using a combination of 2D DARR, 3D NCACX, and 3D NCACB experiments, we have identified 100 spin systems for 12 residue types. The chemical shifts of the assignments were shown in Table 4.1. And the differences between the solid-state and solution NMR chemical shifts were summarized in Table A.1.

Sequence-specific resonance assignments were afforded by the addition of 3D NCOCX spectra in the data analysis, given the excellent resolution shown in the 2D separated by ^{15}N chemical shifts (Figure 4.5). Based on the 3D solid-state NMR spectra alone, we have established sequence-specific assignments for 61 residues. Two representative backbone walks illustrating the assignments for the sequence stretch A105-L111 and L43-A47 is provided in Figure 4.6.

A number of residue types (Arg, Asn, Gln, Glu, Lys, Met, Phe, and Trp) present a major challenge with regard to assignment based on correlations in the 14.1 T solid-state 2D and 3D spectra alone: their C^{α}-C^{β} cross peaks do not appear in isolated regions and generally exhibit substantial spectral overlap. We therefore compared the 3D NCACB and NCACX solid-state data with the solution NMR chemical shifts of the individual N- and C- terminal CA domains. This is a reasonable approach since it has been

demonstrated by us and others for a number of proteins that solution and solid-state chemical shifts generally agree well, provided that the secondary structure (backbone torsion angles) for a given amino acid residue is the same in solution and in the solid [18-20,22-25].

Comparison of the firmly assigned solid-state and solution ^{13}C and ^{15}N chemical shifts by residue type (Ala, Asp, Cys, Gly, His, Ile, Leu, Pro, Ser, Thr, Tyr, and Val) revealed excellent agreement within 0.1, 0.4, and 0.3 ppm for NH, C^{α}, and C^{β}, respectively (Table A.1 in the Appendix). Based on this comparison, we derived tentative sequence-specific assignments for additional residues yielding the following degree of overall assignment: Arg (10/11), Asn (9/11), Gln (13/15), Glu (9/17), Lys (6/11), Met (5/11), Phe (4/4), Trp (5/5). Correlations for the remaining residues were either missing or resided in spectral regions too crowded for unambiguous identification. Based on the comparison between solution and solid-state chemical shifts, an additional 86 residues have been assigned, and tentative intra-residue assignments for 21 more amino acids are available. A summary of all chemical shifts is provided in Table 4.1.

Table 4.1 ^{13}C and ^{15}N Chemical Shift Assignments for the CA Protein in Conical Assemblies.

Residue	N	C^{α}	C^{β}	C'	C^{γ}
P1					
I2	121.3 ± 1	59.6 ± 0.0	37.8 ± 0.1	178.7 ± 0.2*	
V3					
Q4	122.4 ± 1	54.6 ± 0.2	29.6 ± 0.0		
N5	125.1 ± 1	53.5 ± 0.3	39.3 ± 0.4		
I6					
Q7					
G8	107.8 ± 1	44.9 ± 0.2			
Q9	120.5 ± 1	54.3 ± 0.2	29.8 ± 0.0	180.1 ± 0.0	
M10					
V11	120.2 ± 1	58.9 ± 0.1	35.8 ± 0.1		
H12		56.5 ± 0.0	33.3 ± 0.2	174.9 ± 0.2*	
Q13	125.3 ± 1	54.0 ± 0.1	32.8 ± 0.2	176.1 ± 0.3	
A14	125.3 ± 1	51.7 ± 0.1	18.9 ± 0.0	178.9 ± 0.2*	
I15	121.6 ± 1	61.5 ± 0.7	38.4 ± 0.2		
S16	123.4 ± 1	55.7 ± 0.0	63.5 ± 0.0		
P17					
R18	116.3 ± 1	59.2 ± 0.1	29.9 ± 0.1	179.8 ± 0.1	
T19	119.8 ± 1	66.3± 0.7	68.0 ± 0.1		
L20	121.1 ± 1	57.8 ± 0.0	41.2 ± 0.1	179.6 ± 0.2*	
N21	116.1 ± 1	56.1 ± 0.0	38.2 ± 0.1		
A22	122.4 ± 1	55.1 ± 0.1	19.3 ± 0.0	176.3 ± 0.4	
W23	120.2 ± 1	58.4 ± 0.0	30.2 ± 0.1	178.4 ± 0.8	
V24	118.0 ± 1	66.2 ± 0.4	31.6 ± 0.0		

K25	117.8 ± 1	59.4 ± 0.1	32.4 ± 0.2	179.4 ± 0.2*	
V26	119.4 ± 1	66.1 ± 0.1	31.4 ± 0.1		
V27	120.2 ± 1	65.7 ± 0.1	31.6 ± 0.1		
E28					
E29	115.9 ± 1	58.1 ± 0.0	31.0 ± 0.1	183.4 ± 0.3	
K30					
A31					
F32	113.4 ± 1	56.9 ± 0.1	37.6 ± 0.0	176.5 ± 0.2*	
S33					
P34		66.7 ± 0.2*	32.66 ± 0.2*	179.5 ± 0.2*	
E35	115.9 ± 1	59.5 ± 0.2	29.4 ± 0.6		
V36	117.0 ± 1	65.6 ± 0.1	31.9 ± 0.0	177.9 ± 0.1	
I37	119.2 ± 1	67.7 ± 0.0	34.9 ± 0.2*		
P38					
M39	116.2 ± 1	59.1 ± 0.4	32.3 ± 0.1	179.0 ± 0.2	
F40	120.7 ± 1	61.6 ± 0.1	38.5 ± 0.2	179.2 ± 0.1	
S41					
A42	122.7 ± 1	54.4 ± 0.7	18.2 ± 0.2		
L43	116.3 ± 1	56.3 ± 0.1	42.4 ± 0.1	176.5 ± 0.2*	
S44	109.6 ± 1	57.0 ± 0.1	63.6 ± 0.3	176.2 ± 0.2	
E45	125.3 ± 1	58.1 ± 0.3	29.1 ± 0.2	178.3 ± 0.2	
G46	115.6 ± 1	46.1 ± 0.1		174.6 ± 0.2	
A47	119.7 ± 1	53.0 ± 0.5	20.5 ± 0.3	178.8 ± 0.0	
T48	108.6 ± 1	59.6 ± 0.0	69.1 ± 0.1	174.8 ± 0.0	22.3 ± 0.2*
P49					
Q50	116.1 ± 1	59.8 ± 0.3	29.7 ± 0.1	179.9 ± 0.2*	
D51	120.4 ± 1	58.0 ± 0.2*	41.4 ± 0.2	178.8 ± 0.0	
L52	119.9 ± 1	57.8 ± 0.1	41.4 ± 0.4	179.7 ± 0.2*	
N53	118.1 ± 1	56.3 ± 0.0	37.3 ± 0.2		
T54	119.8 ± 1	67.0 ± 0.4	68.1 ± 0.1		

M55				
L56				
N57	116.2 ± 1	56.0 ± 0.2	38.3 ± 0.2	176.7 ± 0.0
T58	111.3 ± 1	63.9 ± 0.5	68.9 ± 0.3	176.9 ± 0.1
V59				
G60	113.8 ± 1	45.4 ± 0.1		
G61	108.9 ± 1	45.3 ± 0.2		
H62		56.2 ± 0.1	28.5 ± 0.0	
Q63	119.6 ± 1	59.8 ± 0.2*	28.3 ± 0.2*	
A64	121.8 ± 1	54.6 ± 0.2	18.3 ± 0.2	181.7 ± 0.2*
A65				
M66				
Q67				
M68				
L69	122.9 ± 1	58.2 ± 0.1	41.3 ± 0.1	178.8 ± 0.2*
K70	119.2 ± 1	60.0 ± 0.1	32.6 ± 0.3	
E71				
T72				
I73	122.0 ± 1	66.4 ± 0.3	37.8 ± 0.1	
N74				
E75				
E76	120.9 ± 1	57.8 ± 0.2*	26.8 ± 0.5	
A77	124.8 ± 1	55.0 ± 0.4	17.8 ± 0.2	179.0 ± 0.2*
A78	119.7 ± 1	54.2 ± 0.1	18.0 ± 0.3	180.1 ± 1.8
E79				
W80	121.0 ± 1	61.0 ± 0.2*	27.5 ± 0.1	
D81	119.4 ± 1	57.0 ± 0.2*	39.7 ± 0.1	
R82	120.8 ± 1	59.3 ± 0.2*	29.3 ± 0.2*	178.7 ± 0.2*
V83				
H84	117.0 ± 1	53.5 ± 0.4	26.7 ± 0.2	171.3 ± 0.1

P85					
V86					
H87	125.4 ± 1	55.4 ± 0.3	30.0 ± 0.3		
A88					
G89					
P90					
I91	122.0 ± 1	59.7 ± 0.3	37.9 ± 0.5		
A92	131.4 ± 1	50.2 ± 0.2*			
P93		61.8 ± 0.2	32.7 ± 0.9	178.8 ± 0.0	
G94	112.2 ± 1	45.4 ± 0.1		172.0 ± 0.1	
Q95	119.6 ± 1	54.1 ± 0.2	29.8 ± 0.4		
M96	122.2 ± 1	54.8 ± 0.0	34.4 ± 0.1		
R97	123.5 ± 1	54.8 ± 0.1	29.7 ± 0.2*		
E98	123.5 ± 1	54.8 ± 0.1	29.7 ± 0.2*	179.2 ± 0.2*	
P99					
R100	121.9 ± 1	54.6 ± 0.6	30.7 ± 0.1		
G101	117.0 ± 1	48.8 ± 0.4		176.4 ± 0.0	
S102					
D103	123.0 ± 1	57.1 ± 0.2	41.3 ± 0.2	180.1 ± 0.0	
I104	126.0 ± 1	65.5 ± 0.0	36.6 ± 0.4		
A105	112.0 ± 1	50.2 ± 0.6	18.6 ± 0.5	178.0 ± 0.1	
G106	102.1 ± 1	45.7 ± 0.3		173.8 ± 0.0	
T107	113.1 ± 1	64.0 ± 0.4	68.6 ± 0.3	177.4 ± 0.0	19.8 ± 0.2*
T108	106.7 ± 1	60.8 ± 0.2	69.4 ± 0.2	173.2 ± 0.2	
S109	112.0 ± 1	54.3 ± 0.5	66.7 ± 0.2	174.8 ± 0.0	
T110	112.4 ± 1	59.9 ± 0.4	71.3 ± 0.4	176.0 ± 0.3	22.5 ± 0.2
L111	122.1 ± 1	58.0 ± 0.2	41.3 ± 0.3		
Q112	115.7 ± 1	59.4 ± 0.2	27.8 ± 0.2		
E113					
Q114	118.0 ± 1	59.1 ± 0.4	27.7 ± 0.2		

I115	118.1 ± 1	65.7 ± 0.1	38.3 ± 0.3	180.2 ± 0.1	
G116	111.4 ± 1	47.4 ± 0.2		175.4 ± 0.4	
W117	121.7 ± 1	61.3 ± 0.8	27.7 ± 0.1	177.5 ± 0.0	
M118	116.2 ± 1	59.5 ± 0.4	35.2 ± 0.3	177.9 ± 0.2	
T119	104.4 ± 1	60.5 ± 0.1	69.8 ± 0.0	173.3 ± 0.2*	
N120	123.4± 1	58.4 ± 0.2	29.0 ± 0.0		
N121					
P122					
P123					
I124	122.9 ± 1	59.2 ± 0.0	38.0 ± 0.1		
P125		62.6 ± 0.2*	28.33 ± 0.2*	179.5 ± 0.2*	
V126	115.8 ± 1	66.3 ± 0.5	30.1 ± 0.1	177.2 ± 0.2*	
G127	106.3 ± 1	47.4 ± 0.3		175.2 ± 0.2	
E128					
I129	121.4 ± 1	64.7 ± 0.0	38.3 ± 0.3	178.9 ± 0.2	
Y130	117.0 ± 0.6	56.0 ± 0.2	37.1 ± 0.3	178.5 ± 0.2*	
K131					
R132	117.9 ± 1	60.1 ± 0.0	29.2 ± 0.1		
W133	122.1 ± 1	58.6 ± 0.2*	29.1 ± 0.1		
I134	119.2 ± 1	65.6 ± 0.3	38.2 ± 0.0	179.0 ± 0.2*	
I135	121.3 ± 1	66.4 ± 0.2*	37.9 ± 0.3		
L136	122.6 ± 1	57.9 ± 0.2*	41.6 ± 0.1	181.1 ± 0.2*	
G137	108.1 ± 1	47.4 ± 0.2		175.5 ± 0.2*	
L138	122.9 ± 1	57.9 ± 0.2*	42.3 ± 0.0	179.5 ± 0.2*	27.2 ± 0.0
N139	117.7 ± 1	57.1 ± 0.2*	38.9 ± 0.2	178.6 ± 0.2*	
K140	119.4 ± 1	59.6 ± 0.2*	32.7 ± 0.1	179.1 ± 0.2*	
I141	120.3 ± 1	65.2 ± 0.2	38.1 ± 0.2	179.6 ± 0.4	
V142	118.6 ± 1	66.1 ± 0.2*	31.4 ± 0.1		
R143	117.9 ± 1	58.8 ± 0.0	29.8 ± 0.0		
M144					

Y145**	115.1 ± 0.5	59.2 ± 0.1			
S146	121.2 ± 1	55.8 ± 0.1	63.7 ± 0.1		
P147					
T148					
S149					
I150	126.0 ± 1	62.1 ± 0.3	39.0 ± 0.0	178.9 ± 0.3	
L151	118.9 ± 1	56.6 ± 0.1	39.7 ± 0.1	177.8 ± 0.2*	
D152	114.4 ± 1	54.6 ± 0.6	41.6 ± 0.0		
I153	122.0 ± 1	60.9 ± 0.3	32.5 ± 0.2		
R154	125.3 ± 1	54.2 ± 0.3	33.0 ± 0.3	175.9 ± 0.1	
Q155	127.8 ± 1	55.5 ± 0.1	26.6 ± 0.6		32.9 ± 0.2*
G156	116.9 ± 1	45.1 ± 0.1		173.3 ± 0.2	
P157	130.5 ± 1	65.8 ± 0.2*	32.7 ± 0.2*	177.4 ± 0.2*	27.8 ± 0.2*
K158	114.9 ± 1	53.7 ± 0.2	32.6 ± 0.2	175.8 ± 0.2*	
E159	125.7 ± 1	53.6 ± 0.2	31.8 ± 0.6		
P160					
F161	128.1 ± 1	62.7 ± 0.1	39.2 ± 0.2	177.2 ± 0.2	
R162	115.7 ± 1	59.4 ± 0.1	29.7 ± 0.2	179.3 ± 0.2*	
D163	118.3 ± 1	56.8 ± 0.2*	39.5 ± 0.1		
Y164	124.6 ± 0.6	59.4 ± 0.2	38.1 ± 0.1	176.1 ± 0.2	
V165	120.2 ± 1	66.6 ± 0.2			
D166	118.3 ± 1	57.5 ± 0.2*		178.0 ± 0.2	
R167	119.1 ± 1	59.3 ± 0.2*	29.6 ± 0.2		
F168	124.2 ± 1	62.1 ± 0.3	38.8 ± 0.2	178.7 ± 0.2*	
Y169	117.2 ± 1	63.3 ± 0.8	38.1 ± 0.3	179.3 ± 0.1	
K170	120.7 ± 1	60.3 ± 0.3	32.4 ± 0.2		
T171	117.5 ± 1	66.7 ± 0.2	68.3 ± 0.1	177.0 ± 0.1	22.1 ± 0.2*
L172	124.1 ± 1	58.5 ± 0.2	41.4 ± 0.2	178.3 ± 0.5	28.9 ± 0.2*
R173	116.3 ± 1	60.3 ± 0.2*			
A174	118.6 ± 1	52.0 ± 0.3	19.3 ± 0.1	176.2	

E175					
Q176	123.9 ± 1	54.3 ± 0.4	27.5 ± 0.2		
A177	126.6 ± 1	51.4 ± 0.1	21.3 ± 0.3	176.3 ± 0.3	
S178	117.5 ± 1	57.6 ± 0.6	65.1 ± 0.5	175.3 ± 0.2*	
Q179	121.3 ± 1	58.5 ± 0.3	28.4 ± 0.0		
E180					
V181	121.2 ± 1	66.4 ± 0.3	31.0 ± 0.1		
K182					
N183	119.1 ± 1	55.8 ± 0.2	37.2 ± 0.0		
W184	122.9 ± 1	61.3 ± 0.1	27.5 ± 0.7		
M185					
T186	112.3 ± 1	66.3 ± 0.3	68.8 ± 0.4	176.5 ± 0.2*	21.9 ± 0.2*
E187	118.7 ± 1	58.5 ± 0.2*	30.2 ± 0.2*		
T188	111.8 ± 1	64.0 ± 0.1	68.7 ± 0.2	177.2 ± 0.5	
L189	121.0 ± 1	57.8 ± 0.1	41.4 ± 0.2		
L190					
V191	115.8 ± 1	66.0 ± 0.5	31.9 ± 0.2		
Q192	116.7 ± 1	59.1 ± 0.7	27.7 ± 0.0		
N193	114.8 ± 1	52.8 ± 0.2	38.2 ± 0.3	175.7 ± 0.2*	
A194	124.6 ± 1	52.0 ± 0.2	19.0 ± 0.2	176.3 ± 0.3	
N195	119.2 ± 1	51.4 ± 0.2	35.1 ± 0.2		
P196		66.9 ± 0.3	32.6 ± 0.3	179.7 ± 0.3	28.3 ± 0.2*
D197	115.4 ± 1	56.8 ± 0.2	40.6 ± 0.3		
C198	116.1 ± 1	63.8 ± 0.1	28.1 ± 0.3	177.2 ± 0.1	
K199	121.8 ± 1	60.0 ± 0.4	32.3 ± 0.5	178.6 ± 0.1	
T200	113.1 ± 1	66.8 ± 0.3	68.9 ± 0.2	176.8 ± 0.8	22.1 ± 0.1
I201	122.8 ± 1	65.0 ± 0.1	38.1 ± 0.1	179.4 ± 0.2*	
L202	118.9 ± 1	57.5 ± 0.3	41.1 ± 0.3		
K203					
A204					

L205	118.3 ± 1	56.4 ± 0.5	42.6 ± 0.0	179.3 ± 0.2*	26.6 ± 0.2*
G206	104.3 ± 1	44.5 ± 0.6		172.6 ± 0.1	
P207					
A208	121.5 ± 1	52.5 ± 0.1	20.0 ± 0.2	178.1 ± 0.1	
A209	122.8 ± 1	52.7 ± 0.3	20.3 ± 0.2	178.7 ± 0.1	
T210	113.2 ± 1	59.6 ± 0.4	71.1 ± 0.3	175.7 ± 0.4	22.3 ± 0.2*
L211	122.6 ± 1	58.1 ± 0.0	41.3 ± 0.1	179.7 ± 0.2	
E212					
E213					
M214	120.0 ± 1	60.1 ± 0.1	32.7 ± 0.0		
M215	116.9 ± 1	59.2 ± 0.2	32.4 ± 0.0		
T216	117.0 ± 1	66.5 ± 0.3	68.2 ± 0.2	176.9 ± 0.5	21.9 ± 0.1
A217	122.8 ± 1	52.8 ± 0.4		177.8 ± 0.1	
C218	110.2 ± 1	59.8 ± 0.1	28.6 ± 0.3	175.6 ± 0.2*	
Q219	122.4 ± 1	59.1 ± 0.8	28.3 ± 0.1		
G220	109.9 ± 1	45.3 ± 0.3			
V221					
G222					
G223	108.1 ± 1	44.3 ± 0.2			
P224					
G225	109.1 ± 1	47.2 ± 0.0		175.0 ± 0.1	
H226					
K227					
A228	125.4 ± 1	52.8 ± 0.6	19.8 ± 0.2		
R229					
V230	122.9 ± 1	61.0 ± 0.2	32.5 ± 0.2		
L231					

* For these residues, chemical shifts were extracted from a single data set, and the errors reflect average errors across all residues for a given atom type.

** For Y145, ^{13}C and ^{15}N chemical shifts exhibit substantial temperature dependences. The chemical shifts reported here were extracted from a 2D NCA spectrum of U-^{13}C,^{15}N-Tyr-labeled CA (conical assembly) acquired at -15 °C.
The errors in the ^{15}N dimension are determined by the digital resolution of the 3D data sets.

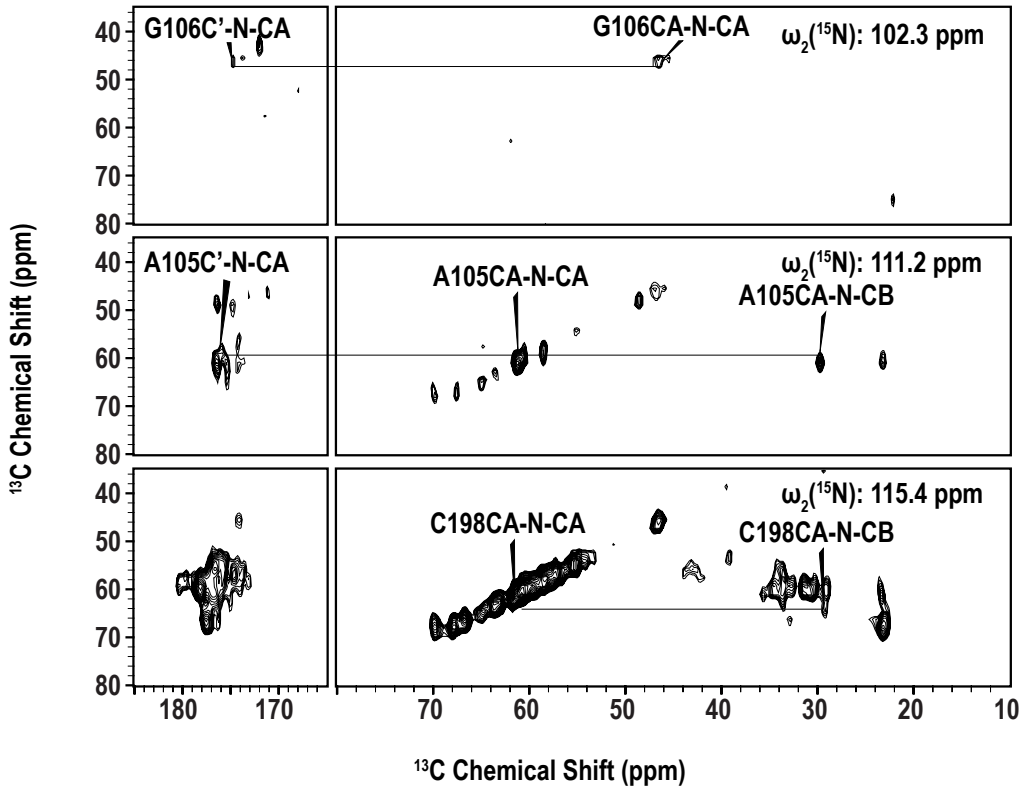
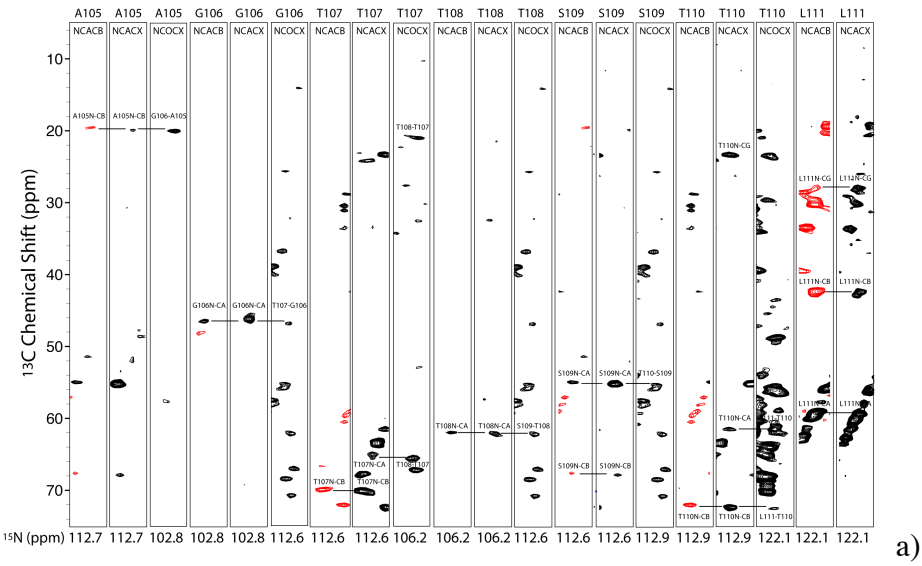


Figure 4.5 Selected 2D planes of the 3D NCACX spectrum of U-^{13}C,^{15}N CA conical assemblies acquired at 14.1 T. Spectral resolution in these planes is high enough to allow assignment of spin systems. The spectrum was acquired as (1000 x 40 x 32) (complex x real x real) matrices with 256 scans and a recycle delay of 2 s. The spectral widths were 50 kHz, 4.167 kHz, and 8.333 kHz in the ω_3, ω_2 and ω_1 dimensions, respectively. SPECIFIC-CP (t_{mix} = 6.2 ms) was employed for the first NCA transfer followed by the DARR mixing (t_{mix} = 20 ms) resulting in predominantly one- and two- bond correlations.



a)

Figure 4.6a Sequential backbone connectivity for the sequence stretch A105-L111 in conical assemblies based on 3D NCOCX, NCACX, and NCACB experiments acquired at 14.1 T. The residue names are shown on top of the spectra at their ^{15}N chemical shift plane. Negative cross-peaks resulting from two-bond N-C^{β} correlations in the NCACB spectra are displayed in red. The NCOCX spectrum was acquired as (1000 x 36 x 40) (complex x real x real) matrices with 280 scans and a recycle delay of 2 s. The spectral widths were 50 kHz, 4.167 kHz, and 6.250 kHz in the ω_3, ω_2, and ω_1 dimensions, respectively. SPECIFIC-CP (t_{mix} = 6.2 ms) was employed for the first NCO transfer followed by the DARR mixing (t_{mix} = 15 ms) resulting in predominantly one- and two-bond correlations. The NCACB spectrum was acquired as (1000 x 32 x 32) (complex x real x real) matrices with 400 scans and a recycle delay of 2 s. The spectral widths were 50 kHz, 4.167 kHz, and 8.333 kHz in the ω_3, ω_2, and ω_1 dimensions, respectively. SPECIFIC-CP (t_{mix} = 6.2 ms) was employed for the first NCA transfer followed by the DREAM mixing (t_{mix} = 3 ms) to establish the C^{α}-C^{β} correlations.

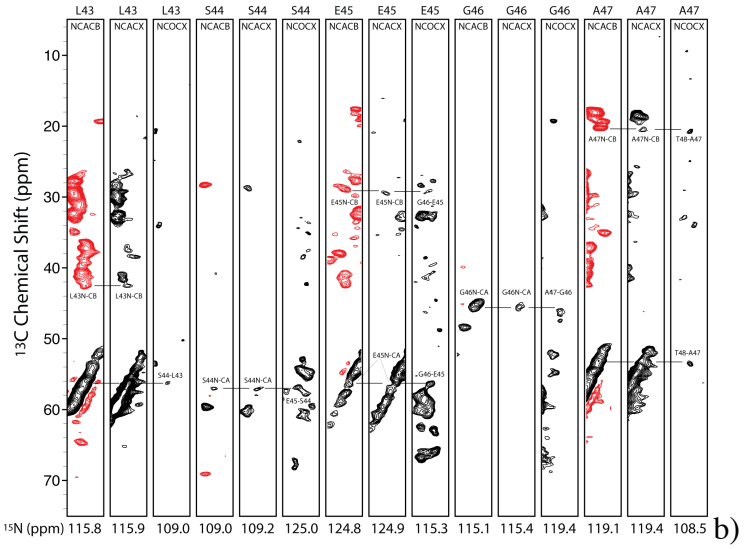


Figure 4.6b Sequential backbone connectivity for the sequence stretch L43-A47 (b) in conical assemblies based on 3D NCOCX, NCACX, and NCACB experiments acquired at 14.1 T. The residue names are shown on top of the spectra at their ^{15}N chemical shift plane. Negative cross-peaks resulting from two-bond N-C^{β} correlations in the NCACB spectra are displayed in red. The NCOCX spectrum was acquired as (1000 x 36 x 40) (complex x real x real) matrices with 280 scans and a recycle delay of 2 s. The spectral widths were 50 kHz, 4.167 kHz, and 6.250 kHz in the ω_3, ω_2, and ω_1 dimensions, respectively. SPECIFIC-CP (t_{mix} = 6.2 ms) was employed for the first NCO transfer followed by the DARR mixing (t_{mix} = 15 ms) resulting in predominantly one- and two-bond correlations. The NCACB spectrum was acquired as (1000 x 32 x 32) (complex x real x real) matrices with 400 scans and a recycle delay of 2 s. The spectral widths were 50 kHz, 4.167 kHz, and 8.333 kHz in the ω_3, ω_2, and ω_1 dimensions, respectively. SPECIFIC-CP (t_{mix} = 6.2 ms) was employed for the first NCA transfer followed by the DREAM mixing (t_{mix} = 3 ms) to establish the C^{α}-C^{β} correlations.

To complete the assignments of conical assemblies, we need to acquire additional data sets, at magnetic fields higher than 14.1 T. This work is ongoing in our laboratory, using improved sample preparation protocols permitting assembly in the absence of PEG. In parallel, we focused our efforts on investigation of tubular assemblies, which give much higher resolution spectra. These studies are discussed in Chapter 6.

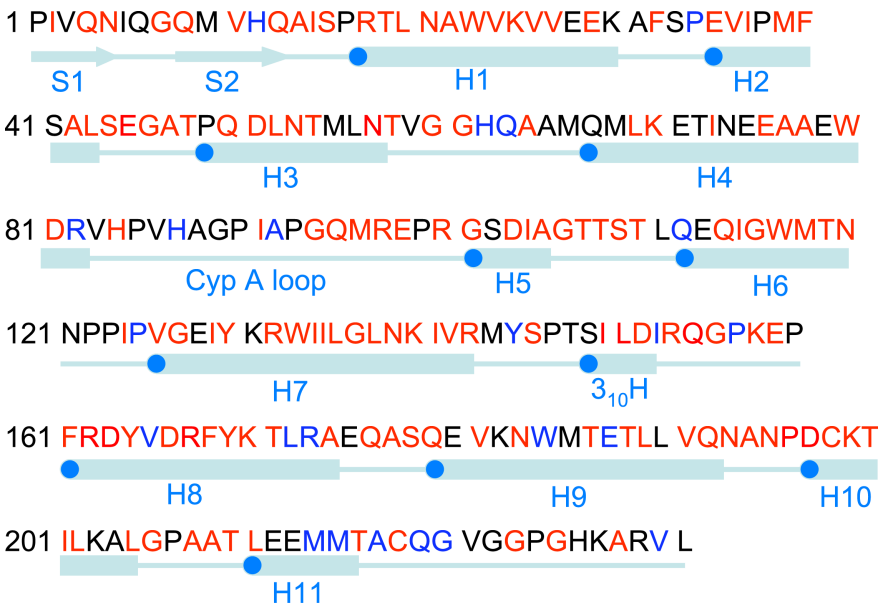


Figure 4.7 Summary of the resonance assignments of CA conical assemblies. The assigned residues are shown in red; the tentatively assigned residues are shown in blue; and the unassigned residues are shown in black.

4.5 Secondary structure analysis (CSI)

For the assigned regions of CA, secondary structure analysis was carried out using the Chemical Shift Index (CSI) based on C^{α} chemical shifts [26,27], using sequence-corrected random coil values for all amino acids [28]. The CSI-derived secondary structural elements for CA were compared between the solid- and solution states (Figure 4.8). The data clearly indicate that the secondary structures derived from solid-state and solution NMR CSI analysis are consistent with the location of helices in the known 3D structure of the individual N- and C-terminal domains [29-32].

