

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/51421346>

# Structure and Dynamics of the P7 Protein from the Bacteriophage $\phi$ 12

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · OCTOBER 2008

Impact Factor: 4.33 · DOI: 10.1016/j.jmb.2008.07.006 · Source: PubMed

---

CITATIONS

13

READS

28

## 9 AUTHORS, INCLUDING:



Ertan Eryilmaz

Boehringer Ingelheim, Ridgefield, CT, USA

16 PUBLICATIONS 94 CITATIONS

[SEE PROFILE](#)



Kaushik Dutta

New York Structural Biology Center

64 PUBLICATIONS 814 CITATIONS

[SEE PROFILE](#)



Paul Gottlieb

City College of New York

34 PUBLICATIONS 296 CITATIONS

[SEE PROFILE](#)



Ranajeet Ghose

City College of New York

62 PUBLICATIONS 975 CITATIONS

[SEE PROFILE](#)



# NIH Public Access

## Author Manuscript

*J Mol Biol.* Author manuscript; available in PMC 2009 October 3.

Published in final edited form as:

*J Mol Biol.* 2008 October 3; 382(2): 402–422. doi:10.1016/j.jmb.2008.07.006.

## Structure and Dynamics of the P7 Protein from the Bacteriophage φ12

Ertan Eryilmaz<sup>1,2,†</sup>, Jordi Benach<sup>3,†,§</sup>, Min Su<sup>3</sup>, Jayaraman Seetharaman<sup>3</sup>, Kaushik Dutta<sup>4</sup>, Hui Wei<sup>5</sup>, Paul Gottlieb<sup>2,5</sup>, John F. Hunt<sup>3</sup>, and Ranajeet Ghose<sup>1,2,\*</sup>

<sup>1</sup>Department of Chemistry, The City College of New York, 160 Convent Avenue, New York, NY 10031.

<sup>2</sup>The Graduate Center of the City University of New York, 365 Fifth Avenue, New York, NY 10016.

<sup>3</sup>Department of Biological Sciences and Northeast Structural Genomics Consortium, 702A Fairchild Center, MC2434, Columbia University, New York, NY 10027.

<sup>4</sup>The New York Structural Biology Center, 89 Convent Avenue, New York, NY 10027.

<sup>5</sup>Sophie Davis School of Biomedical Education, The City University of New York, 160 Convent Avenue, New York, NY 10031.

### Abstract

Cystoviruses are a class of enveloped double-stranded RNA viruses that use a multi-protein polymerase complex (PX) to replicate and transcribe the viral genome. Though the structures of the polymerase and ATPase components of the cystoviral PX are known and their functional behavior understood to a large extent, no atomic-resolution structural information is available for the major capsid protein P1 that defines the overall structure and symmetry of the viral capsid and the essential protein P7. Towards obtaining a complete structural and functional understanding of the cystoviral PX, we have obtained the structure of P7 from the cystovirus φ12 at a resolution of 1.8 Å. The N-terminal core region (1–129) of P7 forms a novel homodimeric α/β-fold with structural similarities with BRCT domains implicated in multiple protein-protein interactions in DNA repair proteins. Our results combined with the known role of P7 in stabilizing the nucleation complex during capsid assembly hints towards its participation in key protein-protein interactions within the cystoviral PX. Additionally, we have found through solution NMR studies that the C-terminal tail of P7 (130–169) that is essential for virus viability, though highly disordered, contains a nascent helix. We demonstrate for the very first time, through NMR titrations, that P7 is capable of interacting with RNA. We find that both the N-terminal core and the dynamic C-terminus tail of P7 play a role in RNA recognition leading to a significant reduction of the degree of disorder in the C-terminal tail. Given the requirement of P7 in maintaining genome packaging efficiency and transcriptional fidelity, our data suggest a central biological role for P7/RNA interactions.

---

© 2009 Elsevier Ltd. All rights reserved.

Address correspondence to: Ranajeet Ghose, Department of Chemistry, The City College of New York, 138<sup>th</sup> Street and Convent Avenue, New York, NY 10031. Phone: (212) 650-6049; Fax: (212) 650-6107; Email: rghose@sci.ccny.cuny.edu.

†These authors contributed equally to this work.

§Present address: ALBA Synchrotron Light Facility, Edifici Ciències. Mòdul C-3 Central Campus Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

### Accession Numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 2Q82.

## Keywords

Bacteriophage; cystovirus; RNA polymerase complex; structural protein; protein-nucleic acid interactions

## Introduction

The cystoviruses ( $\phi$ 6–14)<sup>1</sup> constitute an unique group of enveloped bacteriophages that utilize a three-segmented (S, M and L segments for small, medium and large) double-stranded RNA genome, infect strains of the plant pathogen, *Pseudomonas phaseolicola* and share an overall organization and replicative mechanism analogous to the reoviruses<sup>2</sup>. The nine members of the cystovirus family, discovered so far, have very similar genetic organization and express similar proteins. However, the primary genetic sequences in each dsRNA segment differ considerably allowing division into two major groups in which  $\phi$ 7,  $\phi$ 9,  $\phi$ 10, and  $\phi$ 11 are closely related to  $\phi$ 6, the prototypical (and best studied) cystovirus, while  $\phi$ 8,  $\phi$ 12, and  $\phi$ 13 are distantly related to  $\phi$ 6<sup>1; 3; 4; 5; 6</sup>. These differences allow variety in host specificity<sup>1</sup> and invasion mechanisms<sup>3; 4; 6</sup>.

The innermost layer of the cystoviral virion consists of an arrangement of four proteins<sup>7</sup>: P1 (major capsid protein)<sup>8</sup>, P2 (RNA-directed RNA polymerase)<sup>9; 10; 11</sup>, P4 (packaging ATPase)<sup>12</sup> and P7 (putative assembly co-factor)<sup>13</sup>. These proteins that are encoded by the L-segment of viral genome constitute the polymerase complex (PX) that is responsible for genome replication and transcription<sup>14</sup>. The structures of the PX in the two cystoviruses most distant from each other, namely  $\phi$ 6<sup>15</sup> and  $\phi$ 8<sup>16</sup>, as obtained from recent cryo-EM reconstructions, revealed an overall conservation of structure and organization.

Of the four PX proteins, high-resolution structures are available for the RNA-directed RNA polymerase (RdRp) P2 from  $\phi$ 6<sup>11</sup> and the hexameric ATPase motor P4<sup>12</sup> from  $\phi$ 12. Some low-resolution structural information is also available for the major capsid protein P1 that forms the skeleton of the viral capsid and determines its overall structure and symmetry properties (P1 assembles on a T=13 lattice with overall icosahedral symmetry). Recent 7.5 and 8.5 Å cryo-EM reconstructions of the  $\phi$ 6<sup>15</sup> and  $\phi$ 8<sup>16</sup> PX cores respectively indicated that the P1 protein was highly  $\alpha$ -helical and existed as an asymmetric dimeric species. Other studies have pointed towards a monomeric state for  $\phi$ 6 P1<sup>17</sup> and a tetrameric state for  $\phi$ 8 P1 in solution<sup>18</sup>. However a careful comparison of the electron densities corresponding to the P1 protein seems to imply significant differences in the tertiary structures of the protein from the two species<sup>16</sup>. The least characterized of the PX proteins is the protein P7. Little is known about its structure except that it exists in solution as an elongated dimer<sup>7; 13; 17</sup>. Also unclear is its exact location within the PX with no electron density obviously attributable to P7 found in either the  $\phi$ 6 or the  $\phi$ 8 reconstructions.

Assembly studies carried out by Poranen *et. al.*<sup>17</sup> revealed that in  $\phi$ 6 the incorporation of the P7 protein stabilized the PX and greatly enhanced the assembly rate of PX particles. Thus it was suggested that P7 was most likely an assembly co-factor that stabilized the PX through interactions with P1<sup>17</sup>. This hypothesis was confirmed in  $\phi$ 8 where incorporation of P7 into incomplete particles did not significantly alter the assembly kinetics due to the highly stable P1 tetramers formed as assembly intermediates in that case<sup>18</sup>. Additionally it was shown in  $\phi$ 6 that incomplete procapsid particles (PC) composed of the P1, P2 and P4 proteins (P124 particles) displayed minimal packaging of the positive strand of the viral genome. The P147 particles, however, packaged the positive strands with the same efficiency as the complete PC particles. Further, the specificity of the P147 particles towards the S and L segments was identical to that of the PC while minor differences were noted with the M segment. Thus, in

addition to stabilizing the assembly intermediates, P7 could also be considered a packaging co-factor<sup>13; 19</sup>. It was also demonstrated that P7-null PC particles were transcriptionally deficient and that P124 particles produced RNA transcripts of the incorrect size during semi-conservative positive strand synthesis<sup>13</sup> pointing to a possible role for P7 in regulating the transcriptional activity of the PX. Further it was noted that P7 possessed a flexible C-terminal tail<sup>13</sup> that was essential for virus viability<sup>20</sup>.

Towards our goal of complete structural and functional characterization of all proteins of the cystoviral PX we have obtained a crystal structure of the P7 protein (from the  $\phi$ 12 cystovirus) at 1.8 Å resolution. P7 forms a novel  $\alpha/\beta$ -fold and measurement of solution NMR relaxation rates combined with chemical cross-linking and light scattering studies confirm that it exists as a symmetric homodimer in solution. Solution NMR measurements indicate that the C-terminal tail of P7 though highly disordered possesses residual helical properties. We also provide evidence that P7 is capable of interacting with RNA and this interaction leads to a significant reduction in the degree of disorder in its dynamic C-terminal tail. Based on these observations and existing biological data we speculate a functional role for P7 in the cystoviral polymerase complex.

## Results

### P7 Possess a Highly Flexible C-terminal Tail which Interacts Minimally with the Protein Core

The  $^{15}\text{N}, ^1\text{H}$  HSQC spectrum of fully-protonated full-length P7 (1–169, P7fl) from  $\phi$ 12 revealed a set of highly intense peaks in a background of broad resonances indicative of a highly unstructured region in the protein. These intense resonances were not seen in a construct of P7 (P7 $\Delta$ C, 1–129) lacking the last 40 residues from the C-terminal tail (Figure 1(a)). This is consistent with previous suggestions that the P7 protein in the  $\phi$ 6<sup>20</sup> and  $\phi$ 12<sup>21</sup> cystoviruses possess flexible C-terminal tails.

We obtained complete backbone and sidechain assignments for  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  nuclei for the 41 C-terminal tail residues (129–169) in full-length P7 using standard heteronuclear NMR experiments using fully-protonated P7fl (Figure 1(b)). Additionally, backbone  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}^{\text{s}}$ ,  $^{13}\text{C}^{\alpha}$  and sidechain  $^{13}\text{C}^{\beta}$  assignments for the P7 core-domain (1–129) were obtained using P7 $\Delta$ C fractionally deuterated using the REDPRO approach,<sup>22</sup> and standard TROSY-based triple-resonance NMR experiments<sup>23</sup>. For P7 $\Delta$ C, only 103 of the expected 123 resonances (excluding the N-terminal Met and 5 Pro residues) were observed at 900 MHz, with some very weak peaks seen at the noise level, indicating a significant amount of conformational exchange. Of these 103 observable resonances unambiguous  $^{15}\text{N}$ ,  $^1\text{H}^{\text{N}}$  assignments could be obtained for a total of 98 resonances.

