J|A|C|S
ARTICLES
Published on Web 02/06/2007

Filamentous Phage Studied by Magic-Angle Spinning NMR: Resonance Assignment and Secondary Structure of the Coat Protein in Pf1

Amir Goldbourt,[§] Benjamin J. Gross,[§] Loren A. Day,[‡] and Ann E. McDermott*[,§]

Contribution from the Department of Chemistry, Columbia University, New York, New York 10027, and Public Health Research Institute, 225 Warren Street, Newark, New Jersey 07103

Received September 26, 2006; E-mail: aem5@columbia.edu

Abstract: Assignments are presented for resonances in the magic-angle spinning solid-state NMR spectra of the major coat protein subunit of the filamentous bacteriophage Pf1. NMR spectra were collected on uniformly ^{13}C and ^{15}N isotopically enriched, polyethylene glycol precipitated samples of fully infectious and hydrated phage. Site-specific assignments were achieved for 231 of the 251 labeled atoms (92%) of the 46-residue-long coat protein, including 136 of the 138 backbone atoms, by means of two- and three-dimensional ^{15}N and ^{13}C correlation experiments. A single chemical shift was observed for the vast majority of atoms, suggesting a single conformation for the 7300 subunits in the 36 MDa virion in its high-temperature form. On the other hand, multiple chemical shifts were observed for the Cα, Cβ, and Cγ atoms of T5 in the helix terminus and the Cα and Cβ atoms of M42 in the DNA interaction domain. The chemical shifts of the backbone atoms indicate that the coat protein conformation involves a 40-residue continuous α-helix extending from residue 6 to the C-terminus.

Introduction

The structures of a number of filamentous bacteriophages have been under study for many years by electron microscopy and X-ray and neutron diffraction, as well as by electronic, vibrational, and NMR spectroscopies. Much is understood about their structures, yet, surprisingly, key questions remain unanswered about interactions between the DNA and the protein capsid, about the interactions between subunits within the capsid, and about the conformation of the side chains within the capsid. Pf1 is the one of the best-studied of the filamentous phages from a structural point of view, and it provides a paradigm structure for one of the two symmetry classes of these viruses. Nevertheless, a molecularly detailed structure has still not been established for it. Pf1 was discovered[1] soon after the discoveries of the well-known filamentous phages fd, f1, and M13,[2−4] and, as is characteristic of all such phages,[5,6] its virion is threadlike and contains a circular, single-stranded DNA molecule. The loop of DNA extends from one end to the other and back within a capsid less than 7 nm in diameter made of thousands of copies of a major coat protein, with special proteins at the ends. Pf1 is one of the longest of the phages, its length being close to 2.1 μm. Because of their structures and the ways their genomes can be manipulated, filamentous phages are widely used in DNA cloning and sequencing,[7,8] in a binding selection technique called phage display,[9−11] and in areas of nanotechnology.[12] Also, they orient in magnetic fields, and Pf1, in particular, has been used to impart orientation on other molecules for their study by solution NMR methods.[13,14]

Filamentous phages infect a variety of bacteria, the specific host of Pf1 being *Pseudomonas aeruginosa*, strain K (PAK).[1] Some of the phage/host systems impact on human health, the prime example for which is the phage/prophage CTXΦ, which carries the toxin genes in the virulent strains of *Vibrio cholerae* that causes cholera.[15,16] Another is the phage/prophage Pf4,[17] the host of which is *P. aeruginosa*, strain 01 (PA01), a pathogen in cystic fibrosis and in hospital-acquired infections. Pf4 is a prophage that becomes activated and released as a filamentous phage in dense biofilms of PA01.[17] A third is f237, which is associated with *Vibrio parahemeolyticus* strains causing acute

[§] Columbia University.
[‡] Public Health Research Institute.

(1) Takeya, K.; Amako, K. *Virology* **1966**, *28*, 163−165.
(2) Hofschneider, P. H. *Z. Naturforsch.* **1963**, *B18*, 203ff.
(3) Hoffmann-Berling, H.; Marvin, D. A.; Durwald, H. *Z. Naturforsch.* **1963**, *B18*, 876−883.
(4) Zinder, N. D.; Valentine, R. C.; Roger, M.; Stoeckenius, W. *Virology* **1963**, *20*, 638−640.
(5) Day, L. A.; Marzec, C. J.; Reisberg, S. A.; Casadevall, A. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 509−539.
(6) Day, L. A.; Hendrix, R. W. In *Virus Taxonomy: Eighth Report of the International Committee on the Taxonomy of Viruses*; Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., Ball, L. A., Eds.; Elsevier: Amsterdam, The Netherlands, 2005; p 277.
(7) Sanger, F.; Coulson, A. R.; Barrell, B. G.; Smith, A. J. H.; Roe, B. A. *J. Mol. Biol.* **1980**, *143*, 161−178.
(8) Messing, J. *Methods Mol. Biol.* **2001**, *167*, 13−31.
(9) Kehoe, J. W.; Kay, B. K. *Chem. Rev.* **2005**, *105*, 4056−4072.
(10) Smith, G. P. *Science* **1985**, *228*, 1315−1317.
(11) Arap, W.; Pasqualini, R.; Ruoslahti, E. *Science* **1998**, *279*, 377−380.
(12) Nam, K. T.; Kim, D. W.; Yoo, P. J.; Chiang, C. Y.; Meethong, N.; Hammond, P. T.; Chiang, Y. M.; Belcher, A. M. *Science* **2006**, *312*, 885−888.
(13) Hansen, M. R.; Mueller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, *5*, 1065−1074.
(14) Bax, A.; Grishaev, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 563−570.
(15) Wagner, P. L.; Waldor, M. K. *Infect. Immun.* **2002**, *70*, 3985−3993.
(16) Davis, B. M.; Waldor, M. K. *Curr. Opin. Microbiol.* **2003**, *6*, 35−42.
(17) Webb, J. S.; Lau, M.; Kjelleberg, S. *J. Bacteriol.* **2004**, *186*, 8066−8073.

10.1021/ja066928u CCC: $37.00

dysentery.[18] All three of these viruses are related to Pf1 by sequence homology and/or structure.