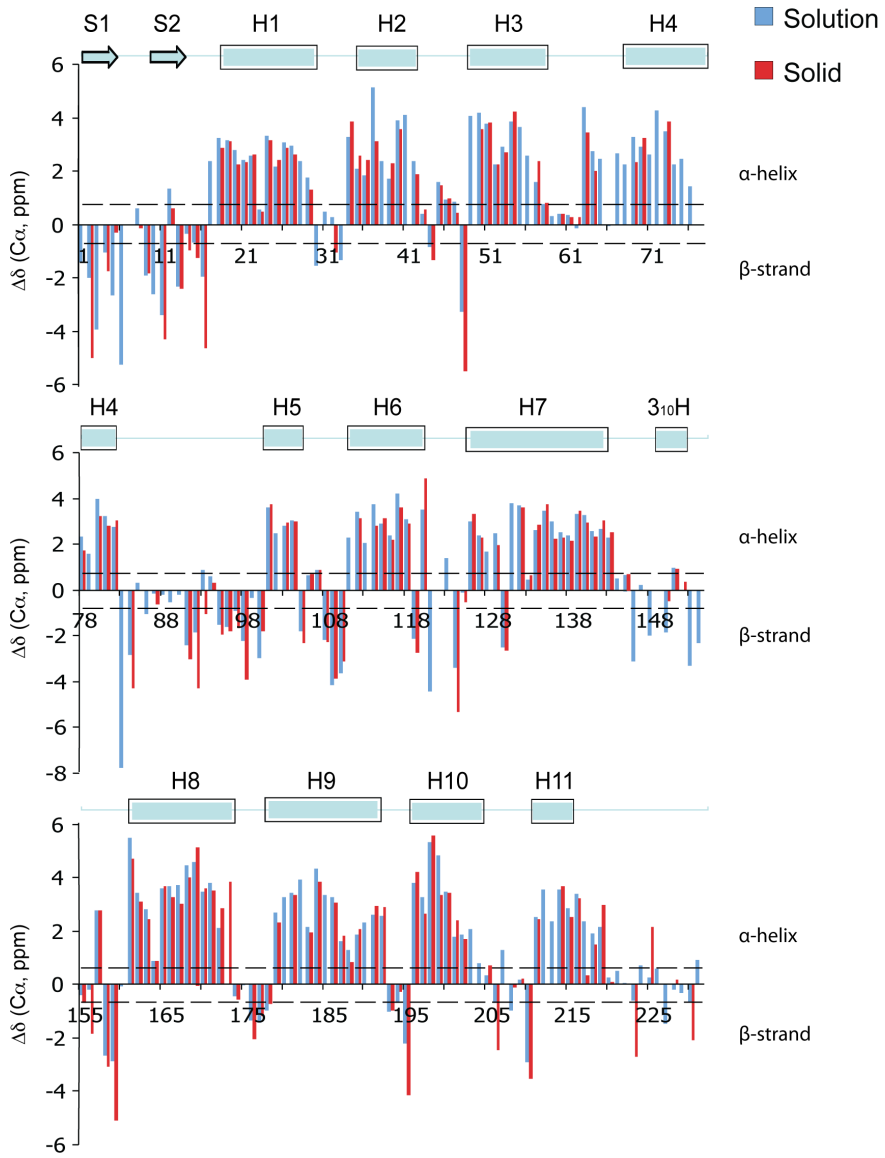


Figure 4.8 Comparison between solid-state and solution secondary chemical shifts for CA conical assemblies. C^{α} shifts that reside in a-helical regions of the protein exhibit positive secondary shifts, while those located in the b-strands and other secondary structure elements are negative.

In addition to C^{α} chemical shifts, the CSI method provides C' chemical shifts as another predictor of the protein secondary structures. For the CA conical assemblies, the results from C' chemical shift analysis as shown in Figure 4.9; these are consistent with the structure derived from other methods and the C^{α} CSI analysis.

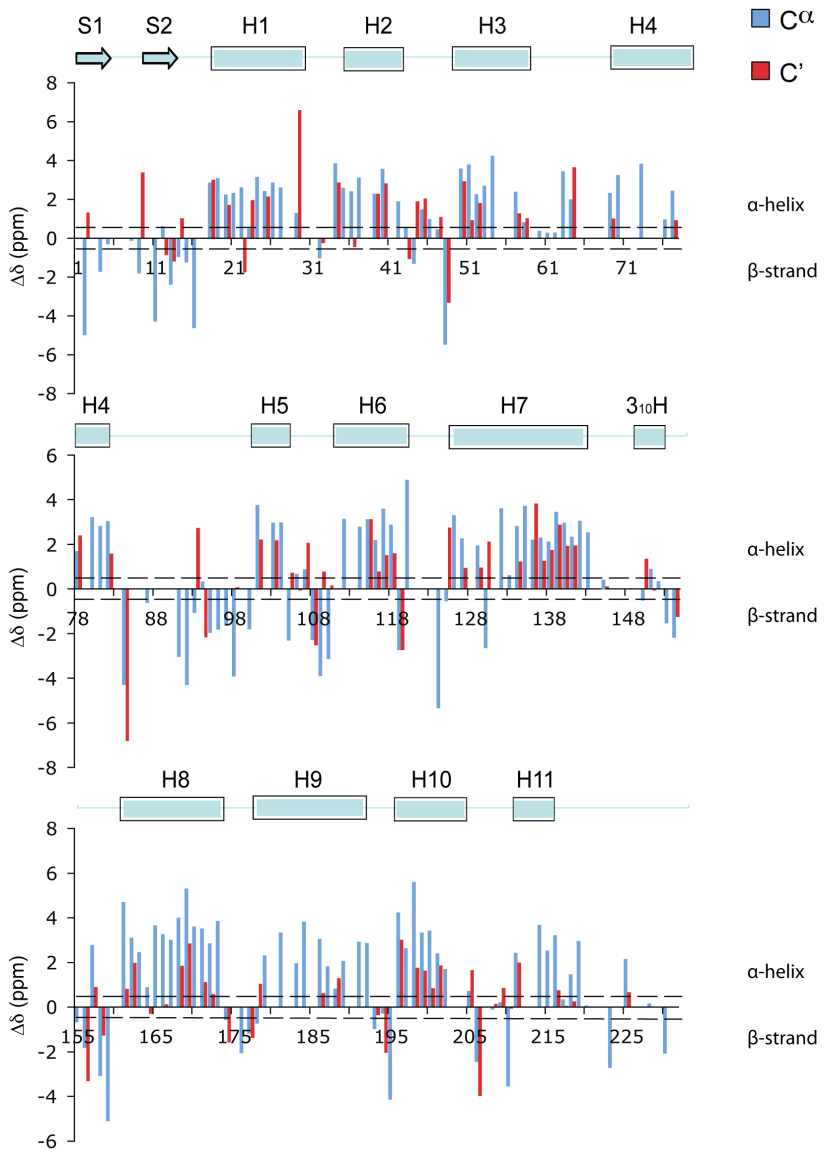


Figure 4.9 Comparison between C^{α} and C' secondary chemical shifts for CA conical assemblies. C^{α} and C' shifts that reside in α-helical regions of the protein exhibit positive secondary shifts, while those located in the β-strands and other secondary structure elements are negative.

References

[1] M. Stremlau; C. M. Owens; M. J. Perron; M. Kiessling; P. Autissier; J. Sodroski. The cytoplasmic body component TRIM5 alpha restricts HIV-1 infection in Old World monkeys. *Nature,* 427(6977): 848-853, 2004.

[2] O. Pornillos; B. K. Ganser-Pornillos; B. N. Kelly; Y. Z. Hua; F. G. Whitby; C. D. Stout; W. I. Sundquist; C. P. Hill; M. Yeager. X-Ray Structures of the Hexameric Building Block of the HIV Capsid. *Cell,* 137(7): 1282-1292, 2009.

[3] I. J. Byeon; G. Hou; Y. Han; C. L. Suiter; J. Ahn; J. Jung; C. H. Byeon; A. M. Gronenborn; T. Polenova. 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(14): 6455-6466, 2012.

[4] I. J. Byeon; X. Meng; J. Jung; G. Zhao; R. Yang; J. Ahn; J. Shi; J. Concel; C. Aiken; P. Zhang; A. M. Gronenborn. Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell,* 139(4): 780-790, 2009.

[5] M. Baldus; A. T. Petkova; J. Herzfeld; R. G. Griffin. Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems. *Mol. Phys.,* 95(6): 1197-1207, 1998.

[6] K. Takegoshi; S. Nakamura; T. Terao. ^{13}C-^{1}H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett.,* 344631-637, 2001.

[7] R. Verel; M. Baldus; M. Ernst; B. H. Meier. A Homonuclear spin-pair filter for solid-state NMR based on adiabatic passage techniques. *Chem. Phys. Lett.,* 287421-428, 1998.

[8] G. Neue; C. Dybowski. Determining temperature in a magic-angle spinning probe using the temperature dependence of the isotropic chemical shift of lead nitrate. *Solid State Nucl. Magn. Reson.,* 7(4): 333-336, 1997.

[9] C. R. Morcombe; K. W. Zilm. Chemical shift referencing in MAS solid state NMR. *J. Magn. Reson.,* 162(2): 479-486, 2003.

[10] A. E. Bennett; C. M. Rienstra; M. Auger; K. V. Lakshmi; R. G. Griffin. Heteronuclear decoupling in rotating solids. *J. Chem. Phys.,* 103(16): 6951-6958, 1995.

[11] D. Marion; K. Wüthrich. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of H1-H1 spin-spin coupling-constants in proteins. *Biochem. Biophys. Res. Commun.,* 113(3): 967-974, 1983.

[12] F. Delaglio; S. Grzesiek; G. W. Vuister; G. Zhu; J. Pfeifer; A. Bax. NMRPipe- a multidimensional spectral processing system based on Unix pipes. *J. Biomol. NMR,* 6(3): 277-293, 1995.

[13] T. D. Goddard; D. G. Kneller. SPARKY 3, University of California, San Francisco.

[14] C. Tang; Y. Ndassa; M. F. Summers. Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. *Nat. Struct. Biol.,* 9(7): 537-543, 2002.

[15] D. K. Worthylake; H. Wang; S. H. Yoo; W. I. Sundquist; C. P. Hill. Structures of the HIV-1 capsid protein dimerization domain at 2.6 angstrom resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.,* 5585-92, 1999.

[16] B. K. Ganser-Pornillos; A. Cheng; M. Yeager. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell,* 131(1): 70-79, 2007.

[17] T. I. Igumenova; A. E. Mcdermott; K. W. Zilm; R. W. Martin; E. K. Paulson; A. J. Wand. Assignments of carbon NMR resonances for microcrystalline ubiquitin. *J. Am. Chem. Soc.,* 126(21): 6720-6727, 2004.

[18] D. Marulanda; M. L. Tasayco; M. Cataldi; V. Arriaran; T. Polenova. Resonance assignments and secondary structure analysis of *E.coli* thioredoxin by magic angle spinning solid-state NMR spectroscopy. *J. Phys. Chem. B,* 109(38): 18135-18145, 2005.

[19] D. Marulanda; M. L. Tasayco; A. Mcdermott; M. Cataldi; V. Arriaran; T. Polenova. Magic angle spinning solid-state NMR spectroscopy for structural studies of protein interfaces. Resonance assignments of differentially enriched *Escherichia coli* thioredoxin reassembled by fragment complementation. *J. Am. Chem. Soc.,* 126(50): 16608-16620, 2004.

[20] A. Mcdermott; T. Polenova; A. Bockmann; K. W. Zilm; E. K. Paulsen; R. W. Martin; G. T. Montelione. Partial NMR assignments for uniformly (C-13, N-15)-enriched BPTI in the solid state. *J. Biomol. NMR,* 16(3): 209-219, 2000.

[21] K. Takegoshi; S. Nakamura; T. Terao. C-13-H-1 dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett.,* 344(5-6): 631-637, 2001.

[22] J. Pauli; M. Baldus; B. Van Rossum; H. De Groot; H. Oschkinat. Backbone and side-chain C-13 and N-15 signal assignments of the alpha-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 tesla. *ChemBioChem,* 2(4): 272-281, 2001.

[23] J. Yang; S. Paramasivan; D. Marulanda; M. Cataidi; M. L. Tasayco; T. Polenova. Magic angle spinning NMR spectroscopy of thioredoxin reassemblies. *Magn. Reson. Chem.,* 45(S1): S73-S83, 2007.

[24] S. Sun; A. Siglin; J. C. Williams; T. Polenova. Solid-state and solution NMR studies of CAP-Gly domain of mammalian dynactin and its interaction with microtubules. *J. Am. Chem. Soc.,* 13110113-10126, 2009.

[25] W. T. Franks; D. H. Zhou; B. J. Wylie; B. G. Money; D. T. Graesser; H. L. Frericks; G. Sahota; C. M. Rienstra. Magic-angle spinning solid-state NMR spectroscopy of the beta 1 immunoglobulin binding domain of protein G (GB1): N-15 and C-13 chemical shift assignments and conformational analysis. *J. Am. Chem. Soc.,* 127(35): 12291-12305, 2005.

[26] D. S. Wishart; B. D. Sykes. The C-13 Chemical-Shift Index- a simple method for the identification of protein secondary structure using C-13 chemical shift data. *J. Biomol. NMR,* 4(2): 171-180, 1994.

[27] D. S. Wishart; B. D. Sykes; F. M. Richards. The Chemical-Shift Index- a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry,* 31(6): 1647-1651, 1992.

[28] S. Schwarzinger; G. J. A. Kroon; T. R. Foss; J. Chung; P. E. Wright; H. J. Dyson. *J. Am. Chem. Soc.,* 1232970-2978, 2001.

[29] J. Lanman; T. T. Lam; S. Barnes; M. Sakalian; M. R. Emmett; M. A.; P. E. J. Prevelige. *J. Mol. Biol.,* 325759-772, 2003.

[30] R. K. Gitti; B. M. Lee; J. Walker; M. F. Summers; S. Yoo; W. I. Sundquist. Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science,* 273(5272): 231-235, 1996.

[31] T. R. Gamble; S. H. Yoo; F. F. Vajdos; U. K. Vonschwedler; D. K. Worthylake; H. Wang; J. P. Mccutcheon; W. I. Sundquist; C. P. Hill. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science,* 278(5339): 849-853, 1997.

[32] T. R. Gamble; F. F. Vajdos; S. H. Yoo; D. K. Worthylake; M. Houseweart; W. I. Sundquist; C. P. Hill. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell,* 87(7): 1285-1294, 1996.

Chapter 5

DYNAMICS STUDIES OF CONICAL ASSEMBLIES OF U-^{13}C, ^{15}N-TYR-LABELED HIV-1 CA

5.1 Introduction

Capsid proteins (CA) of the human immunodeficiency virus 1 (HIV-1) form conical assemblies to protect the viral RNA genome and other complement proteins in the mature virions [1-6]. According to numerous electron microscopy studies on the mature HIV-1 particles, the CA conical assemblies are composed of approximately 250 fullerene hexamers and exactly 12 pentamers [7-9]. Protein backbone dynamics may have defined the conformational space for CA protein to form different lattice units and morphologies with different conformations, especially for the functionally critical residues.

High-resolution structures of the individual N- and C-terminal domains of HIV-1 CA have been determined by solution NMR [10] and X-ray crystallogrphy [11,12], and aided in generating a model of the CA assemblies by cryo-EM [13]. According to those studies, one molecule of CA protein can be divided into two independently folded domains: N-terminal domain (NTD, ~145 aa) and C-terminal domain (CTD, ~90 aa). In solution, the full-length CA forms oligomers, which renders severely line-broadened solution NMR spectra, while NTD exists as a monomer and CTD exists as a dimer. On the basis of those studies, a hypothesis has been put forth that the polymorphism of CA

assemblies is due to the inherent flexibility of the hinge region between the NTD and CTD, allowing for slightly different local arrangement of the two rigid-body domains and leading to formation of different oligomeric structures [14,15]. X-ray and cryo-EM studies to date have predominantly illuminated the structure of 3D and 2D crystalline arrays of CA protein molecules, but were not able to investigate the backbone dynamics behaviors of individual amino acid [10,16-20]. Therefore, we conducted a solid-state NMR study of conformational mobility of CA to investigate the molecular mechanisms giving rise to the conformational heterogeneity of the intermolecular interfaces that may result in the polymorphism of CA assemblies.

At the time we initiated this study, we were unable to work with tubular CA assemblies because they require high salt concentration, and focused our attention on the conical structures. Furthermore, we decided to address the mobility of the hinge region, and specifically of the Tyr-145 residue that occupies a pivotal position in the hinge between the NTD and CTD. According to the functional studies, the hinge region has been demonstrated to be a critical determinant of proper core assembly and stability, and Tyr-145 is a critical residue in this regard [20]. In addition to Tyr-145, there are three other Tyr residues in CA: one (Y130) resides in the Helix 7 of NTD, and the other two (Y164 and Y169) are in Helix 8 of CTD [21]. Therefore, investigating the backbone atoms of those tyrosine residues would permit assessment of the dynamic behavior of each domain and of the critical hinge region. We prepared conical assemblies of CA containing U-^{13}C,^{15}N-labeled Tyr residues.

As discussed in Chapter 2, information about dynamics on the nano- to millisecond time scales is obtained in MAS NMR from backbone ^{1}H–^{15}N, ^{1}H–^{13}C, and

$^{13}C^{\alpha}$–^{15}N dipolar coupling constants, and ^{15}N chemical shift anisotropies [22-25]. The solid-state NMR data presented in this chapter indicate that Y145 undergoes motions on the timescale of milliseconds and is rigid on faster timescales, while all other tyrosine residues are devoid of motions in this time regime. This millisecond time scale motion of the hinge permits the exploitation of multiple conformations in the assembly of the capsids of varied morphologies.

5.2 Materials and methods

5.2.1 Expression and purification of U-^{13}C,^{15}N-Tyr labeled CA

Expression and purification of the U-^{13}C, ^{15}N-Tyr labeled CA protein and CTD control samples were performed by Dr. Jinwoo Ahn and his lab members (University of Pittsburgh).

A cDNA for the Gag polyprotein, $pr55^{gag}$, was obtained from the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH [26]. Regions encoding Pro1-Leu231 (full-length CA), Met144-Leu231 (CTD (144-231)), Ile150-Leu231 (CTD (150-231), Gly137-Leu231 (CTD (137-231)) and Val142-Leu231 (CTD (142-231)) were amplified and subcloned into pET21 (EMD chemicals, Inc. San Diego, CA) using NdeI and XhoI sites. Proteins were expressed in *E. coli* Rosetta 2 (DE3), cultured in Luria-Bertani or modified minimal media, after induction with 0.4 mM IPTG at 23 ºC for 16 h. All proteins were purified over 5 mL Hi-Trap QP columns (GE Healthcare, Piscataway, NJ) in 25 mM sodium phosphate (pH 7.0), 1 mM DTT, 0.02 % sodium azide, and 5 mL Hi-Trap SP columns (GE Healthcare, Piscataway, NJ), using a 0-1 M NaCl gradient in 25 mM sodium phosphate (pH 5.8), 1 mM DTT, 0.02 % sodium

azide. The final purification step comprised gel-filtration over a Hi-Load Superdex 200 16/60 column (GE Healthcare) in 25 mM sodium phosphate (pH 6.5), 100 mM NaCl, 1 mM DTT, 0.02 % sodium azide. Proteins were concentrated to 10 mg/mL using Amicon concentrators (Millipore, Billerica, MA), flash-frozen in liquid N_2 and stored at -80 ºC. Molecular masses of purified CA proteins were confirmed by LC-TOF mass spectrometry (Bruker Daltonics, Billerica, MA). Uniform ^{15}N- and ^{13}C-labeling of the proteins was carried out by growth in modified minimal medium using $^{15}NH_4Cl$ and $^{13}C_6$-glucose as the sole nitrogen and carbon sources, respectively. Uniform ^{2}H-, ^{15}N- and ^{13}C-labeling of the proteins was achieved using $^{2}H_2O$, $^{15}NH_4Cl$ and $^{13}C_6$,$^{2}H_7$-glucose as deuterium, nitrogen and carbon sources. For solid-state NMR experiments, full-length CA and two CTD constructs, CTD (137-231) and CTD (142-231), were labeled with $^{13}C/^{15}N$ tyrosine by adding U-$^{13}C/^{15}N$-tyrosine prior to IPTG induction in modified minimal medium containing $^{15}NH_4Cl$ and $^{12}C_6$-glucose, as reported previously [27].

5.2.2 Preparation of U-^{13}C, ^{15}N-Tyr labeled protein samples for SSNMR

CA samples for solid-state NMR experiments were prepared in 25 mM sodium phosphate buffer pH 5.5, containing 1 mM DTT and 0.02% NaN_3. Conical assemblies were prepared by adding a 20 % w/v solution of PEG-20,000 to 32 mg/ml U-$^{13}C/^{15}N$-Tyr CA protein in a 1:1 ratio, followed by incubation at 37 ºC for 1 hr, as reported previously [28]. The morphology was evaluated by negatively stained TEM and confocal microscopy as described in Chapter 3. CTD (137-231) and CTD (142-231) samples were prepared by controlled precipitation with PEG 3,350, by adding a 60% w/v solution of PEG 3,350 to 32 mg/ml protein in a 1:1 ratio. Samples were pelleted by centrifugation at 6,000 G for 15 minutes, the supernatant was removed, and 15 mg (0.59 mmoles) of full-

length conical CA assemblies were packed into a 3.2 mm MAS rotor. 17 mg (1.61 mmoles) of CTD (137-231) and 18.5 mg (1.83 mmoles) of CTD (142-231) were packed into two 1.8 mm rotors. The rotors were sealed using an upper spacer and a top spinner for subsequent MAS NMR experiments.

5.2.3 SSNMR spectroscopy

All solid-state NMR experiments were carried out on a 14.1 T narrow bore Varian InfinityPlus solid-state NMR spectrometer operating at Larmor frequencies of 599.8, 150.8 and 60.8 MHz for ^{1}H, ^{13}C and ^{15}N, respectively. All solid-state NMR spectra of U-$^{13}C/^{15}N$-Tyr labeled capsid assemblies were obtained with a 3.2 mm Varian triple-resonance T3 probe, and spectra of the CTD samples were acquired with a 1.8 mm MAS triple-resonance probe developed in the Samoson laboratory. The MAS frequency was set at 10 kHz for all NMR experiments, and was controlled to within ± 2 Hz by a Varian MAS controller. $PbNO_3$ was used as temperature sensor to calibrate the sample temperature in both MAS NMR probes [29]. The actual temperature of the sample was maintained to within ± 0.1 ºC using the Varian temperature controller, and was set to the desired value over the range of 238 K to 288 K. All CP, ^{1}H-^{15}N dipolar, ^{15}N CSA, and $^{15}N\{^{13}C\}$ REDOR spectra were collected at 258 K. ^{13}C and ^{15}N chemical shifts were referenced with respect to the external standards adamantane and ammonium chloride, respectively [30].

Intra-residue assignments of tyrosine resonances were obtained using 1D ^{15}N and ^{13}C CP/MAS as well as 2D homonuclear ^{13}C-^{13}C DARR and $R2_1^1$ experiments [31,32], and heteronuclear ^{13}C-^{15}N NCA experiment. For H-X (X = ^{13}C or ^{15}N) cross polarization

(CP), a linearly ramped rf field and a constant-amplitude rf field were applied on the X nucleus and on ^{1}H, respectively. For the NCA experiments, SPECIFIC-CP [33] with a tangent amplitude ramp was used for band selective magnetization transfer from ^{15}N to ^{13}C. All CP and SPECIFIC-CP conditions were optimized experimentally to attain the first-order Hartmann-Hahn matching condition.

In the 2D DIPSHIFT [34] and ^{15}N chemical shift anisotropy (CSA) recoupling experiments, rotor-synchronized $R18_1^7$ symmetry [35] and ROCSA [36] blocks were incorporated during the t_1 period for ^{1}H-^{15}N dipolar and ^{15}N CSA evolution, respectively. The radio frequency field strength applied on the ^{1}H channel during the $R18_1^7$ period was 90 kHz, and that applied on the ^{15}N channel during the ROCSA period was 42.8 kHz. A ^{15}N{^{13}C} [37] rotational echo double resonance (REDOR) experiment [38] was conducted to assess the $^{13}C_{\alpha}$-^{15}N dipolar coupling constants for each tyrosine residue in the three protein samples. A series of REDOR spectra were recorded with mixing times of 1.0 ms, 1.6 ms and 2.0 ms. A 1D CA-N experiment was performed on the full-length conical CA assemblies, whereas SPECIFIC-CP with a tangent amplitude ramp was used to achieve the band selective magnetization transfer from $^{13}C_{\alpha}$ to ^{15}N. All solid-state NMR data were processed with NMRPipe [39] and analyzed with SPARKY3 (version 3.115) [40].

5.2.4 Analysis and simulations of solid-state NMR spectra

This part of work was carried out by Dr. Guangjin Hou in the Polenova group.

Numerical simulations of ^{1}H-^{15}N dipolar and ^{15}N CSA spectra as well as of ^{15}N{^{13}C} REDOR dipolar dephasing curves were performed with the SIMPSON software

package [41]. 986 pairs of {α, β} angles generated according to the ZCW algorithm [42] and 64 γ angles (resulting in a total of 63104 angle triplets) were used to generate a powder average for all simulations. For extracting the best-fit dipolar (ω_D) and CSA parameters (δ_s, η_s), all the experimental and processing parameters (i.e., Larmor frequency, MAS frequency, RF field strength, number of t_1 points, finite pulse lengths, zero-filling, line broadening, etc.) were taken into account in the simulations for each experiment.

5.2.5 CSA tensor convention

We use the Haeberlen convention in which the CSA tensor is characterized by three parameters, isotropic chemical shift (δ_{iso}), reduced anisotropy (δ_s), and asymmetry parameter (η_s) [43]. These parameters are defined in terms of the principal components of the CSA tensor, with the isotropic chemical shift as $\delta_{iso} = (\delta_{11} + \delta_{22} + \delta_{33})/3$. The reduced anisotropy (δ_s) and the asymmetry parameter (η_s) are defined as $\delta_\sigma = \delta_{11} - \delta_{iso}$ and $\eta_\sigma = (\delta_{22} - \delta_{33})/\delta_\sigma$ if $|\delta_{11} - \delta_{iso}| \geq |\delta_{33} - \delta_{iso}| \geq |\delta_{22} - \delta_{iso}|$ (positive anisotropy), or $\delta_\sigma = \delta_{33} - \delta_{iso}$ and $\eta_\sigma = (\delta_{22} - \delta_{11})/\delta_\sigma$ if $|\delta_{33} - \delta_{iso}| \geq |\delta_{11} - \delta_{iso}| \geq |\delta_{22} - \delta_{iso}|$ (negative anisotropy). The asymmetry parameter ranges from 0 to 1.

5.3 Resonance assignments of Tyr residues

To assign the Tyr resonances, we have examined the solution NMR assignments of the U-^{13}C/^{15}N labeled NTD and CTD samples performed by our collaborators, Dr. Angela Gronenborn's group at the University of Pittsburgh [44,45]. As discussed in Chapter 4, the variations between solid-state and solution chemical shifts are small for the

majority of the CA residues. Therefore, we decided that it is reasonable to assign the four Tyr residues using solution NMR chemical shifts as a reference [44,46]. We acquired the 1D ^{13}C and ^{15}N CPMAS spectra of U-^{13}C,^{15}N-Tyr labeled CA conical assemblies and of the two CTD(137-231) and CTD(142-231) nanocrystal control samples; the results are shown Figure 5.1.

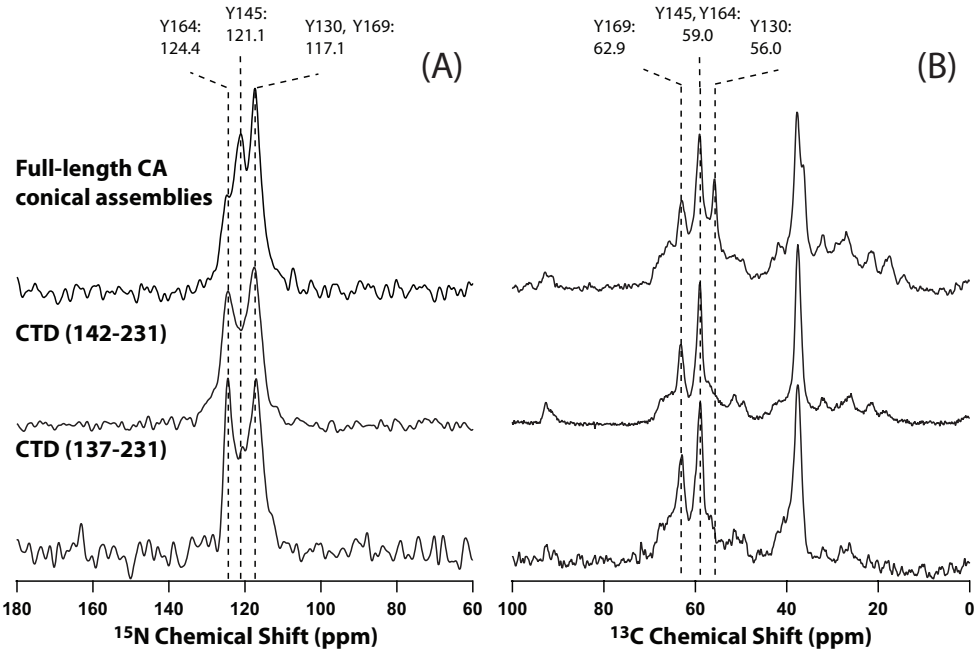


Figure 5.1 (A) ^{15}N CPMAS spectra of full-length CA conical assemblies (top), CTD (142-231) (middle), and CTD (137-231) (bottom). (B) ^{13}C CPMAS spectra of full-length conical CA assemblies (top), CTD (142-231) (middle), and CTD (137-231) (bottom). Resonance assignments are shown and were confirmed in 2D ^{13}C-^{13}C and ^{15}N-^{13}C correlation spectra. The solid-state NMR chemical shifts agree well with solution NMR shifts.

As illustrated in Figure 5.1, the chemical shifts of the tyrosine residues are resolved in ^{15}N and $^{13}C^{\alpha}$ CPMAS spectra. For ^{15}N CP spectra shown in Figure 5.1 (A), three distinct peaks appear: at 124.4 ppm (corresponding to Y164), at 121.1 ppm (corresponding to Y145), and at 117.1 ppm (corresponding to Y130 and Y169, who

exhibiti the same ^{15}N chemical shift). Interestingly, the peak for Y145 peak is well resolved in the full-length CA conical assemblies sample, while it is broadened in the spectra of two CTD samples, especially for the shorter construct (CTD (142-231)), indicating the presence of either conformational heterogeneity or internal backbone motion on the time scale of cross-polarization, or both. In the ^{13}C CPMAS spectra shown in Figure 5.1 (B), there are three resolved peaks: at 62.9 ppm (for Y169), at 56.0 ppm (for Y130), and at 59.0 ppm (the signal for Y145 and Y164 who have identical chemical shifts). No significant differences between the spectra of the full-length CA conical sample and the CTD control samples are observed, except for the expected disappearance of the signal of the NTD residue Y130. To resolve the chemical shift degeneracy of the Y130/Y169 ^{15}N and Y145/Y164 $^{13}C^{\alpha}$, two-dimensional spectra were needed as discussed below.

Figure 5.2 shows the ^{13}C-^{13}C homonuclear correlation spectra acquired with $R2_1^1$ pulse sequence [31] of CA CTD (142-231) (Figure 5.2 A)) and full-length CA conical assemblies (Figure 5.2 B)). In both spectra, Y164 and Y169 showed strong aliphatic-to-aliphatic and carbonyl-to-aliphatic correlations. Y130 in the full-length CA sample showed the same behavior as well. On the other hand, Y145 did not appear in the full-length CA spectrum and showed very weak and split peaks in the CTD (142-231) spectrum, suggesting that multiple conformers exist for Y145 in the CTD sample. The 1D slices extracted from the 2D C-C correlation spectra of the CTD sample indicate that the signals of the Y145 are clearly above the noise level and are real peaks. The most likely reason that Y145 signals are not observed in the full-length CA is that motions

interfere with their detection. As discussed below, presence of such motions is evident in the ^{15}N-^{13}C data sets.