$^{15}\text{N}-^1\text{H}^{\text{N}}$  resonances corresponding to the P7 core in  $^{15}\text{N}, ^1\text{H}$  TROSY spectra of  $^{15}\text{N}$ , REDPRO- $^2\text{H}$ -labeled<sup>22</sup> P7fl and P7 $\Delta$ C were fully superimposable (Figure 1(a)) indicating no significant structural changes in the core upon C-terminal tail-truncation and minimal interaction of the C-terminal tail with the protein core. Additionally, tail truncation had no effect on the stability of the P7 protein with the melting temperatures of P7fl and P7 $\Delta$ C, as determined by differential scanning calorimetry (Del Rio et. al, unpublished results), being 47.5 °C and 48.4 °C respectively.

### Structure of the P7 Core

The P7 core (P7 $\Delta$ C) crystallized in a P3<sub>2</sub>1 space-group, the same space group and similar unit cell parameters as reported in previous preliminary crystallization studies<sup>21</sup>, and we solved its structure to a resolution of 1.8 Å using single wavelength anomalous diffraction (SAD). Details of the data collection, structure refinement and structural statistics are provided in Table

1. P7 $\Delta$ C forms a novel  $\alpha$ - $\beta$  sandwich (Figure 2) with two sets of  $\alpha$ -helices flanking a central five-stranded parallel  $\beta$ -sheet ( $\alpha$ - $\beta$ - $\beta$ - $\alpha$ - $\beta$ - $\alpha$ - $\beta$ - $\alpha$ - $\beta$ - $\alpha$  geometry). No electron density corresponding to the first 13 residues was observed. These residues were found to be highly dynamic in solution with the steady-state  $^{15}\text{N}$ -{ $^1\text{H}$ } NOE values varying from -0.47 to 0.37 (*vide infra*). Additionally, no electron density corresponding to residues Glu56 and Leu57 was seen. Notably,  $^{15}\text{N}$ - $^1\text{H}^\text{N}$  resonances corresponding to these two residues (the stretch between residues 56–59) could not be assigned in the TROSY spectra, most likely the result of exchange broadening due to conformational flexibility on the  $\mu\text{s}$ -ms timescale.

The structure of P7 $\Delta$ C contains 5  $\alpha$ -helices comprising residues 15–3 ( $\alpha$ 1), 49–2 ( $\alpha$ 2), 68–76 ( $\alpha$ 3), 98–103 ( $\alpha$ 4) and 112–127 ( $\alpha$ 5) and 5  $\beta$ -strands comprising residues 25–30 ( $\beta$ 1), 42–45 ( $\beta$ 2), 59–63 ( $\beta$ 3), 84–89 ( $\beta$ 4) and 109–111 ( $\beta$ 5). In addition, two  $3_{10}$  helices, the first between residues 65–67 connecting  $\beta$ 3 to  $\alpha$ 3, and the second comprising residues 78–80 lying between  $\beta$ 3 and  $\alpha$ 4, and two large loops (L1, 32–41) connecting  $\beta$ 1 and  $\beta$ 2 and second (L2, 90–97) connecting  $\alpha$ 4 and  $\beta$ 4 were seen. Multiple helix-helix contacts, involving largely hydrophobic interactions, stabilize the core structure. Some of these key interactions involve Leu19 ( $\alpha$ 1) with Ile119 ( $\alpha$ 5), Tyr22 ( $\alpha$ 1) with Val115 ( $\alpha$ 5), Met49 ( $\alpha$ 2) with Leu70 ( $\alpha$ 3) and Phe74 ( $\alpha$ 3), Ala71 ( $\alpha$ 3) with Val100 ( $\alpha$ 4). The sidechain of Asp118 on the C-terminal helix ( $\alpha$ 5) forms a salt-bridge that of Arg18 on  $\alpha$ 1. Helix  $\alpha$ 5 is rather unique, being relatively long (16 residues) with few stabilizing contacts especially at its C-terminal end and stabilized by crystal contacts (intermolecular contacts with  $\alpha$ 1 and  $\alpha$ 4 helices on neighboring molecules) *in crystallo*, is highly disordered in solution as determined by NMR relaxation studies described below.

It has been suggested that P7 exists as an elongated dimer in the cystoviruses  $\phi$ 6<sup>13</sup>,  $\phi$ 8<sup>18</sup> and  $\phi$ 12<sup>21</sup>. Chemical cross-linking studies (see Materials and Methods) revealed the dimer as the dominant species (> 95 %), both in P7fl and P7 $\Delta$ C indicating that C-terminal tail truncation does not interfere with the ability of P7 to dimerize, consistent with previous studies<sup>21</sup>. Static light scattering studies on P7 $\Delta$ C at protein concentrations ranging from 200  $\mu\text{M}$  to 1.7 mM at pH values of 6.5 and 7.5 (corresponding to the NMR and crystallization conditions respectively, see Materials and Methods) confirmed that P7 $\Delta$ C was dimeric over a wide range of concentrations. Further, NMR relaxation studies (*vide infra*) revealed a correlation time of 19.92 ns consistent with a dimeric P7 $\Delta$ C (see Table 2).

Thus, even though the crystals of P7 $\Delta$ C contained a single molecule per asymmetric unit, it is clear from the studies described in the previous paragraph that it was dimeric. Analysis of the  $^{15}\text{N}$ , $^1\text{H}$  TROSY spectra (Figure 1(a)) revealed a single set of resonances implying that P7 $\Delta$ C exists as a symmetric homodimer in solution. The dimer that produced the largest intermonomer contact surface was generated using one of the 2-fold crystallographic symmetry operations of the P3<sub>2</sub>1 space-group. This putative dimerization surface consists of interactions involving helices  $\alpha$ 2 and  $\alpha$ 3 (and adjoining loops) on each monomer (Figure 3(a)) and is composed of the largely hydrophobic residues - Thr48, Met49, Leu50, Leu53, Tyr72, Leu79, Ala80, Val81 and Gly82 (Figure 3(b)). Inter-monomer hydrogen bonds involving the backbone amides ( $\text{H}^\text{N}$ ) of Met49 and Leu50 and the carbonyl oxygens of Ala80 and Val81 respectively and a salt-bridge involving the sidechains of Asp69 and Arg77 stabilize the dimer interface (Figure 3(c)). Disruption of this salt bridge results in dissociation of the dimer (data shown for an Asp69 to Arg mutation in Figure 3(d)). Dimer formation results in the burial of approximately 10.2 % (674.2  $\text{\AA}^2$  of 6631.7  $\text{\AA}^2$  per subunit) of the solvent exposed surface in the monomer, a value that is quite characteristic of proteins of this size.

A search for structural homologues in the PDB and SCOP archives using SSM (Secondary Structure Matching) (<http://www.ebi.ac.uk/msd-srv/ssm>)<sup>24</sup> yielded several structures with Z-scores > 3.0. The structures with the largest Z scores included the C-terminal BRCT domain from human BRCA1<sup>25</sup> (PDB code: 1OQA, Z-score = 4.1, RMSD = 2.78  $\text{\AA}$ , 52 residues

aligned) and the third BRCT domain of similar to *S. pombe* RAD4+/CUT5+ (PDB code: 1WF6, Z-score = 3.7, RMSD = 3.05 Å, 58 residues aligned) (2004, Nagashima *et. al.* Riken Structural Genomics/Proteomics Initiative). Additional hits with Z-scores > 3.0 included other BRCT domains (PDB: 2D8M, 2COK) and thioredoxins (PDB: 1THX, 2GZZ, 1NW2, 2PVO, 1UVZ and 1W4V). Upon visual inspection of the individual structures it was found that, in spite of key differences, there were some topological and structural similarities between P7ΔC and BRCT domains which have a conserved fold consisting of a central 4-stranded parallel β-sheet flanked by two α-helices on one side and one α-helix on the other side. BRCT domains take part in key interactions involving proteins implicated in DNA damage repair and cell-cycle checkpoint control<sup>26, 27, 28</sup>. Interestingly, BRCT domains usually occur as dimeric tandems with the interaction surface composed primarily of hydrophobic residues on conserved helices of the interacting domains (see for example Figure 2 of Derbyshire *et. al.*<sup>26</sup>) similar to P7ΔC. BRCA1, where these domains were first discovered (BRCT - after the C-terminal domain of the breast cancer susceptibility protein BRCA1), has been shown to play a key role in transcriptional regulation.

### Dynamics of the P7 Core

We measured a complete set of backbone amide <sup>15</sup>N relaxation rates—R<sub>1</sub> ( $0.78 \pm 0.24 \text{ s}^{-1}$ ), R<sub>2</sub> ( $23.83 \pm 6.83 \text{ s}^{-1}$ ) and <sup>15</sup>N-{<sup>1</sup>H} NOE ( $0.58 \pm 0.34$ ) values in P7ΔC at 600 MHz. Data for residues 16, 17, 41, 54, 81, 89 were excluded due to spectral overlap and those for residues 33 and 110 were excluded since the corresponding resonance lineshapes showed “doubling” due to slow exchange. Thus the relaxation rates for a total of 89 residues could be reliably determined. The relaxation rates were analyzed utilizing the program DIFFTENS v2.0<sup>29</sup> using amide unit-vectors generated from the crystal structure of P7ΔC with hydrogen atoms added to the amide positions (without further energy minimization) using the program MOLMOL<sup>30</sup>. The analysis (see Table 2) yielded an axially-symmetric rotational diffusion tensor with a large axial ratio (D<sub>aniso</sub>, see Table 2) of 1.66 and an effective isotropic rotational correlation time of 19.92 ns.

Using the axially-symmetric diffusion tensor determined above we utilized the measured relaxation rates to calculate the micro-dynamic parameters of P7ΔC using the Lipari-Szabo<sup>31, 32</sup> model-free approach as implemented in the DYNAMICS package<sup>33</sup>. As noted above, the residues 1–13 were not present in the crystal structure and the microdynamic parameters for these residues were obtained using an isotropic model and an effective correlation time that was the same as that of the core structure. The S<sup>2</sup> values (Figure 4(a)) for the first 9 residues (varying from  $0.18 \pm 0.01$  to  $0.42 \pm 0.01$ ) were quite low indicating substantial disorder in solution and a possible explanation for the lack of electron density for this region. For the core structure (14–129), S<sup>2</sup> values for the center of α1 were smaller (Arg18:  $0.63 \pm 0.04$ ) than the average along the polypeptide chain ( $0.83 \pm 0.16$ , 10 % trimmed –  $0.87 \pm 0.08$ ) indicating some dynamics on the ps-ns timescale. Interestingly, resonances corresponding to residues 19–25 (the majority of the helix α1) could not be assigned. This could be in part due to significant exchange broadening of all or some of the resonances corresponding to this region suggesting conformational flexibility on both the fast (ps-ns) as well as the slow (μs-ms) timescales. This correlates quite well with the C<sup>α</sup> thermal factors for α1, which were, on an average, significantly higher (C<sup>α</sup> B-factors =  $46.9 \pm 7.5$ ) than other regions of the protein with definite secondary structure (C<sup>α</sup> B-factors =  $30.2 \pm 8.7$ ). Some low order parameters were also seen in the extreme C-terminus at the end of helix α5 though the C<sup>α</sup> thermal factors for this region are relatively low ( $27.2 \pm 4.6$ ). As mentioned previously, this could be the result of this helix being stabilized by crystal contacts that were lacking in solution. The central portion of this helix that is stabilized by intra-molecular interactions had extremely high S<sup>2</sup> values, suggesting fraying of the ends of α5. Notably, though the resonances corresponding to the N-terminal end of this helix consisting of residues 112–116 could not be assigned. It is nevertheless expected that this

region should be highly ordered given that the extensive network of contacts that stabilize it in the crystal structure. However, it is also possible that the resonances in this region are exchange broadened due to extensive conformer sampling and only one of these conformers was trapped in the crystal form. Interestingly, lower  $S^2$  values ( $0.73 \pm 0.07$  and  $0.79 \pm 0.03$  for Glu109 and Ser111 respectively) were seen in  $\beta 5$ . It is possible that the slight disorder in  $\beta 5$  could facilitate rigid body motions of  $\alpha 5$ . The Ser111 position is occupied by threonine residues in the  $\phi 6$  and  $\phi 8$  P7 proteins ( $\phi 13$  also contains a threonine residue that is displaced by one towards the N-terminal end, see Figure 7(a) below). While the inherent backbone flexibility of threonine residues is less than glycine or serine residues, it still lies at the center of the inherent flexibility scale<sup>34</sup>. The only other residues in the P7 $\Delta$ C core that had low  $S^2$  values were Phe74 ( $0.41 \pm 0.02$ ) that faces into the large central cavity at the dimer interface and is in close proximity to two crystallographic waters and Val100 ( $0.56 \pm 0.03$ ) which also points towards an internal cavity.