X-ray fiber diffraction has revealed that the capsids of Ff (fd, M13, f1), If1, and IKe are helical arrays of pentamers of subunits with about 16 Å rise and 36° rotation per pentamer,[19−21] referred to as Class I symmetry. By comparison, Class II capsids, for phages Pf1, Xf, Pf3, and PH75, have been described as helices of about 27 subunits in five turns with a rise of about 3 Å and a rotation of about 67° per monomer,[20,21] but there may be more than one monomer per asymmetric unit.[22−24] Models of virus capsids have been developed for fd, and for Pf1 in its high- and low-temperature forms, based both on X-ray fiber diffraction[20] and on the polarization inversion spin exchange at magic angle (PISEMA) static solid-state NMR technique.[25−28] However, the NMR-based models and X-ray based models are in poor agreement for both Pf1 and fd. An attempt to reconcile the differences for fd through a refinement procedure incorporating both types of data has been reported.[29] Pf1 is the only phage found, so far, to have a simple unit stoichiometry of one protein subunit for each nucleotide in its DNA,[6,30] a fact that led to a virion model having a maximally extended and twisted DNA helix with phosphates at the center.[23] The DNA part of that model has found support from single-molecule physics[31] and from polarized vibrational spectroscopy,[32] but it has not been fully established.[33] At present, there are 11 different models of the Pf1 capsid protein subunit in the Protein Data Bank (PDB).[34] Six entries (1IFM, 2IFM, 3IFM, 4IFM, 1PFI, and 1ZN4) are for the coat protein in the virion at temperatures below 10 °C, where a structural transition occurs, and the other five entries (1IFN, 2IFN, 1QL1, 1QL2, and 1PJF) are for the virus in its high-temperature form above 10 °C. All of these models propose high α-helicity for the subunit, but they differ in starting points and paths of the α-helical backbone, in the positions of the side chains, and in the deformations of the helix that are proposed (bends, kinks, and looped-out regions).

Two of the 11 entries in the PDB result from studies of Pf1 using static solid-state NMR experiments of properly aligned samples[35] using PISEMA-based techniques. Models have been published for other bacteriophages and membrane proteins as well. PISEMA solid-state NMR primarily probes the orientation of the backbone amide bonds with respect to the alignment axis and uses those bond angles as constraints for structural calculations, where assignments either are part of the fitting process[27] or are obtained using specific labeling.[36]

The magic-angle spinning (MAS) solid-state NMR technique does not require long-range order or alignment and can be used to study biological samples in a variety of forms that have been described as nanocrystalline, precipitated, lyophilzed, or aggregated.[37,38] It has recently become possible to obtain site-specific assignments of NMR resonances of proteins in the solid state using multidimensional heteronuclear (^{15}N and ^{13}C) MAS-based techniques.[37,39,40] Site-specific assignments obtained from MAS solid-state NMR (SSNMR) data can greatly facilitate the implementation of a variety of techniques that probe structure and dynamics efficiently throughout the molecule. Many proteins have been so assigned to date, including BPTI,[41] SH3,[42] ubiquitin,[43,44] GB1,[45] thioredoxin,[46] LH2,[47] kaliotoxin,[48] Crh,[49] and HET-s prion fibrils.[50] On the basis of such assignments, partial structural information has been obtained for several proteins,[51,52] and complete tertiary structures have been obtained in three cases.[48,53,54]

In this study, we have assigned most of the resonances in the MAS SSNMR spectra of the high-temperature form of the major coat protein in intact Pf1 filamentous phage using high-field 2D and 3D MAS SSNMR spectroscopy on precipitated, hydrated, infectious virus samples. With site-specific chemical shift assignments in hand, we have performed an initial database-derived conformational analysis and thus have been able to determine the secondary structure of the coat protein, predict potential location of kinks or bends, and determine that a single conformation of the coat protein exists in the virion. The use of a naturally occurring macromolecular assembly led to very high quality data here and indicates promise for atomic-level studies of other symmetric biological assemblies.

(18) Nasu, H.; Iida, T.; Sugahara, T.; Yamaichi, Y.; Park, K.-S.; Yokoyama, K.; Makino, K.; Shinagawa, H.; Honda, T. *J. Clin. Microbiol.* **2000**, *38*, 2156−2161.
(19) Caspar, D. L. D.; Makowski, L. *J. Mol. Biol.* **1981**, *145*, 611−617.
(20) Marvin, D. A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 150−158.
(21) Pederson, D. M.; Welsh, L. C.; Marvin, D. A.; Sampson, M.; Perham, R. N.; Yu, M.; Slater, M. R. *J. Mol. Biol.* **2001**, *309*, 401−421.
(22) Peterson, C.; Winter, W. T.; Dalack, G. W.; Day, L. A. *J. Mol. Biol.* **1982**, *162*, 877−881.
(23) Liu, D. J.; Day, L. A. *Science* **1994**, *265*, 671−674.
(24) Welsh, L. C.; Symmons, M. F.; Sturtevant, J. M.; Marvin, D. A.; Perham, R. N. *J. Mol. Biol.* **1998**, *283*, 155−177.
(25) Thiriot, D. S.; Nevzorov, A. A.; Zagyanskiy, L.; Wu, C. H.; Opella, S. J. *J. Mol. Biol.* **2004**, *341*, 869−879.
(26) Zeri, A. C.; Mesleh, M. F.; Nevzorov, A. A.; Opella, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6458−6463.
(27) Marassi, F. M.; Opella, S. J. *Protein Sci.* **2003**, *12*, 403−411.
(28) Thiriot, D. S.; Nevzorov, A. A.; Opella, S. J. *Protein Sci.* **2005**, *14*, 1064−1070.
(29) Marvin, D. A.; Welsh, L. C.; Symmons, M. F.; Scott, W. R. P.; Straus, S. K. *J. Mol. Biol.* **2006**, *355*, 294−309.
(30) Kostrikis, L. G.; Liu, D. J.; Day, L. A. *Biochemistry* **1994**, *33*, 1694−1703.
(31) Allemand, J. F.; Bensimon, D.; Lavery, R.; Croquette, V. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14152−14157.
(32) Tsuboi, M.; Kubo, Y.; Ikeda, T.; Overman, S. A.; Osman, O.; Thomas, G. J., Jr. *Biochemistry* **2003**, *42*, 940−950.
(33) Welsh, L. C.; Marvin, D. A.; Perham, R. N. *J. Mol. Biol.* **1998**, *284*, 1265−1271.
(34) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(35) Marassi, F. M.; Crowell, K. J. *J. Magn. Reson.* **2003**, *161*, 64−69.
(36) De Angelis, A. A.; Jones, D. H.; Grant, C. V.; Park, S. H.; Mesleh, M. F.; Opella, S. J. *Methods Enzymol.* **2005**, *394*, 350−382.
(37) Straus, S. K. *Philos. Trans. R. Soc. London B* **2004**, *359*, 997−1008.
(38) Tycko, R. *Prog. Nucl. Magn. Reson. Spectrosc.* **2003**, *42*, 53−68.
(39) Baldus, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *41*, 1−47.
(40) McDermott, A. E. *Curr. Opin. Struct. Biol.* **2004**, *14*, 554−561.
(41) McDermott, A.; Polenova, T.; Bockmann, A.; Zilm, K. W.; Paulsen, E. K.; Martin, R. W.; Montelione, G. T. *J. Biomol. NMR* **2000**, *16*, 209−219.
(42) Pauli, J.; Baldus, M.; van Rossum, B.; Groot, H. D.; Oschkinat, H. *ChemBioChem* **2001**, *2*, 272−281.
(43) Igumenova, T. I.; McDermott, A. E.; Zilm, K. W.; Martin, R. W.; Paulson, E. K.; Wand, A. J. *J. Am. Chem. Soc.* **2004**, *126*, 6720−6727.
(44) Igumenova, T. I.; Wand, A. J.; McDermott, A. E. *J. Am. Chem. Soc.* **2004**, *126*, 5323−5331.
(45) Franks, W. T.; Zhou, D. H.; Wylie, B. J.; Money, B. G.; Graesser, D. T.; Frericks, H. L.; Sahota, G.; Rienstra, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 12291−12305.
(46) Marulanda, D.; Tasayco, M. L.; Cataldi, M.; Arriaran, V.; Polenova, T. *J. Phys. Chem. B* **2005**, *109*, 18135−18145.
(47) van Gammeren, A. J.; Hulsbergen, F. B.; Hollander, J. G.; de Groot, H. J. M. *J. Biomol. NMR* **2005**, *31*, 279−293.
(48) Lange, A.; Becker, S.; Seidel, K.; Giller, K.; Pongs, O.; Baldus, M. *Angew. Chem., Int. Ed.* **2005**, *44*, 2089−2092.
(49) Bockmann, A.; Lange, A.; Galinier, A.; Luca, S.; Giraud, N.; Juy, M.; Heise, H.; Montserret, R.; Penin, F. O.; Baldus, M. *J. Biomol. NMR* **2003**, *27*, 323−339.
(50) Siemer, A. B.; Ritter, C.; Steinmetz, M. O.; Ernst, M.; Riek, R.; Meier, B. H. *J. Biomol. NMR* **2006**, *34*, 75−87.
(51) Wylie, B. J.; Franks, T.; Graesser, D. T.; Rienstra, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 11946−11947.
(52) Ladizhansky, V.; Jaroniec, C. P.; Diehl, A.; Oschkinat, H.; Griffin, R. G. *J. Am. Chem. Soc.* **2003**, *125*, 6827−6833.
(53) Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H. *Nature* **2002**, *420*, 98−102.
(54) Zech, S. G.; Wand, A. J.; McDermott, A. E. *J. Am. Chem. Soc.* **2005**, *127*, 8618−8626.