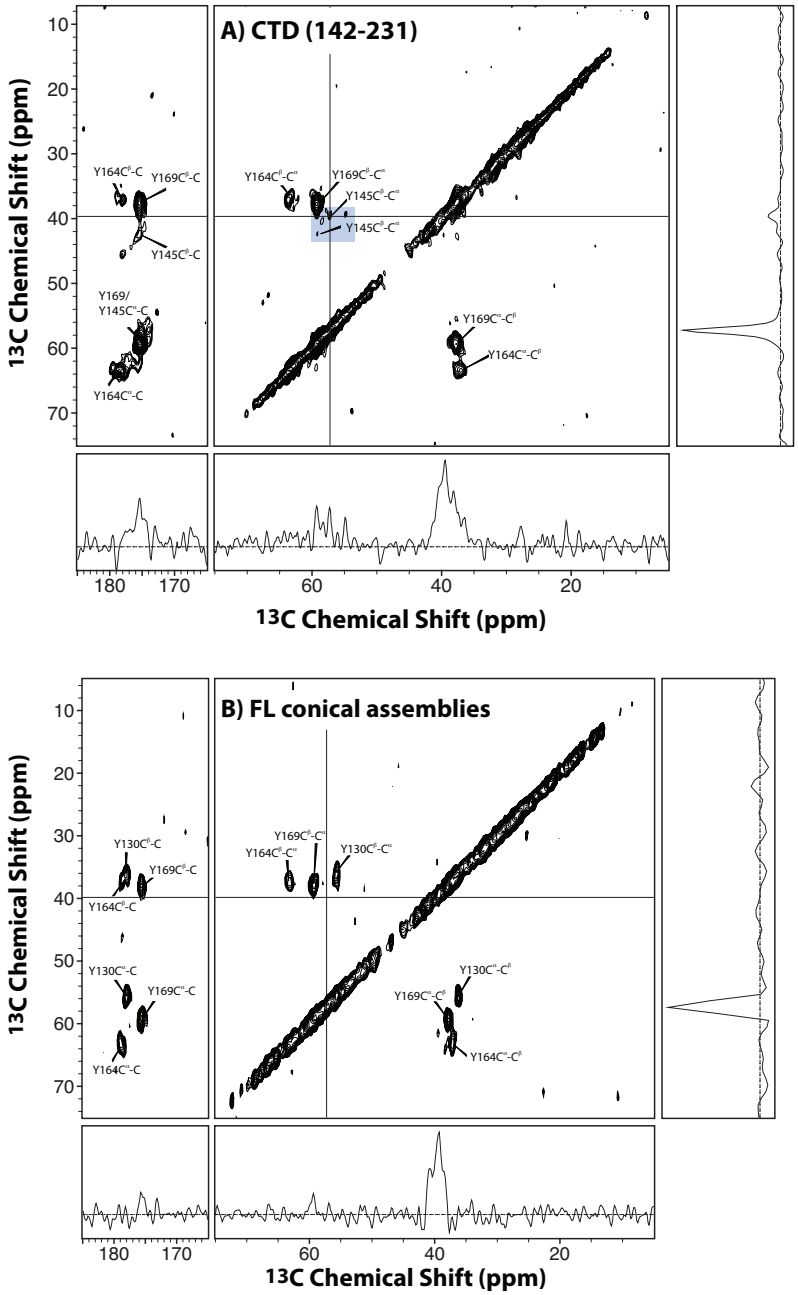


Figure 5.2 ^{13}C-^{13}C $R2_1^1$ correlation spectra of CTD (142-231) (A) and full-length conical CA assemblies (B) for the aliphatic and carbonyl regions. 1D traces extracted along the ω_2 and ω_1 frequency dimensions are shown to illustrate that weak cross peaks corresponding to the different Y145 conformers are clearly above the noise level in the CTD (142-231) spectra. No Y145 resonances were detected for full-length CA assemblies.

5.4 Nano- to microsecond timescale dynamics in U-^{13}C, ^{15}N-Tyr CA

Motions on nano- to microsecond timescales are ubiquitous in proteins, especially for residues in the functional hinge regions [47,48]. We hypothesized that such motions may occur in the hinge region between CTD and NTD of CA, and conducted MAS NMR experiments to test the presence of such motions.

5.4.1 ^{1}H-^{15}N dipolar lineshapes

As discussed in Chapter 2, ^{1}H-^{15}N dipolar couplings are sensitive to motions occurring on nano-to microsecond time scales. We have employed a 2D $R18_7^1$ pulse sequence shown in Figure 2.4 to acquire the ^{1}H-^{15}N line shapes of the full-length U-^{13}C,^{15}N-Tyr CA conical assemblies and the two CTD (CTD (137-231) and CTD (144-231)) control samples. By numerical simulations, we were able to extract the specific values of the dipolar coupling constant, ω_D. The results are illustrated in Figure 5.3 and reveal that the ^{1}H-^{15}N dipolar lineshapes are strikingly similar for all tyrosine residues in the three samples. The rigid limit value for ω_D is 11.34 kHz, and the dipolar order parameters of the Y145 range from 0.84 to 0.91 while those for other tyrosine residues range from 0.88 to 0.93, indicating that the H-N bond in all the tyrosine residues is nearly rigid on the time scales faster than ca. 80 μs. This conclusion is supported by the fact that the dipolar line shapes are consistent with an axial dipolar tensor.

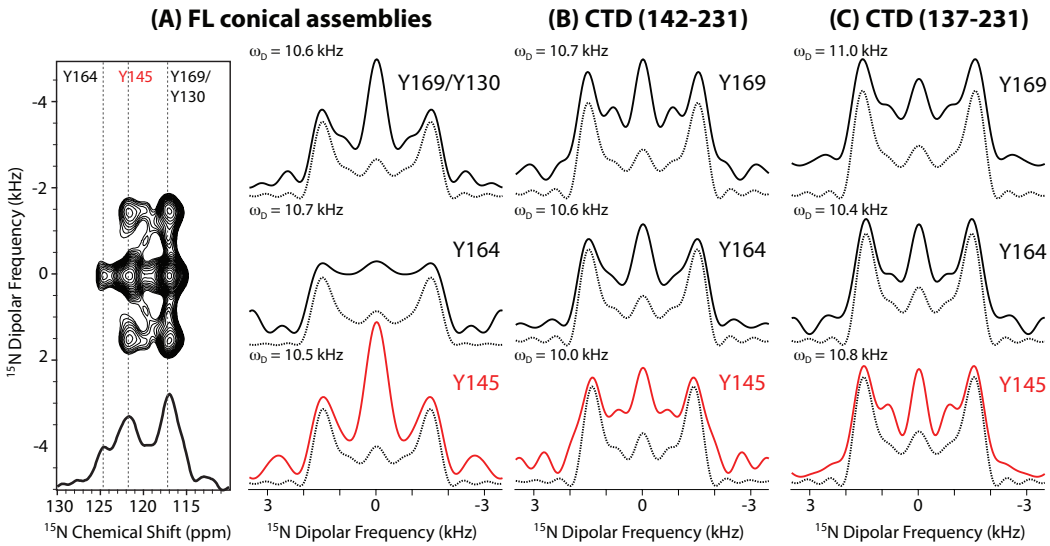


Figure 5.3 ^{1}H-^{15}N dipolar lineshapes for tyrosine residues in (A) full-length CA conical assemblies, (B) CTD (142-231), and (C) CTD (137-231). For full-length assembled CA, the 2D ^{1}H-^{15}N DIPSHIFT spectrum is shown to illustrate the high quality of the data. Experimental dipolar lineshapes and the best-fit lineshapes are shown in solid lines and dashed lines, respectively. The experimental lineshapes for Y145 are highlighted in red. For each residue, the ^{1}H-^{15}N dipolar coupling constant, w_D, extracted by numerical simulations, is listed. Note that all tyrosine residues exhibit essentially rigid-limit behavior on the timescale of the DIPSHIFT experiment.

5.4.2 ^{15}N CSA lineshapes

2D ^{15}N-detected ROCSA experiment has been utilized to probe the ^{15}N chemical shift anisotropies of all the tyrosine residues. The principles were discussed in Chapter 2, and the pulse sequence was shown in Figure 2.5. The rigid limit of the ^{15}N reduced chemical shift anisotropy value is 110 ppm (165 ppm for full CSA interaction strength, or 10.0 kHz at 14.1 T). This value is very close to the rigid-limit value of the ^{1}H-^{15}N dipolar coupling constant (11.34 kHz). Therefore, both experiments are sensitive to the motions occurring on the time scales of micro- to nanoseconds.

The ^{15}N CSA tensor parameters, $\delta_{\circ}$ and $\eta_{\circ}$ are strikingly similar for all tyrosine residues in all samples and reveal rigid limit values, as shown in Figure 5.4. Taken

together, the ^{1}H-^{15}N dipolar and the ^{15}N CSA experiments indicate that all tyrosine residues, including Y145, in all the samples are rigid or mostly rigid on the microsecond or faster time scales.

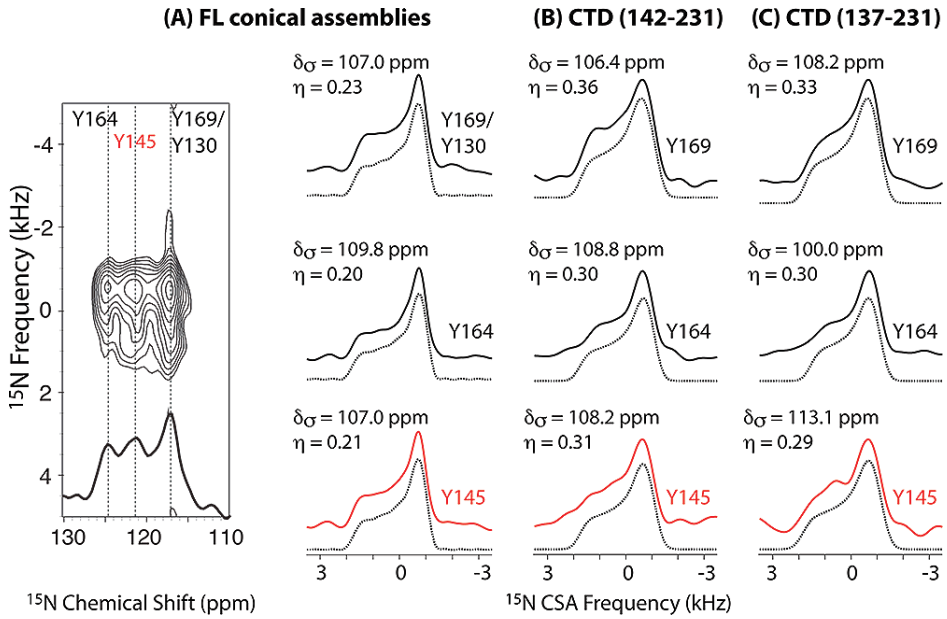


Figure 5.4 ^{15}N CSA line shapes for tyrosine residues in (A) full-length CA conical assemblies, (B) CTD (142-231), and (C) CTD (137-231). For full-length CA, the 2D ^{15}N ROCSA spectrum is shown to illustrate the high quality of the data. Experimental and best-fit CSA lineshapes are shown in solid and dashed lines, respectively. The experimental lineshapes for Y145 are highlighted in red. For each residue, the ^{15}N CSA tensor parameters, δ_s and η_s, extracted from numerical simulations, are listed. Note that all tyrosines residues exhibit essentially rigid-limit parameters on the timescale of the ROCSA experiment.

5.5 Millisecond timescale dynamics in CA

The absence of internal mobility on nano- to microsecond timescales in the hinge region of assembled CA prompted us to examine the motions occurring on slower, millisecond times scales. For this, we have turned our attention to the analysis of ^{13}C-^{15}N dipolar couplings, qualitatively (through temperature-dependent NCA and CAN

spectroscopy) and quantitatively (through the explicit measurement of the dipolar coupling constants in a REDOR experiment).

5.5.1 Temperature dependence of heteronuclear 2D NCA spectra

The rigid-limit ^{15}N-$^{13}C^{\alpha}$ dipolar coupling constant value is ω_D=984 Hz. Figure 5.5 depicts the 2D NCA spectra acquired on full-length CA conical assemblies and on CA CTD (142-231) used as a control. In the 2D NCA spectra, Y130, Y164 and Y169 in the full-length CA sample showed strong signals. The experiment temperature was varied from +5 °C to -35 °C, with 5-degree increments. No variations of the chemical shifts were observed for those residues at these different temperatures. At -35 °C, the Y169 peak disappeared, suggesting that under these conditions, motions may be occurring on appropriate time scale and interfere with the signal detection. We also note that at -35 °C, the peak intensities for all tyrosine residues decreased.

Interestingly, Y145 signal did not show up in these 2D NCA spectra at any temperature. It is likely that motional averaging of the Y145 N-C dipolar coupling causes the loss of the Y145 cross peak in the NCA spectra.

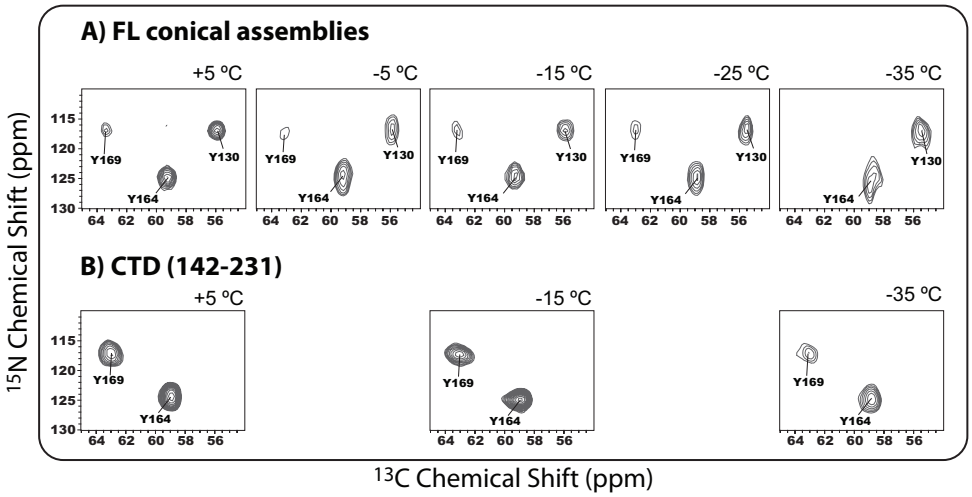


Figure 5.5 Temperature dependence of the solid-state NCA spectra of (A) full-length CA conical assemblies and (B) CTD (142-231). The spectra were acquired at 14.1 T, and the MAS frequency was 10 kHz. The temperature at which every spectrum was collected is indicated next to the corresponding data set. Note that Y145 cross peak is not present at any temperature.

5.5.2 CA-N spectra

Since the Y145 N-C^{α} signal was missing in the NCA 2D spectra, we decided to turn our attention to a similar experiment conducted in a one-dimensional format and called CA-N. The information content of the CA-N spectrum is similar to that of the NCA except that the magnetization is transferred from the C^{α} carbon to the N, and since the ^{15}N peak for Y145 is resolved in the 1D ^{15}N CPMAS spectrum, the CA-N experiment does not require the second indirect dimension and can be conducted as a one-dimensional measurement. The advantage of this approach is that the indirect dimension period is eliminated, therefore preventing any potential loss of magnetization that may be associated with short T_2* rather than the motional averaging of the ^{13}C-^{15}N dipolar

interaction for Y145. Therefore, the CA-N experiment is a complementary and more direct reporter on the possible motional averaging of the ^{13}C-^{15}N dipolar interaction in Y145.

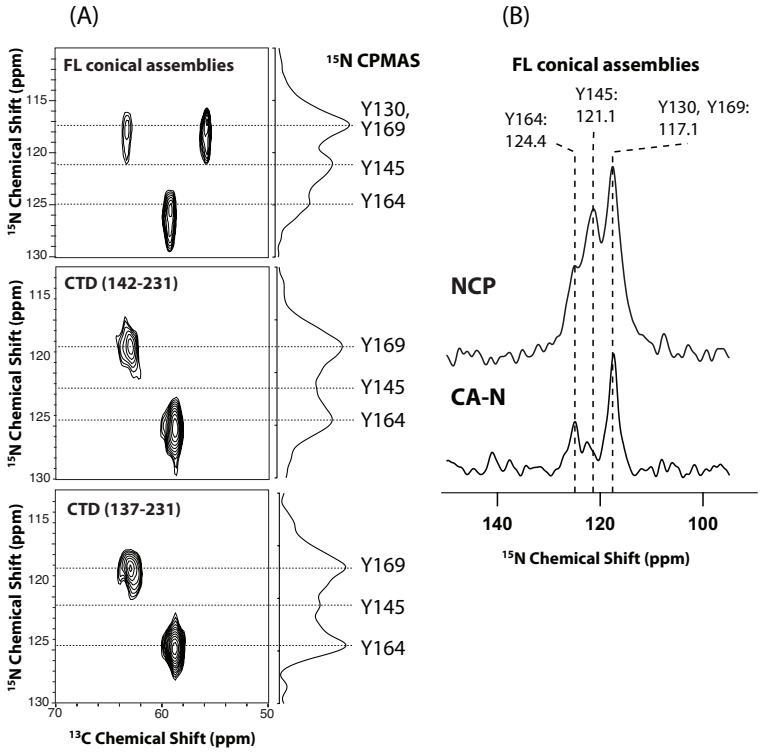


Figure 5.6 (A) Two-dimensional NCA spectra of full-length conical CA assemblies (top), CTD (142-231) (middle), and CTD (137-231) (bottom) at 258 K. On the right, the corresponding 1D ^{15}N CPMAS spectra are displayed. Note that the Y145 peak is present in all ^{15}N CPMAS spectra, while the corresponding Y145 cross peaks are missing in the 2D NCA spectra, suggesting that the ^{15}N-^{13}C dipolar couplings are motionally averaged. (B) CA-N (bottom) and ^{15}N CPMAS (top) spectra of full-length CA conical assemblies, illustrating that the Y145 peak intensity is severely attenuated in the CA-N data, most likely due to motional averaging of the corresponding ^{15}N-^{13}C dipolar coupling.

As shown in Figure 5.6 (B), the Y145 signal that is well resolved and strong in the ^{15}N CPMAS spectrum (top) , is got severely attenuated in the corresponding CA-N spectrum (bottom), indicating that motional averaging of the corresponding dipolar interaction is taking place.

5.5.3 ^{13}C-^{15}N REDOR dipolar dephasing curves

The experimental evidence from NCA and CA-N spectra presented above suggests that the ^{13}C-^{15}N dipolar coupling is motionally averaged for Y145. To probe the extent of the motional averaging, we have acquired the $^{15}N\{^{13}C\}$ REDOR dephasing curves. As illustrated in Figure 5.7, the dephasing profiles for Y130, Y164 and Y169 in the full-length CA conical assemblies (and for Y164 and Y169 in the two CTD control samples) are indistinguishable within the experimental error and the corresponding ^{15}N-$^{13}C^{\alpha}$ dipolar couplings extracted from the curves are 500±200 Hz, slightly lower than the rigid limit value of 984 Hz (the deviations from the strict rigid-limit value are likely due to the experimental imperfections for the 180-degree dephasing pulses, which are invevitable with our current hardware). In contrast, the Y145 ^{15}N-$^{13}C^{\alpha}$ dipolar coupling is completely averaged out to zero and undetectable within the experimental error (no dipolar dephasing is observed, Figure 5.7), indicating the presence of motions on a timescale of 10 ms or faster.

Taken together, the ^{15}N-^{13}C REDOR, ^{1}H-^{15}N dipolar and ^{15}N CSA MAS NMR data discussed in this chapter clearly indicate motions in the millisecond regime for Y145 in both the conical CA assemblies and the control CTD CA samples. These millisecond time scale motions in the hinge region detected by MAS solid-state NMR open up the conformational space in principle accessible to CA for the assembly into capsids of different morphologies. The solution NMR work of our collaborators, Drs. In-Ja Byeon and Angela Gronenborn of the University of Pittsburgh, further revealed that there is a precise molecular switch mechanism that controls the number of conformers (two) and their populations and through which CA may assemble into distinct oligomeric structures

(pentamers and hexamers) that form the closed conical capsids. The results of our joint hybrid solid-state/solution NMR study have been recently reported in the *Journal of the American Chemical Society* [49].

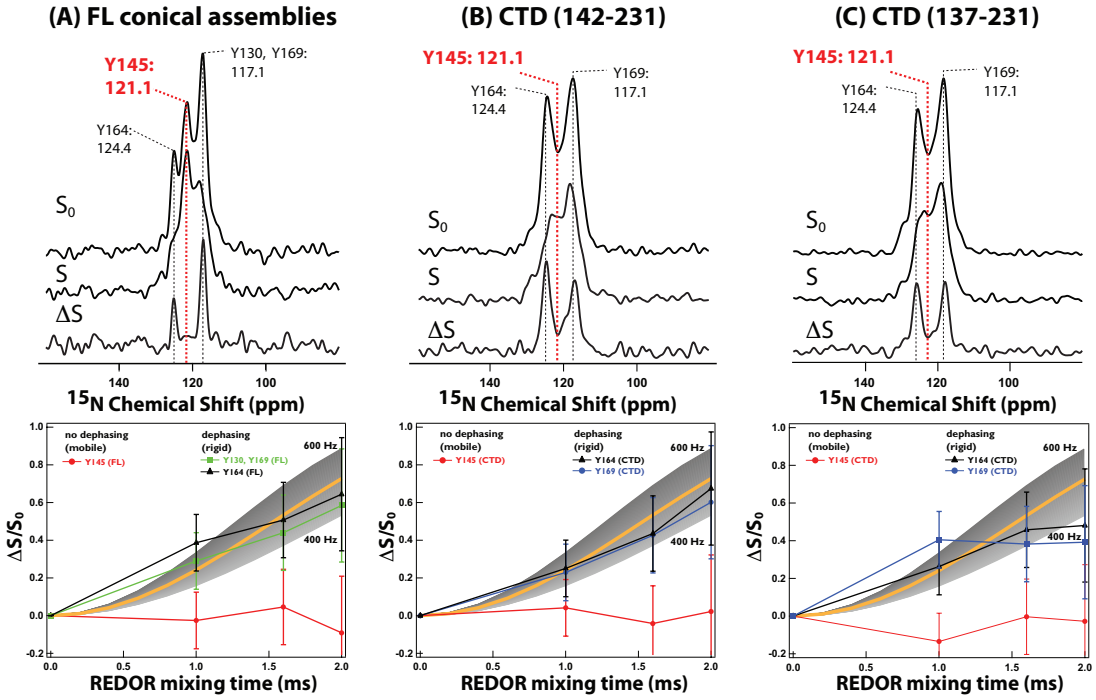


Figure 5.7 ^{15}N-^{13}C REDOR spectra (top) and dephasing curves (bottom) of (A) full-length CA conical assemblies, (B) CTD (142-231), and (C) CTD (137-231). In the REDOR spectra, S_0 and S are the peak intensities in the absence and presence of REDOR dephasing pulses, respectively. DS = S_0-S is the intensity difference. In the REDOR dephasing curves, the REDOR fraction, DS/S_0, is plotted against the mixing time. Note that for all three proteins, Y145 does not exhibit REDOR dephasing caused by motional averaging of the ^{15}N-^{13}C dipolar coupling on the timescale of the REDOR experiment (~10 ms).

References

[1] P. Ashorn; T. J. Mcquade; S. Thaisrivongs; A. G. Tomasselli; W. G. Tarpley; B. Moss. An inhibitor of the protease blocks maturation of human and simian immunodeficiency viruses and spread of infection. *Proc. Natl. Acad. Sci. U. S. A.,* 87(19): 7472-7476, 1990.

[2] N. E. Kohl; E. A. Emini; W. A. Schleif; L. J. Davis; J. C. Heimbach; R. A. Dixon; E. M. Scolnick; I. S. Sigal. Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. U. S. A.,* 85(13): 4686-4690, 1988.

[3] H. G. Gottlinger; J. G. Sodroski; W. A. Haseltine. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A.,* 86(15): 5781-5785, 1989.

[4] F. Barre-Sinoussi; J. C. Chermann; F. Rey; M. T. Nugeyre; S. Chamaret; J. Gruest; C. Dauguet; C. Axler-Blin; F. Vezinet-Brun; C. Rouzioux; W. Rozenbaum; L. Montagnier. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science,* 220(4599): 868-871, 1983.

[5] R. C. Gallo; S. Z. Salahuddin; M. Popovic; G. M. Shearer; M. Kaplan; B. F. Haynes; T. J. Palker; R. Redfield; J. Oleske; B. Safai; Et Al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science,* 224(4648): 500-503, 1984.

[6] S. Campbell; V. M. Vogt. Self-assembly *in vitro* of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. *J. Virol.,* 69(10): 6487-6497, 1995.

[7] B. K. Ganser; S. Li; V. Y. Klishko; J. T. Finch; W. I. Sundquist. Assembly and analysis of conical models for the HIV-1 core. *Science,* 283(5398): 80-83, 1999.

[8] S. Li; C. P. Hill; W. I. Sundquist; J. T. Finch. Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature,* 407(6802): 409-413, 2000.

[9] O. Pornillos; B. K. Ganser-Pornillos; B. N. Kelly; Y. Hua; F. G. Whitby; C. D. Stout; W. I. Sundquist; C. P. Hill; M. Yeager. X-ray structures of the hexameric building block of the HIV capsid. *Cell,* 137(7): 1282-1292, 2009.

[10] R. K. Gitti; B. M. Lee; J. Walker; M. F. Summers; S. Yoo; W. I. Sundquist. Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science,* 273(5272): 231-235, 1996.

[11] T. R. Gamble; S. H. Yoo; F. F. Vajdos; U. K. Vonschwedler; D. K. Worthylake; H. Wang; J. P. Mccutcheon; W. I. Sundquist; C. P. Hill. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science,* 278(5339): 849-853, 1997.

[12] T. R. Gamble; F. F. Vajdos; S. H. Yoo; D. K. Worthylake; M. Houseweart; W. I. Sundquist; C. P. Hill. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell,* 87(7): 1285-1294, 1996.

[13] B. K. Ganser-Pornillos; A. Cheng; M. Yeager. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell,* 131(1): 70-79, 2007.

[14] M. Yeager; B. J. Nicholson. Structure of gap junction intercellular channels. *Curr. Opin. Struct. Biol.,* 6(2): 183-192, 1996.

[15] H. R. Gelderblom; E. H. S. Hausmann; M. Ozel; G. Pauli; M. A. Koch. Fine-structure of human-immunodeficiency-virus (HIV) and immunolocalization of structural proteins. *Virology,* 156(1): 171-176, 1987.

[16] C. Tang; Y. Ndassa; M. F. Summers. Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. *Nat. Struct. Biol.,* 9(7): 537-543, 2002.

[17] T. R. Gamble; F. F. Vajdos; S. Yoo; D. K. Worthylake; M. Houseweart; W. I. Sundquist; C. P. Hill. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell,* 87(7): 1285-1294, 1996.

[18] C. Momany; L. C. Kovari; A. J. Prongay; W. Keller; R. K. Gitti; B. M. Lee; A. E. Gorbalenya; L. Tong; J. Mcclure; L. S. Ehrlich; M. F. Summers; C. Carter; M. G. Rossmann. Crystal structure of dimeric HIV-1 capsid protein. *Nat. Struct. Biol.,* 3(9): 763-770, 1996.

[19] D. Ivanov; O. V. Tsodikov; J. Kasanov; T. Ellenberger; G. Wagner; T. Collins. Domain-swapped dimerization of the HIV-1 capsid C-terminal domain. *Proc. Natl. Acad. Sci. U. S. A.,* 104(11): 4353-4358, 2007.

[20] I. J. Byeon; X. Meng; J. Jung; G. Zhao; R. Yang; J. Ahn; J. Shi; J. Concel; C. Aiken; P. Zhang; A. M. Gronenborn. Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell,* 139(4): 780-790, 2009.

[21] J. Jiang; S. D. Ablan; S. Derebail; K. Hercik; F. Soheilian; J. A. Thomas; S. Tang; I. Hewlett; K. Nagashima; R. J. Gorelick; E. O. Freed; J. G. Levin. The interdomain linker region of HIV-1 capsid protein is a critical determinant of proper core assembly and stability. *Virology,* 421(2): 253-265, 2011.

[22] V. Chevelkov; U. Fink; B. Reif. Quantitative analysis of backbone motion in proteins using MAS solid-state NMR spectroscopy. *J. Biomol. NMR,* 45(1-2): 197-206, 2009.

[23] N. Giraud; M. Blackledge; M. Goldman; A. Bockmann; A. Lesage; F. Penin; L. Emsley. Quantitative analysis of backbone dynamics in a crystalline protein from nitrogen-15 spin-lattice relaxation. *J. Am. Chem. Soc.,* 127(51): 18190-18201, 2005.

[24] A. Krushelnitsky; D. Reichert. Complex $^{1}H,^{13}C$-NMR relaxation and computer simulation study of side-chain dynamics in solid polylysine. *Biopolymers,* 78(3): 129-139, 2005.

[25] J. Yang; M. L. Tasayco; T. Polenova. Dynamics of reassembled thioredoxin studied by magic angle spinning NMR: snapshots from different time scales. *J. Am. Chem. Soc.,* 131(38): 13690-13702, 2009.

[26] S. Erickson-Viitanen; J. Manfredi; P. Viitanen; D. E. Tribe; R. Tritch; C. A. Hutchison, 3rd; D. D. Loeb; R. Swanstrom. Cleavage of HIV-1 Gag polyprotein

synthesized *in vitro*: sequential cleavage by the viral protease. *AIDS Res. Hum. Retroviruses,* 5(6): 577-591, 1989.

[27] S. Sun; S. Paramasivam; S. Yan; Y. Han; C. Lightcap; J. C. Williams; I. J. L. Byeon; J. Ahn; A. M. Gronenborn; T. Polenova In Protein NMR Techniques; Shekhtman, A., Burz, D., Eds.; Humana Press: Louisville, 2011.

[28] Y. Han; J. Ahn; J. Concel; I. J. Byeon; A. M. Gronenborn; J. Yang; T. Polenova. Solid-state NMR studies of HIV-1 capsid protein assemblies. *J. Am. Chem. Soc.,* 132(6): 1976-1987, 2010.

[29] G. Neue; C. Dybowski. Determining temperature in a magic-angle spinning probe using the temperature dependence of the isotropic chemical shift of lead nitrate. *Solid State Nucl. Magn. Reson.,* 7(4): 333-336, 1997.

[30] C. R. Morcombe; K. W. Zilm. Chemical shift referencing in MAS solid state NMR. *J. Magn. Reson.,* 162(2): 479-486, 2003.

[31] G. J. Hou; S. Yan; S. J. Sun; Y. Han; I. J. L. Byeon; J. Ahn; J. Concel; A. Samoson; A. M. Gronenborn; T. Polenova. Spin diffusion driven by R-symmetry sequences: applications to homonuclear correlation spectroscopy in MAS NMR of biological and organic solids. *J. Am. Chem. Soc.,* 133(11): 3943-3953, 2011.

[32] K. Takegoshi; S. Nakamura; T. Terao. ^{13}C-^{1}H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett.,* 344631-637, 2001.

[33] M. Baldus; A. T. Petkova; J. Herzfeld; R. G. Griffin. Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems. *Mol. Phys.,* 95(6): 1197-1207, 1998.

[34] M. Munowitz; W. P. Aue; R. G. Griffin. Two-dimensional separation of dipolar and scaled isotropic chemical-shift interactions in magic angle NMR-spectra. *J. Chem. Phys.,* 77(4): 1686-1689, 1982.