While the interpretation of relaxation rates in terms of local dynamics using the Lipari-Szabo model free approach is quite standard in globular proteins, it is to be remembered that the microdynamic parameters have a highly non-linear dependence on the relaxation rates. The influence of this non-linearity on model selection and on the determined order parameters is greatly exacerbated in the present context due to the highly anisotropic overall rotational diffusion. Additionally, the influence of local motional anisotropy<sup>35</sup> and residue-by-residue variations in the <sup>15</sup>N chemical shift anisotropy<sup>36; 37</sup> on the microdynamic parameters could also be enhanced due to the motional anisotropy possibly leading to erroneous values. It is also to be noted that explicit separation of local and overall motion is no longer possible when the overall rotational diffusion is anisotropic<sup>31; 32</sup>. In order to confirm our results from the model-free analysis on fast, ps-ns timescale along the protein backbone, we used the reduced spectral density function approach<sup>38; 39</sup>, that has a linear dependence on the measured relaxation rates, to evaluate the spectral density function around the <sup>1</sup>H Larmor frequency  $J(0.87\omega_H)$  (Figure 4(b)). The spectral density function at this frequency is most sensitive to sub-nanosecond dynamics. Though the spectral density functions at zero and the <sup>15</sup>N Larmor frequency were also determined, these are dominated by the effects of the overall rotational diffusion, which being highly anisotropic in the present case, complicates their direct interpretation<sup>40</sup>.  $J(0.87\omega_H)$ , on the other hand, is dominated by local effects and may be interpreted directly. The 10 % trimmed average value of  $J(0.87\omega_H)$  was found to be  $20.9 \pm 11.5$  ps/rad for the core region (14–129) of P7 $\Delta$ C. The same set of residues, as in the Lipari-Szabo model-free analysis described above, namely Phe74 ( $70.0 \pm 21.5$  ps/rad) and Val100 ( $37.0 \pm 4.4$  ps/rad) showed large values of  $J(0.87\omega_H)$  suggesting enhanced dynamics on the sub-nanosecond timescale. Additionally, Ile85, that lies in close spatial proximity to Phe74 also displayed a higher than average value of  $42.1 \pm 5.7$  ps/rad for  $J(0.87\omega_H)$ .

Several residues displayed medium to large  $R_{ex}$  values (Figure 5(a)) that varied from  $1.87 \pm 0.73$  s<sup>-1</sup> for Gln123 to  $12.50 \pm 2.71$  s<sup>-1</sup> for Val60, implying slow conformational dynamics on the  $\mu$ s-ms timescale. The largest  $R_{ex}$  values occurred in the hydrophobic core (Figure 5(b)) including two of the residues that showed the largest  $R_{ex}$  values – Val60 and Leu63 ( $12.11 \pm 2.88$  s<sup>-1</sup>) (red in Figure 5) and in the inter-monomer interface (Leu79  $8.21 \pm 1.87$  s<sup>-1</sup>). Both these clusters of significant  $R_{ex}$  values occur near large internal cavities where slow, breathing motion is not unexpected. Additional smaller  $R_{ex}$  values were also seen at the C-terminal end of  $\alpha 5$ . As pointed out earlier, the C-terminal end of  $\alpha 5$  is stabilized by intermolecular contacts in the P7 $\Delta$ C crystal and disorder on both the fast, ps-ns timescale (low  $S^2$  values, see above) and slow,  $\mu$ s-ms timescale is commonly observed. Additionally, resonances corresponding to residues 56–59 could not be assigned and this coupled with the fact that no electron density could be attributed to residues 56 and 57, suggest that this region in P7 $\Delta$ C could be dynamic on the  $\mu$ s-ms timescale. This region contains a proline residue that is strictly conserved in

cystoviruses (see Figure 7(a) below), a slow sampling of alternate conformations by this residue cannot be ruled out.

### Residual Structure in the C-terminal Tail of P7

As previously mentioned, C-terminal tail residues in full-length P7 protein were highly dynamic. The  $^{15}\text{N}-\{\text{H}\}$  NOE values for the C-terminal tail resonances were very low indicative of extensive disorder on the ps-ns timescale ( $0.03 \pm 0.32$ ) (see Figure 6(a)). However, the residues comprising the stretch between Leu136 and Gln145 (excluding Gln132) on the N-terminal end of the tail had significantly higher  $^{15}\text{N}-\{\text{H}\}$  NOE values ( $0.31 \pm 0.03$ ) indicating a higher degree of local order. Thus, in order to determine the presence of residual structure, if any, in the residues 136–145, we measured the difference in the deviations of the  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$  chemical shifts from their random coil values (see Figure 6(b))<sup>41</sup>. A set of positive values for the residues Leu136 through Gln145 hints towards the presence of a residual helical structure in this region. Additionally, we found medium-range ( $i,i+3$  and  $i,i+4$ ) and  $\text{H}^\alpha-\text{H}^\alpha$  connectivities<sup>42</sup> corresponding to the region 136–140 (Figure 6(c)) confirming the helix-like nature of this region. Notably,  $^3J(\text{H}^\alpha\text{H}^\alpha)$  values for this region were marginally smaller than the average ( $5.24 \pm 0.38$  Hz compared with  $7.02 \pm 1.67$  Hz), these values are usually  $< 3-4$  Hz for well-structured helices<sup>43</sup>. Thus, the region comprising residues Leu136 through Lys140 contains some residual helical structure. It is worth noting, that a comparison of the far-UV CD spectra (data not shown) of P7 and P7 $\Delta$ C revealed a change of approximately 4 % in helical content (as determined by the ellipticity value at 222 nm). This may be qualitatively interpreted as corresponding to a loss of about 6 helical residues upon tail truncation. It is possible that this region may form a well-ordered helix upon interactions with other components of the viral PX most notably the P1 protein which is known to interact with all of the PX proteins<sup>20</sup>. This speculation is further supported by Raman difference spectra obtained by Benevides et. al.<sup>44</sup> which seem to indicate a higher  $\alpha$ -helix content for full-length P7 in the fully assembled PX in  $\phi$ 6. It was suggested by the authors that this additional helicity was a result of P7/P1 interactions.

### Sequence Conservation Among Various Cystoviral P7 Proteins

An investigation of the P7 sequences from the  $\phi$ 6,  $\phi$ 8,  $\phi$ 12 and  $\phi$ 13 cystoviruses revealed that they could be broadly classified into two groups based on their sequences (see Figure 7(a)). As will be clarified in the discussion below. The hydrophobic residues at the dimer interface are largely conserved among all the cystovirus P7 proteins (Figure 7(a)). Additionally, as mentioned above, the dimer interface in  $\phi$ 12 P7 is stabilized by salt-bridges between residues Asp69 on one monomer and Arg77 on the other. Both these residues that lie on  $\alpha$ 3 form a D-(X)<sub>7</sub>-R sequence with X denoting any (mostly hydrophobic) residue. A similar sequence – E-(X)<sub>7</sub>-R can also be identified in  $\phi$ 8 P7 displaced by two residues towards the N-terminus of the protein. For  $\phi$ 6 and  $\phi$ 13 P7 proteins on the other hand, sequences given by R-(X)<sub>7</sub>-E can be identified at the dimer interface. These sequences lie on the  $\alpha$ 2 helix (as opposed to the  $\alpha$ 3 helix as in  $\phi$ 8 and  $\phi$ 12) and the adjoining loop. Notably, the dimer interface is bounded by  $\alpha$ 2 on the N-terminal end and  $\alpha$ 3 and the second  $3_{10}$  helix at the C-terminal end. This arrangement thus suggests a reversed “polarity” for the dimer interface for P7 proteins in  $\phi$ 6 and  $\phi$ 13 compared with  $\phi$ 8 and  $\phi$ 12. Nevertheless, based on the sequence analysis and expected structural features (see Figure 2(d)) it is expected that the overall nature of the dimer interface is likely to be largely similar for P7 proteins in all the cystoviruses. Additionally, as shown earlier, disruption of the salt bridge at the intermonomer interface results in dissociation of the P7 dimer (Figure 3(d)). It has been proposed that in  $\phi$ 6, dimeric P7 is required for assembly onto the nucleation complex formed by P4 hexamers and P1 dimers as an initial step in procapsid formation<sup>17</sup>. Thus inhibiting dimer formation by P7 is expected to have an adverse effect on capsid assembly.

The most strictly conserved residues in P7 lie on  $\beta$ 3 and form a P-(X)<sub>3/5</sub>-V-(X)-G motif. Of these residues only Pro58 (in  $\phi$ 12) is surface exposed (note that no electron density was seen for residues 56–57 and the degree of surface exposure could be modified in their presence) while the other two Val62 and Gly64 are buried and most likely play a role in stabilizing the overall P7 fold. Additional strictly conserved residues are seen at the N-terminal end of  $\beta$ 1 (Leu26 in  $\phi$ 12), in the loop between  $\alpha$ 4 and  $\beta$ 5 (Gly105) that facilitates formation of a tight  $\beta$ -turn between residues 104–107 and also most likely plays a structural role in stabilizing the P7 fold. Notably, the most highly conserved surface exposed residues (Glu23, Leu26, Leu53 and Pro58) form a single patch on the surface of P7 (Figure 7(b)). Interestingly, Leu26 ( $4.6 \pm 1.4 \text{ s}^{-1}$ ) and Leu53 ( $6.9 \pm 2.1 \text{ s}^{-1}$ ) show significant  $R_{\text{ex}}$  values and as previously stated residues 56–57 had no electron density attributable to them. Additionally, backbone amide resonances corresponding to the region 19–25 (that includes Glu23) and 56–59 could not be assigned (Figure 7(c)). These observations considered together are indicative of the presence of slow dynamics in the  $\mu$ s-ms timescale in this region of P7. It is notable that interaction surfaces in proteins often display slow conformational sampling forming “hot-spots” for protein-protein interactions<sup>45, 46</sup>. Also given the high degree of conservation of these residues it is highly likely that this surface patch takes part in key protein-protein interactions with other components of the cystoviral PX, most likely P1.

Beyond the core domain of P7, the general degree of sequence conservation is quite low. The only strictly conserved residue in the C-terminal tail region is an arginine residue (Arg141 in  $\phi$ 12). It is to be noted however, that the extreme C-terminal tail of P7 is highly acidic in all the sequenced cystoviral P7 proteins.