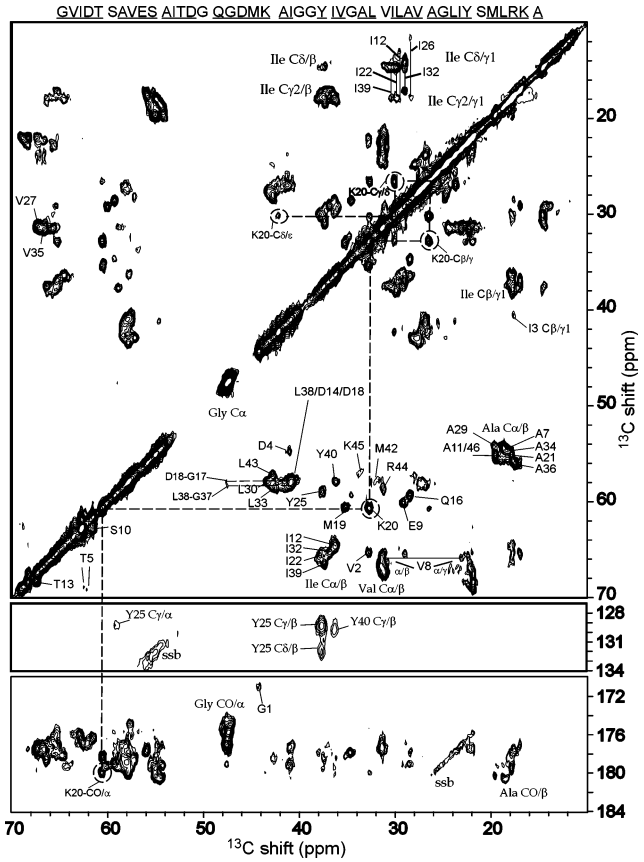

Figure 1. 2D $^{13}C-^{13}C$ DARR spectrum of the intact Pf1 virion (750 MHz, 10 ms mixing), dominated by contributions from the multicopy coat protein. The amino acid sequence of the coat protein is shown on top. Only 39 residues (underlined) of the total 46 are indicated in this particular figure (although all residues were assigned). The aliphatic (top) and carbonyl (bottom) regions were processed with a Lorentz-to-Gauss transformation in the acquisition dimension and a cosine bell in the indirect dimension. The aromatic region (center) was processed with 50 Hz exponential broadening in both dimension. C$\alpha-C\beta$ cross-peaks are indicated by the amino acid and residue number (e.g., K45). Other types of contacts are indicated explicitly. The complete "side-chain walk" of Lys20 is marked with a dotted line connecting the CO−Cα peak to the C$\alpha-C\beta$, C$\beta-C\gamma$, C$\gamma-C\delta$, and C$\delta-C\epsilon$ peaks. Representative peaks are shown for two-bond (several Ala CO−Cβ, Tyr25 C$\gamma-$Cα, Val8 C$\alpha-C\gamma$) and sequential (Gly17 and Gly37) contacts. The Ile side chains corresponding to C$\delta-C\gamma$1 and Cγ2−Cγ1 are explicitly indicated.

Results

Assignment of Resonances for the Intact Virion of Pf1 to Specific Atoms of Its Major Coat Protein. Resonance assignments of the Pf1 virion NMR spectra were established by means of 2D homonuclear $^{13}C-^{13}C$ correlation experiments[55,56] combined with 3D heteronuclear $^{15}N-^{13}C-^{13}C$ correlation experiments,[44,57] all conducted on uniformly ^{13}C- and ^{15}N-labeled virus particles. Sequential and intra-residue correlations were obtained by means of selective transfer between the backbone ^{15}N and either the ^{13}C of the carbonyl of the preceding residue or the Cα of the same residue, respectively.