[35] X. Zhao; M. Eden; M. H. Levitt. Recoupling of heteronuclear dipolar interactions in solid-state NMR using symmetry-based pulse sequences. *Chem. Phys. Lett.,* 342(3-4): 353-361, 2001.

[36] J. C. C. Chan; R. Tycko. Recoupling of chemical shift anisotropies in solid-state NMR under high-speed magic-angle spinning and in uniformly C-13-labeled systems. *J. Chem. Phys.,* 118(18): 8378-8389, 2003.

[37] M. Hologne; V. Chevelkov; B. Reif. Deuterated peptides and proteins in MAS solid-state NMR. *Prog. Nucl. Magn. Reson. Spectrosc.,* 48(4): 211-232, 2006.

[38] T. Gullion; J. Schaefer. Rotational-echo double-resonance NMR. *J. Magn. Reson.,* 81(1): 196-200, 1989.

[39] F. Delaglio; S. Grzesiek; G. W. Vuister; G. Zhu; J. Pfeifer; A. Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR,* 6(3): 277-293, 1995.

[40] T. D. Goddard; D. G. Kneller; 3.110 ed.; University of California: San Francisco, 2004, http://www.cgl.ucsf.edu/home/sparky/.

[41] M. Bak; J. T. Rasmussen; N. C. Nielsen. SIMPSON: A general simulation program for solid-state NMR spectroscopy. *J. Magn. Reson.,* 147(2): 296-330, 2000.

[42] V. B. Cheng; H. H. Suzukawa; Wolfsber.M. Investigatinos of a nonrandom numerical method for multidimensional integration. *J. Chem. Phys.,* 59(8): 3992-3999, 1973.

[43] U. Haeberlen In Advances in Magnetic Resonance; Waugh, J. S., Ed.; Academic Press: Orlando, 1976; Vol. Supplement 1.

[44] Y. Han; J. Ahn; J. Concel; I. J. L. Byeon; A. M. Gronenborn; J. Yang; T. Polenova. Solid-state NMR studies of HIV-1 capsid protein assemblies. *J. Am. Chem. Soc.,* 132(6): 1976-1987, 2010.

[45] J. Jung; I. J. L. Byeon; J. Ahn; J. Concel; A. M. Gronenborn. H-1, N-15 and C-13 assignments of the dimeric C-terminal domain of HIV-1 capsid protein. *Biomol. NMR Assign.,* 4(1): 21-23, 2010.

[46] A. C. S. Saphire; M. D. Bobardt; P. A. Gallay. Cyclophilin a plays distinct roles in human immunodeficiency virus type 1 entry and postentry events, as revealed by spinoculation. *J. Virol.,* 76(9): 4671-4677, 2002.

[47] K. A. Henzler-Wildman; M. Lei; V. Thai; S. J. Kerns; M. Karplus; D. Kern. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature,* 450(7171): 913-916, 2007.

[48] D. E. Shaw; P. Maragakis; K. Lindorff-Larsen; S. Piana; R. O. Dror; M. P. Eastwood; J. A. Bank; J. M. Jumper; J. K. Salmon; Y. B. Shan; W. Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science,* 330(6002): 341-346, 2010.

[49] G. Hou; S. Paramasivam; I. J. L. Byeon; A. M. Gronenborn; T. Polenova. Determination of relative tensor orientations by gamma-encoded chemical shift anisotropy/heteronuclear dipolar coupling 3D NMR spectroscopy in biological solids. *Phys. Chem. Chem. Phys.,* 12(45): 14873-14883, 2010.

Chapter 6

RESONANCE ASSIGNMENTS AND CONFORMATIONAL ANALYSIS OF TUBULAR ASSEMBLIES OF HIV-1 CA AND CA-SP1 PROTEINS

6.1 Introduction

Being a potential drug target for maturation inhibitors, the SP1 peptide in HIV-1 Gag and in intermediate Gag cleavage products has received recent attention [1,2]. The mechanism of how the inhibitors such as the clinically effective Bevirimat (BVM) block the virus replication remains unknown, largely due to lack of conformation information of assembled CA-SP1, especially the SP1 region.

Secondary structure predictions have suggested that the SP1 peptide may adopt an α-helical conformation [3,4]. An X-ray analysis of CA(CTD)-SP1 revealed a disordered region encompassing the final 12 residues in the crystal structure [3,4]. Solution NMR investigations have demonstrated that in pure H_2O, SP1 peptide has only a slight propensity for the formation of an a-helix in the region comprised of SP1, the last 13 residues of CA, and the first 13 residues of NC [5]. Recent solution NMR studies on two synthetic peptides encompassing the SP1 region of Gag have suggested that the SP1 region adopts an α-helical conformation when examined in solutions containing 30% trifluoroethanol (TFE). TFE is known to promote the formation of α-helices in regions that have a high propensity toward α-helix formation. In this same investigation, one of the two synthetic peptides showed little propensity for α-helix formation in pure H_2O and

the other peptide did not provide analyzable data due to the formation of a gel [5]. Up till now, no reports have appeared on the SP1 structure in full-length assembled CA-SP1, and the hypothesis that SP1 may act as a molecular switch by adopting an α-helical conformation upon assembly remains largely unsubstantiated.

As discussed throughout this thesis, the challenges in structural investigations of HIV capsid protein assemblies at atomic-level resolution have been largely due to the inherent limitations of the mainstream structural biology techniques: the requirement for solubility (solution NMR) or for diffraction-quality crystals (X-ray crystallography) or the intrinsic lack of atomic-level structural details (cryoEM). Solid-state NMR spectroscopy is void of these limitations and can yield atomic-level detail in insoluble non-crystalline and large systems, and in the recent years, significant progress has been made in using magic angle spinning (MAS) NMR to characterize large protein assemblies [6-10]. Our laboratory has demonstrated that HIV-1 CA protein assemblies of conical morphologies are amenable to detailed characterization by MAS NMR spectroscopy [11]. Around the same time, Tycko and colleagues reported solid-state NMR investigation of CA assemblies of tubular morphologies [9]. While this pioneering study established the feasibility of characterization of the tubular assemblies of CA, the resolution and sensitivity of these original data sets were limited, and even with sparse labeling of the protein, the spectra were highly congested with broad unresolved peaks. With the advent of Low-E [12], Scroll Coil [7,13], and the most recent Enhanced-Design-Low-E probes it became possible to perform MAS NMR experiments on CA tubular assemblies.

In this chapter, we investigated the structure and dynamics of CA and CA-SP1 assemblies of tubular morphologies, and in particular we examined the conformation of the spacer peptide SP1. At 21.1 T, these tubular CA and CA-SP1 assemblies give rise to unprecedented-quality homo- and heteronuclear correlation spectra, from which site-specific assignments of key residues can be readily obtained. According to the chemical shifts assigned for SP1 peptide residues and the lack of signals in the through-space experiments, it is clear that the spacer peptide is in random coil conformation in the assembled CA-SP1. Our findings do not support the prior hypothesis on the role of SP1 as a molecular switch that triggers the assembly process by adopting α-helical conformation. The outstanding sensitivity and resolution of these ultrahigh field NMR spectra will enable further atomic-resolution structural and dynamics investigations of HIV-1 protein assemblies.

6.2 Experiments and Methods

6.2.1 Sample preparation

Two series of tubular assemblies of HIV-1 CA and CA-SP1 protein constructs were prepared and analyzed: an HXB2 variant [14] (which will be referred to as "WT") and the A92E mutant of the NL4-3 variant [15] (which will be referred to as "A92E"). The primary sequence for these two variants is discussed below. The detailed procedure for expression and purification were described in detail in Chapter 3 and in [10,16-18].

The assembly procedures were described in Chapter 3.

For SSNMR experiments, eighteen milligrams of the CA A92E and twenty-eight milligrams of CA-SP1 A92E tubular assemblies were packed into 3.2 mm Bruker rotors,

and sealed using a top spinner. Twelve milligrams of the CA WT and eleven milligrams of CA-SP1 WT tubular assemblies were packed into 3.2 mm Agilent rotors, sealed using an upper spacer and a top spinner, and used in subsequent solid-state NMR experiments.

6.2.2 Transmission electron microscopy and confocal laser scanning microscopy

The detailed procedure for TEM and confocal laser scanning microscopy characterization of CA assemblies was reported by us previously [10,17] and described in Chapter 3.

6.2.3 NMR spectroscopy

The SSNMR spectra of CA WT and CA-SP1 WT series of constructs were acquired at the National High Magnetic Field Laboratory (NHMFL) at 21.1 T (900 MHz) on a ultrawide bore instrument outfitted with a 3.2 mm ^{1}H/X/Y Enhanced-Design Low-E MAS Probe designed and built at NHMFL. Larmor frequencies were 899.1 MHz for ^{1}H, 226.1 MHz for ^{13}C, and 91.1 MHz for ^{15}N. The temperature was 4.0 °C and was maintained to within ±1 °C throughout the experiments by Bruker temperature controller. Spectra were collected at the MAS frequency of 14.000±0.001 kHz regulated by a Bruker MAS controller. ^{13}C and ^{15}N chemical shifts were referenced with respect to DSS and adamantane used as external referencing standards [19]. The typical pulse lengths were 2.5 μs (^{1}H), 3 μs (^{13}C), and 5 μs (^{15}N). ^{1}H- ^{13}C(^{15}N) cross polarization employed a linear amplitude ramp; the ^{1}H radio frequency (rf) field was 62 kHz with the center of the ramp Hartmann-Hahn matched to the first spinning sideband. The ^{1}H-^{13}C and ^{1}H-^{15}N contact times were 1 ms. The ^{15}N-^{13}C DCP contact time was 5.0 ms. The ^{1}H field strength during DARR was 14 kHz, and the DARR mixing time was 50 ms. Small phase

incremental alternation (SPINAL-64) ^{1}H decoupling with a rf field strength of 100 kHz was used during both the direct and indirect acquisition periods [20]. Recycle delays in all experiments were 2 s. Time-proportional phase incrementation (TPPI) [21] was used for phase-sensitive detection in the indirect dimensions.

The SSNMR spectra of CA A92E and CA-SP1 A92E constructs were acquired at the University of Delaware on a Bruker 19.9 T spectrometer outfitted with a 3.2 mm BioMAS HXY probe. Larmor frequencies were 850.400 MHz for ^{1}H, 213.855 MHz for ^{13}C, and 86.181 MHz for ^{15}N. Spectra were collected at the MAS frequency of 13.000±0.001 kHz regulated by a Bruker MAS controller, and the temperature was maintained at 4±0.5 °C throughout the experiments using a Bruker temperature controller. The typical pulse lengths were 3.3 μs (^{1}H), 3.8 μs (^{13}C), and 4.2 μs (^{15}N), and ^{1}H-^{13}C cross polarization was performed with a linear amplitude ramp (80-100%), having the ^{1}H radio frequency field of 80 kHz matched to the second spinning sideband. The ^{1}H-^{13}C and ^{1}H-^{15}N contact times were 1 and 1.5 ms, respectively. SPINEL-64 decoupling (80 kHz) was applied during the evolution and acquisition periods [20]. The ^{1}H field strength during DARR was 13 kHz, and the DARR mixing time was 50 ms. Under these conditions predominantly one-bond correlations are observed in the spectra. States-TPPI protocol [22] was used for frequency discrimination in the indirect dimension.

Other relevant experimental conditions are indicated in the figure captions and in the Appendix.

6.2.4 NMR data processing and analysis

All spectra were processed with NMRpipe [23] and analyzed using Sparky [24]. For 2D data sets, 60° or 30° shifted sine bell apodization followed by a Lorentzian-to-

Gaussian transformation was applied in both dimensions. Forward linear prediction to twice the number of the original data points was used in the indirect dimension in some data sets followed by zero filling to twice the total number of points. Detailed information about processing parameters is provided in the Appendix Table B.1.

6.3 Characterization of CA and CA-SP1 tubular assemblies

The conditions for preparation of tubular CA WT assemblies suitable for solid-state NMR studies were established by us previously, and we have shown that these assemblies are stable upon magic angle spinning [11]. The conditions for generating CA A92E and CA-SP1 A92E assemblies of tubular morphology are exactly the same, and high-quality tubes form in the presence of 0.5 M NaCl (concentration in the sample), as illustrated in Figure 6.1. The dimensions of CA and CA-SP1 tubes are similar. One should pay special attention to the temperature to preserve the samples because tubes decompose after being frozen and thawed (as discussed in Chapter 3).

The CA WT protein forms both single-walled and multi-walled tubular assemblies, and tends to form tube bundles in solution. CA-SP1 WT sample forms not only single and multi-walled tubular assemblies, but also some cone shapes according to the TEM images. In the confocal images, both regular tubes and small fluorescent dots were observed, indicating morphologies other than tubes. Despite the small variations in morphologies, assemblies of all the four protein samples yield outstanding-quality MAS NMR spectra as discussed below.

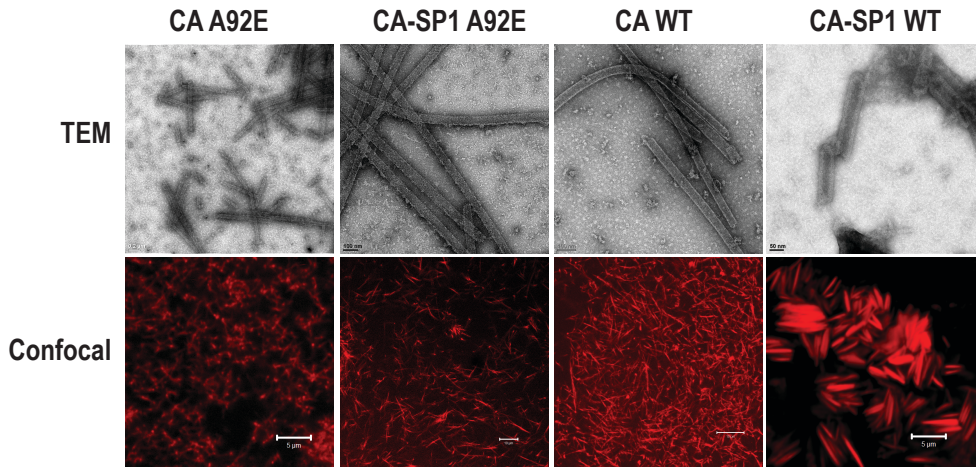


Figure 6.1 From left to right: TEM (top) and confocal (bottom) images of CA A92E, CA-SP1 A92E, CA WT and CA-SP1 WT tubular assemblies. For the CA-SP1 WT sample, conical assemblies were observed in addition to tubes by both TEM and confocal methods (not shown).

As discussed above, the WT and A92E CA/CA-SP1 protein constructs are from two functionally different strains of HIV-1 virus: NL4-3 and HXB2. In Figure 6.2, the primary sequences are for these two constructs under investigation. NL4-3 strain encodes all known HIV-1 proteins, whereas HXB-2 strain is defective for three auxiliary genes: vpr, vpu, and nef [21,25]. The protein sequence differs by 4 residues: I/L6, V/L83, N/H120 and A/G208. Residue A92 in the NL4-3 strain was mutated to glutamic acid as this mutation was found to give rise to more homogenous assemblies according to the cryo-EM results [26]. The SP1 peptide is a 14-residue peptide located at the C-terminus of CA (see Figure 6.2).

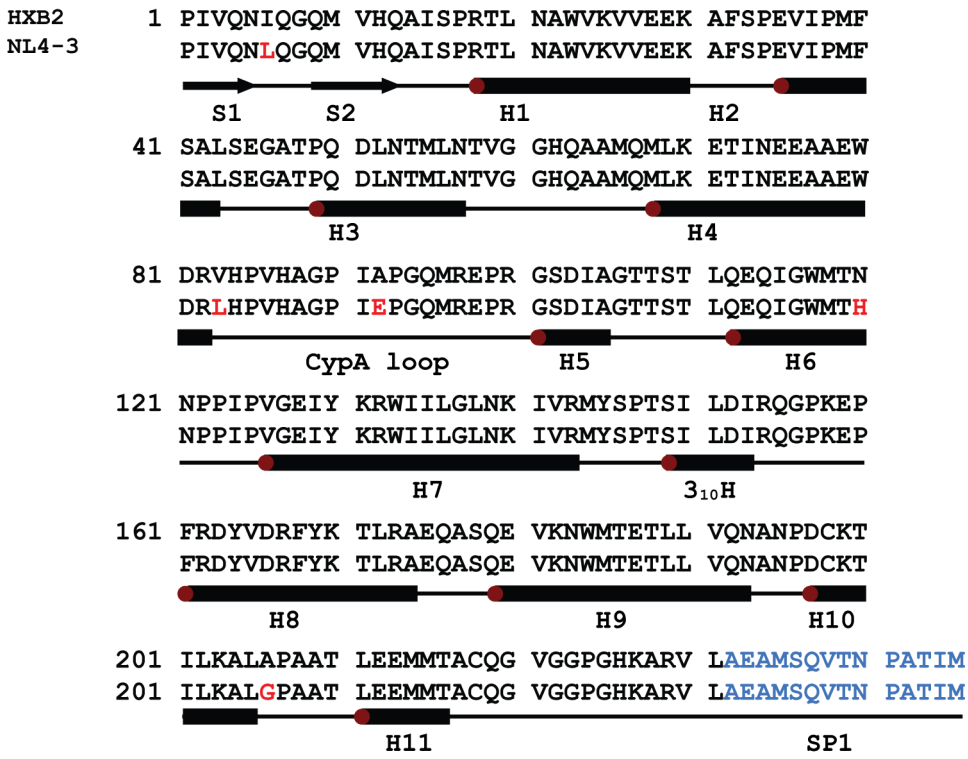


Figure 6.2 Primary sequences of the HXB2 and NL4-3 (A92E mutant) strains used in our investigations of CA and CA-SP1 assemblies of the tubular morphology [27]. The variations between the amino acid residues in HXB2 and NL4-3 strains are shown in red, and the final fourteen amino acids of the construct belong to the SP1 spacer peptide tail, shown in blue.

6.4 Tubular assemblies yield ultrahigh resolution spectra

The DARR and NCA spectra (Figure 6.3 and 6.4) reveal surprisingly high resolution and sensitivity, with quality superior to that observed by us in conical assemblies and by others in a prior report on tubular assemblies [9,11].

Figure 6.3 depicts the heteronuclear correlation (NCA) spectra obtained in A92E CA and CA-SP1 tubular assemblies acquired at 21.1 T. From the NCA data sets, 146 resolved peaks were counted in CA spectrum (of a total of 231 expected peaks), and 172 peaks were counted in CA-SP1 spectrum (of a total of 245 expected peaks). The 1D

traces extracted from the 2D spectra at 112.4 ppm are shown and reveal that the line-widths of the resolved peaks are between 0.2 to 0.5 ppm.

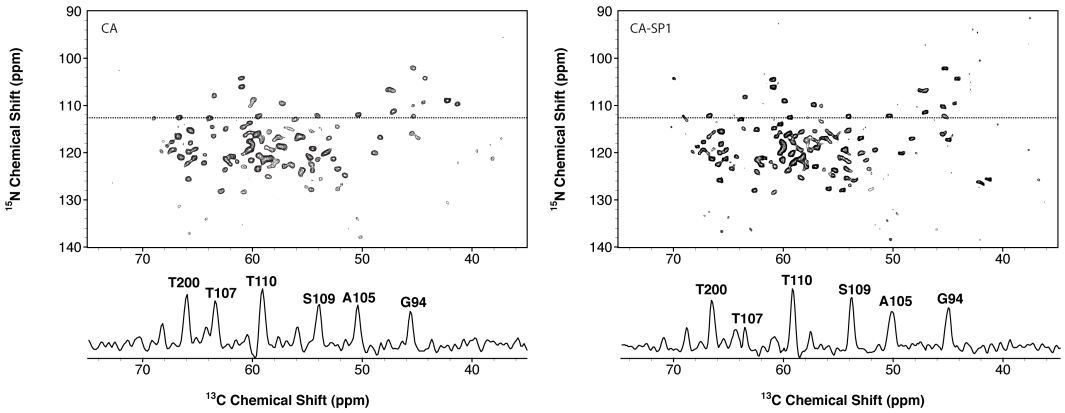


Figure 6.3 The 21.1 T 2D NCA spectra of U-^{13}C,^{15}N HIV-1 CA (left) and CA-SP1 (right). The 1D traces in the ω_2 dimension corresponding to the resolved cross-peaks at ^{15}N shift 112.4 ppm are shown and illustrate the remarkable spectral sensitivity and resolution. The acquisition parameters for the spectra of U-^{13}C,^{15}N HIV-1 CA and CA-SP1 tubular assemblies are given in Table B.1 in the Appendix.

6.5 Resonance Assignments

Because the CA and CA-SP1 tubular assemblies yielded remarkably ultra-high resolution spectra, it was possible to perform resonance assignments on the basis of the 2D and 3D data sets.

6.5.1 2D spectra assignments

Figure 6.4 is an example of the sequential assignments of tubular assemblies of CA-SP1 A92E (NL4-3 strain) completed with 2D data sets (NCACX, NCOCX, and DARR). The 1D trace along the ^{15}N dimension extracted at 112.4 ppm is shown as well. From the trace, it is clear that this sample exhibited narrow line widths. This ultra-high resolution facilitates greatly the resonance assignments.

The resonance assignments followed the general procedures outlined in Chapter 4. We used the resolved peaks in the Ser-Thr region as the starting point. Using the known assignments from solution NMR studies of the NTD [11,28], the chemical shift of the backbone nitrogen of T110 was tentatively assigned to be 112.4 ppm, and the corresponding C^{α}, C^{β} and C' cross peaks at this nitrogen shift were identified in the NCACX spectrum at 59.4, 71.2 and 175.8 ppm respectively. Meanwhile, the cross peaks corresponding to C^{α} and C^{β}, C^{α} and C' in the DARR spectrum were found at both sides of the diagonal, confirming that those chemical shifts are from the same spin system. For the sequential assignments, the NCOCX 2D spectrum was used, and ^{13}C peaks corresponding to the ^{15}N shift of 112.4 ppm (T110) were found at 54.0, 66.8 and 174.1 ppm, which are the typical shifts for serine residues. The obtained chemical shifts indicate the Ser-Thr amino acid pair in the protein. There is only one Ser-Thr pair in the protein sequence: S109-T110. Therefore, the previous assumption that those signals are from T110 is proved correct, and consistent with the solution NMR results [10,29]. Taking T110 assignment as an anchor point, we have assigned the stretch of residues R100-T110 using the 2D data sets, as shown in the Figure 6.4. Benefiting from the extraordinarily high quality of the spectrum, we identified not only the distinct C^{α} and C^{β} shifts of those residues, but also the C' and some of the side chain signals, providing more reliable CSI analysis for the secondary structure determination. More assignments obtained with this procedure and these are summarized in Table 6.1.

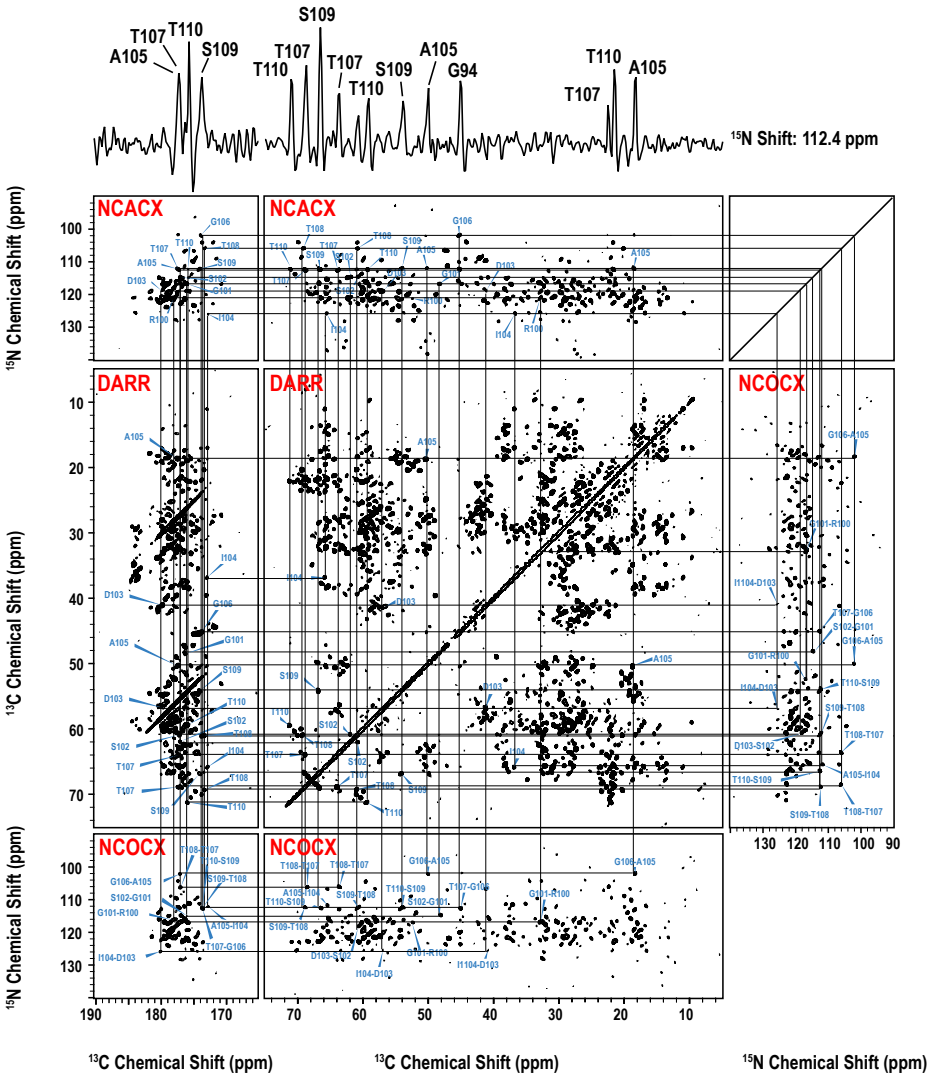


Figure 6.4 Backbone walk for the stretch of residues R100-T110 of U-^{13}C,^{15}N tubular assemblies of HIV-1 CA-SP1 A92E (NL4-3 strain) acquired at 19.9 T and 4 °C. The 1D trace at 112.4 ppm along the ^{15}N dimension of the NCACX spectrum is shown on top. The following 2D data sets were used to establish this backbone walk: NCACX (top), NCOCX (middle right and bottom), and DARR (middle left and middle center). The DARR mixing time was 50 ms. The acquisition parameters for the spectra of U-^{13}C,^{15}N HIV-1 CA-SP1 tubular assemblies are given in Table B.1 in the Appendix.

6.5.2 3D backbone walks

Despite the outstanding quality of the 2D spectra acquired at 19.9 T, the 2D data sets alone were not sufficient for us to be able to complete the resonance assignments.

Additional 3D spectra have proven to be necessary. These were acquired in tubular assemblies of CA WT (HXB2 strain) protein. A combination of 3D and 2D spectra was subsequently employed for the resonance assignments: 3D NCACX, 3D NCOCX, 2D NCA, and 2D DARR (with the mixing times of 20 ms, 50 ms and 250 ms). With these data sets in hand, we have performed resonance assignments following the protocols outlined in Chapter 2.2.1. Figure 6.5 is a representative backbone walks of the residue span R100-G116. The chemical shifts of those residues are identical with the values obtained from the 2D data sets. The chemical shifts for the assigned residues are summarized in Table 6.1.

To date, we have completed assignments of the CTD domain and partially assigned the chemical shifts of the NTD domain of CA in the tubular assemblies. The summary of chemical shifts is presented in Table 6.1.

The secondary structure was predicted by the consensus chemical shift indices of C^{α}, C^{β} and C' and is in agreement with the prior solution and solid-state reports for the unassembled CA and the conical CA assemblies [10,30].

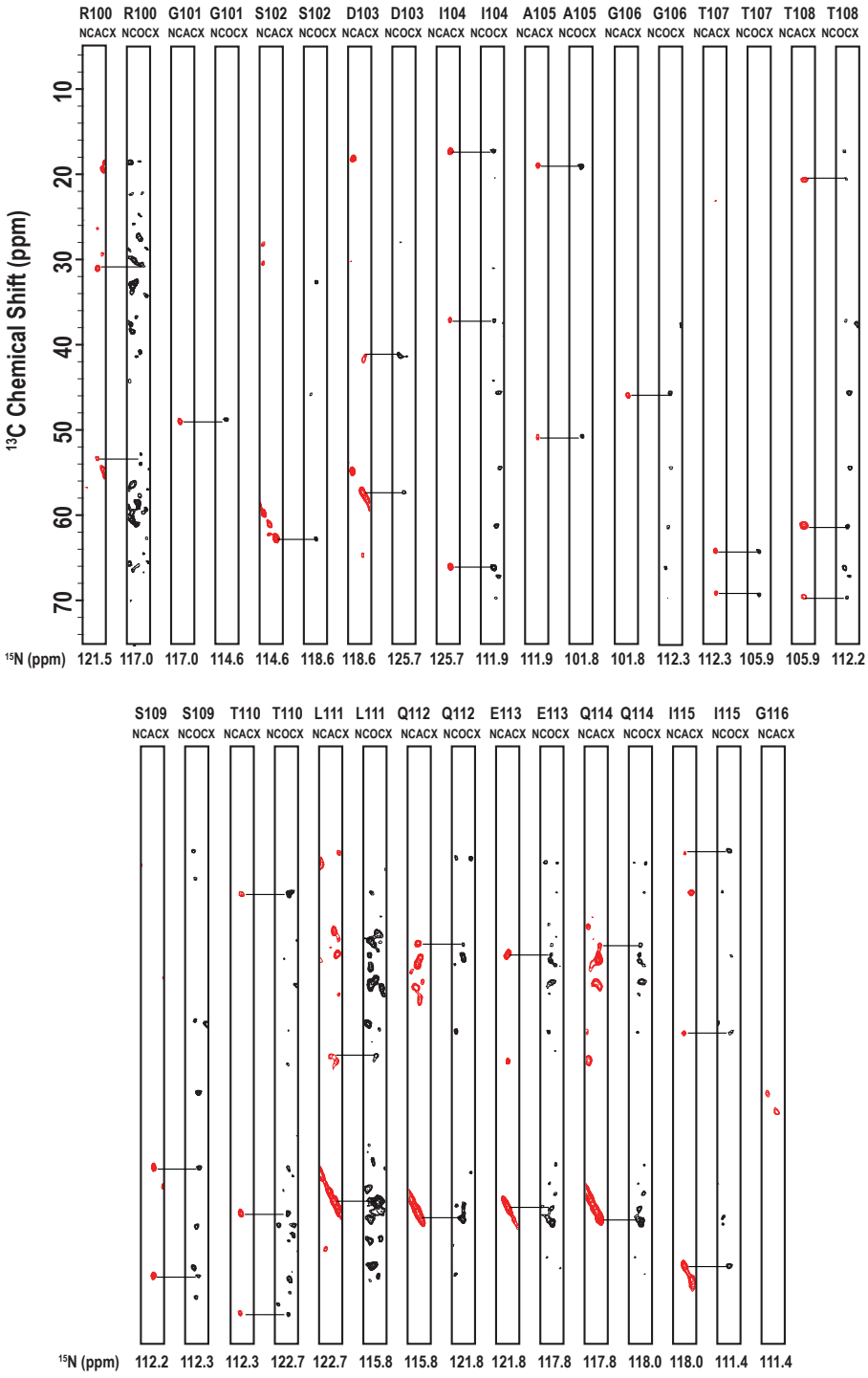


Figure 6.5 Backbone walk of the stretch of residues R100-T108 (top) and S109-G116 (buttom) of tubular assemblies of U-^{13}C,^{15}N HIV-1 CA WT (HXB2 strain). The NCACX and NCOCX spectra are shown in red and black, respectively.