### Possible Role of P7 in Viral RNA Recognition

We next considered the intriguing possibility whether P7 could play a role in RNA recognition. It has been shown that P7 negative particles in  $\phi$ 6 package the positive strands of the viral genome with greatly reduced efficiency<sup>19</sup>. This effect is most likely not due to structural instability of the P7 negative particles since it has been shown that P14 and P124 particles are stable in  $\phi$ 6<sup>47</sup>. In order to test this hypothesis, we recorded a series of <sup>15</sup>N,<sup>1</sup>H HSQC spectra of fully-protonated, full-length P7 containing varying amounts of a 5-nt RNA constructs corresponding to the 5'-ends of the plus-strands of the S (Oligo1), M (Oligo2) or L (Oligo3) segments of the  $\phi$ 12 genome (see Materials and Methods). Upon analysis of the dynamic C-terminal tail resonances (data for Oligo1 is shown in Figure 8), significant chemical shift changes were seen in the presence of RNA. Majority of the residues (for the tail) that show large chemical shift perturbations were located near the C-terminal end of the protein (159–163) (see Figure 8(b)). Notably, some of the residues that showed substantial changes in chemical shift were negatively charged (notably Glu135, Glu143 and Asp160). These changes were unlikely due to the ionic strength effects of the added RNA since several other negatively charged residues located in this disordered region did not show chemical shift changes. Additionally (data not shown), analysis of the <sup>15</sup>N-{<sup>1</sup>H} NOE values of the C-terminal tail region revealed ordering on the ps-ns timescale in the presence of sub-stoichiometric amounts RNA (1:0.7 P7fl:Oligo1) (data not shown) with the largest changes displayed by His161 ( $\Delta\text{NOE} = 0.63 \pm 0.18$ ) and by Glu143 ( $\Delta\text{NOE} = 0.22 \pm 0.11$ ). Small but significant changes were also seen for Asp160, Gly162 and Ala163. The resonance corresponding to Phe159 was completely broadened out even at this P7fl: RNA ratio and could not be analyzed.

In order to test whether P7 could interact with RNA in the absence of the dynamic C-terminal tail, experiments similar to those described in the previous paragraph were carried out using <sup>15</sup>N,<sup>2</sup>H-REDPRO-labeled P7 $\Delta$ C containing varying amounts of Oligo2 and Oligo3. Substantial chemical shift changes were seen in the N-terminal core (results for Oligo2 are shown in Figure 9). These changes were localized in two specific regions – at the C-terminus

(near where the dynamic C-terminal tail would have been in the full-length construct suggesting perhaps that helix  $\alpha$ 5, also extremely dynamic in solution and the dynamic C-terminal tail play a role in interacting or sensing RNA) and at the opposite face of the protein (Figure 9(c)). Some changes were also seen at the extreme N-terminus (the N- and C-termini of P7 $\Delta$ C are highly dynamic, note that electron density corresponding to residues 1–13 were not seen). Given the extensive nature of chemical shift changes seen in the N-terminal core, and the small size of the RNA construct used, two possible scenarios can be imagined – (1) RNA binding by the core domain induces order in the dynamic C-terminal region or (2) RNA binds at two sites, the first located on the N-terminal core and a second that involves the C-terminal tail. These two scenarios are analyzed below (see Discussion section). In order to obtain an estimate of the affinity of P7 towards RNA, we fitted the changes in chemical shift with increasing RNA concentration for a select group of resonances in the fast exchange regime (those with chemical shift changes but no appreciable line-broadening) for both the C-terminal tail in P7fl as well as in P7 $\Delta$ C to a simple quadratic equation (corresponding to a one-site binding model)<sup>48</sup> and obtained  $K_d$  values between 50–80  $\mu$ M. However, it should be stressed here that this number may not provide a true measure of the affinity of P7/RNA interactions given the possibility of multiple-site binding and/or allosteric effects. The exact nature of these changes and the affinity of P7/RNA interactions can only be ascertained by further biochemical and biophysical studies and upon availability of the structure of P7 in full-length form bound to RNA. These studies are currently in progress in the laboratory.

## Discussion

The similarities between the  $\alpha/\beta$  folds of the homodimeric P7 core (P7 $\Delta$ C) with tandem BRCT domains hints towards a structural role for P7 in the cystoviral polymerase complex similar to that seen for BRCT domains<sup>28</sup> in DNA repair proteins. This is supported by *in vitro* assembly studies in  $\phi$ 6 which reveal that incorporation of P7 greatly accelerates formation of the P1–P4 nucleation and complete procapsids<sup>17</sup>. It has been suggested that this stabilization occurs through interactions between P1 and P7<sup>18</sup>. P1–P7 interactions in the fully assembled procapsid can also be envisaged to play a role during genome packaging. P7 could facilitate procapsid expansion through motional coupling with P1 segments. It has been suggested that procapsid expansion during packaging occurs due to rigid body movements of P1 domains<sup>15</sup>. This may be facilitated by disorder to order transitions in the C-terminal tail of P7 through direct interactions with P1 as indicated by the higher helix content in P7 in the context of the assembled PX<sup>44</sup>. Additionally, given the sensing/recognition of RNA by P7, and seemingly extensive conformational changes in the N-terminal core in the presence of RNA, P7 could serve as a secondary sensor of the positive strands of the genomic RNA during packaging although the primary recognition of the packaging signals (*pac* sequences on the 5' ends of the S, M and L segments) occurs through P1<sup>47</sup>. This role for P7 would be consistent with the fact that that  $\phi$ 6 procapsids lacking P7 package the positive strands of the viral genome with greatly reduced efficiency<sup>19</sup>. Given that chemical shift changes are seen for P7 $\Delta$ C, P7 does not require the C-terminal tail to interact with RNA. However, what is clear, from the widespread chemical shift changes seen, the N-terminal core region seems to undergo a large conformational change in the presence of RNA. This conformational change could be transmitted to the C-terminal tail which becomes more rigid either through interactions with the core region in its new conformational state either directly or through the bound RNA. These conformational changes could in turn facilitate genome packaging by allowing expansion of the empty capsids. In a second scenario, there could be multiple sites of P7/RNA interactions – at the C-terminus including the dynamic tail and at the opposite face of the protein. It is, however, quite clear from the present results that the recognition of RNA does not occur solely through electrostatic interactions of the polyanionic RNA with positively charged residues on P7 (Figure 10). Preliminary experiments monitoring the change of fluorescence anisotropy of 5'-fluorescein-tagged RNA in the presence of P7 seem to indicate that P7 in full-length form seems to bind

RNA with a higher affinity than P7ΔC (Eryilmaz et. al. unpublished results). Further, the affinity of RNA towards both P7fl and P7ΔC increases in the presence of Mg<sup>+2</sup>. Given that several residues that show large chemical shift changes in the presence of RNA both for the N-terminal core as well as the C-terminal tail are acidic, recognition of RNA by acidic residues mediated by Mg<sup>+2</sup> ions is also possible<sup>49</sup>. Clearly, additional experiments are needed to fully understand the nature of P7/RNA interactions and the role of the resultant conformational changes in the context of the PX. As mentioned earlier, P7 has structural similarities with BRCT domains which have been recently shown to bind DNA<sup>50</sup> in addition to serving protein-protein interaction and phosphoserine-binding modules.

Additionally, P7 negative particles have been shown to be deficient in positive strand synthesis producing transcripts of the incorrect size in ϕ6 while having normal replicase (negative strand synthesis) and polymerase activity (compared with the complete PC particles)<sup>13</sup>. Thus, P7 may play a role in regulating transcription perhaps by sensing the 3' ends of the positive strands being extruded through the P4 hexamers (which is supposed to act as a passive portal) during semi-conservative transcription<sup>7</sup>, signaling the end of the transcription process for a particular genomic segment and allowing correctly sized transcripts to be produced. This provides another possible biological function for P7/RNA interactions. It has been suggested that P7 is located along the 5-fold axes of the cystoviral PX<sup>13</sup>. P4 hexamers are also known to be located on the 5-fold axes, forming a symmetry mismatch<sup>15; 16</sup>. This would place P4 in close proximity to P7. A preliminary single particle reconstruction of the ϕ12 PX using electron cryo-microscopy indeed points to the presence of P7 along the 5-fold axes proximal to P4 hexamers (Morgan, *et. al.*, manuscript in preparation). On the other hand P7 could affect the polymerase activity of P2 through direct physical interactions with it. A recent cryo-EM reconstruction of the PX from ϕ6 suggests that while P2 is located close to the 3-fold axes in empty procapsids, it would be shifted closer to the 5-fold axes as expansion occurs during packaging bringing it closer to P4 and possibly P7<sup>51</sup>.

While we have completely characterized the structure and dynamics of P7 in full-length form using a combination of several biophysical techniques and provided evidence that it is capable of interacting with RNA, several questions still remain: What is the nature of the interactions of P7 with other components of the cystoviral PX most notably P1? Is the interaction 21 of P7 with RNA specific and does it depend on the length of the construct? Does P7 possess a single RNA-binding site and association with RNA triggers a large conformational change and ordering of the C-terminus or there are two sites that bind RNA? What is the exact mechanism that induces order in the flexible C-terminal tail in the presence of RNA? What is the physiological role of the P7/RNA interaction in controlling the efficiency of packaging the genomic plus strands<sup>19</sup> and regulating the fidelity of transcription<sup>13</sup> as evidenced by the available biochemical data? We are currently performing additional experiments to answer these questions.

## Materials and Methods

### Overexpression and Purification of P7

The ϕ12 P7 expression plasmid (pPG29) was prepared by PCR amplification of the P7 gene from pP12L1 template with recombinant Pfu DNA polymerase (Stratagene) using the following oligonucleotides 5'-GGTAACCATGGACTTCATTACTGAC-3' and 5'-AGGATCCTTATTCGTCGGCATCATCGAT -3' as upstream and downstream primers, respectively. The PCR fragment was digested with NcoI and BamHI and inserted into pET-21-d vector (Novagen) at NcoI-BamHI sites. Full-length (1–169) P7 was expressed in *E. coli* BL21 (DE3/pPG29) in ampicillin (100 ng/mL) supplemented LB medium. After the cells were grown overnight at 37 °C, 5 ml of overnight culture was used to inoculate 500 ml M9 medium containing <sup>15</sup>NH<sub>4</sub>Cl and <sup>13</sup>C<sub>6</sub>-glucose and grown at 37 °C to an A<sub>600</sub> of 0.6 followed by

induction with 0.5 mM of isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG) and growth at 20 °C for 14 hours. The cells were centrifuged and resuspended in 30 ml of TNM buffer (20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 7.5 mM MgCl<sub>2</sub>) containing 2 tablets of Complete Mini EDTA free protease inhibitors (Roche Applied Science). The suspension was lysed by sonication and the lysate was centrifuged at 18000 rpm for 30 min. at 4 °C. Supernatant fractions were loaded onto a HiTrap Heparin agarose HP column (Amersham Biosciences). The P7-containing flow-through was applied to a HiTrap Q-Sepharose HP column (Amersham Biosciences) and eluted with a linear 0.1–1 M NaCl gradient buffered with 20 mM Tris-HCl pH 7.5, containing 7.5 mM MgCl<sub>2</sub>. Fractions containing P7 were pooled, filtered, concentrated and injected onto a gel-filtration column (Superdex-200 26/60, Amersham Biosciences). The purified protein was dialyzed into NMR buffer containing 20 mM bis-tris, 50 mM NaCl, 7.5 mM MgCl<sub>2</sub> at pH 6.5.

For the production of the <sup>15</sup>N, <sup>13</sup>C, REDPRO-<sup>2</sup>H-labeled <sup>22</sup> protein, cells from the overnight culture were diluted 50 times and grown at 37 °C in unlabeled M9 minimal medium (H<sub>2</sub>O), until an A<sub>600</sub> of 0.6. Cells were collected by centrifugation, washed with phosphate-buffered saline, re-suspended in M9 minimal medium (containing <sup>15</sup>NH<sub>4</sub>Cl and <sup>13</sup>C<sub>6</sub>-glucose in > 99 % D<sub>2</sub>O) and induced with 0.5 mM of IPTG followed by overnight aeration at 20 °C. Further purification was carried out as described above. This procedure yielded 200 mg/L of purified triple-labeled (REDPRO) full-length P7 protein (P7fl).