A typical 2D $^{13}C-^{13}C$ dipolar assisted rotational resonance (DARR) spectrum, shown in Figure 1, illustrates assignments of many amino acid residue types according to their chemical shifts. Since ~93% of the virion mass is due to the 7300 subunits of the major coat protein,[5] the observed peaks in our spectra primarily represent those subunits. Minor proteins comprise ~1% of the total mass and are undetectable, and the DNA nucleic acids are normally weak and mostly resonate at different chemical shifts than those of the amino acids. In order to achieve site-specific assignments for backbone and side-chain atoms, we have performed 3D experiments, the results of which appear in Figure 2. Figure 2a shows two sets of strip plots connecting the backbone atoms of residues 8−15 (Val-Glu-Ser-Ala-Ile-Thr-Asp-Gly). Every strip depicts part of the $^{13}C-^{13}C$ plane cut at the ^{15}N shift where the peak for the indicated residue is strongest. The strips drawn in red are taken from NCACX spectra showing primarily intra-residue correlations, and those drawn in blue are from NCOCX spectra showing sequential correlations. For both experiments, in the top strips, the vertical and horizontal axes represent the Cα and CO shifts (in ppm), respectively. The ^{15}N shift of a particular plane and the CO shift (the center of the horizontal axis) are indicated at the bottom of those strips. The solid horizontal lines link strips of adjacent residues; thus, they connect the cross-peaks $N_i-CO_{i-1}-C\alpha_{i-1}$ (blue) and $N_i-C\alpha_i-CO_i$ (red), as has been done in previous studies.[58] The 3D experiment resolves congested regions because of relatively good dispersion in the ^{15}N dimension. In Figure 2b, a projection down the indirect CO dimension in the NCOCA spectrum reveals contacts between the ^{15}N shifts of a given residue and the CO and Cα shifts of the preceding residue. Such shifts are often diagnostic of the nature of both residues. For example, in the Pf1 homonuclear spectra (Figure 1), the Gly peaks are largely unresolved. However, as shown in Figure 2b, peaks belonging to Gly residues are clearly resolved on the basis of their ^{15}N shifts. Once site-specifically assigned, Gly (and Ser) residues served as starting points for the sequential assignment process. Several of the site-specific assignments of the Gly residues from the 3D experiments have been confirmed by sequential correlations in the 2D experiments.

The highly α-helical nature of the coat protein causes overlap in the 3D spectra. When backbone $N_i-CO_{i-1}-C\alpha_{i-1}$ (or $N_i-C\alpha_i-CO_i$) cross-peaks were not resolved in our spectra (noted by "NR" on top of the strip plots), peak congestion was relieved by analyzing two- and three-bond transfers in the 3D spectra. For many of the residues in the coat protein, the sequential correlation experiments exhibited $N_i-CO_{i-1}-C\beta_{i-1}$ cross-peaks, which were very useful for resolving degenerate peaks. These key peaks are indicated in the blue subspectra in Figure 2a (bottom), where the ^{15}N and CO shifts match those of the strip plots, and the vertical axis is centered on the corresponding Cβ value (indicated above every subspectrum). In two cases, S10→E9 and G15→D14, the auto-correlation peaks $N_i-CO_{i-1}-CO_{i-1}$ were better resolved and had a larger signal-to-noise ratio, and thus they were plotted in the figure with axes centered on the CO ^{13}C shift, as indicated in the corresponding strip plot. In a similar manner, $N_i-C\alpha_i-C\beta_i$ cross-peaks from the NCACX spectra are shown as well (red subspectra), with the Cβ shifts along the horizontal axis and the vertical axis centered on the Cα value. C$\alpha-C\beta$ peaks for most of the backbone residues have been detected in the 2D spectra and are indicated site-

(55) Morcombe, C. R.; Gaponenko, V.; Byrd, R. A.; Zilm, K. W. *J. Am. Chem. Soc.* **2004**, *126*, 7196−7197.
(56) Takegoshi, K.; Nakamura, S.; Terao, T. *Chem. Phys. Lett.* **2001**, *344*, 631−637.
(57) Straus, S. K.; Bremi, T.; Ernst, R. R. *J. Biomol. NMR* **1998**, *12*, 39−50.
(58) Bartels, C.; Xia, T. H.; Billeter, M.; Guntert, P.; Wuthrich, K. *J. Biomol. NMR* **1995**, *6*, 1−10.