Table 6.1 ^{13}C and ^{15}N chemical shift assignments (ppm) for the CA protein in tubular assemblies. The CSI column is the secondary structure predicted by CSI method: H-α helix, B-β sheet, C-unstructured coil.

Residue	N	C^{α}	C^{β}	C'	C^{γ}	CSI
P1						
I2						
V3						
Q4				176.8		C
N5		55.0	37.9	178.2		H
I6	124.0	63.6	38.3	177.0		C
Q7	116.7	56.5	32.0	179.2		C
G8	106.9	47.1				H
Q9						
M10						
V11	121.6	59.0	36.0	180.1	21.9	B
H12	121.6	56.2	34.4	177.3		C
Q13	125.5	54.2	33.0	176.1		B
A14		52.2	19.0	175.7		C
I15	121.1	61.7	38.1	177.0		C
S16	123.8	56.3	63.4			C
P17						
R18						
T19	116.0	66.5	68.9		21.7	H
L20						
N21						

A22					
W23					
V24					
K25					
V26					
V27					
E28					
E29					
K30					
A31				177.7	H
F32	114.7	62.4	38.3	176.6	H
S33	116.5	57.1	64.6		B
P34					
E35					
V36					
I37					
P38					
M39	115.3	58.2	32.7	178.0	H
F40	123.0	61.0	42.7	178.2	H
S41	112.3	60.8	63.7	174.8	H
A42	121.9	54.2	19.1	180.8	H
L43	116.6	55.8	42.0	176.2	C
S44	110.0	57.4	63.7	174.8	C
E45	125.3	58.3	27.2	179.0	H
G46	116.1	45.4		174.6	C
A47	120.0	53.2	20.4	179.1	C
T48	108.7	60.2	69.7		B

P49					
Q50					
D51					
L52					
N53	116.9	56.3	37.2	178.3	H
T54	119.8	67.4	68.3		H
M55					
L56		58.0	41.1	178.2	H
N57	116.8	55.9	37.3	176.4	H
T58	112.6	63.5	69.5	174.5	C
V59	121.0	64.4	32.5	176.8	C
G60	117.6	45.3		177.0	C
G61	106.9	47.5			H
H62		56.2	28.6		C
Q63					
A64					
A65					
M66					
Q67					
M68					
L69					
K70					
E71					
T72	117.6	67.4	68.4	176.4	H
I73	122.6	66.0	38.1	177.1	C
N74	116.7	55.5	38.0		H
E75					

E76						
A77						
A78						
E79						
W80	121.6	62.2	28.4	180.2		H
D81	120.2	57.3	40.0	178.3		H
R82	120.2	59.5	30.9	178.8		H
V83	112.5	62.4	32.3	175.4		C
H84	118.7	53.2	27.0			C
P85						
V86	121.5	62.5	33.0	176.5	21.0	C
H87	125.3	55.7	29.4	175.3		C
A88	127.8	52.1	19.0	178.6		C
G89	109.6	44.5				C
P90						
I91	121.5	60.1	38.5	176.0	17.8	C
A92	131.8	50.7	17.7			C
P93						
G94	112.4	45.3		174.6		C
Q95	119.5	54.4	30.0	175.2		C
M96	120.5	55.0	34.8	177.0		C
R97	122.9	55.3	29.6	174.9		C
E98	123.9	54.5	29.7	178.8		C
P99						
R100	121.3	52.3	31.4	178.8		C
G101	117.1	48.4		176.7		H
S102	114.8	60.8	62.4	175.3		H

D103	118.8	57.0	41.2	177.7	180.9	H
I104	125.9	65.8	36.7	172.8		H
A105	112.0	50.3	18.7	177.8		C
G106	102.0	45.3		174.0		C
T107	112.5	63.8	69.0	177.7	22.3	C
T108	106.0	61.1	69.3	173.5		C
S109	112.3	54.0	66.8	174.1		C
T110	112.4	59.4	71.2	175.8		C
L111	122.8	57.9	41.1	178.1		H
Q112	116.0	59.6	27.9	180.1		H
E113	122.0	59.0	29.1	177.9		H
Q114	117.9	60.0	28.0	178.0		H
I115	118.0	65.6	38.3	179.6		H
G116	111.5	47.1		175.2		H
W117	121.7					
M118	116.0	59.7	34.9	177.8		H
T119	104.7	60.8	70.2	173.7	21.8	B
N120	122.7	53.3	39.4	173.2		C
N121	116.7	49.7	40.1			B
P122						
P123						
I124						
P125						
V126						
G127						
E128						
I129						

Y130	117.1	54.2	37.0		C
K131					
R132					
W133					
I134					
I135					
L136					
G137					
L138					
N139					
K140					
I141					
V142					
R143					
M144					
Y145					
S146					
P147					
T148	117.6	60.2	67.3	178.2	C
S149	109.9	56.6	63.4	177.6	C
I150	128.7	60.6	39.4	175.5	B
L151	118.9	56.9	41.5	177.8	H
D152	114.8	54.2	42.0	175.9	C
I153	121.9	60.1	38.3	175.4	B
R154	125.7	54.3	32.7	175.9	B
Q155	128.2	55.9	27.2		C
G156	117.5	44.9			

P157					
K158	115.3	54.3	32.5	176.1	
E159	126.1	53.7	32.2		
P160					
F161	128.2	62.6	39.3	177.4	H
R162	115.2				
D163					
Y164	123.7	58.4	38.2	178.3	C
V165	121.5	66.8	32.4		C
D166		58.2	42.2	176.0	C
R167	121.7	60.9	32.6	177.3	H
F168	123.8	62.4	38.4	177.1	H
Y169	116.8	62.4	37.4	175.9	H
K170	121.7	60.1	42.0	179.3	H
T171	117.9	65.9	67.6		H
L172		58.2	42.5	178.6	H
R173	117.4	60.3	29.7	178.9	H
A174	119.3	54.5	17.8	181.7	H
E175	121.6	58.6	29.2	180.1	H
Q176	121.6	58.2	29.1	180.2	H
A177	121.9	54.2	19.0	180.9	H
S178	116.8	57.5	64.9	177.8	B
Q179	121.8	59.0	28.3	179.8	H
E180	117.8	60.2	29.7	180.1	H
V181	121.6	65.6	31.1	178.0	H
K182	119.6	60.1	32.1	178.7	H
N183	120.2	57.2	37.6	175.7	H

W184	123.0	62.0	28.3	180.6		H
M185	120.2	59.4	30.9	178.8		H
T186	112.3	66.8	69.1	179.1		H
E187	117.5	59.3	28.0	178.1		H
T188	112.1	64.7	69.2			C
L189	119.2	58.8	41.7	178.3		H
L190	121.6	58.9	41.8	176.1		C
V191	123.6	64.8	29.6	174.8	18.4	H
Q192	121.5	58.5	28.6	177.1		H
N193	114.4	52.8	38.3	175.3		C
A194	120.0	52.8	18.6	175.1		C
N195	119.6	51.7	35.5			C
P196						
D197		58.5	40.3			C
C198		63.0	26.6	177.4		H
K199	121.6	62.5	32.5	175.5		C
T200	118.2	67.1	68.1	175.9		H
I201	122.2	65.4	38.1	178.6		H
L202	119.5	58.6	41.9	178.8		H
K203	119.3	58.4	30.1	178.1		H
A204	122.0	54.2	18.1	176.7		H
L205	117.7	56.5	42.8	179.0		H
G206	104.2	44.5				C
P207						
A208	125.3	51.5	18.6	175.0		C
A209	123.3	52.5	20.0	178.5		H
T210	114.3	60.2	70.6			B

L211	121.6	58.7	41.6	176.3	C
E212	118.2	60.0	29.6	179.1	H
E213	119.7	60.2	29.7	178.6	H
M214	118.9	60.3	32.8	176.9	H
M215	115.9	59.1	32.7	177.0	H
T216	116.0	66.4	68.8	176.4	H
A217	123.7	56.2	18.8		C
C218		60.3	28.6	179.9	H
Q219	122.7	59.0	27.3	181.7	H
G220	106.9	47.2		176.2	H
V221	121.7	62.5	32.9	179.4	C
G222	116.1	45.2		177.2	C
G223	102.3	45.2			C
P224		59.0		178.7	C
G225	98.7	44.3			C
H226					
K227					
A228					
R229					
V230					
L231					

6.6 SP1 is unstructured in assembled CA-SP1

As discussed above, SP1 is a biologically important 14-residue spacer peptide that is cleaved off at the final stage of the HIV-1 virus maturation. To study the structure of SP1 in the CA-SP1 assembled state, we acquired through-space and through-bond SSNMR spectra of CA and CA-SP1. Analysis of the extra peaks and chemical shift perturbations in these spectra yields information regarding the SP1 conformation.

6.6.1 SP1 signals in the through-space correlation SSNMR spectra

Our initial expectation was that any new peaks emerging in CA-SP1 spectra as compared to those of CA should be arising from the SP1 peptide. We have acquired through-space correlation spectra that are based on direct dipolar polarization transfers, spin diffusion, and Nuclear Overhauser Effect (NOE) cross-relaxation transfers. The results of these experiments are presented below.

6.6.1.1 Direct dipolar-based polarization transfer and spin diffusion experiments

The pulse sequences used for dipolar- and spin-diffusion based experiments, NCA, NCACX, NCO, NCOCX, and DARR, are presented in Figure 4.4. To our surprise, the NCA and DARR spectra of CA A92E and CA-SP1 A92E tubular assemblies are identical as shown in Figure 6.6. No new peaks were observed, indicating that the SP1 tail is either unstructured or mobile, the latter resulting in dynamic averaging of the NMR signals.

On the other hand, comparison of the NCA spectra WT CA and CA-SP1 tubular assemblies reveals that a number of signals are shifted and spectral congestion is present in the CA-SP1 data set; at the same time, no distinct new peaks appear. In the DARR

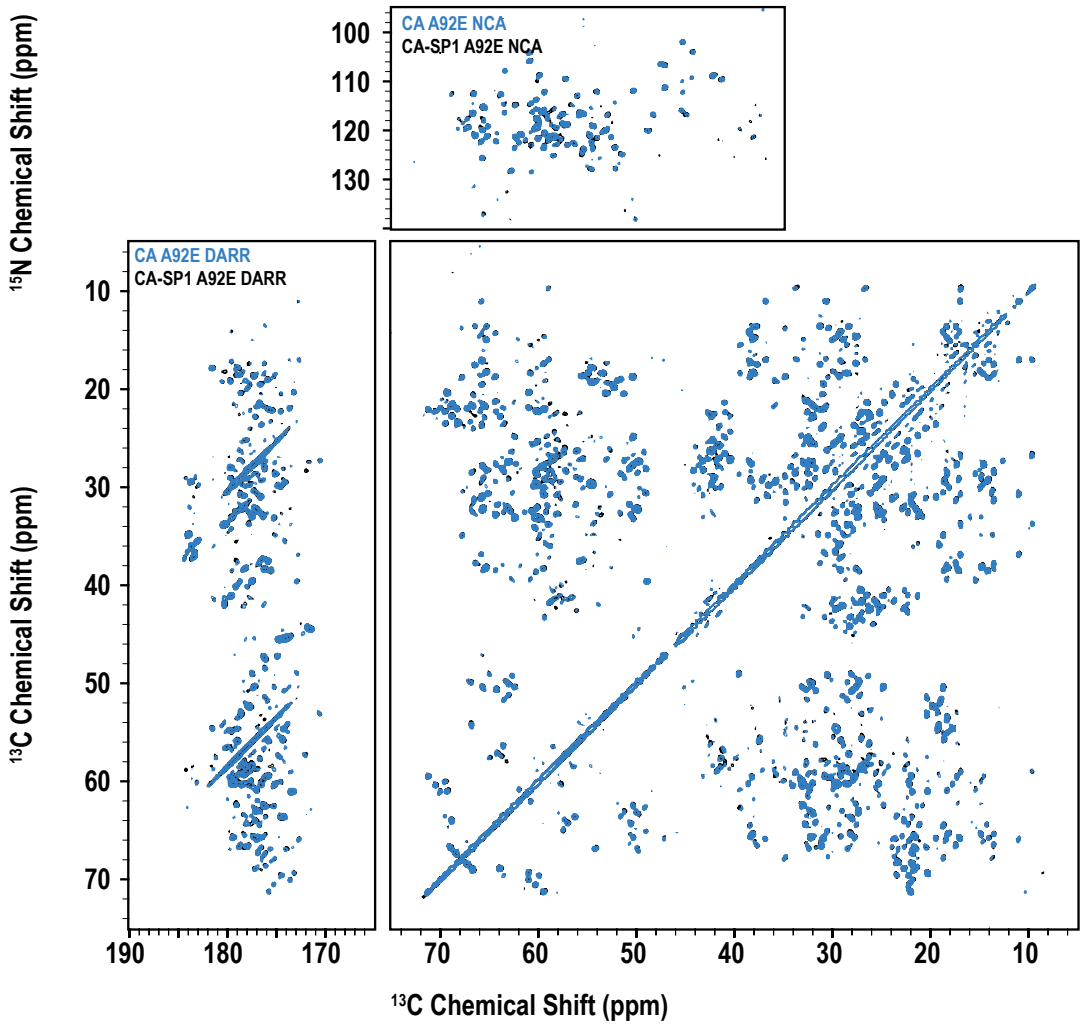


Figure 6.6 The superposition of the a) NCA and b) C^{α}-C' region, c) aliphatic region of DARR spectra of tubular assemblies acquired at 19.9 T and 4 °C. CA A92E is in blue and CA-SP1 A92E is in black. The DARR mixing time was 50ms. The acquisition parameters for the spectra of U-^{13}C,^{15}N HIV-1 CA-SP1 tubular assemblies are given in Table B.1 in the Appendix.

spectrum, two new strong peaks are present: one is in the alanine C^{α}-C^{β} cross peak region and the other is in the threonine C^{β}-C^{γ} cross peak region. According to the SP1 sequence (see Figure 6.2), there are two Ala and two Thr residues in the peptide. Therefore, it is

possible that these new peaks belong to the SP1 residues. With assistance of other types of spectra, these peaks were indeed assigned to the residues in the SP1 peptide, as discussed below.

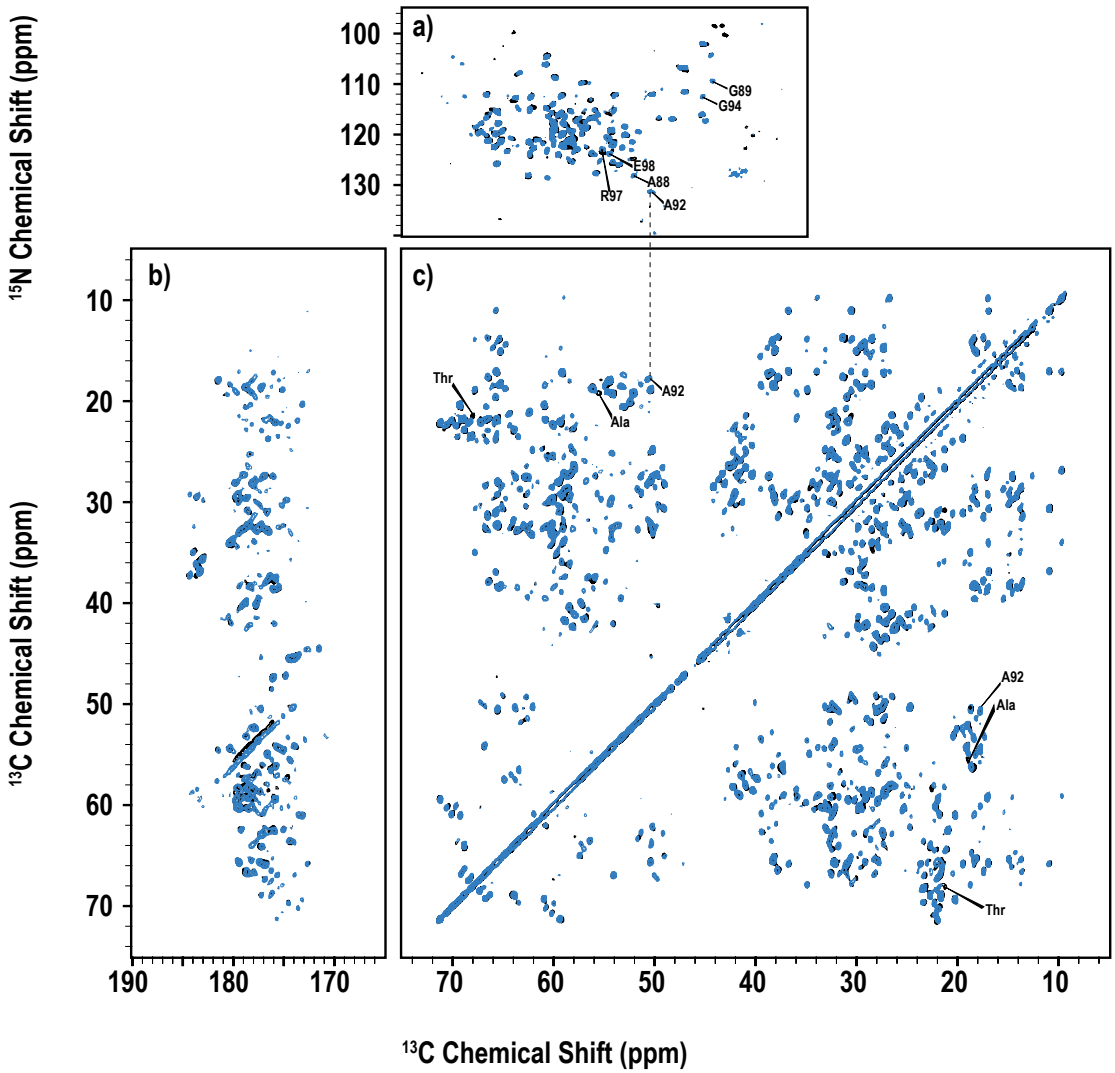


Figure 6.7 The superposition comparison of the a) NCA and b) C^{α}-C' region, c) aliphatic region of DARR spectra acquired at 21.1 T and 4 °C. CA WT is in blue and CA-SP1 WT is in black. The DARR mixing time was 50ms. The acquisition parameters for the spectra of U-^{13}C,^{15}N HIV-1 CA-SP1 tubular assemblies are given in Table B.1 in the Appendix.

As discussed above, a number of chemical shifts are different in the NCA spectra of the WT CA and CA-SP1. We have assigned most of the shifted peaks to the residues residing in the CypA loop region, such as R97 and E98. Interestingly, there are substantial variations in the peak intensities between the CA and CA-SP1 spectra, with the number of residues exhibiting strong peaks in the CA spectrum and the corresponding peaks severely attenuated or completely missing in the CA-SP1 data set, notably corresponding to the residues comprising the CypA loop- G89, A92, and G94. The A92 signals of the CA-SP1 assembly are also missing in the DARR spectrum.

The above observation of the peak intensity variations for the CypA loop residues is intriguing. As discussed earlier, CypA is a functionally important region of CA. It binds the cellular protein Cyclophilin A during the virus replication, and the mutations in the loop result in defective virus particles [31-33]. Our findings suggest that SP1 tail may make the CypA loop region more flexible in the CA-SP1 WT tubular assemblies compared to the corresponding CA assemblies. It is surprising that the SP1 tail has such significant effects on the residues that are 150 amino acids away in the primary sequence and structurally belonging to the two different domains. Even more surprising is the fact that for the CA A92E construct, no such destabilization of the CypA loop by the SP1 tail was observed. One possible reason for this is that the A92E mutation in the CypA loop stabilizes the loop so that the SP1 tail no longer influences its dynamics. This hypothesis is corroborated by the recent findings in our laboratory that the CypA loop dynamics is profoundly different in the various CA and CA-SP1 constructs, with the A92E mutant showing greatly attenuated mobility on the nano to microsecond timescales (to be reported elsewhere).

6.6.1.2 Nuclear Overhauser cross relaxation based experiments

The primary polarization transfer mechanism in the Nuclear Overhauser Effect Spectroscopy (NOESY) [34] is cross relaxation. NOESY-based transfer can be used in the cases where motional averaging is suspected to result in signal disappearance, such as in our studies of the CA-SP1 assemblies, where the SP1 tail may be dynamically disordered. We have therefore pursued the NOESY-DARR experiments in our attempt to detect the SP1 signals. The NOESY-DARR pulse sequence adapted to the MAS NMR conditions is shown Figure 6.8.

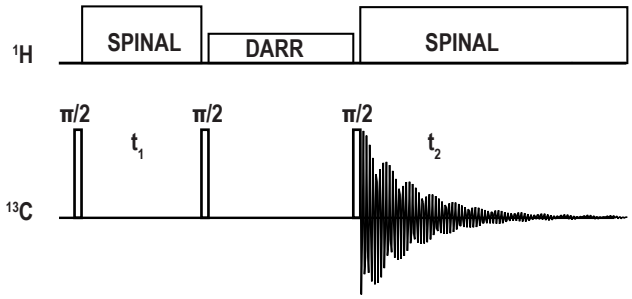


Figure 6.8 A scheme of NOESY-DARR pulse sequence. The carbon nuclei are first excited with a 90-degree pulse and the correlations were established through DARR irradiation during the mixing time [35].

As illustrated in Figure 6.9 showing the superposition of the NOESY-DARR spectra of CA and CA-SP1 A92E tubular assemblies, several new peaks emerge in the spectra of CA-SP1. All of these peaks correspond to the residue types that exist in the SP1 peptide, and the total count of new peaks reveals: 2 Thr (of 2 present in SP1), 1 Ala (of 2 present in SP1), 1 Ser (of 1 present in SP1), and 1 Asn (of 1 present in SP1). These new peaks are strong and appear on both sides of the diagonal. Besides these new peaks, no other differences in the spectra could be identified, corroborating the conclusion that

the CA protein exhibits the same or very similar structure in the tubular assemblies of A92E CA and CA-SP1.

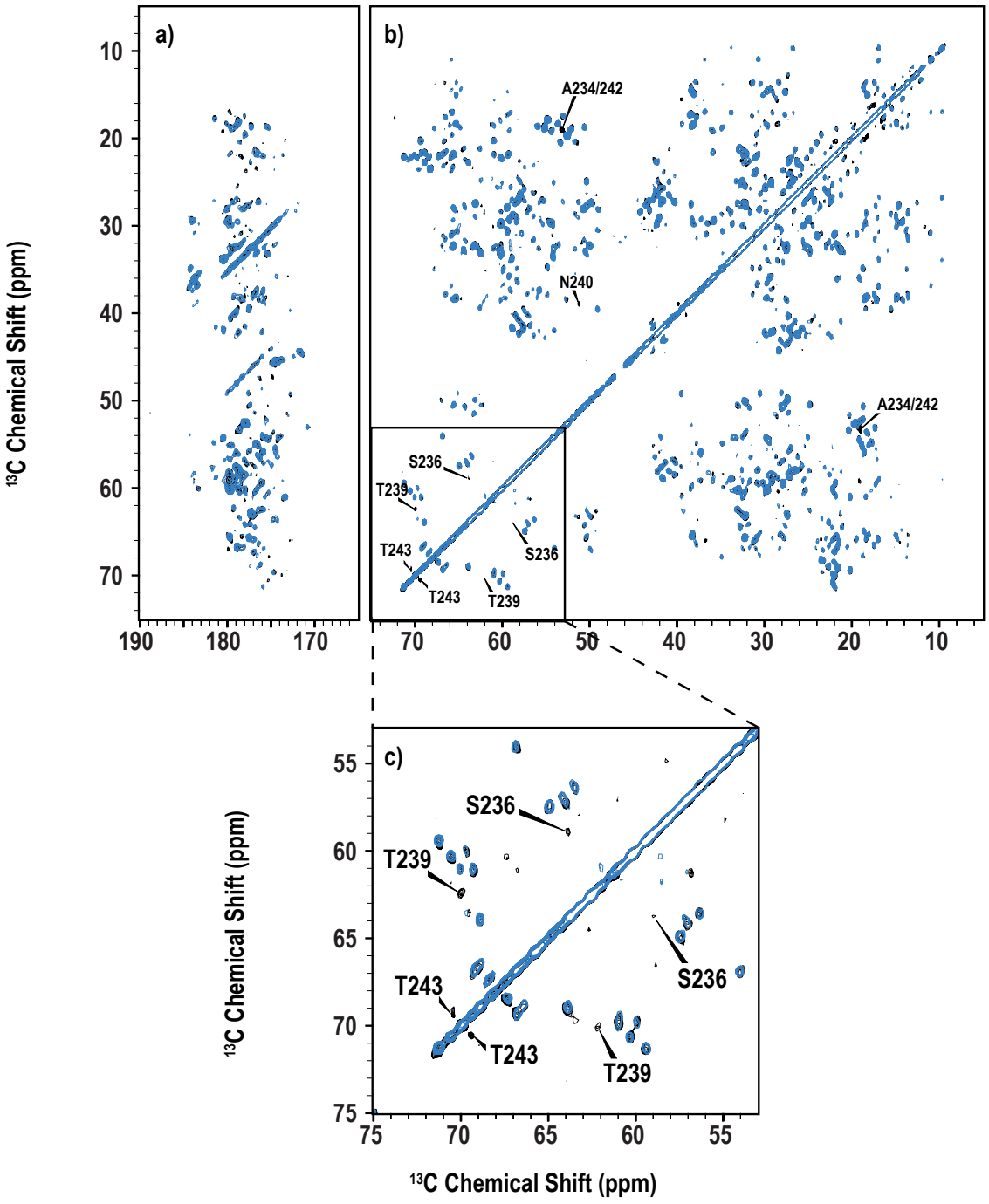


Figure 6.9 The superposition of the NOESY-DARR spectra of the tubular assemblies of CA A92E (blue) and CA-SP1 A92E (black) acquired at 19.9 T and 4 °C. a) The C^{α}-C' region of the spectra; b) the aliphatic region of the spectra; c) expansion of the aliphatic region of Thr-Ser residue types. The DARR mixing time was 50 ms. The acquisition parameters for the spectra are given in Table B.1 in the Appendix.

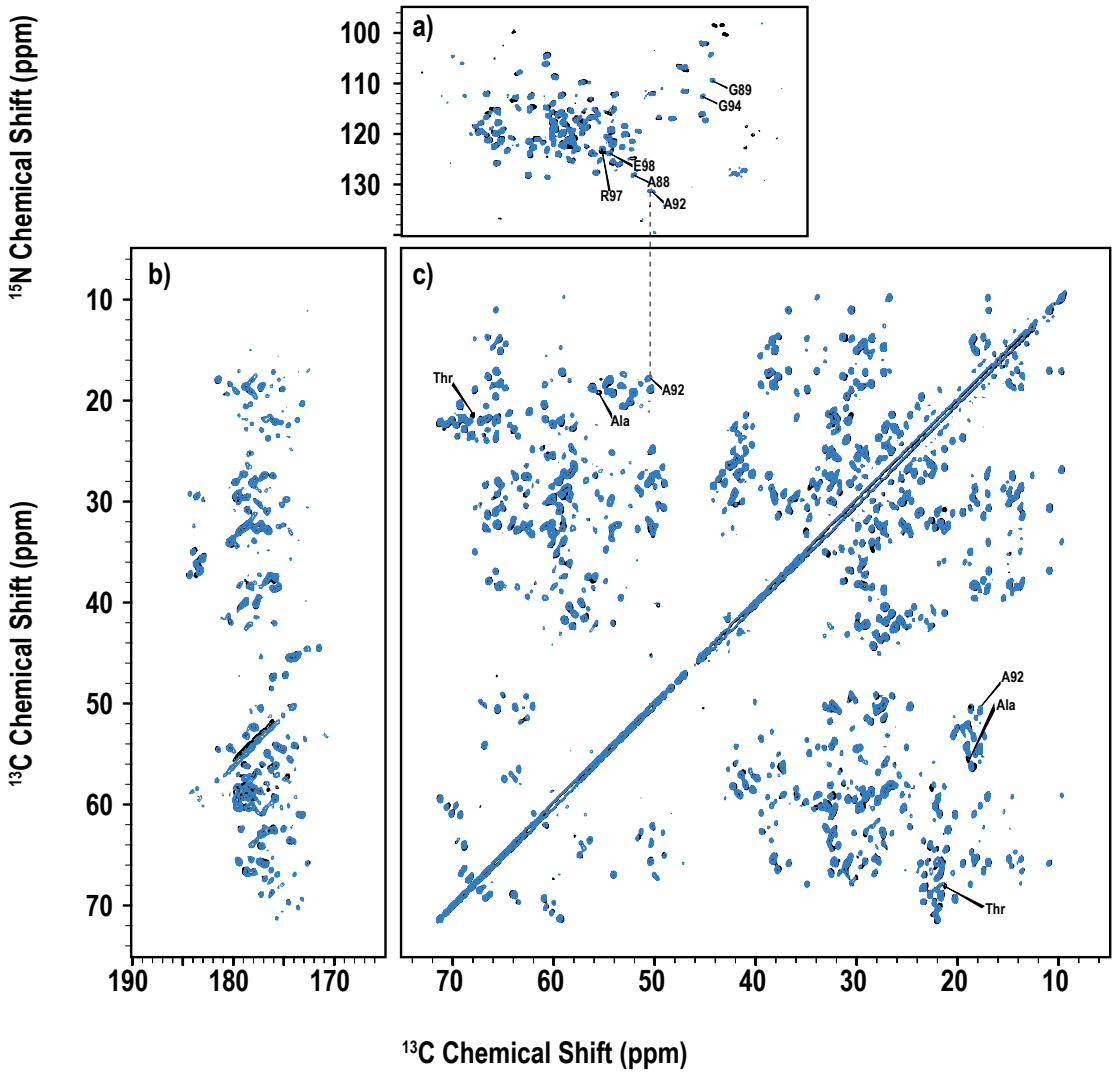


Figure 6.10 The superposition of the DARR (blue) and NOESY-DARR (black) spectra of CA-SP1 A92E acquired at 19.9 T and 4 °C. a) The C^{α}-C' region of the spectra; b) the aliphatic region of the spectra; c) expansion of the aliphatic region of Thr-Ser residue types. The DARR mixing times were 50 ms. The acquisition parameters for the spectra are given in Table B.1 in the Appendix.

We note that the new peaks belonging to SP1 identified in the NOESY-DARR spectra did not appear in the conventional CP-based DARR spectrum of CA-SP1 A92E, as shown in Figure 6.10. Even though most of the signals superimpose well, the NOESY-DARR spectrum exhibits more cross peaks in the C^{α}-C^{β} region than the DARR spectrum, especially in the C^{α}-C^{β} region of Ala. The removal of the cross polarization step in the pulse sequence makes it possible to detect these residues from the SP1 tail, indicating that they are intrinsically mobile. To corroborate this finding, we have next turned our attention to the scalar- (J-) based correlation experiments.

6.6.2 J-based correlation SSNMR spectra of HIV-1 CA and CA-SP1 assemblies

In scalar (J-coupling) based experiments, polarization is transferred through covalent bonds rather than through space. This transfer mechanism is much less sensitive to the presence of motions, and has been employed by us and by others [36-39] in the recent years to acquire MAS NMR homo- and heteronuclear correlation in a variety of systems. In our study of HIV-1 capsid protein assemblies, we employ a particular J-based experiment dubbed INADEQUATE to elucidate conformation of the SP1 peptide.

6.6.2.1 INADEQUATE and CP-INADEQUATE spectra of CA and CA-SP1 A92E tubular assemblies

INADEQUATE (Incredible Natural Abundance Double-Quantum Transfer Experiment) is one of the commonly used homonuclear pulse sequences in solution NMR to recouple the J-coupling between covalently bonded ^{13}C nuclei [40-42]. More recently, INADEQUATE was explored for solid-state NMR studies of organic and biological systems [43]. The INADEQUATE pulse sequence adapted for MAS NMR experiments

is shown in Figure 6.11. In the INADEQUATE spectra, each bonded pair of nuclei yields a pair of peaks, which have the vertical coordinate being the sum of the chemical shifts of the two nuclei (because of the double-quantum polarization transfer), and have horizontal coordinates being the respective chemical shifts of each nucleus.