Truncated P7 (P7ΔC, 1–129) was prepared in a fashion similar to full-length P7 (with the 3' PCR primer hybridized before the end of the encoding sequence) and was obtained with a similar yield.

Se-methionine labeled truncated P7 was expressed in the methionine-auxotroph *E. coli* strain B834 (DE3) in defined-M9 medium containing 100 ng/mL of ampicillin. Individual amino acids, one of which was Se-labeled methionine were added into the M9 medium (0.1 g each for a 2 L culture). The insoluble amino acids, tryptophan and tyrosine (0.5 g each), were dissolved in 10 mL 2 M HCl diluted to 100ml and 20ml of the solution was added to the medium. The expression was started with an overnight growth in LB medium (37 °C) that was then diluted 100-fold for overnight growth (37 °C) in M9 medium 20 mL of which was used to inoculate two liters of defined-M9 medium. The cells were grown at 37 °C to an A<sub>600</sub> of 0.6. At this point the cells were induced by 1 mM IPTG and grown for an additional 16 hours at 16 °C. The purification protocol was similar to full-length P7 except that gel-filtration column was not used, instead fractions collected from a Q-Sepharose HP column were reloaded in the same column and eluted with a NaCl gradient buffered with 20 mM bis-Tris at pH 6.0. The fractions were then exchanged into buffer containing 10 mM tris, 100 mM NaCl, pH 7.5.

### Chemical Crosslinking Studies

Chemical cross-linking experiments were performed at the NMR concentrations (~750 μM – 1.2 mM) to determine the oligomeric states of P7fl and P7ΔC. Purified P7fl/P7ΔC was diluted to 800 μM in TNM buffer, and 20 μL of samples were mixed with 2 μL of glutaraldehyde (up to 1.3 % final glutaraldehyde concentration). After 15 minutes of incubation, 20 μL of 2 X sample buffer was added and the samples were boiled immediately and analyzed on an SDS-PAGE gel.

### Differential Scanning Calorimetry

DSC experiments were performed using a VP-DSC microcalorimeter (MicroCal) with a scan rate of 1.5 deg/min. Samples were filtered and centrifuged at 14000 rpm prior to the measurements. Calorimetric cells (operating volume = 0.5 mL) were kept under an excess pressure of 207 kPa to prevent degassing during the scan. In all measurements, the buffer from

the final dialysis step was used in the reference cell of the calorimeter. The concentration of the P7fl and P7 $\Delta$ C were 1 mg/mL.

### CD Spectroscopy

Spectra were collected on a Jasco 810 spectropolarimeter at 25 °C using a cuvette with a 1 cm path-length. The protein concentration was 10  $\mu$ M, in PBS, containing 50 mM NaCl at pH 6.5. Spectra were acquired with a resolution of 1 nm from 260 to 200 nm, with 4 s acquisition time and signal-averaged over 10 scans. Estimates of the helical content were obtained using the following equation<sup>52</sup>

$$f_H = \frac{\theta_{222} - 2220 + 53T}{\theta_H - 2220 + 53T}$$

$$\theta_H = (-44000 + 250T) \left(1 - \frac{3}{N}\right) \quad (1)$$

where  $f_H$  is the fractional helical content at temperature  $T$  (in °C),  $\theta_{222}$  is the molar ellipticity measured at 222 nm,  $\theta_H$  is molar ellipticity of a pure  $\alpha$ -helix, and  $N$  is the number of residues in the protein.

### Interferometric Refractometry and Static Light Scattering Measurements

An Äkta Purifier chromatography system (Amersham Biotech, Piscataway, NJ) was used with a stainless steel Bio-Sil 250 guard column (BioRad, Hercules, CA) and an 8 mm × 300 mm Shodex Protein KW-802.5 (Waters Corporation, Milford, MA) analytical gel-filtration column. Both columns were packed with silica-based TSK-Gel resins manufactured by Tosoh (Tokyo, Japan). Buffers were degassed using an in-line ERC-3415a degasser (Showa Denko, Tokyo, Japan), and all tubing in the system was either Teflon or PEEK. The eluate from the column was monitored successively by the optical absorbance detector on the Äkta system (typically at 280 nm), followed by a Dawn EOS 14-angle SLS detector, and finally an Optilab DSP interferometric refractometer or RI detector (both from Wyatt Technologies, Santa Barbara, CA). All data were analyzed using the ASTRA software from Wyatt. The chromatography system and UV detector were housed in a cold box so that chromatography could be performed at 4 °C. The SLS and RI detectors were located outside of the cold box. SLS measurements were conducted at ambient temperature (~25 °C), but the sample chamber in the RI detector was maintained at 35 °C to minimize instability due to thermal drift. 100  $\mu$ l aliquots of sample containing the protein at 200  $\mu$ M, 500  $\mu$ M and 2.5 mM (200  $\mu$ M, 500  $\mu$ M, 1.7 mM) concentrations were injected onto the analytical gel filtration column running at a flow-rate of 0.5 ml/min. The HPLC column was equilibrated with NMR (crystallography) buffer; 20 mM Tris-HCl, 50 mM NaCl, 7.5 mM MgCl<sub>2</sub>, pH 7.5 (20 mM Tris-HCl, 100 mM NaCl, pH 7.5). The molecular mass of the protein was determined based on the SLS and RI signals from the peak using the ASTRA software (Wyatt Technologies, Santa Barbara, CA) based on a Debye plot of R( $\Theta$ )/c vs.  $\sin^2(\Theta/2)$ .

### NMR Spectroscopy

NMR spectra were recorded using <sup>15</sup>N- or <sup>15</sup>N,<sup>13</sup>C-labeled samples for P7fl and <sup>15</sup>N,REDPRO-<sup>2</sup>H- or <sup>15</sup>N,<sup>13</sup>C,REDPRO-<sup>2</sup>H-labeled samples for P7 $\Delta$ C on either Varian Inova spectrometer operating at 600 MHz (equipped with a triple-resonance probe capable of applying pulse field gradients along z-axis) or Bruker Avance spectrometers at 700, 800 or 900 MHz (all equipped with cryoprobes capable of applying pulse field gradients along z-axis). All experiments were performed at 25 °C, processed using NMRPipe<sup>53</sup> and analyzed using either NMRPipe<sup>53</sup> or NMRView<sup>54</sup> software suites.

## Backbone Resonance Assignments

For sequential backbone assignment of the flexible (128–169) C-terminal tail of P7fl, standard triple-resonance experiments HNCACB/CBCA(CO)NH, HN(CA)CO/HNCO and HNCA/HN(CO)CA<sup>55</sup> were used on fully-protonated <sup>15</sup>N,<sup>13</sup>C-labeled P7fl. Sidechain <sup>13</sup>C and <sup>1</sup>H assignments were obtained from (H)C(CO)NH and H(CCO)NH experiments respectively. <sup>3</sup>J(H<sup>N</sup>H<sup>a</sup>) coupling constants were obtained from an HNHA experiment<sup>43</sup> using a mixing time of 26.1 ms and an empirical correction factor of 0.9 to account for the differential relaxation of the antiphase and pure multiple quantum coherences. In addition a <sup>15</sup>N-edited NOESY with a 150 ms mixing time was used to locate regions of residual structure in the tail (128–169) of P7fl.

Sequential backbone assignments were obtained for <sup>15</sup>N,<sup>13</sup>C,REDPRO-<sup>2</sup>H-labeled P7ΔC using a standard combination of TROSY-based HN(CA)CO/HNCO, HNCA/HN(CO)CA and HNCACB/HN(CO)CACB experiments<sup>23</sup>.

## Backbone Relaxation Measurements

All relaxation experiments were performed at 600 MHz at 25 °C. <sup>15</sup>N-{<sup>1</sup>H} NOE measurements were made for fully-protonated <sup>15</sup>N-labeled P7fl and <sup>15</sup>N,<sup>13</sup>REDPRO-<sup>2</sup>H-labeled P7ΔC using standard pulse sequences with a period of 3 s with and without proton saturation and an additional delay of 1 s to allow locking. Additionally, a lower resolution <sup>15</sup>N-{<sup>1</sup>H} NOE dataset was collected with a 6.5 s period with and without proton-saturation (and an additional lock delay of 1 s) to evaluate the effects of incomplete equilibration of the magnetization on the NOE value. The two datasets produced results that were consistent within experimental error. R<sub>1</sub> and R<sub>2</sub> rates were measured in <sup>15</sup>N,<sup>13</sup>REDPRO-<sup>2</sup>H-labeled P7ΔC using recycle delays of 1.5 s and the following relaxation delays – for R<sub>1</sub> : 7.0 (x2), 102, 202, 302, 502, 702, 1002, 1202 and 1502 ms; for R<sub>2</sub> – 10 (x2), 30, 50, 70, 90, 110 (x2) and 130 ms. All data were processed using the NMRpipe suite<sup>53</sup> and relaxation rates were obtained using in-house software that utilized the ODRPACK<sup>56</sup> library. Errors for the <sup>15</sup>N-{<sup>1</sup>H} NOE were obtained by simple propagation of the uncertainties in peak intensities. Errors for the R<sub>1</sub> and R<sub>2</sub> (68.3 % confidence bounds) were obtained from the respective inverse covariance matrices of the fits.

## Determination of the Hydrodynamic Properties of P7ΔC in Solution from NMR Relaxation Data

The hydrodynamic properties of P7ΔC were determined using the program DIFFTENS v2.0 utilizing methods<sup>29</sup> described previously using the measured relaxation rates and the crystal structure of P7ΔC (after addition of protons using the program MOLMOL<sup>30</sup> and no further structure minimization). Errors (68.3 % confidence limits) in the principal values and orientation of the rotational diffusion tensor were obtained from the analytically determined inverse covariance matrix of the fits. Selection of models between the fully-anisotropic, axially-symmetric and isotropic models was performed using the statistical F-test. Probabilities (P %) indicating the possibility that the improvement in fits on increasing model complexity were obtained by chance, were calculated for each of the pairs of models – fully-anisotropic/axially-symmetric and axially-symmetric/isotropic. Values of P > 1 % were not considered to be statistically significant.

## Determination of Backbone Microdynamic Parameters

An analysis of the micro-dynamic motional parameters using the Lipari-Szabo framework<sup>31</sup>; <sup>57</sup> was performed utilizing the DYNAMICS package<sup>33</sup> using the measured R<sub>1</sub>, R<sub>2</sub> and <sup>15</sup>N-{<sup>1</sup>H} NOE values for P7ΔC. Errors (68.3 % confidence limits) in the micro-dynamic parameters were obtained from a Monte-Carlo analysis using synthetic datasets generated using

the experimental uncertainties in the measured relaxation rates. Computational approaches and motional models utilized in the analyses were as those described previously<sup>33</sup>.

The reduced spectral density function near the proton Larmor frequency,  $J(0.87\omega_H)$ , was evaluated using the following expression

$$J(0.87\omega_H) = - \left( \frac{\mu_0}{4\pi} \frac{\gamma_H \gamma_N \hbar}{2r_{\text{NH}}^3} \right)^{-2} \frac{\gamma_N}{5\gamma_H} (1 - \text{NOE}) R_1 \quad (2)$$

where the symbols have their usual meaning.

### **Crystallization of P7ΔC**

Monoclinic-shaped crystals grew at 20 °C in 1 + 1 μl hanging-drop vapor-diffusion reactions over a reservoir containing 23 % PEG 10000, 0.1 M sodium acetate, and 10 % n-Hexyl-β-D-glucoside. Crystals were flash-frozen in liquid propane using 23 % PEG 10000, 0.1 M sodium acetate, and 10 % n-Hexyl-β-D-glucoside and 25% PEG400 as cryoprotectant.