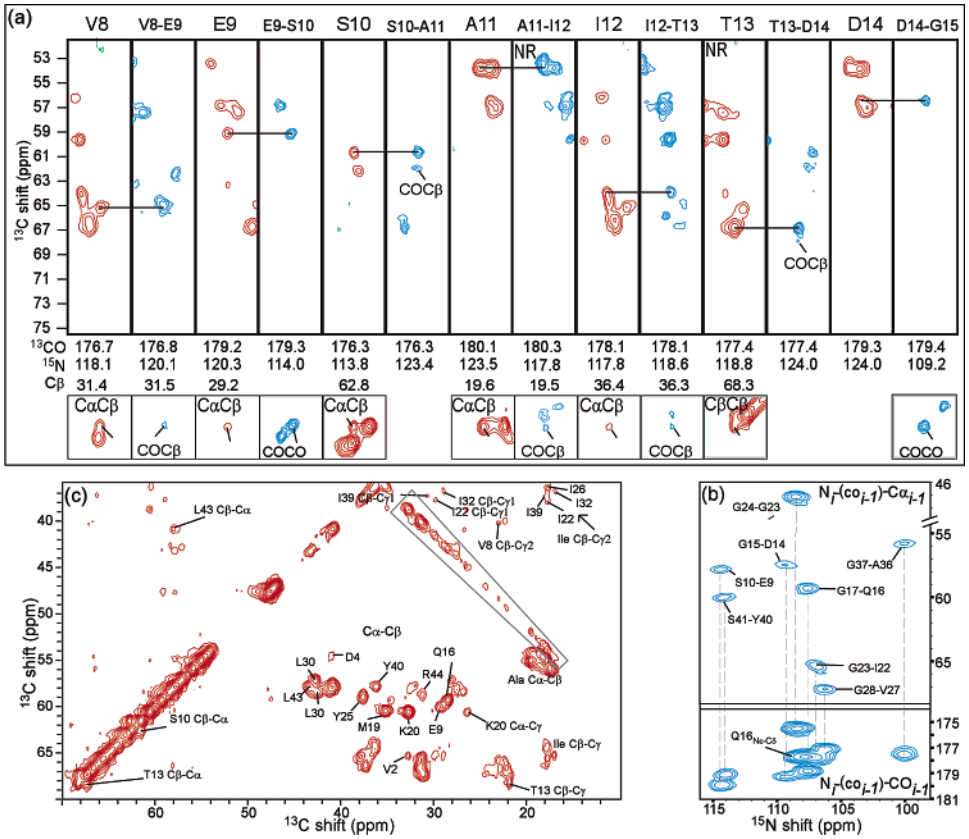

Figure 2. 3D heteronuclear correlation experiments on the Pf1 virion (750 MHz). (a) Strip plots for residues 8−15 show 2D contours in planes cut at the indicated ^{15}N frequency, with ^{13}CO chemical shifts on the horizontal axes and ^{13}Cα shifts on the vertical axis. Red and blue spectra correspond to intra-residue NCACX and sequential NCOCX experiments, respectively. Horizontal bars link the strips from ^{15}N planes of residues i and $i-1$ that have in common the CO−Cα peaks of residue $i-1$. The ^{15}N shifts and the center of the CO axis (width 4.5 ppm) corresponding to the particular strip are indicated at the bottom of the strip. The label "NR" indicates cases for which a peak was not resolved in one of the three dimensions. In such cases, alternative peaks were used for the assignment, e.g., the CO−Cβ peak of A11 (at the ^{15}N shift of S12) shown in one of the squares below the strips. The squares below the strips represent subspectra corresponding to 2D planes cut at the same ^{15}N shift as the strip above and display N−CO−Cβ or N−Cα−Cβ cross-peaks; the vertical axis is centered on either CO or Cβ for subspectra from the NCOCX experiment and is centered on the Cα chemical shift for the NCACX experiment. A width of 4.5 ppm in all subspectra was used in both dimensions. The lowest contours in the NCOCX spectrum were 8 times the average noise level and were incremented in multiplicative steps of 1.4. The lowest contours in the NCACX spectrum were 3 times higher than those for NCOCX, corresponding to the average difference in signal intensity between the two 3D spectra. (b) The projection down the CO dimension of the NCOCX sequential experiment is shown. The spectra show regions corresponding to the cross-peaks between the ^{15}N shifts of Gly and Ser residues and the CO−Cα peaks of the preceding residues. These contacts demonstrate many starting points for the sequential assignment process. In addition, the Nϵ_i−Cδ_i−Cδ_i peak belonging to Gln16 was identified (Nϵ_i−Cδ_i−Cγ_i observed but not shown). (c) The projection down the ^{15}N dimension of the NCACX is shown, with many individual assignments marked. Although this dimension is collapsed, the projection is a useful representation of the 3D spectrum and shows many of the cross-peaks resulting from two- and three-bond transfers. The gray box indicates the aliased auto-correlation peaks resulting from N−Cβ and N−Cγ transfers.

specifically in Figure 1, the 2D homonuclear spectrum, as well as Figure 2c, the projection of the NCACX 3D dataset down the ^{15}N dimension. This particular projection provides a useful summary of completeness, line shape, and peak congestion of the 3D spectrum. In addition, peaks resulting from two- and three-bond heteronuclear transfers in the NCACX experiments can be clearly identified. The longer-distance heteronuclear transfers were achieved using a 4 ms ^{15}N−^{13}C tangent-shaped cross-polarization element.[59] The resulting cross-peaks were aliased into the indirect dimension in our sampling scheme. Peaks emerging from N_i−Cβ_i(Cγ_i)−CX correlations have also proved beneficial for the assignment of this highly congested system. For example, many such correlations were observed for the Ile residues, as indicated in the projection spectrum, Figure 2c. This resolution of these bulky hydrophobic groups, as indicated specifically in Figure 1 for Cγ1−Cγ2 and Cγ1−Cδ cross-peaks, can provide useful long-range contacts because they are distributed throughout the protein, and thus are important for further structural analysis of the protein. Two-bond homonuclear correlations, such as Cα−Cγ correlations, are also manifested in the projection plot but are not visible in the strip plots. The NCACX experiment also provided many sequential contacts resulting from magnetization transfer from N to Cα of the same residue, followed by a two-bond transfer to the preceding residue, i.e., N_i−Cα_i−CO$_{i-1}$. Two such contacts (Ser41-Tyr40 and Ser10-Glu9) are illustrated in Figure S1 of the Supporting Information.

Completeness of the Coat Protein Assignment. Overall, 92% (231/251) of all nitrogen and carbon atoms in the protein were assigned to specific resonances. From a total of 138 backbone atoms, only two, the Met42 carbonyl peak (unresolved) and the N-terminal Gly1 amide, were not assigned. The complete list of assignments includes all of the residues, as detailed in Table S1 of the Supporting Information. The assignments are summarized schematically in Figure 3a. The vast majority of the assigned atoms show a single chemical shift,

(59) Baldus, M.; Geurts, D. G.; Hediger, S.; Meier, B. H. *J. Magn. Reson. A* **1996**, *118*, 140−144.

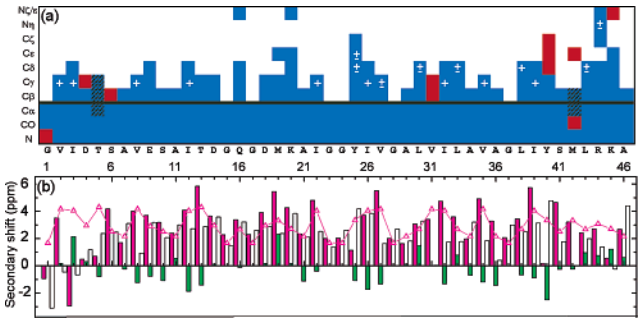

Figure 3. (a) Summary of the assignments, with a black horizontal bar separating the backbone from the side chains. Blue and red colors indicate the assigned and unassigned atoms, respectively. In order to account for the existence of duplicate atom labels (Cγ in Ile and Val, Cδ in Leu and Tyr, Cϵ in Tyr, and Nϵ in Arg), + and ± signs are used to designate resolution and assignment of both vs identification of a single resonance attributed to the functional group, respectively. The black shaded areas indicate the existence of multiple conformations for that atom. (b) Deviations of the experimental shifts from random coil values are shown as bars for Cα (magenta), Cβ (green), and CO (white). Triangles designate Cα average secondary shifts. The positive shifts of CO and Cα from residue 6 to the C-terminus suggest that the coat protein is mostly helical.

with typical backbone line widths of 80−150 Hz, suggesting a single conformation of the coat protein in the virion. Nevertheless, it is interesting that a few site-specifically assigned atoms belonging to two residues exhibit multiple shifts and broader, weaker peaks; these cases are indicated using slanted lines in Figure 3a. These structurally heterogeneous sites include Cα/β/γ of Thr5 in the helix termination and Cα/β of Met42 in the DNA binding domain. Characteristic spectra illustrating the basis for assignment for those resonances are explicitly depicted in Figure S1 of the Supporting Information.