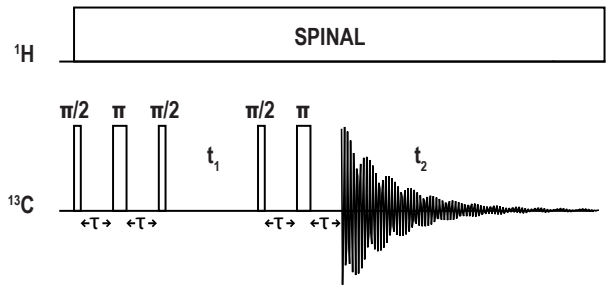


Figure 6.11 The INADEQUATE pulse sequence [43]. The carbon nuclei are first excited with a 90-degree pulse and double quantum coherences are then established through a composite of π and π/2 pulses.

The INADEQUATE spectra of CA and CA-SP1 A92E tubular assemblies were acquired with the same parameters (see Table B.1 in the Appendix). The superposition of the spectra is shown in Figure 6.12. Signals from the SP1 peptide emerged from the CA-SP1 spectrum as new peaks. The peaks corresponding to Ala, Asn, Ser, and Thr of the SP1 tail have identical chemical shifts to those in the NOESY-DARR spectra. In the INADEQUATE spectra, we have assigned all the new peaks to SP1 residues. The corresponding chemical shifts are summarized in Table 6.2. The secondary structure analysis according to the CSI method reveals that these chemical shifts correspond to unstructured random coil. Other residues in the SP1 tail, such as Gln, Glu, and Met, could not be assigned site-specifically on the basis of the INADEQUATE spectra because their chemical shifts are similar. However, there are four new peaks in this region of the spectrum, which is exactly the number of residues in the SP1 peptide sequence: one Glu,

one Gln, and two Met. Therefore, we attribute these four new peaks to the corresponding amino acids in the SP1 tail as well. Furthermore, the CSI analysis of these peaks indicates that their consensus CSI values correspond to unstructured random coil, consistent with what we found for Ala, Asn, Ser, and Thr. Other new peaks in the INADEQUATE spectrum of CA-SP1 are unambiguously assigned to S236, V238 and I244 in the SP1 tail, on the basis of their unique chemical shifts.

Despite the fact that there are 3 Ala residues in the SP1 tail, only one new peak in the corresponding region of the INADEQUATE spectra was observed. This new peak corresponds very well with the new peak in the NOESY-DARR spectrum. The absence of peaks corresponding to the other two Ala residues suggests that extensive local motions may have averaged the signals out. To search for these peaks, the spectra would need to be collected in a broad range of temperatures. However, as discussed in the previous chapters, tubular assemblies are stable only in a narrow temperature range above 0 ºC, precluding further investigations of the temperature dependence of the spectra.

Another interesting observation was made from the INADEQUATE spectra. For N240, the INADEQUATE spectra reveal two signals separated by 0.3 ppm, while single peaks appear in the NOESY-DARR spectra at the same position. Since the signals are strong and distinct and there is only one Asn residue in the sequence, those peaks are assigned to two N240 conformers.

Additional information can be obtained about the chemical exchange rates between the two conformers and the time scale of motions. The ^{13}C-^{13}C J-coupling is around 35 Hz [44], and the cross polarization rf field required to yield signals in the DARR spectra is around 30 kHz [45,46]. N240 appears as a broadened pair of peaks in

the INADEQUATE spectra, as a single peak in the NOESY-DARR spectra (Figure 6.12), and is completely missing in the DARR spectra (Figure 6.10). The difference in the resonance frequencies of the two conformers is ca. 60 Hz. Therefore, the INADEQUATE spectra detected the exchange between the two conformers, whereas the NOESY-DARR spectra detected the weighted average signal of the two conformers. For the absence of signals in the DARR spectra, it is possible that there are motions at the time scale of tens of microsecond that averaged out the ^{1}H-^{13}C dipolar couplings in the protein backbone. As discussed later, this motion is also confirmed by the lack of signals in the CP-INADEQUATE spectra, which include the cross polarization step in the sequence.

In the case of Thr, there are three resolved new peaks in the Thr C^{α} region of the CA-SP1 INADEQUATE spectra, two of which are double peaks separated by 0.5 ppm (~100 Hz). In comparison, a single peak appears in the corresponding region of the NOESY-DARR spectra, indicating chemical exchange between two conformers, which is similar to the finding for N240 discussed above. Those peaks are completely missing in the DARR spectra just like N240 signals are. Since T239 is residue that precedes N240 and the double peaks exhibited same behavior in the different spectra, we tentatively assigned the double peaks to the two conformers of T239 instead of T243. Correspondingly, the pair of peaks at 63.0 ppm and 71.0 ppm was assigned to the C^{α} and C^{β} of T243.

For other residues in the SP1 tail, most of the C^{α}-C^{β} coherence signals were observed as new peaks in the INADEQUATE spectra of CA-SP1. One exception is P241, whose C^{δ}-C^{γ} coherence was observed. Since P241 is the residue that follows

residue N240, we speculate that motions of the backbone atoms of this residue may have averaged the C^{α}-C^{β} signals out.

Interestingly, the INADEQUATE spectra of CA and CA-SP1 A92E shown in Figure 6.12 reveal that distinct peaks are present in the CA spectrum that do not appear in the CA-SP1 data set. Assigning those peaks is challenging on the basis of the 2D data sets alone. However, analysis of the chemical shifts with respect to the amino acid topologies indicates that these new peaks are from the stretch of residues H226-L231. This finding suggests that not only is the SP1 tail dynamic itself but it also modulates the dynamics of the nearby residues.

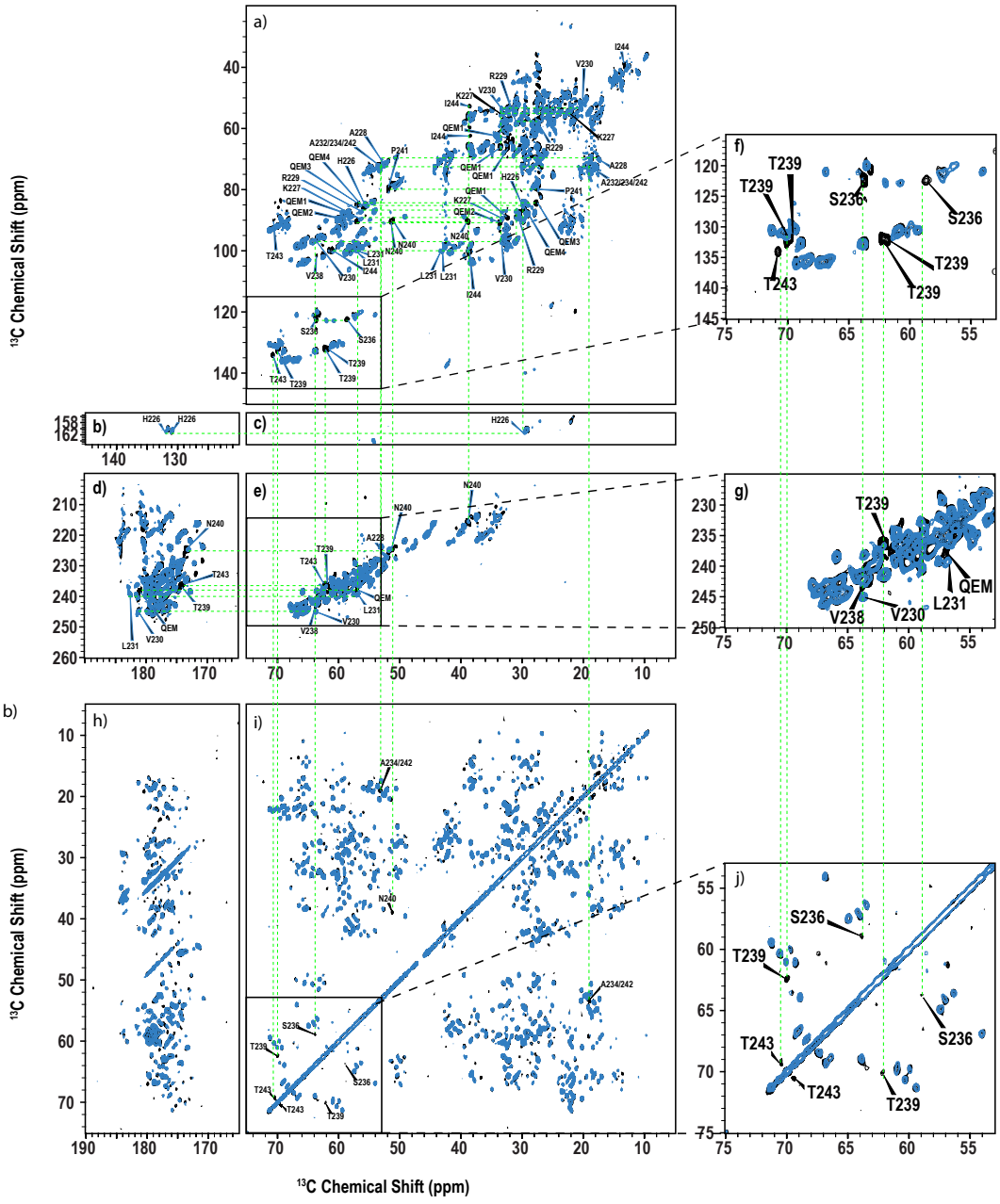


Figure 6.12 The superposition of the INADEQUATE and NOESY-DARR spectra of CA A92E (blue) and CA-SP1 A92E (black) acquired at 19.9 T and 4 °C. a) the aliphatic region; b) and c) the C^{β}-C^{γ} region of His residues; d) and e) the C^{α}-C' region; f) and g) the expansion of Thr, Ser regions; h) and i) the superposition of NOESY-DARR spectra that correspond well with those in the INADEQUATE spectra; j) the expansion in the Thr, Ser region of NOESY-DARR spectra. The delay τ in the INADEQUATE pulse sequence was optimized before each experiment was set up and the experiment time was kept the same for both samples. The acquisition parameters for the spectra are given in Table B.1 in the Appendix.

Table 6.2 Chemical shift assignments of the SP1 tail and nearby residues of CA and CA-SP1 A92E tubular assemblies. The CSI column is the secondary structure predicted by CSI method: H-α helix, B-β sheet, C-unstructured coil.

Residue	C^{α}	C^{β}	C^{γ}	C^{δ}	C'	CSI
G225						
H226	55.6	29.8	131.6			C
K227	56.5	33.3	22.0			C
A228	52.9	17.3			173.4	C
R229	56.3	30.9	27.4			C
V230	63.6	33.4	20.0		181.2	C
L231	56.7	43.4			182.4	H
A232	53.2	19.0				C
E233	57.2	33.5				C
	56.9	32.4				
	56.8	27.9				
	56.1	29.7				
A234	53.2	19.0				C
M235	57.2	33.5				C
	56.9	32.4				
	56.8	27.9				
	56.1	29.7				
S236	58.8	63.7				C
Q237	57.2	33.5				C
	56.9	32.4				
	56.8	27.9				
	56.1	29.7				
V238	63.6	37.9	179.8		179.8	C
T239	61.8	70.3			174.3	C
	62.3	69.6			173.8	
N240	51.2	38.9	177.2		173.2	C
	51.5	39.0				
P241			27.6	51.9		
A242	53.2	19.0				C
T243	69.4	70.5	19.3			C
I244	61.1	38.6	14.1			C
M245	57.2	33.5				C
	56.9	32.4				
	56.8	27.9				
	56.1	29.7				

In light of the observations discussed above, namely that we could detect the peaks corresponding to the SP1 tail and that the tail appears to be dynamic, we have decided to perform an additional CP-INADEQUATE experiment. In this experiment, a cross polarization period is incorporated prior to the INADEQUATE step [47]. For rigid residues, this additional ^{1}H-^{13}C CP period results in sensitivity enhancement. On the other hand, if residues are mobile on the time scale of ^{1}H-^{13}C dipolar interactions (~30 kHz), i.e. tens of microseconds, the CP does not work, and no signal can be detected.

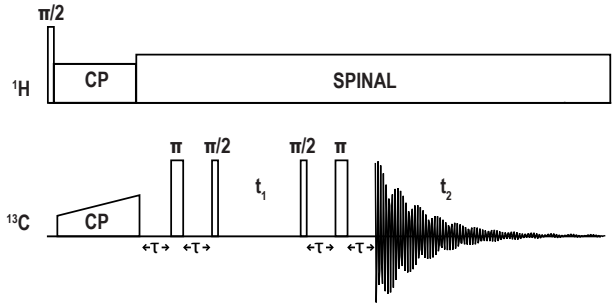


Figure 6.13 A scheme of CP-INADEQUATE pulse sequence. The carbon nuclei are first polarized by cross polarization, and the double quantum coherences are established through the INADEQUATE step.

Figure 6.14 shows the overlay of the CP-INADEQUATE and INADEQUATE spectra of the CA A92E tubular assemblies. The stretch of residues H226-L231 present in the INADEQUATE spectrum is missing in the CP-INADEQUATE data set. This finding, together with the lack of these signals in the DARR spectra corroborates our hypothesis that these residues are mobile on the timescales of the ^{1}H-^{13}C dipolar couplings, i.e. tens of microseconds. Furthermore, as discussed above and shown in Figure 6.12, the signals from the H226-L231 stretch of residues do not appear in the INADEQUATE spectra of CA-SP1 A92E either. It therefore appears that the attachment

of SP1 tail introduced slow motions to H226-L231, which occur at the time scale of J-coupling experiments, i.e. tens of milliseconds. It is not surprising because slow millisecond timescale motions were also observed in the SP1 residues, such as T239, N240 etc.

In summary, 12 out of 14 residues in the SP1 peptide have appeared in the INADEQUATE spectra of the tubular assemblies of CA-SP1 A92E. The secondary structure determined using the CSI method is random coil. The SP1 tail in the CA-SP1 A92E is mobile on the timescale of tens of microseconds. Some of the residues in the SP1 tail, such as T239 and N240, are mobile on the timescale of tens of milliseconds. Furthermore, the SP1 tail also affects the mobility of the nearby residues: residues H226-L231 in the CA-SP1 are mobile on both microsecond and millisecond timescales, while the same residues are mobile only on microsecond timescales in CA.

The CP-INADEQUATE spectra of the CA and CA-SP1 A92E samples did not show any differences (Figure 6.15), similar to the DARR spectra. Neither the signals from SP1 tail nor the signals from the H226-L231 stretch appeared in the CP-INADEQUATE spectra, confirming the mobility of those residues on the timescale of tens of microseconds.

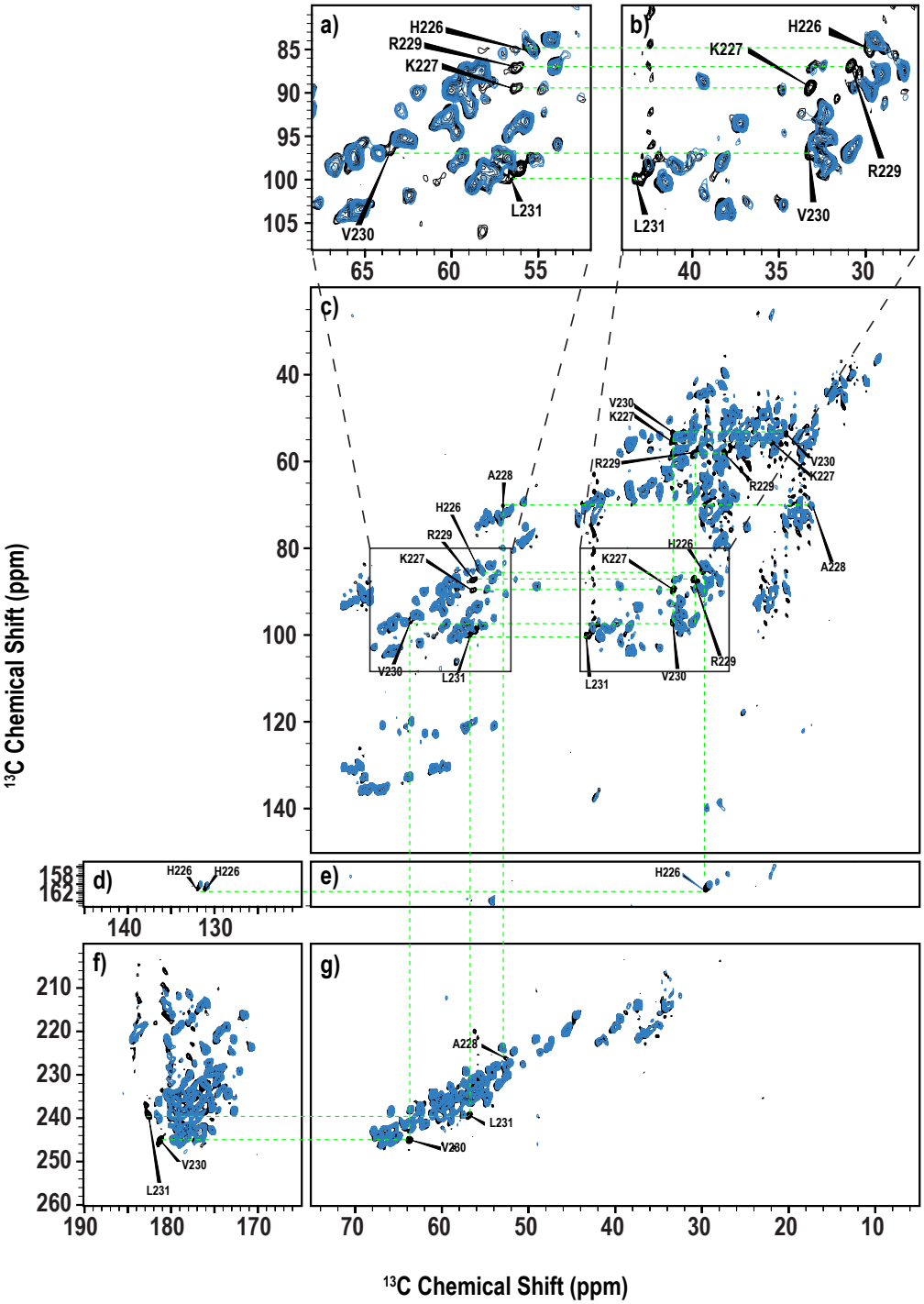


Figure 6.14 The superposition of the CP-INADEQUATE (blue) and INADEQUATE spectra (black) of CA A92E tubular assemblies acquired at 19.9 T and 4 °C. a) and b) The expansion of the region that extra peaks arise; c) the aliphatic region; d) and e) the C^{β}-C^{γ} region of His residues; f) and g) the C^{α}-C' region. The delay τ in the CP-INADEQUATE pulse sequence was optimized in each experiment, and the experiment time was kept the same for both samples. The acquisition parameters for the spectra are given in Table B.1 in the Appendix.

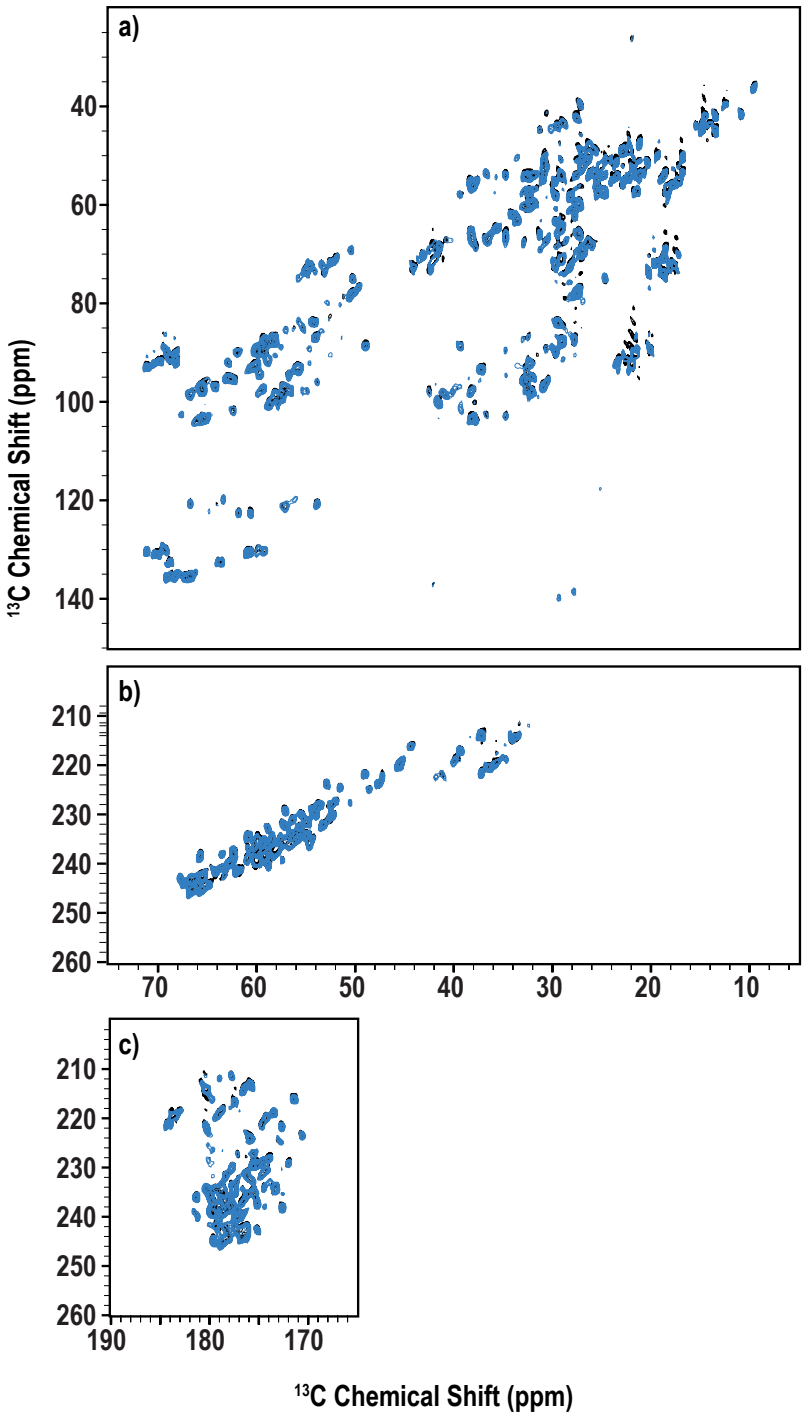


Figure 6.15 The superposition of the CP-INADEQUATE spectra of CA A92E (blue) and CA-SP1 A92E (black) acquired at 19.9 T and 4 °C. a) The aliphatic region; b) and c) the C_{α}-C' region. The signal-to-noise ratios of the spectra are the same in both samples. The acquisition parameters are given in Table B.1 in the Appendix.

6.6.2.2 INADEQUATE and CP-INADEQUATE spectra of CA WT and CA-SP1 WT tubular assemblies

As discussed earlier in this chapter and shown in Figure 6.7, signals from the SP1 peptide emerged in the through-bond correlation spectra of the CA-SP1 WT assemblies. To get additional information on the conformation of the SP1 tail, we have acquired INADEQUATE and CP-INADEQUATE spectra (see Figures 6.11 and 6.13 for the corresponding pulse sequences).

The INADEQUATE spectra of CA and CA-SP1 tubular assemblies superimpose very well (Figure 6.16), indicating that the overall structure is very similar. Unlike the CA-SP1 A92E tubular assemblies, only a few signals from SP1 residues appear in the CA-SP1 WT sample. The resonance assignments are summarized in Table 6.3. One set of new peaks appearing at the chemical shifts of (C^{α} 57.2 ppm, C^{β} 33.6 ppm, C^{γ} 33.7 ppm, C^{δ} 177.4 ppm) corresponds to E233 or Q237 in the SP1 tail. Another set of Thr peaks appears at the same position as T239 in the CA-SP1 A92E protein assemblies, and this set arises from two conformers of T239. The signals from other residues in the SP1 tail may have been averaged out by the motions or shifted to the congested areas in the spectra. According to the CSI method, the chemical shifts of the assigned residues in the SP1 tail correspond to the unstructured random coil conformation, which is consistent with our findings for the CA-SP1 A92E tubular assemblies (Table 6.3).

The signals of H226, K227, R229 and L231 appeared in the CA (but not CA-SP1) WT spectra at exactly the same positions as those in the CA A92E spectra, and we assigned these peaks using assignments for the CA A92E sample (Table 6.3). In the stretch of residues H226-L231 in the CA protein sequence, the signals of A228 and V230

did not appear at the expected positions, which indicate that these signals may have been averaged out by motions. According to the assignments, the conformation of SP1 tail and nearby residues is the same as in the A92E sample but the dynamics is different and warrants further exploration in the future.

Similar to the A92E samples, we have acquired and compared the CP-INADEQUATE spectra of CA and CA-SP1 WT tubular assemblies. As illustrated in Figure 6.17, the spectra agree with each other well for the majority of the peaks, especially in the Thr and Ser regions. However, some differences could be seen in certain regions of the two data sets (see Figure 6.17), suggesting variations in the structure and dynamics of the two protein constructs. It is not possible to specifically analyze those fine structure variations with current data sets, and additional measurements will be performed in the future.

As shown in Figure 6.18, the peaks of R229, K227 and L231 appear exclusively in the INADEQUATE (but not CP-INADEQUATE) spectra of CA, similar to our findings for the CA A92E sample. As discussed above, this observation indicates that these residues are mobile on the time scale of tens of microseconds. Interestingly, the peak corresponding to H226 appeared in both CP-INADEQUATE and INADEQUATE spectra of the CA WT sample, in contrast to CA A92E. Moreover, the A228 and V230 CP-INADEQUATE signals of CA WT sample did not appear at the same positions as in the CA A92E spectra. Therefore, the A92E mutation and sequence variability (HXB2 vs. NL4-3) have pronounced effects on the structure and dynamics of those residues.

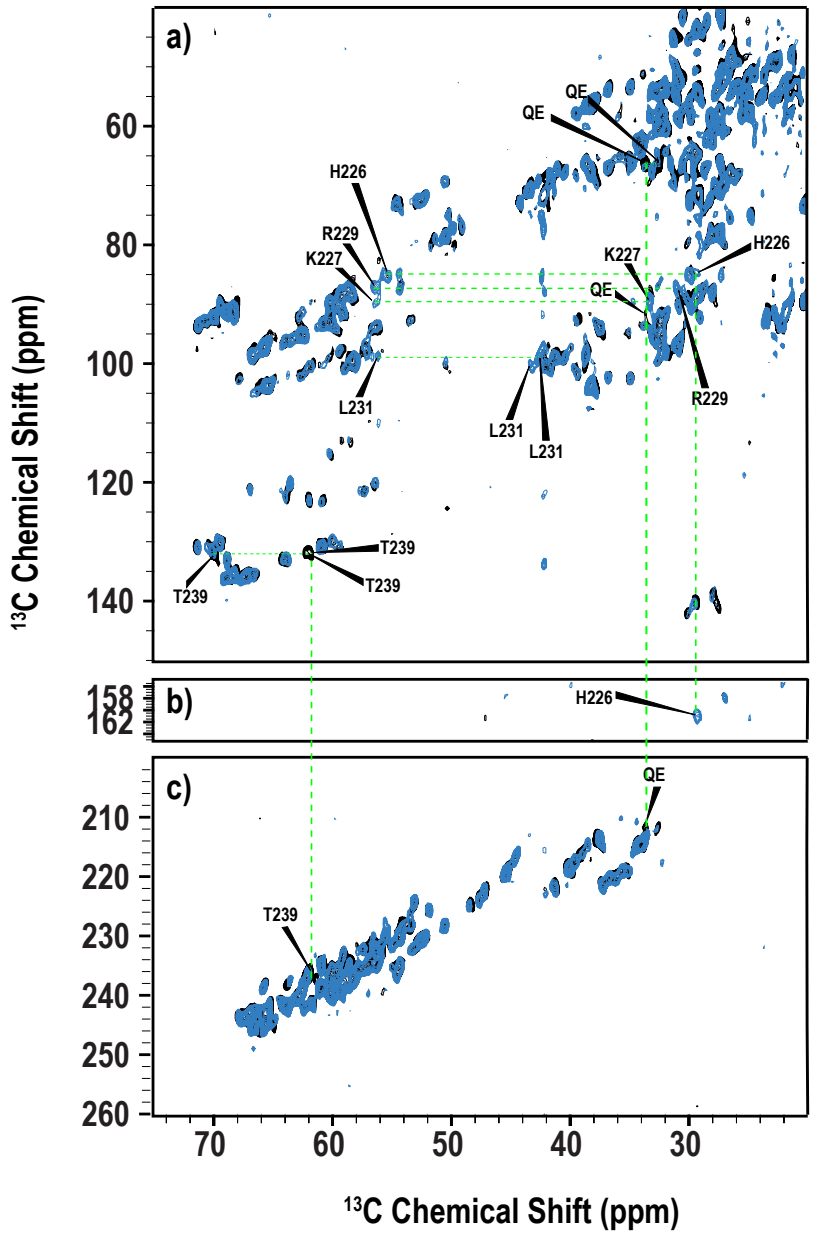


Figure 6.16 The comparison of the INADEQUATE spectra of CA WT (blue) and CA-SP1 WT (black) acquired at 21.1 T and 4 °C. a) The aliphatic region; b) the C^{β}-C^{γ} region of His residues; c) the C^{α}-C' region. The delay in the INADEQUATE pulse sequence was optimized before each experiment was set up and the experiment time was kept the same for both samples. The acquisition parameters for the spectra are given in Table B.1 in the Appendix.

Table 6.3 Chemical shifts of the SP1 tail and nearby residues of CA-SP1 WT tubular assemblies. The CSI column is the secondary structure predicted by CSI method: H-α helix, B-β sheet, C-unstructured coil.