### **X-Ray Data Collection**

Single-wavelength anomalous diffraction data were collected from a single crystal at the selenium edge on a Quantum-4 CCD detector (ADSC, San Diego, CA) on beamline X12B at the National Synchrotron Light Source in a single sweep of  $400 \times 1^\circ$ , 8 s oscillations from a crystal maintained at ~100 K using 0.97947 Å radiation.

### **Data Processing, Model Building and Refinement**

Data were processed using HKL2000 and SCALEPACK<sup>58</sup>. For each data set, the scaling thermal B factors of the final frames were within 3 Å<sup>2</sup> of the initial frames, indicating that there was minimal crystal decay due to radiation damage. Given one molecule of P7ΔC per asymmetric unit in the P3<sub>2</sub>21 lattice, the packing density in the crystal was 2.7 Å<sup>3</sup>/Da, which was in the most probable range for proteins<sup>59</sup>. 7 of the 8 Se atoms in the asymmetric unit were identified using BnP<sup>60</sup>, yielding a map that was used for density modification (with 54% solvent content) and automated iterative model building in AUTO\_RESOLVE<sup>61</sup>. This program identified 60% of the backbone and 40% of the side chains in the final model and produced a map that enabled the structure to be built by hand using COOT<sup>62</sup>. Completion of the structure required iterative cycles of refinement in CNS<sup>63</sup> and manual rebuilding using standard stereochemical restraints<sup>64</sup>. The structure refinement was monitored using a randomly selected R<sub>free</sub> set containing 10% of the reflections. B-factors were refined using standard vicinal restraints (1.5–2.0 Å<sup>2</sup> for main-chain atoms and 2.0–2.5 Å<sup>2</sup> for side-chain atoms). Water-molecule sites were selected automatically using CNS and checked for consistency with 2F<sub>o</sub>-F<sub>c</sub> electron-density and hydrogen-bonding criteria. Structural quality was assessed using the program PROCHECK<sup>65</sup>. The biological unit was generated using the following symmetry operation(1-x,1-x+y,2/3-z)

### **NMR Titrations with RNA**

5'-nt RNA constructs corresponding to the 5' ends of the ϕ12 S (Oligo1 - 5'-rGrArArUrA-3'), M (Oligo 2 - 5'-rGrArArUrU-3') and (Oligo3 - 5'-rArCrArArU-3') were commercially synthesized (Integrated DNA Technologies). A set of <sup>1</sup>H, <sup>15</sup>N HSQC spectra were recorded (at 600 MHz) using <sup>15</sup>N-labeled fully-protonated P7fl (conc. 300 μM) in the presence of Oligo1 or Oligo2 (final protein to RNA ratios of 1:0.1, 1:0.2, 1:0.3, 1:0.4, 1:0.5, 1:0.6 and 1:1). All titrations were carried out in the NMR buffer (see above) in the presence of commercial RNase

inhibitor (RNAasein from Promega 10  $\mu$ l of 40 units/ $\mu$ l in a 300  $\mu$ l NMR sample). The scaled chemical shift changes ( $\Delta\delta$ ) were calculated using the following equation

$$\Delta\delta = \sqrt{(\Delta\omega_H)^2 + (0.11\Delta\omega_N)^2} \quad (3)$$

Where  $\Delta\omega_H$  and  $\Delta\omega_N$  are the chemical shift changes for the  $^1\text{H}$  and  $^{15}\text{N}$  dimensions respectively in the presence of RNA<sup>66</sup>. For Oligo3 HSQC spectra were recorded for final protein to RNA ratios of 1:0.25, 1:0.5 and 1:6.0. In addition, { $^1\text{H}$ }- $^{15}\text{N}$  steady-state NOE values were also measured (at 600 MHz) for  $^{15}\text{N}$ -labeled fully-protonated P7 containing Oligo1 at a sub-stoichiometric ratio of 1:0.7 using the same acquisition parameters above. Chemical shifts in the presence of RNA for the N-terminal core were monitored using  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectra (at 600 MHz) of  $^{15}\text{N}$ ,  $^2\text{H}$ -REDPRO-labeled P7 $\Delta$ C in the presence of Oligo2 and Oligo3 to final P7 $\Delta$ C:RNA ratios of 1:0.5, 1:1, 1:2, 1:4 and 1:0.25 and 1:4 respectively in the NMR buffer (see above).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

The authors thank Aswin Natarajan and Dr. Andrea Piserchio for help with protein purification and characterization; and Dr. Armando Del Rio for preliminary work on the biophysical characterization of P7; Dr. Maria-Luisa Tasayco for help with the DSC experiments and Dr Mark Arbing (Columbia University) for help with the Se-Met labeling protocol. This work has been supported by the following grants: MCB-0347100 and DBI-0619224 from the National Science Foundation to RG; 5G12 RR03060 towards support for the NMR facilities at the City College of New York, P41 GM-66354 to the New York Structural Biology Center and NIGMS-P50-GM62413 from the Protein Structure Initiative to the Northeast Structural Genomics Consortium from the National Institutes of Health.

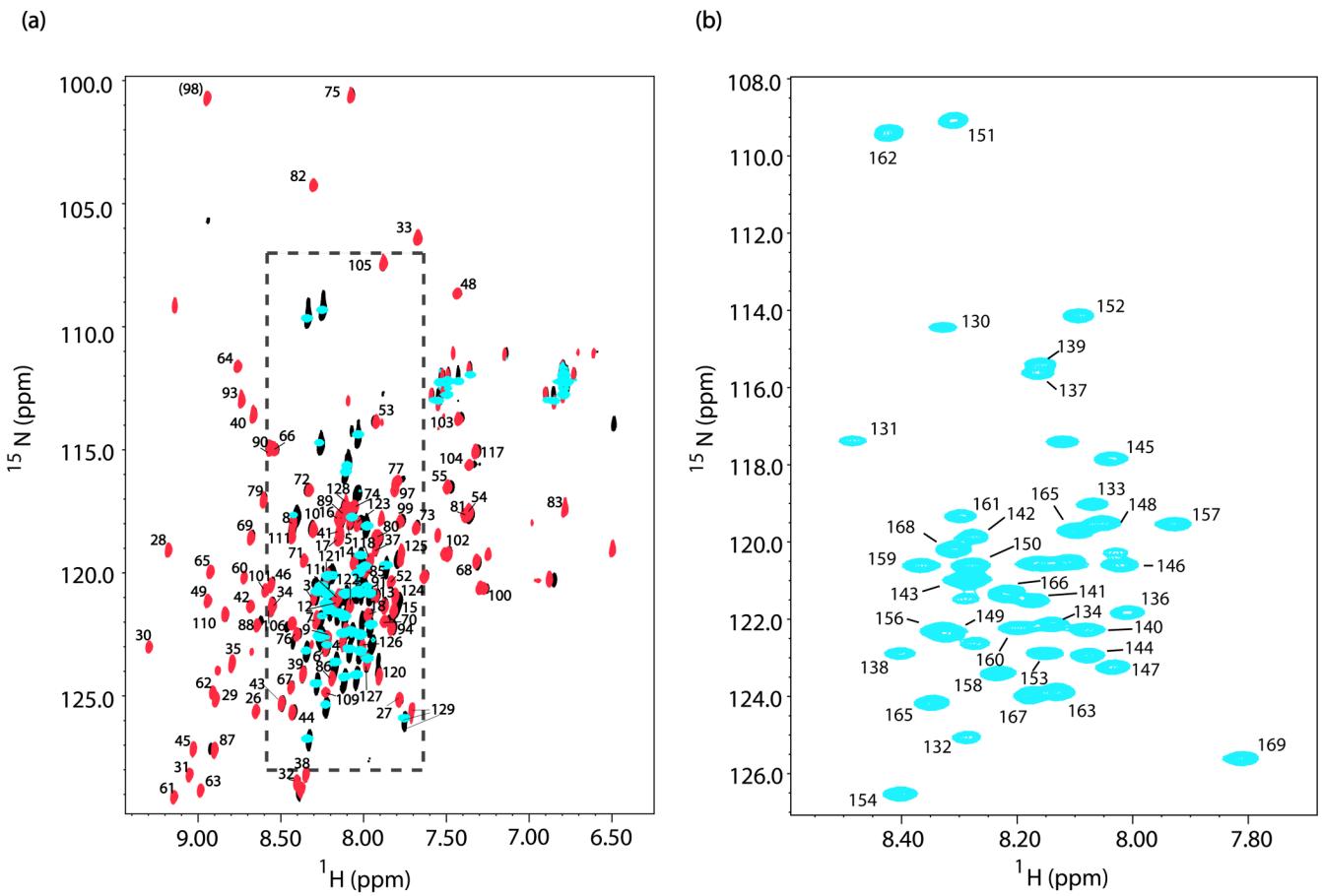
## References

1. Mindich L, Qiao X, Qiao J, Onodera S, Romantschuk M, Hoogstraten D. Isolation of additional bacteriophages with genomes of segmented double-stranded RNA. *J. Bacteriol* 1999;181:4505–4508. [PubMed: 10419946]
2. Mertens P. The dsRNA viruses. *Virus Res* 2004;101:3–13. [PubMed: 15010213]
3. Hoogstraten D, Qiao X, Sun Y, Hu A, Onodera S, Mindich L. Characterization of  $\phi$ 8, a bacteriophage containing three double-stranded RNA genomic segments and distantly related to  $\phi$ 6. *Virology* 2000;272:218–224. [PubMed: 10873764]
4. Qiao X, Qiao J, Onodera S, Mindich L. Characterization of  $\phi$ 13, a bacteriophage related to  $\phi$ 6 and containing three dsRNA genomic segments. *Virology* 2000;275:218–224. [PubMed: 11017801]
5. Gottlieb P, Potgieter C, Wei H, Toporovsky I. Characterization of  $\phi$ 12, a bacteriophage related to  $\phi$ 6: nucleotide sequence of the large double-stranded RNA. *Virology* 2002;295:266–271. [PubMed: 12033785]
6. Gottlieb P, Wei H, Potgieter C, Toporovsky I. Characterization of  $\phi$ 12, a bacteriophage related to  $\phi$ 6: nucleotide sequence of the small and middle double-stranded RNA. *Virology* 2002;293:118–124. [PubMed: 11853405]
7. Kainov DE, Lisal J, Bamford DH, Tuma R. Packaging motor from double-stranded RNA bacteriophage  $\phi$ 12 acts as an obligatory passive conduit during transcription. *Nucleic Acids Res* 2004;32:3515–3521. [PubMed: 15247341]
8. Butcher SJ, Dokland T, Ojala PM, Bamford DH, Fuller SD. Intermediates in the assembly pathway of the double-stranded RNA virus  $\phi$ 6. *EMBO J* 1997;16:4477–4487. [PubMed: 9250692]
9. Makeyev EV, Bamford DH. Replicase activity of purified recombinant protein P2 of double-stranded RNA bacteriophage  $\phi$ 6. *EMBO J* 2000;19:124–133. [PubMed: 10619851]