Discussion

Previous studies[23,25,28,60,61] have shown that the coat protein in the virion of Pf1 has a helical conformation with a central kink or bend. Some models also suggest that the surface-exposed N-terminus is non-helical. More specifically, for the high-temperature form of the coat protein subunit there are three recent PDB depositions: (i) 1PJF, a refined model derived from static SSNMR data;[25] (ii) 1QL1, a recent refined model derived from X-ray fiber diffraction;[61] and (iii) 1QL2, based on the same set of fiber diffraction data as 1QL1, but incorporating a different symmetry.[61] The asymmetric unit for 1QL1 consists of a single subunit, whereas the asymmetric unit for 1QL2 consists of three subunits differing slightly in their conformations. The root-mean-square deviations (rmsd) for any pair of corresponding backbone atoms of the monomers of the trimers from 1QL2 are approximately 0.1 Å, while the rmsd for such pairs of side-chain atoms are larger, 1.2−1.5 Å. Trimers were introduced to account for near-meridional intensity on layer 16 of the X-ray diffraction pattern, intensity that was disallowed for a capsid symmetry of 27 units (monomers) in five turns but allowed for a capsid with 9 units (trimers) of five turns. The issue of capsid symmetry is crucial to structure determinations and hence understanding of capsid organization. Here we have tested the proposition of the presence of a trimer (three conformationally distinct copies of the coat protein per asymmetric unit) by examining these high-resolution NMR spectra for dispersion vs degenerate chemical shifts. If the virion contains three different conformers of the coat protein in equimolar amounts, multiple distinct chemical shifts for some of the atoms in the coat protein should be observed in these spectra. We have considered the magnitude of the differences expected in the chemical shifts for the three copies. To predict the NMR spectrum for the trimer in 1QL2, proposed for Pf1 in fibers, and compare the prediction with our data for the hydrated, precipitated samples, we used the program SHIFTX[62] to predict backbone and Cβ chemical shifts. The results (Table S2 of the Supporting Information) indicate that, for 26 residues, at least one atom (N, CO, Cα, or Cβ) is expected to show a chemical shift difference larger than 0.4 ppm for corresponding atoms in a pair within the trimer. Calculations with another program, SHIFTS,[63] gave a qualitatively similar result, with 37 residues (70 atoms) predicted to have shifts within two of the copies in the trimer that are larger than 0.4 ppm. As expected, the largest deviations are observed for the ^{15}N shifts, since they can be significantly affected by the side-chain conformations, where the rmsd is the largest.[63,64] We therefore posit that, if a trimer of the type described in the recent structure 1QL2 were present, many of the sites of the coat protein would give rise to peaks with apparent splittings, or multiple peaks that are partially resolved. Such dispersions (0.4 ppm or greater) could have been readily resolved in our spectra, if it were present. In stark contrast, our data exhibit narrow lines and single peaks for the vast majority of atoms. For only two particular cases (M42 and T5) do we observe multiple shifts. Thus, our data do not support the existence of the trimers described in 1QL2 for the high-temperature form of Pf1.

The chemical shifts can offer insight regarding the secondary structure of the protein. Since there remains debate about the details of the non-helical regions in this protein, we have predicted the likely secondary structure context for the amino acids in the Pf1 coat protein on the basis of our chemical shifts, and we have compared the results to the backbone conformations of 1QL1 and 1PJF. To highlight the differences in current structures, the 1QL1 and 1PJF model structures were aligned using the SwissPDB viewer[65] and found to have an rmsd of 2.84 Å for backbone atoms in the two structures (Figure 4c). They both are bent helices, as is clear from a comparison with an ideal helix (black).

We have used publicly available computational tools[66,67] to analyze our experimental chemical shifts and address whether previously proposed secondary structure models for the major capsid protein of Pf1 are likely to be correct. Secondary shifts, defined as the differences between observed shifts and random coil values taken from a database for backbone and Cβ chemical shifts,[68] have been shown to correlate to the three major secondary structure motifs: α-helix, β-sheet, and a coil.[69,70] A combination of secondary shifts from all assigned atoms

(60) Gonzales, A.; Nave, C.; Marvin, D. A. *Acta Crystallogr. D* **1995**, *51*, 792−804.
(61) Welsh, L. C.; Symmons, M. F.; Marvin, D. A. *Acta Crystallogr. D* **2000**, *56*, 137−150.
(62) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J. Biomol. NMR* **2003**, *26*, 215−240.
(63) Xu, X.-P.; Case, D. A. *J. Biomol. NMR* **2001**, *21*, 321−333.
(64) Dedios, A. C.; Pearson, J. G.; Oldfield, E. *Science* **1993**, *260*, 1491−1496.
(65) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714−2723.
(66) Wishart, D. S.; Case, D. A. *Methods Enzymol.* **2001**, *338*, 3−34.
(67) Hunter, C. A.; Packer, M. J.; Zonta, C. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *47*, 27−39.
(68) Wang, Y. J.; Jardetzky, O. *Protein Sci.* **2002**, *11*, 852−861.
(69) Spera, S.; Bax, A. *J. Am. Chem. Soc.* **1991**, *113*, 5490−5492.
(70) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311−333.

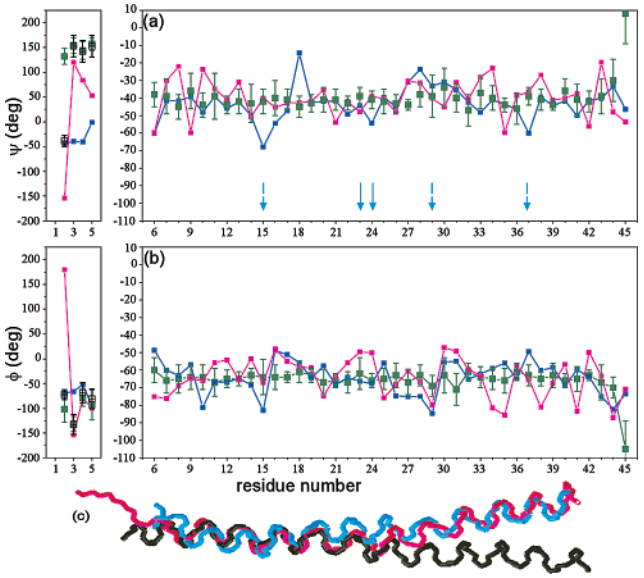

Figure 4. Comparison of torsion angles ψ (a) and ϕ (b), calculated from the Pf1 coat protein high-temperature models 1QL1 (magenta) and 1PJF (blue), to predictions by TALOS (green, error bars at one standard deviation) and PREDITOR (black), which are based on the observed SSNMR chemical shifts. PREDITOR data are shown only for residues 2−5 for clarity; however, they are similar to the results reported by TALOS throughout the backbone. Residues 2−5 are drawn on a different scale and, according to the predicted angles, correspond to a non-helical region. Residues 6−46 suggest a continuous helical conformation, although arrows indicate region of possible bends in the helix and are based on a calculation made with re-referenced chemical shifts and the program PSSI (see Discussion). (c) Models 1QL1 and 1PJF are structurally aligned (using SwissPDB viewer) and compared to a hypothetical straight helix (black), which was structurally fit to residues 6−20 of 1QL1.