Residue	C^{α}	C^{β}	C^{γ}	C^{δ}	C'	CSI
G225						
H226	55.3	29.4	131.5			C
K227	56.5	33.3				C
A228						
R229	56.3	30.9				C
V230						
L231	56.7	43.4				C
	56.2	42.5				
A232	55.6	19.0				C
E233	57.2	33.6	33.7	177.4		C
A234	48.4	16.8				
M235						
S236						
Q237	57.2	33.6	33.7	177.4		C
V238						
T239	61.8	70.3			174.3	C
	62.3	69.6			174.3	
N240						
P241						
A242	55.6	19.0				C
T243						
I244						
M245						

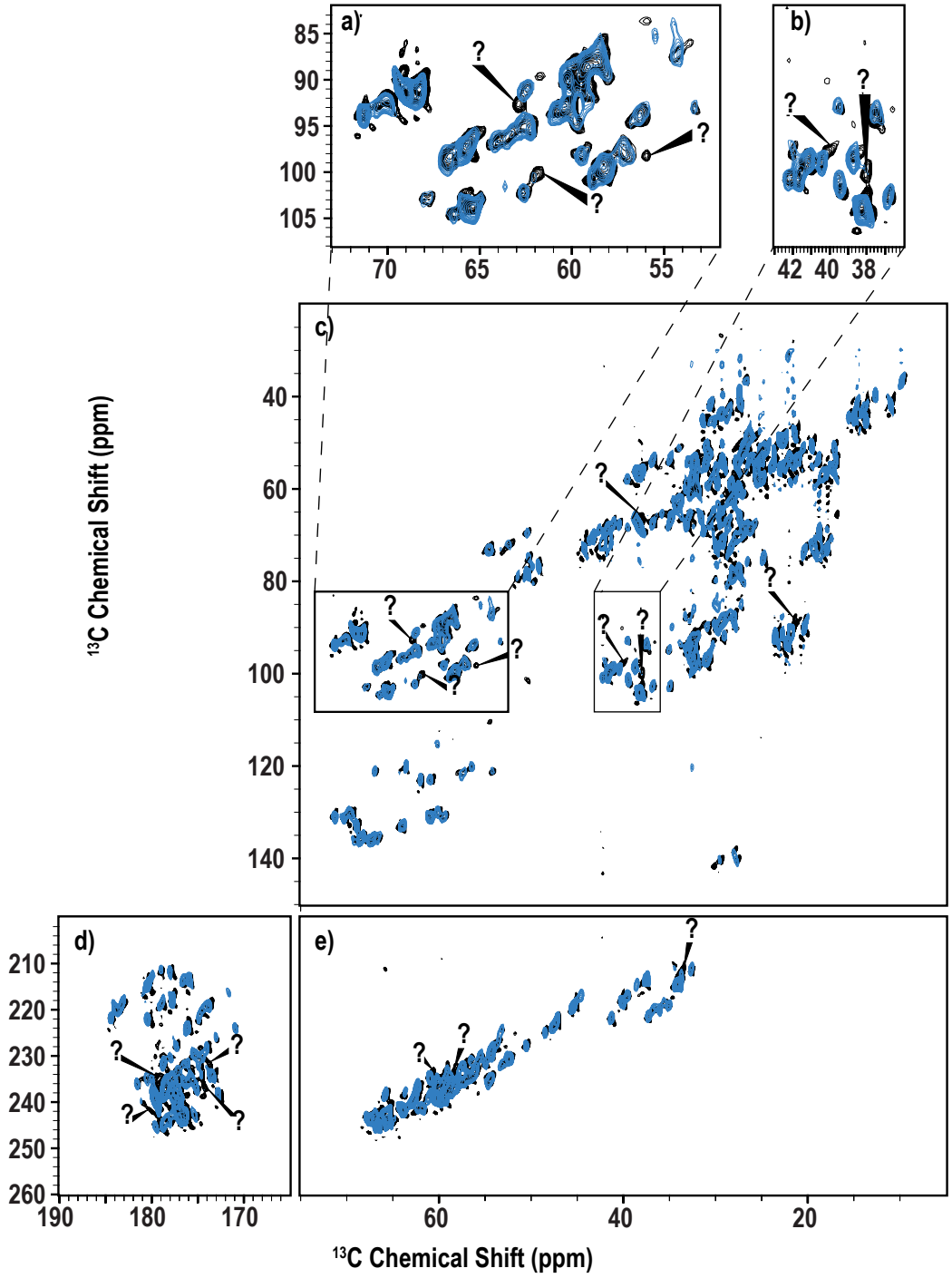


Figure 6.17 The superposition of the CP-INADEQUATE spectra of CA WT (blue) and CA-SP1 WT (black) acquired at 21.1 T and 4 °C. a) and b) The expansion of the region that extra peaks arise; c) the aliphatic region; d) and e) the C^{α}-C' region. The delay in the CP-INADEQUATE pulse sequence was optimized in each experiment, and the experiment time was the same for both samples. The acquisition parameters for the spectra are given in Table B.1 in the Appendix.

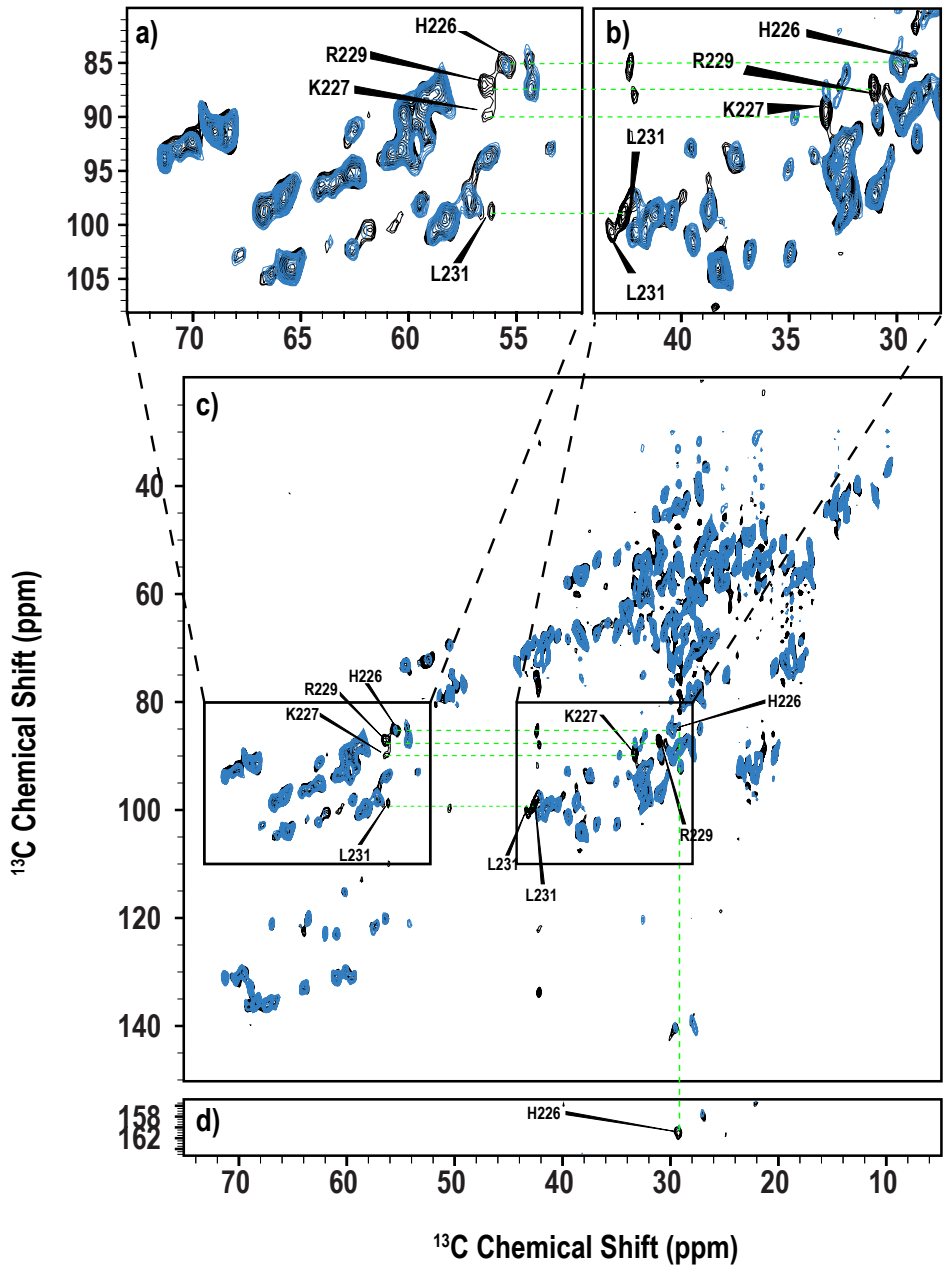


Figure 6.18 The superposition of the CP-INADEQUATE (blue) and INADEQUATE (black) spectra of CA WT tubular assemblies acquired at 21.1 T and 4 °C. a) and b) The expansion of the region that extra peaks arise; c) the aliphatic region; d) the C^{β}-C^{γ} region of His residues. The signal-to-noise ratios of the spectra are the same for both samples. The acquisition parameters for the spectra are given in Table B.1 in the Appendix.

6.7 The effect of mutation and variability in the protein sequence on the chemical shifts

As discussed in the previous chapters and shown in Figure 6.2, the protein constructs from HXB2 and NL4-3 strains of HIV-1 are different by four residues (I/L6, V/L83, N/H120, and A/G208), and additionally A92E mutation is introduced into the NL4-3 strain. The logical question is whether this makes any difference to the structure and dynamics of the corresponding proteins. To address this question, we compared the spectra of the CA WT (HXB2 strain) and CA A92E (NL4-3 strain) tubular assemblies. In the NCA and NCACX spectra of the two CA samples, the signals corresponding to A92 (HXB2 strain) and E92 (NL4-3 strain) appear as well-resolved strong peaks, as illustrated in Figure 6.19. Similarly, the signals corresponding to the variable residues are distinct and well resolved in the respective spectra, except for the signal of A208 that is missing. We have carried out the resonance assignments with sequential backbone walks, and the chemical shifts agree well with the solution NMR assignments of the separate domains of the corresponding constructs. It is interesting that in addition to the mutated and variation sites, other shifted peaks are also observed. According to the results of the sequential assignments, these shifted peaks correspond to the residues in the CypA loop belonging to the NTD and the loop preceding Helix 8 in the CTD. With respect to the CypA loop residues, since A92E and V/L83 both reside in this region, it is not surprising that the chemical environment has changed due to the mutation and sequence variability. On the other hand, the observation of chemical shift perturbations for the residues located in the loop region preceding Helix 8 is quite surprising as the nearest variation residues are N/H120 and A/G208, which are ca. 50 residues away in the

primary sequence. At the same time, it is clear from the 3D structure model shown in Figure 6.19 that the variation site A/G208 is close in space to this loop region preceding Helix 8. Therefore, the substitution of Ala to Gly in residue 208 modulates the conformation of this loop by changing the local environment in the vicinity of the loop.

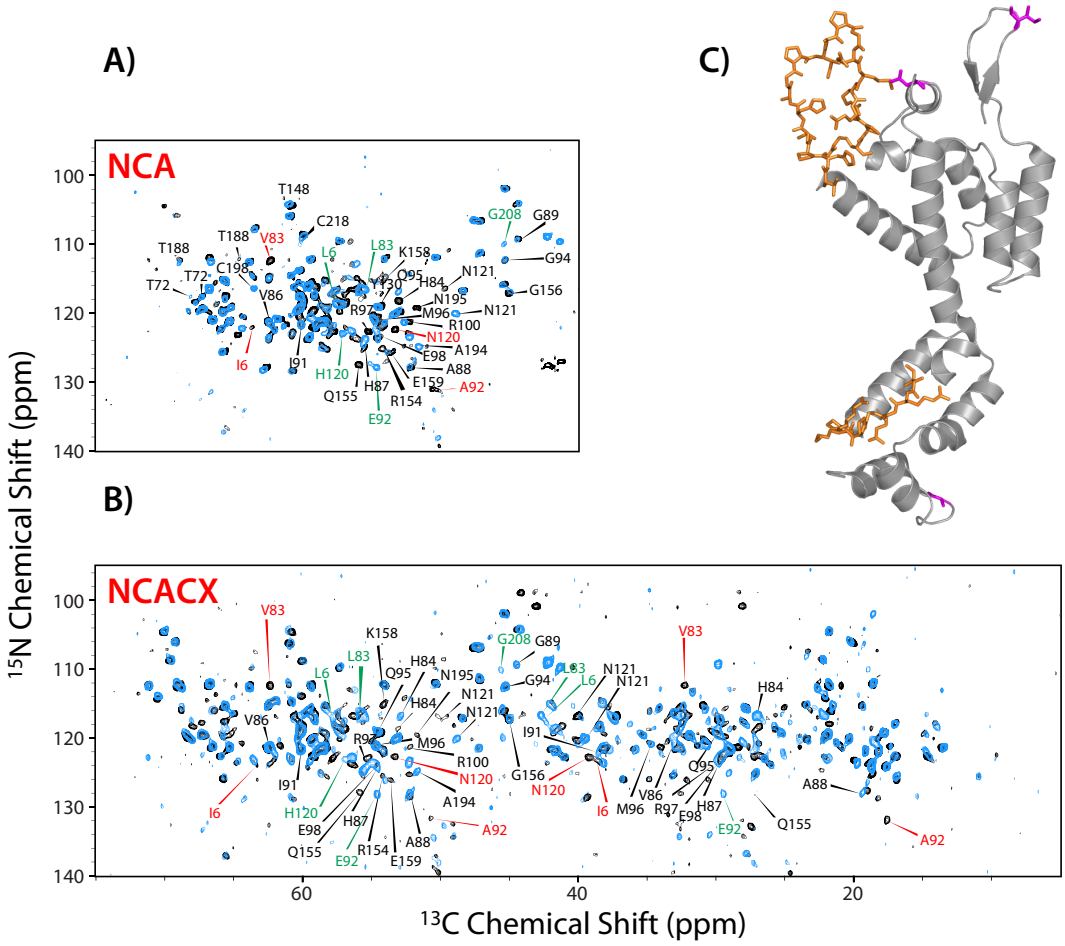


Figure 6.19 The superposition of the NCA (A) and NCACX (B) spectra of CA NL4-3 A92E (blue) and CA HXB-2 (black). The green and red labels represent signals from the distinct residues in the NL4-3 A92E and HXB-2 variants, respectively. (C) A view of the 3D structure of CA (HXB2 variant, PDB file 3NTE). Colored magenta are the residues that vary between the two constructs; in orange shown are the residues that are identical in the two variants but whose chemical shifts are perturbed.

6.8 Conclusions

Tubular assemblies of HIV-1 capsid (CA) and capsid-spacer peptide 1 (CA-SP1) proteins yielded unprecedented high-quality homo- and heteronuclear MAS NMR correlation spectra at ultrahigh magnetic fields, enabling their structural and dynamics investigations at atomic resolution. The analysis of the through-space and through-bond correlation spectra unequivocally indicates that the fourteen residues comprising the SP1 tail of CA-SP1 A92E and CA-SP1 WT tubular assemblies are unstructured in the assembled state, and that the SP1 tail as well as the nearby residues are mobile. The analysis of ^{13}C and ^{15}N chemical shifts indicates that the variation in the protein sequence has significant effect on the CA conformation. These findings warrant further comprehensive investigations into the conformation and dynamics (rates and amplitudes of motions) on multiple time scales of CA and CA-SP1 assemblies.

References

[1] C. S. Adamson; E. O. Freed. Recent progress in antiretrovirals--lessons from resistance. *Drug Discov. Today,* 13(9-10): 424-432, 2008.

[2] C. S. Adamson; K. Salzwedel; E. O. Freed. Virus maturation as a new HIV-1 therapeutic target. *Expert Opin. Ther. Targets,* 13(8): 895-908, 2009.

[3] C. Liang; J. Hu; R. S. Russell; A. Roldan; L. Kleiman; M. A. Wainberg. Characterization of a putative alpha-helix across the capsid-SP1 boundary that is critical for the multimerization of human immunodeficiency virus type 1 Gag. *J. Virol.,* 76(22): 11729-11737, 2002.

[4] D. K. Worthylake; H. Wang; S. Yoo; W. I. Sundquist; C. P. Hill. Structures of the HIV-1 capsid protein dimerization domain at 2.6 A resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.,* 55(Pt 1): 85-92, 1999.

[5] J. L. Newman; E. W. Butcher; D. T. Patel; Y. Mikhaylenko; M. F. Summers. Flexibility in the P2 domain of the HIV-1 Gag polyprotein. *Protein Sci.,* 13(8): 2101-2107, 2004.

[6] E. R. Andrew; A. Bradbury; R. G. Eades. Nuclear Magnetic Resonance Spectra from a Crystal Rotated at High Speed. *Nature,* 182(4650): 1659-1659, 1958.

[7] J. A. Stringer; C. E. Bronnimann; C. G. Mullen; D. H. H. Zhou; S. A. Stellfox; Y. Li; E. H. Williams; C. M. Rienstra. Reduction of RF-induced sample heating with a scroll coil resonator structure for solid-state NMR probes. *J. Magn. Reson.,* 173(1): 40-48, 2005.

[8] A. Bockmann; A. Lange; A. Galinier; S. Luca; N. Giraud; M. Juy; H. Heise; R. Montserret; F. Penin; M. Baldus. Solid state NMR sequential resonance assignments and conformational analysis of the 2 x 10.4 kDa dimeric form of the Bacillus subtilis protein Crh. *J. Biomol. NMR,* 27(4): 323-339, 2003.

[9] B. Chen; R. Tycko. Structural and dynamical characterization of tubular HIV-1 capsid protein assemblies by solid state nuclear magnetic resonance and electron microscopy. *Protein Sci.,* 19(4): 716-730, 2010.

[10] Y. Han; J. Ahn; J. Concel; I. J. Byeon; A. M. Gronenborn; J. Yang; T. Polenova. Solid-state NMR studies of HIV-1 capsid protein assemblies. *J. Am. Chem. Soc.,* 132(6): 1976-1987, 2010.

[11] Y. Han; J. Ahn; J. Concel; I. J. L. Byeon; A. M. Gronenborn; J. Yang; T. Polenova. Solid-state NMR studies of HIV-1 capsid protein assemblies. *J. Am. Chem. Soc.,* 132(6): 1976-1987, 2010.

[12] P. L. Gor'kov; E. Y. Chekmenev; C. G. Li; M. Cotten; J. J. Buffy; N. J. Traaseth; G. Veglia; W. W. Brey. Using low-E resonators to reduce RF heating in biological samples for static solid-state NMR up to 900 MHz. *J. Magn. Reson.,* 185(1): 77-93, 2007.

[13] S. A. Mcneill; P. L. Gor'kov; K. Shetty; W. W. Brey; J. R. Long. A low-E magic angle spinning probe for biological solid state NMR at 750 MHz. *J. Magn. Reson.,* 197(2): 135-144, 2009.

[14] L. Ratner; A. Fisher; L. L. Jagodzinski; H. Mitsuya; R. S. Liou; R. C. Gallo; F. Wong-Staal. Complete nucleotide sequences of functional clones of the AIDS virus. *AIDS Res. Hum. Retroviruses,* 3(1): 57-69, 1987.

[15] A. Adachi; H. E. Gendelman; S. Koenig; T. Folks; R. Willey; A. Rabson; M. A. Martin. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J. Virol.,* 59(2): 284-291, 1986.

[16] S. Sun; S. Paramasivam; S. Yan; Y. Han; C. Lightcap; J. C. Williams; I. J. L. Byeon; J. Ahn; A. M. Gronenborn; T. Polenova In Protein NMR Techniques; Shekhtman, A., Burz, D., Eds.; Humana Press: Louisville, 2011.

[17] S. Sun; Y. Han; S. Paramasivam; S. Yan; A. E. Siglin; J. C. Williams; I. J. Byeon; J. Ahn; A. M. Gronenborn; T. Polenova. Solid-state NMR spectroscopy of protein complexes. *Methods Mol. Biol.,* 831303-331,

[18] I. J. L. Byeon; X. Meng; J. W. Jung; G. P. Zhao; R. F. Yang; J. W. Ahn; J. Shi; J. Concel; C. Aiken; P. J. Zhang; A. M. Gronenborn. Structural Convergence between Cryo-EM and NMR Reveals Intersubunit Interactions Critical for HIV-1 Capsid Function. *Cell,* 139(4): 780-790, 2009.

[19] T. I. Igumenova; A. E. Mcdermott; K. W. Zilm; R. W. Martin; E. K. Paulson; A. J. Wand. Assignments of carbon NMR resonances for microcrystalline ubiquitin. *J. Am. Chem. Soc.,* 126(21): 6720-6727, 2004.

[20] G. Comellas; J. J. Lopez; A. J. Nieuwkoop; L. R. Lemkau; C. M. Rienstra. Straightforward, effective calibration of SPINAL-64 decoupling results in the enhancement of sensitivity and resolution of biomolecular solid-state NMR. *J. Magn. Reson.,* 209(2): 131-135, 2011.

[21] D. Marion; K. Wüthrich. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of H1-H1 spin-spin coupling-constants in proteins. *Biochem. Biophys. Res. Commun.,* 113(3): 967-974, 1983.

[22] D. Marion; M. Ikura; R. Tschudin; A. Bax. Rapid Recording of 2d NMR-Spectra without Phase Cycling - Application to the Study of Hydrogen-Exchange in Proteins. *J. Magn. Reson.,* 85(2): 393-399, 1989.

[23] F. Delaglio; S. Grzesiek; G. W. Vuister; G. Zhu; J. Pfeifer; A. Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR,* 6(3): 277-293, 1995.

[24] T. D. Goddard; D. G. Kneller. SPARKY 3.0. 1998.

[25] G. Neue; C. Dybowski. Determining temperature in a magic-angle spinning probe using the temperature dependence of the isotropic chemical shift of lead nitrate. *Solid State Nucl. Magn. Reson.,* 7(4): 333-336, 1997.

[26] I. J. Byeon; X. Meng; J. Jung; G. Zhao; R. Yang; J. Ahn; J. Shi; J. Concel; C. Aiken; P. Zhang; A. M. Gronenborn. Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell,* 139(4): 780-790, 2009.

[27] J. Lanman; T. T. Lam; S. Barnes; M. Sakalian; M. R. Emmett; A. G. Marshall; P. E. Prevelige. Identification of novel interactions in HIV-1 capsid protein assembly

by high-resolution mass spectrometry (vol 325, pg 759, 2003). *J. Mol. Biol.,* 334(5): 1133-1133, 2003.

[28] H. C. Wong; R. Shin; N. R. Krishna. Solution structure of a double mutant of the carboxy-terminal dimerization domain of the HIV-1 capsid protein. *Biochemistry,* 47(8): 2289-2297, 2008.

[29] F. Delaglio; S. Grzesiek; G. W. Vuister; G. Zhu; J. Pfeifer; A. Bax. NMRPipe- a multidimensional spectral processing system based on Unix pipes. *J. Biomol. NMR,* 6(3): 277-293, 1995.

[30] J. Jung; I. J. L. Byeon; J. Ahn; J. Concel; A. M. Gronenborn. H-1, N-15 and C-13 assignments of the dimeric C-terminal domain of HIV-1 capsid protein. *Biomol. NMR Assign.,* 4(1): 21-23, 2010.

[31] T. D. Goddard; D. G. Kneller. SPARKY 3, University of California, San Francisco.

[32] S. Sun; A. Siglin; J. C. Williams; T. Polenova. Solid-state and solution NMR studies of CAP-Gly domain of mammalian dynactin and its interaction with microtubules. *J. Am. Chem. Soc.,* 13110113-10126, 2009.

[33] W. T. Franks; D. H. Zhou; B. J. Wylie; B. G. Money; D. T. Graesser; H. L. Frericks; G. Sahota; C. M. Rienstra. Magic-angle spinning solid-state NMR spectroscopy of the beta 1 immunoglobulin binding domain of protein G (GB1): N-15 and C-13 chemical shift assignments and conformational analysis. *J. Am. Chem. Soc.,* 127(35): 12291-12305, 2005.

[34] D. S. Wishart; B. D. Sykes. The C-13 Chemical-Shift Index- a simple method for the identification of protein secondary structure using C-13 chemical shift data. *J. Biomol. NMR,* 4(2): 171-180, 1994.

[35] J. Jeener; Broekaer.P. Nuclear magnetic resonance in solids - thermodynamic effects of a pair of Rf pulses. *Phys. Rev.,* 157(2): 232-&, 1967.

[36] L. L. Chen; J. M. Kaiser; J. F. Lai; T. Polenova; J. Yang; C. M. Rienstra; L. J. Mueller. J-based 2D homonuclear and heteronuclear correlation in solid-state proteins. *Magn. Reson. Chem.,* 45S84-S92, 2007.

[37] L. Chen; J. M. Kaiser; T. Polenova; J. Yang; C. M. Rienstra; L. J. Mueller. Backbone assignments in solid-state proteins using J-based 3D Heteronuclear correlation spectroscopy. *J. Am. Chem. Soc.,* 129(35): 10650-10651, 2007.

[38] W. C. Lai; N. Mclean; A. Gansmuller; M. A. Verhoeven; G. C. Antonioli; M. Carravetta; L. Duma; P. H. M. Bovee-Geurts; O. G. Johannessen; H. J. M. De Groot; J. Lugtenburg; L. Emsley; S. P. Brown; R. C. D. Brown; W. J. Degrip; M. H. Levitt. Accurate measurements of C-13-C-13 J-couplings in the rhodopsin chromophore by double-quantum solid-state NMR spectroscopy. *J. Am. Chem. Soc.,* 128(12): 3878-3879, 2006.

[39] P. Schanda; M. Huber; R. Verel; M. Ernst; B. H. Meier. Direct detection of (3h)J(NC') hydrogen-bond scalar couplings in proteins by solid-state NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.,* 48(49): 9322-9325, 2009.

[40] D. S. Wishart; B. D. Sykes; F. M. Richards. The Chemical-Shift Index- a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry,* 31(6): 1647-1651, 1992.

[41] J. Lanman; T. T. Lam; S. Barnes; M. Sakalian; M. R. Emmett; M. A.; P. E. J. Prevelige. *J. Mol. Biol.,* 325759-772, 2003.

[42] T. R. Gamble; F. F. Vajdos; S. H. Yoo; D. K. Worthylake; M. Houseweart; W. I. Sundquist; C. P. Hill. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell,* 87(7): 1285-1294, 1996.

[43] A. Lesage; C. Auger; S. Caldarelli; L. Emsley. Determination of through-bond carbon-carbon connectivities in solid-state NMR using the INADEQUATE experiment. *J. Am. Chem. Soc.,* 119(33): 7867-7868, 1997.

[44] N. E. Kohl; E. A. Emini; W. A. Schleif; L. J. Davis; J. C. Heimbach; R. A. Dixon; E. M. Scolnick; I. S. Sigal. Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. U. S. A.,* 85(13): 4686-4690, 1988.

[45] H. G. Gottlinger; J. G. Sodroski; W. A. Haseltine. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A.,* 86(15): 5781-5785, 1989.

[46] F. Barre-Sinoussi; J. C. Chermann; F. Rey; M. T. Nugeyre; S. Chamaret; J. Gruest; C. Dauguet; C. Axler-Blin; F. Vezinet-Brun; C. Rouzioux; W. Rozenbaum; L. Montagnier. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science,* 220(4599): 868-871, 1983.

[47] D. Ivanov; O. V. Tsodikov; J. Kasanov; T. Ellenberger; G. Wagner; T. Collins. Domain-swapped dimerization of the HIV-1 capsid C-terminal domain. *Proc. Natl. Acad. Sci. U. S. A.,* 104(11): 4353-4358, 2007.

Chapter 7

CONCLUSIONS AND FUTURE OUTLOOK

The global pandemic of AIDS pushes the scientific community to develop drugs and vaccines against this currently incurable disease. It has now been well recognized that HIV-1 virus is the causative reagent of AIDS. Recently, it was discovered that the Rhesus Monkey TRIM5α rendered this species immune to AIDS [1]. Another recent finding demonstrates that a clinically effective drug Bevirimat (BVM) blocks the virus replication by preventing the late step in Gag cleavage cascade, i.e. the cleavage of CA-SP1 precursor to CA [2]. These discoveries suggest novel approaches to AIDS therapies. To develop these new therapies, it is necessary to gain atomic-resolution structural information on the target proteins of the HIV virus, such as Gag and capsid (CA), in their biologically relevant assembled states. As demonstrated in this thesis, solid-state NMR techniques are very well suited to probe atomic-level structural and dynamics information in HIV-1 protein assemblies.

The work detailed in this thesis makes several significant contributions in the field of HIV-1 structural biology. First, we have established conditions for preparing the CA assemblies of spherical, conical and tubular morphologies that are amenable for high-resolution solid-state NMR characterization. In the process of sample preparation, I have developed a novel confocal method to visualize the CA protein assemblies, which has proven particularly useful samples of spherical and tubular morphologies. Second, we

have investigated the structures of CA conical assemblies by solid-state NMR, and discovered that the overall conformation is consistent with the results derived from solution NMR and X-ray crystallography using unassembled CA. Third, we have characterized the conformational dynamics of the hinge region connecting the CTD and NTD and demonstrated the relationship between structural polymorphism in CA assemblies and millisecond timescale motions present in the hinge region. Fourth, we have elucidated the conformation of SP1 peptide, the target of a maturation inhibitor BVM, in the CA-SP1 tubular assemblies. We have discovered that SP1 is dynamically flexible random coil. We have also discovered that the mutation and variations in the CA protein primary sequence affect the local and long-range structure of CA in the assembled state.

An important technical accomplishment is the resonance assignments of an extensive portion of CA in conical and tubular assemblies, which have been made on the basis of multiple 2D and 3D experiments. These assignments will be essential for the next immediate step of this work: gaining long-range restraints to derive the atomic-level structure of CA and CA-SP1 assemblies. Looking into the future, it will be important to investigate the interactions between CA and cellular host factors, such as TRIM5α, as well as the binding modes of various small-molecule inhibitors, such as BVM and its derivatives. There is likely still a long road ahead before complete cure of AIDS can be established, and the work presented in this thesis paves the way for future comprehensive structural analysis of HIV-1 protein assemblies alone and in the context of the host cell factors, by solid-state NMR.

References

[1] M. Stremlau; C. M. Owens; M. J. Perron; M. Kiessling; P. Autissier; J. Sodroski. The cytoplasmic body component TRIM5 alpha restricts HIV-1 infection in Old World monkeys. *Nature,* 427(6977): 848-853, 2004.

[2] C. S. Adamson; E. O. Freed. Novel approaches to inhibiting HIV-1 replication. *Antiviral Res,* 85(1): 119-141, 2010.

Appendix A

Table A.1: Chemical shift differences between solid-state and solution for the assigned residues in the conical assemblies of HIV-1 CA.