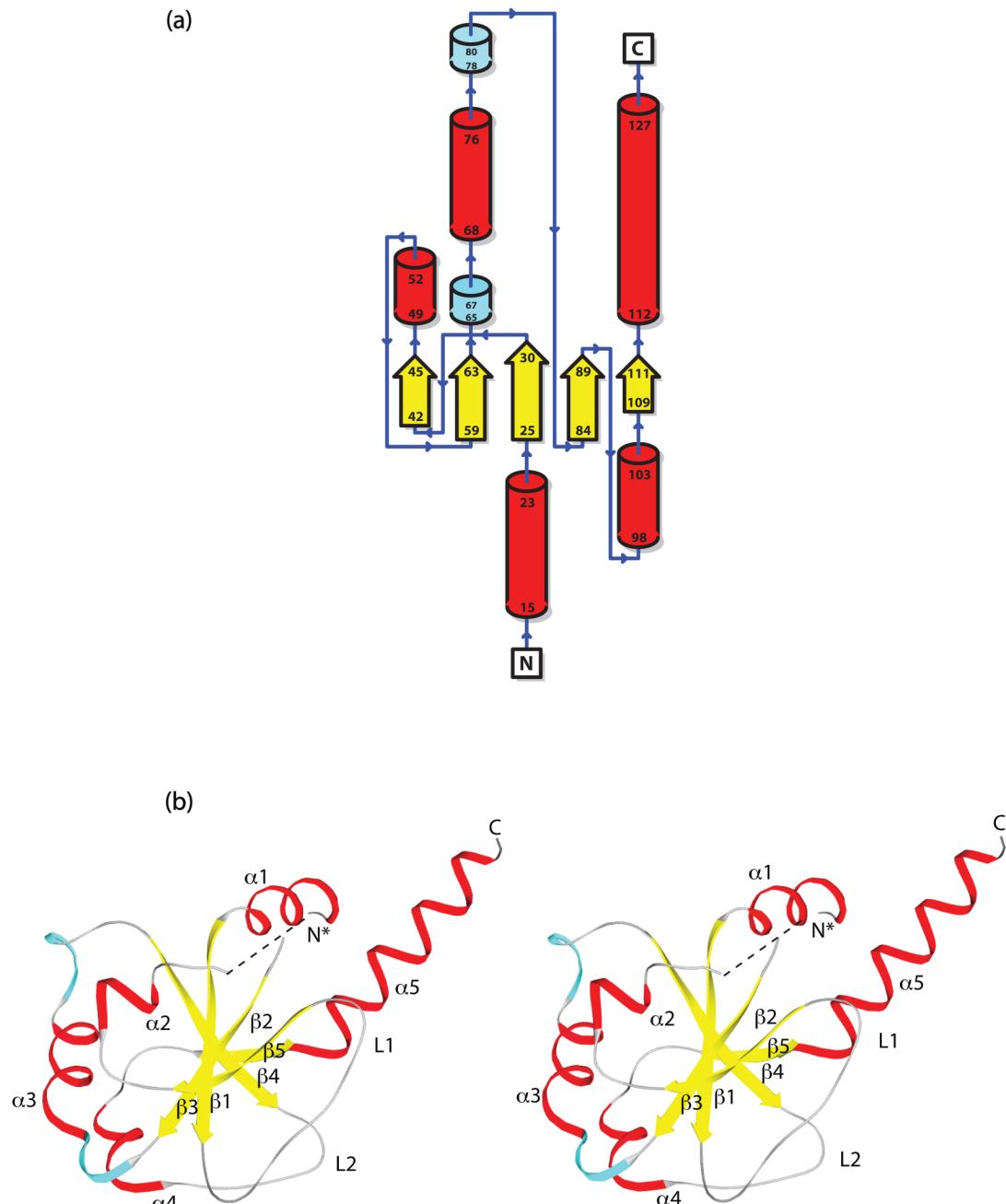
10. Yang H, Makeyev EV, Bamford DH. Comparison of polymerase subunits from double-stranded RNA bacteriophages. *J. Virol* 2001;75:11088–11095. [PubMed: 11602748]
11. Butcher SJ, Grimes JM, Makeyev EV, Bamford DH, Stuart DI. A mechanism for initiating RNA-dependent RNA polymerization. *Nature* 2001;410:235–240. [PubMed: 11242087]
12. Mancini EJ, Kainov DE, Grimes JM, Tuma R, Bamford DH, Stuart DI. Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation. *Cell* 2004;118:743–755. [PubMed: 15369673]
13. Juuti JT, Bamford DH. Protein P7 of phage  $\phi$ 6 RNA polymerase complex, acquiring of RNA packaging activity by in vitro assembly of the purified protein onto deficient particles. *J. Mol. Biol* 1997;266:891–900. [PubMed: 9086268]
14. Mindich L, Nemhauser I, Gottlieb P, Romantschuk M, Carton J, Frucht S, Strassman J, Bamford DH, Kalkkinen N. Nucleotide sequence of the large double-stranded RNA segment of bacteriophage  $\phi$ 6: genes specifying the viral replicase and transcriptase. *J. Virol* 1988;62:1180–1185. [PubMed: 3346944]
15. Huiskonen JT, de Haas F, Bubeck D, Bamford DH, Fuller SD, Butcher SJ. Structure of the bacteriophage  $\phi$ 6 nucleocapsid suggests a mechanism for sequential RNA packaging. *Structure* 2006;14:1039–1048. [PubMed: 16765897]
16. Jaalinoja HT, Huiskonen JT, Butcher SJ. Electron cryomicroscopy comparison of the architectures of the enveloped bacteriophages  $\phi$ 6 and  $\phi$ 8. *Structure* 2007;15:157–167. [PubMed: 17292834]
17. Poranen MM, Paatero AO, Tuma R, Bamford DH. Self-assembly of a viral molecular machine from purified protein and RNA constituents. *Mol. Cell* 2001;7:845–854. [PubMed: 11336707]
18. Kainov DE, Butcher SJ, Bamford DH, Tuma R. Conserved intermediates on the assembly pathway of double-stranded RNA bacteriophages. *J. Mol. Biol* 2003;328:791–804. [PubMed: 12729755]
19. Juuti JT, Bamford DH. RNA binding, packaging and polymerase activities of the different incomplete polymerase complex particles of dsRNA bacteriophage phi 6. *J Mol Biol* 1995;249:545–554. [PubMed: 7783210]
20. Poranen MM, Tuma R. Self-assembly of double-stranded RNA bacteriophages. *Virus Res* 2004;101:93–100. [PubMed: 15010220]
21. Kainov DE, Simonov V, Bamford DH, Tuma R, Gottlieb P, Wei H, Walsh MA, Belrhali H, Merckel MC. Crystallization and preliminary X-ray diffraction analysis of bacteriophage  $\phi$ 12 packaging factor P7. *Acta Crystallogr. D Biol. Crystallogr* 2004;60:2368–2370. [PubMed: 15583391]
22. Shekhtman A, Ghose R, Goger M, Cowburn D. NMR structure determination and investigation using a reduced proton (REDPRO) labelling strategy for proteins. *FEBS Lett* 2002;524:177–182. [PubMed: 12135763]
23. Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. USA* 1998;95:13585–13590. [PubMed: 9811843]
24. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr* 2004;60:2256–2268. [PubMed: 15572779]
25. Gaiser OJ, Ball LJ, Schmieder P, Leitner D, Strauss H, Wahl M, Kuhne R, Oschkinat H, Heinemann U. Solution structure, backbone dynamics, and association behavior of the C-terminal BRCT domain from the breast cancer-associated protein BRCA1. *Biochemistry* 2004;43:15983–15995. [PubMed: 15609993]
26. Derbyshire DJ, Basu BP, Serpell LC, Joo WS, Date T, Iwabuchi K, Doherty AJ. Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor. *EMBO J* 2002;21:3863–3872. [PubMed: 12110597]
27. Joo WS, Jeffrey PD, Cantor SB, Finnin MS, Livingston DM, Pavletich NP. Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brca1 BRCT structure. *Genes Dev* 2002;16:583–593. [PubMed: 11877378]
28. Glover JN, Williams RS, Lee MS. Interactions between BRCT repeats and phosphoproteins: tangled up in two. *Trends Biochem. Sci* 2004;29:579–585. [PubMed: 15501676]
29. Ghose R, Fushman D, Cowburn D. Determination of the rotational diffusion tensor of macromolecules in solution from NMR relaxation data with a combination of exact and approximate methods -

- application to the determination of interdomain orientation in multidomain proteins. *J. Magn. Reson* 2001;149:204–217. [PubMed: 11318619]
30. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph* 1996;14:51–55. 29–32. [PubMed: 8744573]
  31. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc* 1982;104:4546–4559.
  32. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules 2. *J. Am. Chem. Soc* 1982;104:4559–4570.
  33. Fushman D, Cahill S, Cowburn D. The main chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: analysis of  $^{15}\text{N}$  relaxation with monomer/dimer equilibration. *J. Mol. Biol* 1997;266:173–194. [PubMed: 9054979]
  34. Huang F, Nau WM. A conformational flexibility scale for amino acids in peptides. *Angew. Chem. Int. Ed. Engl* 2003;42:2269–2272. [PubMed: 12772159]
  35. Fischer MWF, Zeng L, Pang Y, Hu W, Majumdar A, Zuiderweg ERP. Experimental characterization of models for backbone picosecond dynamics in proteins. Quantification of NMR auto- and cross-correlation relaxation mechanisms involving different nuclei of the peptide plane. *J. Am. Chem. Soc* 1997;119:12629–12642.
  36. Fushman D, Tjandra N, Cowburn D. Direct measurement of  $^{15}\text{N}$  chemical shift anisotropy in solution. *J. Am. Chem. Soc* 1998;120:10947–10952.
  37. Kroenke CD, Rance M, Palmer AGI. Variability of the  $^{15}\text{N}$  chemical shift anisotropy in Escherichia coli ribonuclease H in solution. *J. Am. Chem. Soc* 1999;121:10119–10125.
  38. Farrow N, Zhang O, Szabo A, Torchia D, Kay L. Spectral density function mapping using  $^{15}\text{N}$  relaxation data. *J. Biomol. NMR* 1995;6:153–162. [PubMed: 8589604]
  39. Dayie KT, Wagner G, Lefevre JF. Theory and practice of nuclear spin relaxation in proteins. *Annu. Rev. Phys. Chem* 1996;47:243–282. [PubMed: 8930100]
  40. Pfeiffer S, Fushman D, Cowburn D. Simulated and NMR-derived backbone dynamics of a protein with significant flexibility: a comparison of spectral densities for the beta ARK PH domain. *J. Am. Chem. Soc* 2001;123:3021–3036. [PubMed: 11457013]
  41. Wishart DS, Sykes BD. Chemical shifts as a tool for structure determination. *Meth. Enzymol* 1994;239:363–392. [PubMed: 7830591]
  42. Wüthrich, K. NMR of proteins and nucleic acids. New York: John Wiley and Sons; 1986.
  43. Vuister GW, Bax A. Quantitative J correlation: a new approach for measuring homonuclear three-bond  $\text{J}(\text{H}^{\text{N}}\text{H}^{\alpha})$  coupling constants in  $^{15}\text{N}$ -enriched proteins. *J. Am. Chem. Soc* 1993;115:7772–7777.
  44. Benevides JM, Juuti JT, Tuma R, Bamford DH, Thomas GJ Jr. Characterization of subunit-specific interactions in a double-stranded RNA virus: Raman difference spectroscopy of the phi6 procapsid. *Biochemistry* 2002;41:11946–11953. [PubMed: 12356294]
  45. Ma B, Wolfson HJ, Nussinov R. Protein functional epitopes: hot spots, dynamics and combinatorial libraries. *Curr. Opin. Struct. Biol* 2001;11:364–369. [PubMed: 11406388]
  46. DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol* 2002;12:14–20. [PubMed: 11839484]
  47. Mindich L. Packaging, replication and recombination of the segmented genome of bacteriophage  $\phi$ 6 and its relatives. *Virus Res* 2004;101:83–92. [PubMed: 15010219]
  48. Lian, LY.; Roberts, GCK. NMR of Macromolecules: A Practical Approach. Roberts, GCK., editor. New York: Oxford Univ. Press; 1993.
  49. Lee SJ, Richardson CC. Acidic residues in the nucleotide-binding site of the bacteriophage T7 DNA primase. *J. Biol. Chem* 2005;280:26984–26991. [PubMed: 15917241]
  50. Kobayashi M, Figaroa F, Meeuwenoord N, Jansen LE, Siegal G. Characterization of the DNA binding and structural properties of the BRCT region of human replication factor C p140 subunit. *J. Biol. Chem* 2006;281:4308–4317. [PubMed: 16361700]
  51. Sen A, Heymann JB, Cheng N, Qiao J, Mindich L, Steven AC. Initial location of the RNA-dependent RNA polymerase in the bacteriophage  $\phi$ 6 procapsid determined by cryo-electron microscopy. *J. Biol. Chem* 2008;283:12227–12231. [PubMed: 18287088]