comprises the chemical shift index (CSI),[71] which is used widely for secondary structure prediction in solution NMR studies. More recently, a probabilistic approach was proposed for a similar analysis and compiled into the program PSSI.[68] Torsion angle values can be predicted using TALOS[72] or PREDITOR (or SHIFTOR)[73] programs, which rely on a comparison of the observed shifts to a database of proteins with high-resolution X-ray structures and known solution NMR chemical shifts. Application of TALOS to ^{15}N and ^{13}C shifts from SSNMR studies has been reported with encouraging results for studies of structurally characterized systems, despite the fact that proton shifts are not included.[48,53,54]

The observed secondary shifts from our data are shown in Figure 3b for Cα (purple bars), CO (white), and Cβ (green). Average Cα shifts[68] for a helix are also shown (triangles) and are positive for all residues. Predicted torsion angles ψ and ϕ, based on chemical shifts, are presented in Figure 4a and b, respectively, and are overlaid with calculated torsion angles from 1QL1 (magenta) and 1PJF (blue). Both TALOS (green with error bars indicating one standard deviation) and PREDITOR (black) predict a long helix extending from residue 6 to the C terminus, and a non-structured N-terminus. Results from PREDITOR are similar to the TALOS predictions throughout the protein. We have also analyzed our data using the program PSSI,[68] and the results are similar: residues 1−5 are nonstructured and residues 6−46 are helical. Interestingly, the prediction for the secondary structure of Lys45 at the C-terminus gives a negligible probability for a helix, suggesting some sort of unique conformation of that residue in the DNA binding domain. When the resonances are subjected to a recently proposed referencing protocol,[74] comparable results are obtained by all three methods, except that certain resides in the center of the helix are predicted to have deviations from a helix, including A29 and G24 (see Table S3 of the Supporting Information for full details). These deviations are sufficient to cause a bend, as previously suggested.[23,25] However, a significant non-helical loop between residues I12 and M19, breaking the helix,[75] is not in accordance with our data. The regions where a bend is most likely to occur according to these predictions are indicated by arrows in Figure 4.

Conclusions

We present here the first assignment by MAS SSNMR of a coat protein of an intact virus. This success is significant in that we show here that α-helical segments of ∼40 residues can be assigned at moderate to high magnetic fields, despite formidable spectral congestion. Moreover, this protein was studied in the context of a native, infectious virus of nanoscale dimension and mega-Dalton mass, rather than in an artificial crystalline array. The infectivity was intact before and after the experiments.

In the context of this naturally occurring structure, quite uniform shifts and narrow line widths were observed, suggesting a repeating monomer unit throughout the virion. This result suggests reconsiderations of analyses of X-ray fiber diffraction data with respect to the basic symmetry assigned to the high-temperature form of Pf1. Any differences between subunits must be restricted to effects local to a few atoms.

Secondary structure prediction tools based on our chemical shifts suggest that the structure, with the exception of the first five residues in the N-terminus, is a long helix, the longest assigned thus far by MAS SSNMR methods. A gentle bend in the helix, as suggested in previous models, would be compatible with the data. Such a bend is likely to occur at the pair of Gly residues at the center of the protein or around residue A29.

The resonance assignment of the Pf1 coat protein spectra, including in particular the assignments for most of the side-chain atoms, serves as an initial and essential step toward obtaining detailed atomic-level structure and dynamics of the virion. This work demonstrates that MAS solid-state NMR has become a valuable tool not only for the study of globular proteins, membrane proteins, and protein aggregates but also for the study of natural whole biological assemblies, such as the family of filamentous phage.

Materials and Methods

^{13}C- and ^{15}N-Labeled Pf1 Phage. The phage and its host *P. aeruginosa*, strain K (PAK), were originally obtained from Prof. K. Amako.[1] They were maintained as stocks at the Public Health Research Institute. Standard bacteriological techniques were followed in the purification and assaying of the phage. A single colony of the host was grown for several generations in a minimal nutrient medium having $^{15}NH_4Cl$ and ^{13}C-glucose as the only sources of nitrogen and carbon. The medium contained 42 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.5 mM

(71) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171−180.
(72) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 289−302.
(73) Neal, S.; Berjanskii, M.; Zhang, H. Y.; Wishart, D. S. *Magn. Reson. Chem.* **2006**, *44*, S158−S167.
(74) Wang, Y. J.; Wishart, D. S. *J. Biomol. NMR* **2005**, *31*, 143−148.
(75) Nambudripad, R.; Stark, W.; Opella, S. J.; Makowski, L. *Science* **1991**, *252*, 1305−1308.

NaCl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 20 μM $FeSO_4$, and 10 μg of thiamine hydrochloride (vitamin B1), as well as 4 g/L uniformly labeled D-^{13}C-glucose and 1.6 g/L $^{15}NH_4Cl$ (both >99% isotopic label from Cambridge Isotope Laboratories). Cultures of labeled PAK were infected at $\sim 3 \times 10^7$ cfu/mL with previously labeled Pf1 at $\sim 3 \times 10^8$ pfu/mL. Vigorous aeration for 10−12 h yielded phage titers $>10^{12}$ pfu/mL. Bacteria were pelleted at 9400 relative centrifugal forces (RCF) for 20 min, and the phage in the supernatant solution was precipitated by the additions of polyethylene glycol (PEG) 8000 (Fluka) to 3 g/100 mL and NaCl to 3 g/100 mL. The precipitates were pelleted at 9400 RCF for 15 min and resuspended in 10 mM Tris-HCl buffer, pH 8.4. After centrifugation at 12 000 RCF for 15 min, the clarified phage solutions were brought to a density of 1.31 g/cm^3 with CsCl and centrifuged to equilibrium at 37 000 rpm for ~48 h in Beckman SW41 rotors. Liquid crystal bands formed at buoyant density 1.31 g/mL, as compared to 1.27 g/mL for unlabeled virus. The viscous band of virus was separated by pipet from minor bands above and below it and then subjected to an additional CsCl density gradient equilibrium banding. The solution of recovered virus was diluted to 0.2 M CsCl, after which the virus was precipitated by the addition of PEG 8000 to 3 g/100 mL, pelleted, and resuspended in 10 mM Tris-HCl buffer, pH 8.4. The virus had a UV absorbance spectrum with a maximum at 271 nm, a minimum at 245 nm, and an OD_{271}/OD_{245} ratio of 1.24 ± 0.01. Based on an extinction coefficient of 2.1 A mg^{-1} mL,[30] and a molecular weight of 36 MDa, approximately 40% ± 20% of the Pf1 physical particles gave plaques in assays of infectivity on its specific host PAK.