Residue	N	C^{α}	C^{β}	C^{γ}
P1				
I2	0.0	-0.9	-0.1	
V3				
Q4	0.1	-0.6	-1.1	
N5	0.1	2.5		
I6				
Q7				
G8	-0.1	-0.7		
Q9	0.3	0.0	0.1	
M10				
V11	0.1	-0.8	0.1	
H12		-0.7	0.8	
Q13	-0.4	0.1	0.3	
A14	-0.2	-0.5	-0.1	
I15	-0.1	-0.5	-0.2	
S16	-0.3	-0.7	-0.1	
P17				
R18	0.1	-0.2	-0.1	
T19	0.0	0.0	0.0	
L20	0.0	-0.7	-0.2	
N21	0.0	0.0	0.0	
A22	-0.1	-0.1	1.2	
W23	-0.1	0.1	-0.1	

V24	-0.4	-0.3	-0.1
K25	0.0	0.5	0.3
V26	0.0	0.0	-0.9
V27		-0.3	0.7
E28			
E29	-0.1	-0.5	0.3
K30			
A31			
F32	0.0	-1.3	0.0
S33			
P34		0.5	0.9
E35	-0.1	0.8	-0.4
V36	0.0	0.7	0.0
I37	0.0	-0.1	-0.2
P38			
M39	0.1	0.8	0.3
F40	-0.1	-0.2	0.2
S41			
A42	0.5	-0.5	-0.1
L43	-0.2	0.2	0.1
S44	0.2	-0.5	0.3
E45	-0.1	-0.3	0.0
G46	-0.1	0.1	
A47	-0.2	-0.3	0.2
T48	-0.5	-0.4	-0.3
P49			

Q50	0.0	-0.5	0.7
D51	-0.1	0.1	0.1
L52	-0.2	-0.3	0.9
N53	0.1	-0.2	-0.1
T54	-0.1	0.0	0.1
M55			
L56			
N57	0.0	0.8	0.1
T58	0.0	0.0	-0.7
V59			
G60	0.0	0.0	
G61	0.1	-0.1	
H62		0.5	-0.3
Q63	0.1	-0.8	0.0
A64	0.0	-0.7	0.6
A65			
M66			
Q67			
M68			
L69	0.0	-0.8	-0.5
K70	-0.1	0.3	0.9
E71			
T72			
I73	0.0	0.3	-0.1
N74			
E75			

E76	0.2	-0.4	-0.5
A77	0.6	0.1	0.1
A78	0.3	-0.7	0.3
E79			
W80	0.0	-0.8	-0.5
D81	-0.3	-0.5	0.0
R82	-0.1	0.2	-1.6
V83			
H84	0.0	0.5	-0.4
P85			
V86			
H87	0.0	-0.3	0.2
A88			
G89			
P90			
I91	-0.1	-0.4	-0.7
A92	0.5	-0.5	
P93		-2.0	1.1
G94	0.0	-0.1	
Q95	0.0	-0.6	-0.1
M96	1.1	-0.2	-0.2
R97	-0.1	-0.6	-0.3
E98	0.3	0.3	-0.2
P99			
R100	0.0	1.4	-0.1
G101	0.0	0.2	

S102			
D103	0.0	0.2	0.0
I104	0.1	-0.2	-0.2
A105	-0.4	-0.5	0.0
G106	-0.1	0.1	
T107	-0.1	0.1	0.0
T108	0.6	-0.1	0.2
S109	-0.2	0.2	0.1
T110	-0.3	0.4	0.0
L111			
Q112	-0.1	-0.3	0.1
E113			
Q114	0.0	-0.9	-0.2
I115	0.0	0.2	-0.1
G116	0.3	-0.1	
W117	0.0	-0.7	0.1
M118	-0.1	-0.3	0.4
T119	-0.1	-0.6	-0.1
N120	0.0	1.3	0.3
N121			
P122			
P123			
I124	0.1	0.1	-0.2
P125		-0.2	-0.5
V126	0.0	0.4	-0.7
G127	0.0	0.0	

E128			
I129	0.1	-0.5	0.2
Y130	0.1	-0.5	0.5
K131			
R132	0.0	0.0	-0.2
W133	0.1	0.3	-0.2
I134	0.1	0.2	
I135	0.0	0.3	-0.2
L136	-0.1	-0.8	0.2
G137	0.0	-0.2	
L138	0.0	-0.2	0.2
N139	0.0	0.1	-0.2
K140	-0.1	-0.4	0.1
I141	0.0	0.0	-0.4
V142	0.0	0.4	-0.3
R143	0.1	0.2	-0.4
M144			
Y145		1.8	-1.5
S146	0.0	0.6	-0.2
P147			
T148			
S149			
I150	0.0	1.3	-0.7
L151	0.0	-0.1	-1.2
D152	0.0	0.4	-0.4
I153	-0.1	1.6	-1.4

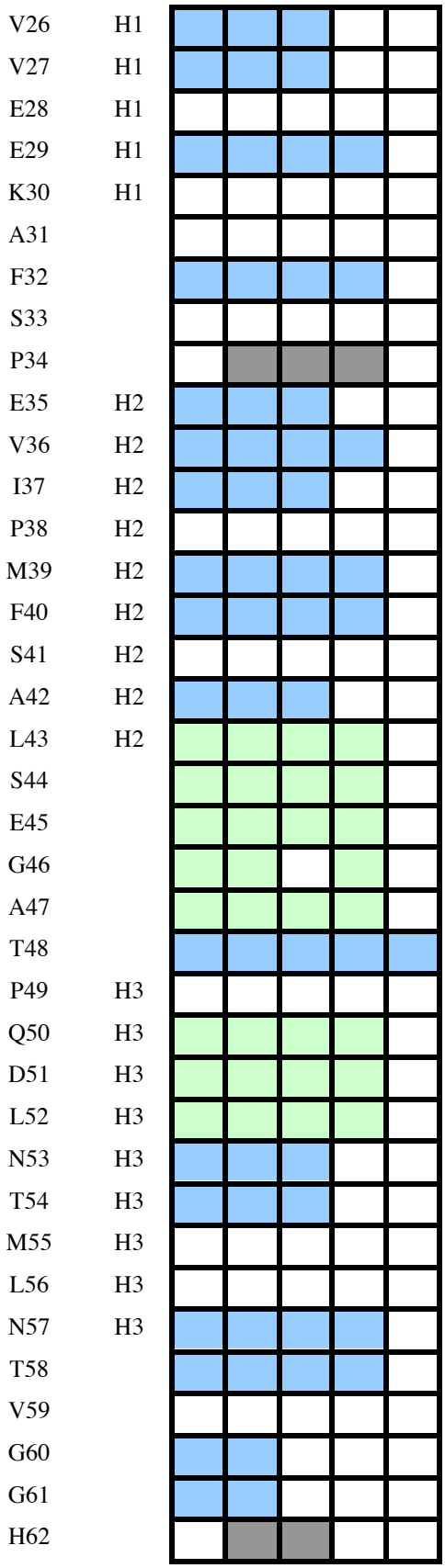
R154	0.1	0.2	0.1	
Q155	0.3	-0.3	-0.5	0.4
G156	0.1	0.3		
P157		0.1	0.8	0.3
K158	0.0	-0.3	0.3	
E159	0.1	-0.3	-0.1	
P160				
F161	-0.2	-0.6	-0.2	
R162	0.0	-0.5	-0.1	
D163		-0.3	-0.3	
Y164	0.4	-0.6	0.0	
V165	0.0	-0.1		
D166	0.0	-0.4		
R167	0.0	-0.8	-1.0	
F168	0.0	-0.4	-0.2	
Y169	-0.9	-0.4	0.7	
K170	0.0	0.1	0.3	
T171	0.0	-0.2	0.1	
L172	-0.1	0.7	-1.1	2.7
R173				
A174	-0.3	-0.1	0.1	
E175				
Q176	0.0	-0.7	0.0	
A177	2.1	0.3	0.6	
S178	1.4	0.2	0.5	
Q179	0.0	-0.5	0.4	

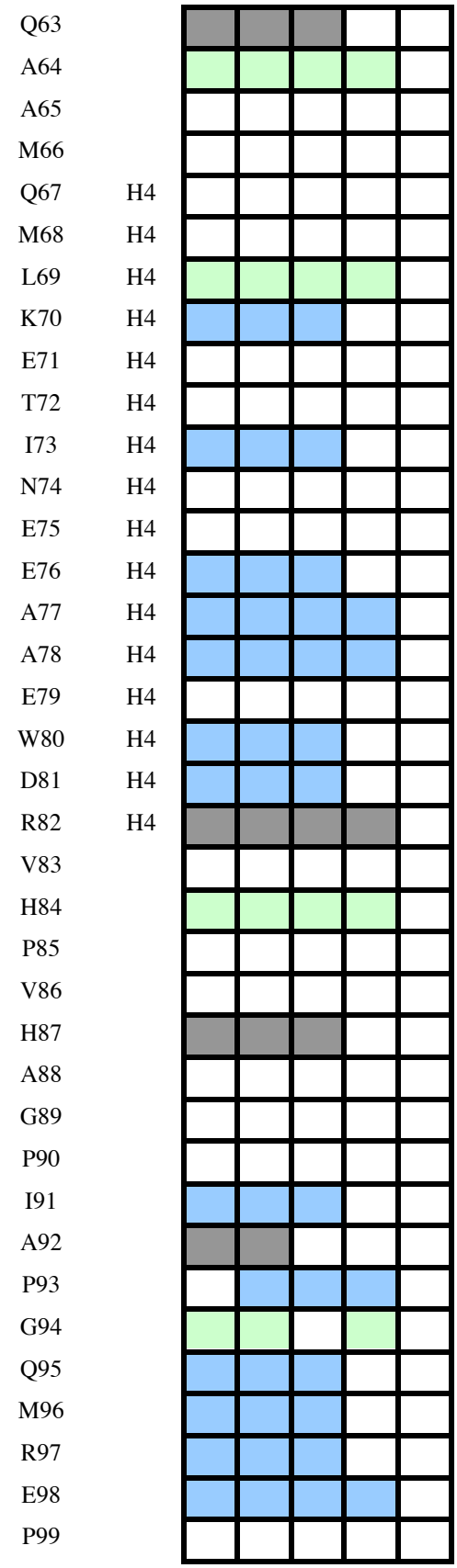
E180				
V181	0.1	-0.1	0.3	
K182				
N183	-0.1	-0.1	-0.2	
W184	0.0	-0.9	-0.9	
M185				
T186	-0.1	-0.1	0.1	
E187	0.0	0.2	-0.1	
T188	0.0	-0.5	-0.7	
L189	0.0	0.1	-0.2	
L190				
V191	-0.1	0.3	0.2	
Q192	0.0	0.3	-0.3	
N193	-0.1	0.1	0.0	
A194	-0.1	0.2	0.5	
N195	-0.1	0.0	-0.1	
P196		0.2	0.6	
D197	-0.1	-0.4	-0.3	
C198	-0.5	0.1	0.5	
K199	0.0	-1.5	0.2	
T200	0.1	0.2	0.0	
I201	0.4	0.6	-0.1	
L202	0.0	-0.1	-0.8	
K203				
A204				
L205	0.2	0.3	0.1	0.5

G206	-0.3	0.2	
P207			
A208	0.0	1.1	1.7
A209	-0.1	0.0	0.4
T210	-0.2	-0.6	-0.3
L211	0.0	-0.2	0.3
E212			
E213			
M214	-0.2	-0.1	-0.3
M215	-0.1	-0.2	-0.2
T216	0.1	0.0	-0.3
A217	0.0	-2.0	
C218	-0.7	-0.5	0.5
Q219	-0.1	0.7	0.0
G220	-0.1	0.0	
V221			
G222			
G223	-0.1	-0.2	
P224			
G225	0.0	2.0	
H226			
K227			
A228	-0.1	0.4	0.7
R229			
V230	-0.1	-1.4	0.0
L231			

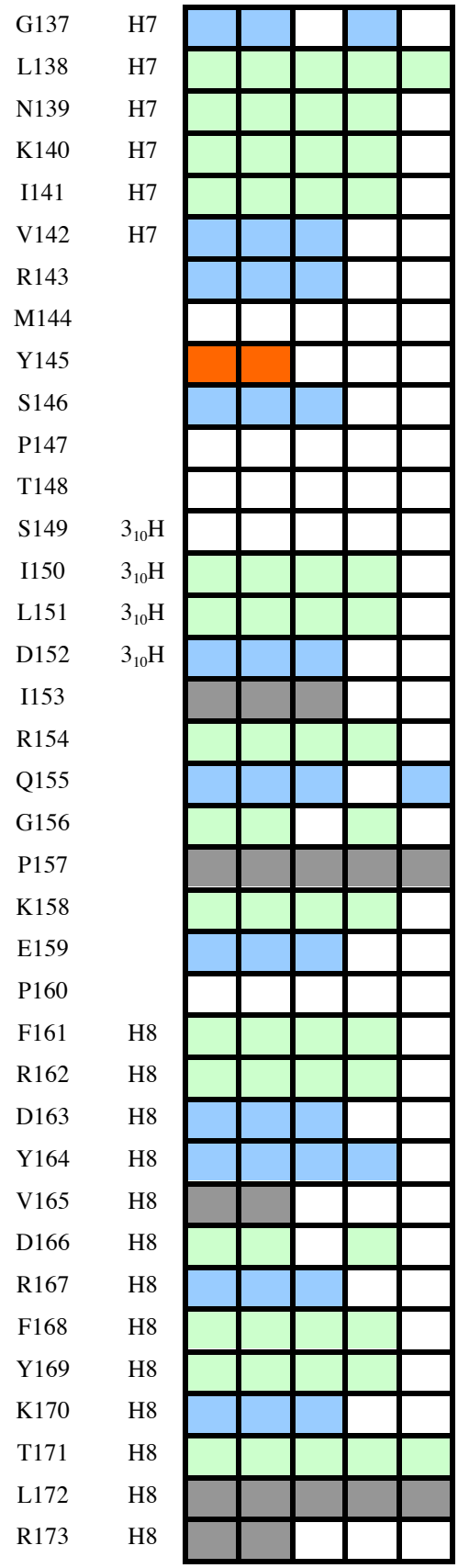
Table A.2: Summary of the resonance assignments for CA conical assemblies. Primary sequence and secondary structure are indicated (H: a-helix, S: b-sheet, 3_{10}: 3_{10}-helix). Residues for which intra-residue and sequential assignments are available are represented by green boxes; residues for which intra-residue resonance assignments were derived based on solution chemical shifts for N- terminal and C-terminal domains (NTD and CTD, respectively) are represented by blue boxes. For Y145, the assignments of N and C^a atoms were derived based on the 2D NCA spectrum of the U-^{13}C,^{15}N-Tyr labeled CA acquired at -15 °C, and these atoms are represented in orange. Residues for which only tentative assignments are available are represented by grey boxes. For the remaining residues, assignments could not be derived from the 14.1 T data sets due to extensive spectral overlap.

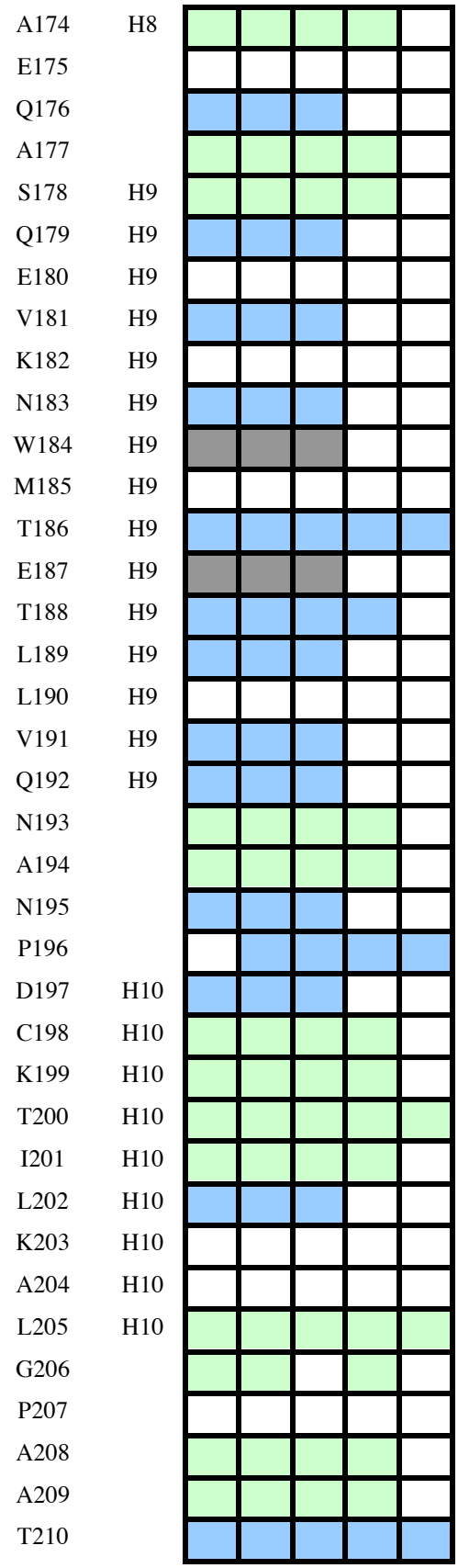
Residue	Secondary structure	N	C^α	C^β	C'	C^γ
P1	S1					
I2	S1					
V3	S1					
Q4	S1					
N5						
I6						
Q7	S2					
G8	S2					
Q9	S2					
M10	S2					
V11	S2					
H12	S2					
Q13	S2					
A14						
I15						
S16						
P17						
R18	H1					
T19	H1					
L20	H1					
N21	H1					
A22	H1					
W23	H1					
V24	H1					
K25	H1					


V26 H1
V27 H1
E28 H1
E29 H1
K30 H1
A31
F32
S33
P34
E35 H2
V36 H2
I37 H2
P38 H2
M39 H2
F40 H2
S41 H2
A42 H2
L43 H2
S44
E45
G46
A47
T48
P49 H3
Q50 H3
D51 H3
L52 H3
N53 H3
T54 H3
M55 H3
L56 H3
N57 H3
T58
V59
G60
G61
H62


Q63
A64
A65
M66
Q67 H4
M68 H4
L69 H4
K70 H4
E71 H4
T72 H4
I73 H4
N74 H4
E75 H4
E76 H4
A77 H4
A78 H4
E79 H4
W80 H4
D81 H4
R82 H4
V83
H84
P85
V86
H87
A88
G89
P90
I91
A92
P93
G94
Q95
M96
R97
E98
P99

R100		
G101	H5	
S102	H5	
D103	H5	
I104	H5	
A105	H5	
G106		
T107		
T108		
S109		
T110		
L111		
Q112	H6	
E113	H6	
Q114	H6	
I115	H6	
G116	H6	
W117	H6	
M118	H6	
T119	H6	
N120	H6	
N121		
P122		
P123		
I124		
P125		
V126	H7	
G127	H7	
E128	H7	
I129	H7	
Y130	H7	
K131	H7	
R132	H7	
W133	H7	
I134	H7	
I135	H7	
L136	H7	


G137 H7
L138 H7
N139 H7
K140 H7
I141 H7
V142 H7
R143
M144
Y145
S146
P147
T148
S149 3_{10}H
I150 3_{10}H
L151 3_{10}H
D152 3_{10}H
I153
R154
Q155
G156
P157
K158
E159
P160
F161 H8
R162 H8
D163 H8
Y164 H8
V165 H8
D166 H8
R167 H8
F168 H8
Y169 H8
K170 H8
T171 H8
L172 H8
R173 H8


A174 H8
E175
Q176
A177
S178 H9
Q179 H9
E180 H9
V181 H9
K182 H9
N183 H9
W184 H9
M185 H9
T186 H9
E187 H9
T188 H9
L189 H9
L190 H9
V191 H9
Q192 H9
N193
A194
N195
P196
D197 H10
C198 H10
K199 H10
T200 H10
I201 H10
L202 H10
K203 H10
A204 H10
L205 H10
G206
P207
A208
A209
T210

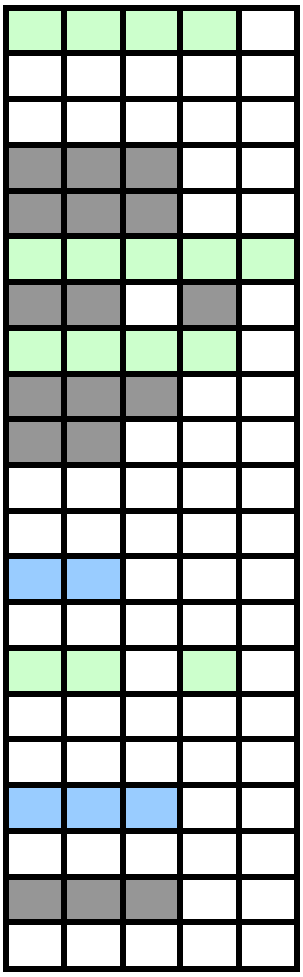
L211	H11	
E212	H11	
E213	H11	
M214	H11	
M215	H11	
T216	H11	
A217		
C218		
Q219		
G220		
V221		
G222		
G223		
P224		
G225		
H226		
K227		
A228		
R229		
V230		
L231		

Table A.3: Acquisition and processing parameters for the solid-state NMR experiments probing CA conical assemblies.

Experiment	Acquisition			Processing		
	ω3	ω2	ω1	ω3	ω2	ω1
14.1 T						
Uniformly labeled conical assemblies						
DARR (τ_{mix}=10 ms)		1000 complex; SW = 50 kHz, 200 scans	400 real, SW = 50 kHz		30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 512 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
NCACX (1)	1000 complex; SW = 50 kHz 128 scans	48 real; (TPPI) SW =5 kHz	48 real; (TPPI) SW=15 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell	60-degree sinebell
NCACX (2)	1000 complex; SW = 50 kHz 128 scans	48 real; (TPPI) SW =5 kHz	48 real; (TPPI) SW=15 kHz	60-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell	60-degree sinebell
NCACX (3)	1000 complex; SW = 50 kHz 128 scans	48 real; (TPPI) SW =5 kHz	48 real; (TPPI) SW=15 kHz	90-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell	60-degree sinebell

NCACX (4)	1000 complex; SW = 50 kHz 128 scans	48 real; (TPPI) SW = 5 kHz	48 real; (TPPI) SW = 15 kHz	90-degree sinebell; Lorentzian-to-Gaussian transformation	90-degree sinebell	90-degree sinebell
NCACB (1)	1000 complex; SW = 50 kHz; 400 scans	32 real; (TPPI) SW = 4.17 kHz	36 real; (TPPI) SW = 8.33 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	30-degree sinebell; Lorentzian-to-Gaussian transformation	30-degree sinebell; Lorentzian-to-Gaussian transformation
NCACB (2)	1000 complex; SW = 50 kHz 400 scans	32 real; (TPPI) SW = 4.17 kHz	36 real; (TPPI) SW = 8.33 kHz	60-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell; Lorentzian-to-Gaussian transformation
NCACB (3)	1000 complex; SW = 50 kHz 400 scans	32 real; (TPPI) SW=4.17 kHz	36 real; (TPPI) SW = 8.33 kHz	60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 64 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 72 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation

NCACB (4)	1000 complex; SW = 50 kHz 400 scans	32 real; (TPPI) SW = 4.17 kHz	36 real; (TPPI) SW = 8.33 kHz	90-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell; Lorentzian-to-Gaussian transformation
NCACB (5)	1000 complex; SW = 50 kHz 400 scans	32 real; (TPPI) SW = 4.17 kHz	36 real; (TPPI) SW = 8.33 kHz	90-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 64 points; 90-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 72 points; 90-degree sinebell; Lorentzian-to-Gaussian transformation
NCOCX (1)	1000 complex; SW = 50 kHz 280 scans	36 real; (TPPI) SW = 4.17 kHz	36 real; (TPPI) SW = 6.25 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	30-degree sinebell; Lorentzian-to-Gaussian transformation	30-degree sinebell; Lorentzian-to-Gaussian transformation

NCOCX (2)	1000 complex; SW = 50 kHz 280 scans	36 real (TPPI) SW = 4.17 kHz	36 real (TPPI) SW = 6.25 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 64 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation
NCOCX (3)	1000 complex; SW = 50 kHz 280 scans	36 real (TPPI) SW = 4.17 kHz	36 real (TPPI) SW = 6.25 kHz	60-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell; Lorentzian-to-Gaussian transformation
NCOCX (4)	1000 complex; SW=50 kHz 280 scans	36 real (TPPI) SW = 4.17 kHz	36 real (TPPI) SW = 6.25 kHz	60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 72 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 72 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation

NCOCX (5)	1000 complex; SW=50 kHz 280 scans	36 real (TPPI) SW = 4.17 kHz	36 real (TPPI) SW = 6.25 kHz	90-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell; Lorentzian-to-Gaussian transformation
NCOCX (6)	1000 complex; SW=50 kHz 280 scans	36 real (TPPI) SW = 4.17 kHz	36 real (TPPI) SW=6.25 kHz	90-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 72 points; 90-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 72 points; 90-degree sinebell; Lorentzian-to-Gaussian transformation
Ala/Thr labeled CA conical assemblies						
DARR (τ_{mix}=500 ms)		1000 complex; SW = 50 kHz, 384 scans	256 real (TPPI) SW = 20 kHz		90-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 512 points; 90-degree sinebell; Lorentzian-to-Gaussian transformation
His/Tyr labeled CA conical assemblies						

DARR (τ_{mix}=10 ms)	1000 complex; SW = 50 kHz, 640 scans	128 real (TPPI) SW = 33.3 kHz	90-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 256 points; 90-degree sinebell; Lorentzian-to-Gaussian transformation
Uniformly labeled spherical assemblies				
DARR (τ_{mix}=10 ms)	1000 complex; SW = 5 0kHz, 256 scans	512 real (TPPI) SW = 50 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 1024 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation

Appendix B

Table B.1: Acquisition and processing parameters for the solid-state NMR experiments probing CA tubular assemblies.

Experiment	Acquisition			Processing		
	ω3	ω2	ω1	ω3	ω2	ω1
19.9 T						
Uniformly labeled CAA92E tubular assemblies						
DARR (τ_{mix}=50 ms)		1024 complex; SW = 64 kHz, 128 scans	512 real, SW = 45 kHz		30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 512 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
NOESY-DARR(τ_{mix}=50 ms)		1024 complex; SW =64 kHz, 192 scans	400 real, SW = 45 kHz		30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 512 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
NCACX		1024 complex; SW =85.2 kHz, 2896	64 complex; (States) SW=6.5 kHz		30-degree sinebell; Lorentzian-to-Gaussian	forward linear prediction of 64 points; 30-degree sinebell;

NCACX	scans	64 complex; (States) SW=6.5 kHz	transformation	Lorentzian-to-Gaussian transformation
NCA	1024 complex; SW =85.2 kHz, 320 scans	44 complex; (TPPI) SW=6.5 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 128 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
NCOCX	1024 complex; SW =64 kHz, 1360 scans	40 complex; (TPPI) SW=5 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 128 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
CP-INADEQUATE	1024 complex; SW =64 kHz, 416 scans	400 real; (TPPI) SW=81 kHz	60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 1024 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation

INADEQUATE	1024 complex; SW =64 kHz, 1376 scans	400 real; (TPPI) SW=81 kHz	60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 1024 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation
Uniformly labeled CA-SP1 A92E tubular assemblies				
DARR (τ_{mix}=50 ms)	1024 complex; SW = 64 kHz, 128 scans	512 real, SW = 45 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 512 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
NOESY-DARR(τ_{mix}=50 ms)	1024 complex; SW =64 kHz, 192 scans	400 real, SW = 45 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 512 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
NCACX	1024 complex; SW =85.2 kHz, 2048 scans	40 complex; (States) SW=6.5 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 64 points; 30-degree sinebell; Lorentzian-to-Gaussian

				transformation
NCA	1024 complex; SW =64 kHz, 320 scans	44 complex; (states) SW=6.5 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 128 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
NCOCX	1024 complex; SW =85 kHz, 2048 scans	40 complex; (TPPI) SW=65 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 128 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
CP-INADEQUATE	1024 complex; SW =64 kHz,512 scans	400 real; (TPPI) SW=81 kHz	60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 1024 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation
INADEQUATE	1024 complex; SW =64 kHz, 1376 scans	400 real; (TPPI) SW=81 kHz	60-degree sinebell; Lorentzian-to-Gaussian	forward linear prediction of 1024 points; 60-degree sinebell;

INADEQUATE	1024 complex; SW =64 kHz, 1376 scans	400 real; (TPPI) SW=81 kHz	transformation	Lorentzian-to-Gaussian transformation
21.1 T				
Uniformly labeled CA WT tubular assemblies				
DARR (τ_{mix}=50 ms)	1536 complex; SW =100 kHz, 96 scans	400 real, SW = 45 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 400 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
NCA	1024 complex; SW =100 kHz, 360 scans	96 real; (TPPI) SW=7 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 256 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
NCACX	1024 complex; SW =100 kHz, 1200 scans	64 real; (TPPI) SW=5 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 64 points; 30-degree sinebell;

						Lorentzian-to-Gaussian transformation
NCACX (1)	1024 complex; SW =100 kHz, 1200 scans	42 real; (TPPI) SW=5 kHz	40 real; SW=7 kHz	60-degree sinebell; Lorentzian-to-Gaussian transformation	60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 64 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation
NCACX (2)	1024 complex; SW =100 kHz, 144 scans	42 real; (TPPI) SW=5 kHz	40 real; SW=7 kHz	90-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 64 points; 90-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 64 points; 90-degree sinebell; Lorentzian-to-Gaussian transformation
NCOCX (1)	1024 complex; SW =100 kHz, 1200 scans	43 real; (TPPI) SW=5 kHz	24 real; SW=5 kHz	60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 64 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 64 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation
NCOCX (2)	1024 complex; SW =100 kHz, 1200	43 real; (TPPI) SW=5 kHz	24 real; SW=5 kHz	90-degree sinebell; Lorentzian-to-Gaussian	forward linear prediction of 64 points; 90-degree sinebell;	forward linear prediction of 64 points; 90-degree sinebell;

NCOCX (2)	scans	43 real; (TPPI) SW=5 kHz	24 real; SW=5 kHz	transformation	Lorentzian-to-Gaussian transformation	Lorentzian-to-Gaussian transformation
CP-INADEQUATE		1536 complex; SW =100 kHz, 400 scans	207 real; (TPPI) SW=81 kHz		60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 512 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation
INADEQUATE		1536 complex; SW =100 kHz, 1200 scans	240 real; (TPPI) SW=81 kHz		60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 512 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation
21.1 T						
Uniformly labeled CA-SP1 WT tubular assemblies						
DARR (τ_{mix}=50 ms)		1536 complex; SW =100 kHz, 96 scans	400 real, SW = 45 kHz		30-degree sinebell; Lorentzian-to-Gaussian	forward linear prediction of 400 points; 30-degree sinebell;

DARR (τ_{mix}=50 ms)	1536 complex; SW =100 kHz, 96 scans	400 real, SW = 45 kHz	transformation	Lorentzian-to-Gaussian transformation
NCA	1024 complex; SW =100 kHz, 128 scans	80 real; (TPPI) SW=5kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 256 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
NOESY-DARR(τ_{mix}=50 ms)	1536 complex; SW =100 kHz, 288 scans	176complex, SW = 45 kHz	30-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 256 points; 30-degree sinebell; Lorentzian-to-Gaussian transformation
CP-INADEQUATE	1536 complex; SW =100 kHz, 384 scans	400 real; (TPPI) SW=81 kHz	60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 1024 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation

INADEQUATE	1536 complex; SW =100 kHz, 1248 scans	240 real; (TPPI) SW=81 kHz	60-degree sinebell; Lorentzian-to-Gaussian transformation	forward linear prediction of 1024 points; 60-degree sinebell; Lorentzian-to-Gaussian transformation

Appendix C

PERMISSION LETTERS

ELSEVIER LICENSE
TERMS AND CONDITIONS

Feb 14, 2013

This is a License Agreement between Yun Han ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Yun Han
Customer address	136 Warren ST Norfolk, VA 23505
License number	3084450116425
License date	Feb 08, 2013
Licensed content publisher	Elsevier
Licensed content publication	Drug Discovery Today
Licensed content title	Recent progress in antiretrovirals – lessons from resistance
Licensed content author	Catherine S. Adamson,Eric O. Freed
Licensed content date	May 2008
Licensed content volume number	13
Licensed content issue number	9–10
Number of pages	9
Start Page	424
End Page	432
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

Order reference number	None
Title of your thesis/dissertation	Solid-State NMR Studies of Structure and Dynamics of HIV-1 Capsid (CA) Protein Assemblies
Expected completion date	Feb 2013
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	**0.00 USD**
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.
16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
17. **Author website** for journals with the following additional clauses:
All content posted to the web site must maintain the copyright information line on the bottom of each image, and the permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx . As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com
20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**:

v1.6

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK500953223. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Feb 14, 2013

This is a License Agreement between Yun Han ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3084341279767
License date	Feb 08, 2013
Licensed content publisher	Nature Publishing Group
Licensed content publication	Nature Reviews Microbiology
Licensed content title	How retroviruses select their genomes
Licensed content author	Victoria D'Souza, Michael F. Summers
Licensed content date	Aug 1, 2005
Volume number	3
Issue number	8
Type of Use	reuse in a thesis/dissertation
Requestor type	academic/educational
Format	print
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 1 | General features of the orthoretroviral replication cycle.
Author of this NPG article	no
Your reference number	None
Title of your thesis / dissertation	Solid-State NMR Studies of Structure and Dynamics of HIV-1 Capsid (CA) Protein Assemblies
Expected completion date	Feb 2013
Estimated size (number of pages)	200
Total	**0.00 USD**

Terms and Conditions

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use, subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should

ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version. Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run). NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
   For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   Note: For republication from the *British Journal of Cancer*, the following credit lines apply.
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication) For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   Note: For adaptation from the *British Journal of Cancer*, the following credit line applies.
   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information. Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

   Note: For translation from the *British Journal of Cancer*, the following credit line applies.
   Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference

citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.
Special Terms:
v1.1

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK500953111. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Feb 14, 2013

This is a License Agreement between Yun Han ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Yun Han
Customer address	136 Warren ST Norfolk, VA 23505
License number	3087930632719
License date	Feb 14, 2013
Licensed content publisher	Elsevier
Licensed content publication	Current Opinion in Structural Biology
Licensed content title	The structural biology of HIV assembly
Licensed content author	Barbie K Ganser-Pornillos,Mark Yeager,Wesley I Sundquist
Licensed content date	April 2008
Licensed content volume number	18
Licensed content issue number	2
Number of pages	15
Start Page	203
End Page	217
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

Order reference number	None
Title of your thesis/dissertation	Solid-State NMR Studies of Structure and Dynamics of HIV-1 Capsid (CA) Protein Assemblies
Expected completion date	Feb 2013
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	**0.00 USD**
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.
16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
17. **Author website** for journals with the following additional clauses:
All content posted to the web site must maintain the copyright information line on the bottom of each image, and the permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx . As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com
20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**:

v1.6

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK500957311. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.