NMR Sample Preparation. Purified virus solutions at ~1 mg/mL were made 30% (v/v) in ethylene glycol as a cryo-protectant, and then made 22% (w/w) in PEG 8000. Upon the addition of $MgCl_2$ to 5 mM, the virus was quantitatively precipitated. Pelletings and transfers by centrifugation yielded hydrated Pf1 samples of 7−15 mg of virus in pellets of approximately 30 μL volume in the rotors for the NMR experiments. The low salt precipitation conditions minimized heating effects in the NMR experiments. Upon completion of the NMR data collection, the virus was recovered and found to have retained its infectivity.

NMR Experiments. Site-specific assignments were obtained by performing three-dimensional (3D) intra-residue NCACX and sequential NCOCX experiments. The pulse sequences are identical to those reported in Figure 2B of Igumenova et al.[44] and are based on the work of Straus et al.[57] Initial ^{15}N magnetization was created by a ramped cross-polarization (CP) process,[76] followed by a ^{15}N evolution period (t1). Magnetization transfer between ^{15}N and ^{13}C was achieved by a ramped double-CP scheme[59,77,78] that results in band selective carbon excitation. In the intra-residue NCACX experiment, mostly the backbone Cα carbons were excited, with partial excitation of side-chain Cβ and Cγ atoms, as discussed in the Results section. The selective transfer was achieved using radio frequency (RF) power levels of 2.5 times the spinning frequency (ν_r) for the ^{15}N channel and 1.5 ν_r for the ^{13}C channel. This condition simultaneously satisfies the Hartman−Hahn transfer condition in the solid state[79] and prevents recoupling due to resonance conditions between the rotor spinning frequency and the RF power.[78] In the sequential NCOCX experiment, band excitation of the carbonyl carbons was obtained using RF power levels of 2.5 ν_r for ^{13}C and 1.5 ν_r for ^{15}N. The higher power on the ^{13}C channel in this case prevented the carbonyl CSA broadening of the rotational resonance minima, thus reducing CP efficiency. The CP process was followed by a ^{13}C evolution period (t2). ^{13}C−^{13}C magnetization transfer was induced during the mixing time by using the dipolar assisted rotational resonance (DARR)[56] condition (or RF assisted diffusion, RAD[55]): ^{13}C magnetization was stored parallel to the magnetic field direction, followed by proton irradiation at a RF power matching the sample spinning frequency, and finally, ^{13}C magnetization was rotated back to the transverse plane for detection (t3). Two-dimensional (2D) homonuclear ^{13}C−^{13}C correlation experiments were carried out at mixing times of 1, 6, and 10 ms, employing the DARR scheme between two ^{13}C evolution periods. All experiments were performed at temperatures that ensure that the virion is in its high-temperature form at all times (see Table S4). 2D and 3D data were collected on Varian InfinityPlus 600 MHz and Bruker 750 MHz spectrometers with spinning frequencies of 11−15 kHz (see also Table S4). Spinning frequencies were chosen to minimize line broadening due to CO−Cα rotational resonance[80] and to avoid overlap of the secondary carbonyl sideband and the alanine CO−Cβ cross-peaks. Typical proton decoupling RF fields were ~85 kHz. DARR mixing at 1 ms was performed on a 400 MHz Varian spectrometer and was used mainly to differentiate between Asp Cβ−Cγ peaks and Leu CO−Cβ peaks by limiting the spectra to peaks emerging from one-bond transfers only. Detailed experimental and processing parameters are given in Table S4 of the Supporting Information.

Data Analysis. All spectra were processed using NMRPipe[81] and assigned using Sparky.[82] During the process of assignment, aliased peaks in the 3D NCACX experiment were given their correct chemical shift values by using features in Sparky. The chemical shift table generated by Sparky reported standard deviations of mostly 0.05−0.1 ppm for the ^{13}C shifts across the protein and <0.2 ppm for the ^{15}N shifts. ^{13}C and ^{15}N chemical shifts were referenced externally to the adamantane CH_2 peak at 40.48 ppm[83] and the NH_4Cl line at 39.27 ppm,[84] respectively.

Acknowledgment. This work was supported by funds from PHRI (L.A.D.) and the Othmer Institute for Interdisciplinary Research at the Polytechnic University (L.A.D.) and grants from NSF, MCB 0316248 (A.E.M.). A.E.M. is a member of the New York Structural Biology Center (NYSBC). Support for the NYSBC has been provided by NIH/NIGMS through grant P41 GM66354.

Supporting Information Available: Figure describing the assignment of Met42 and Thr5; list of assignments and line widths; and tables of chemical shift predictions of the trimers in the Pf1 structure 1QL2, torsion angle predictions, and experimental and processing parameters. This material is available free of charge via the Internet at http://pubs.acs.org.

JA066928U

(76) Hediger, S.; Meier, B. H.; Ernst, R. R. *Chem. Phys. Lett.* **1995**, *240*, 449−456.
(77) Schaefer, J.; McKay, R. A.; Stejskal, E. O. *J. Magn. Reson.* **1979**, *34*, 443.
(78) Baldus, M.; Petkova, A. T.; Herzfeld, J.; Griffin, R. G. *Mol. Phys.* **1998**, *95*, 1197−1207.
(79) Stejskal, E. O.; Schaefer, J.; Waugh, J. S. *J. Magn. Reson.* **1977**, *28*, 105−112.
(80) Raleigh, D. P.; Levitt, M. H.; Griffin, R. G. *Chem. Phys. Lett.* **1988**, *146*, 71−76.
(81) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(82) Goddard, T. D.; Kneller, D. G. *SPARKY 3.110*; University of California: San Francisco, CA, 2004.
(83) Morcombe, C. R.; Zilm, K. W. *J. Magn. Reson.* **2003**, *162*, 479−486.
(84) McDermott, A. E.; Gu, Z. In *Encyclopedia of NMR*; Grant, D. M., Harris, R., Eds.; Wiley: New York, 1996; Vol. 2, pp 1137−1147.