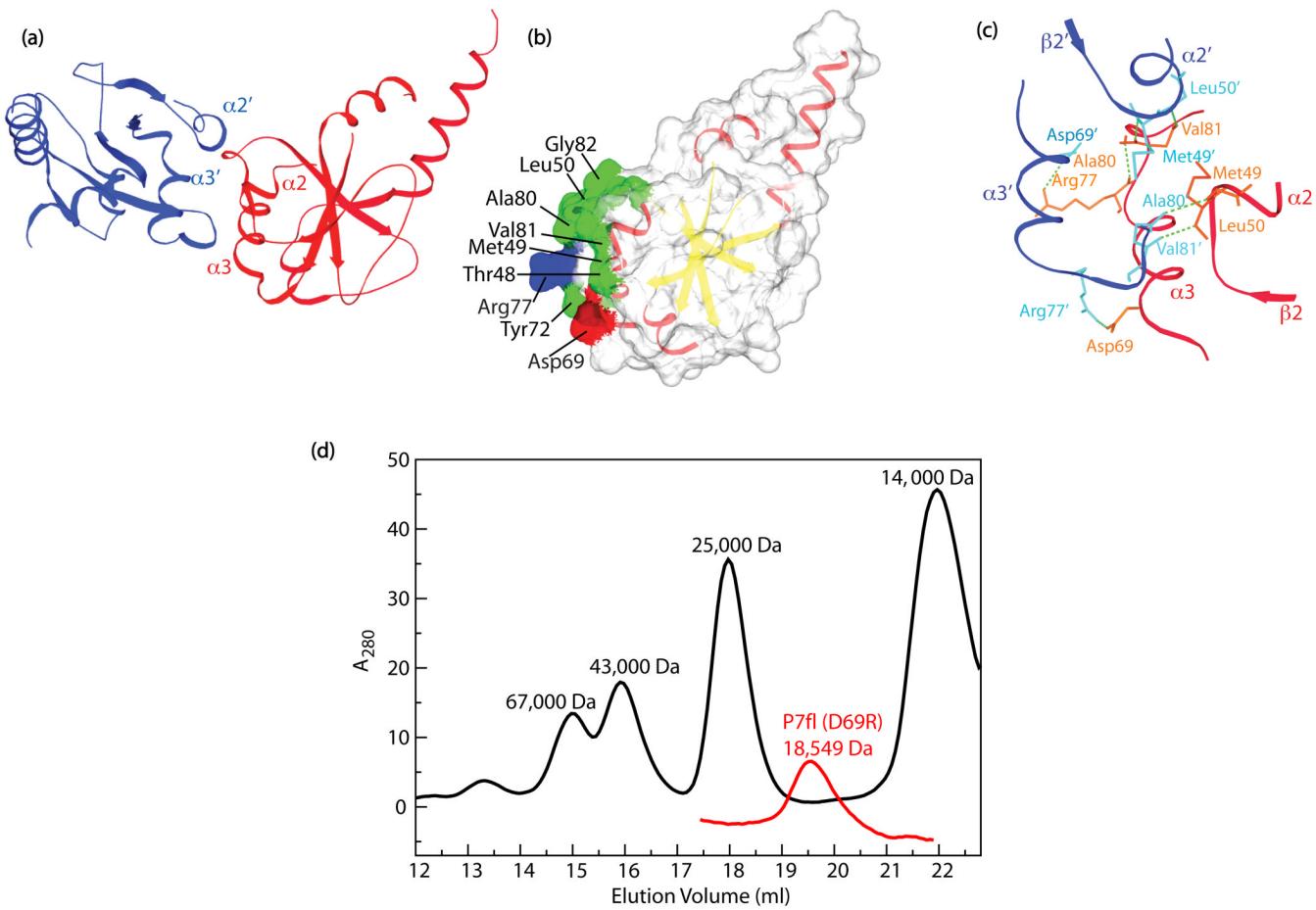
52. Luo P, Baldwin RL. Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* 1997;36:8413–8421. [PubMed: 9204889]
53. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 1995;6:277–293. [PubMed: 8520220]
54. Johnson BA. Using NMR View to visualize and analyze the NMR spectra of macromolecules. *Meth. Mol. Biol.* 2004;278:313–352.
55. Sattler M, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectr.* 1999;34:93–158.
56. Boggs PT, Donaldson JR, Byrd RH, Schnabel RB. ODRPACK software for weighted orthogonal distance regression. *ACM Trans. Math. Software* 1989;15:348–364.
57. Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* 1990;112:4989–4936.
58. Otwinowski Z, Minor W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Meth. Enzymol.* 1997;276:307–326.
59. Matthews BW. Solvent content of protein crystals. *J. Mol. Biol.* 1968;33:491–497. [PubMed: 5700707]
60. Weeks CM, Blessing RH, Miller R, Mungee R, Potter SA, Rappleye J, Smith GD, Xu H, Furey W. Towards automated protein structure determination: BnP, the SnB-PHASES interface. *Z. Krystallo* 2002;217:686–693.
61. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 1999;55:849–861. [PubMed: 10089316]
62. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 2004;60:2126–2132. [PubMed: 15572765]
63. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 1998;54:905–921. [PubMed: 9757107]
64. Engh R, Huber R. Accurate bond and angle parameters. *Acta Crystallogr. A* 1991;47:392–400.
65. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystall.* 1993;26:283–291.
66. Shekhtman A, Ghose R, Wang D, Cole PA, Cowburn D. Novel mechanism of regulation of the non-receptor protein tyrosine kinase Csk: insights from NMR mapping studies and site-directed mutagenesis. *J. Mol. Biol.* 2001;314:129–138. [PubMed: 11724538]
67. Laskowski RA. Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature. *Bioinformatics* 2007;23:1824–1827. [PubMed: 17384425]
68. Laskowski RA, Chistyakov VV, Thornton JM. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res.* 2005;33:D266–D268. [PubMed: 15608193]
69. Laskowski RA. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.* 2001;29:221–222. [PubMed: 11125097]
70. Hutchinson EG, Thornton JM. PROMOTIF--a program to identify and analyze structural motifs in proteins. *Protein Sci.* 1996;5:212–220. [PubMed: 8745398]
71. Drenth, J. Principles of protein crystallography. New York: Springer-Verlag; 1994.

**Figure 1.**

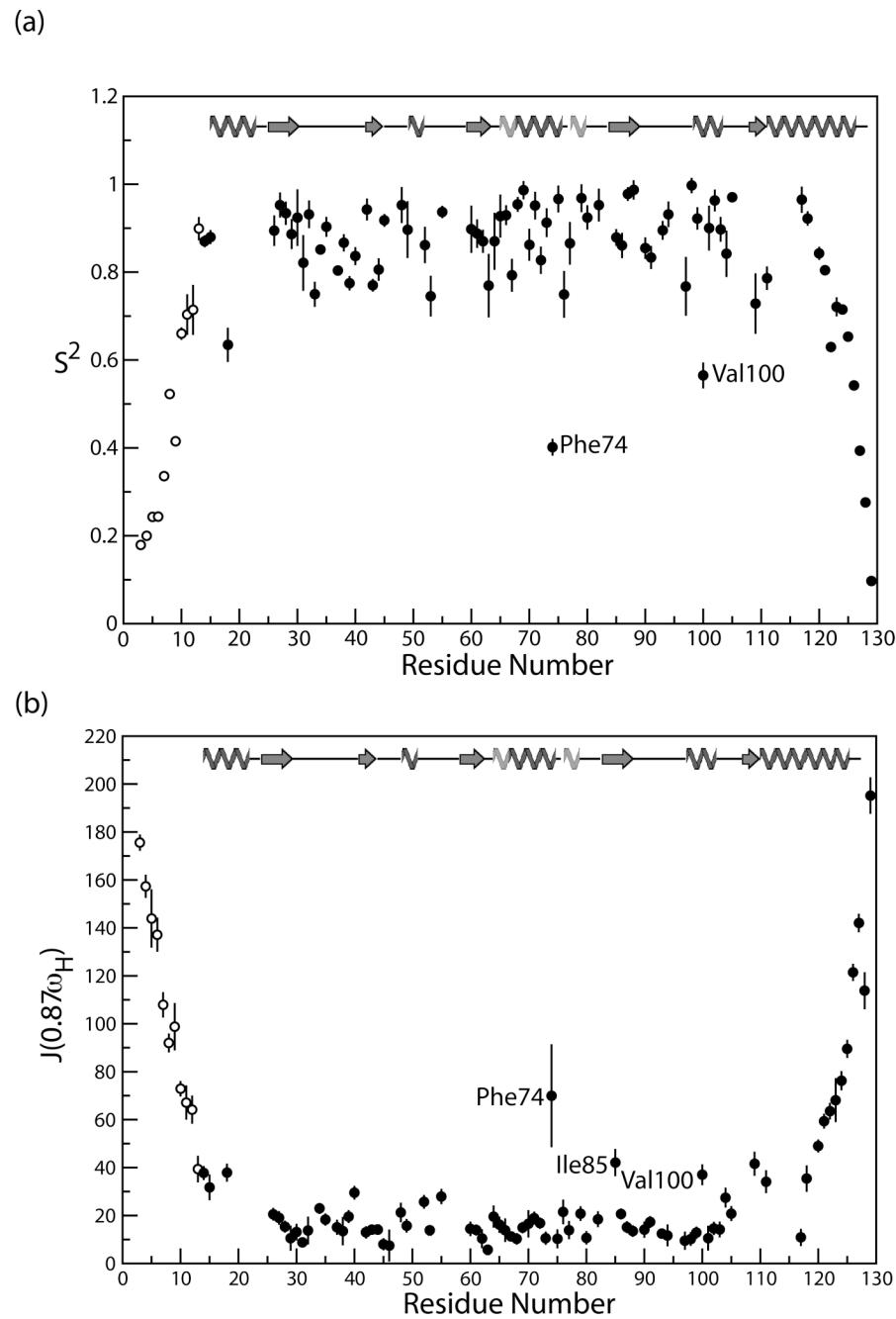
The flexible C-terminal tail of P7 interacts minimally with the protein core. (a) Overlay of  $^{15}\text{N}, ^1\text{H}$  TROSY spectra (800 MHz) of  $^{15}\text{N}$ ,REDPRO- $^2\text{H}$ -labeled P7fl (black) and P7 $\Delta$ C (red). The Ala98 resonance is aliased. Also shown are the intense resonances corresponding to the C-terminal tail residues in a  $^{15}\text{N}, ^1\text{H}$  HSQC spectrum (600 MHz, cyan) of fully-protonated P7fl. The broad resonances corresponding to the protein core are not seen when the spectrum is suitably contoured. The C-terminal tail resonances were independently assigned in the context of P7fl. Peaks have been shifted to compensate for the  $^1\text{J}_{\text{NH}}$  couplings in both dimensions. The boxed region is expanded in (b). (b) Resonance assignments of the flexible C-terminal tail resonances in fully-protonated full-length P7 (P7fl).

**Figure 2.**

Structure of the P7 core (P7 $\Delta$ C). (a) Topological diagram for P7 $\Delta$ C generated using the PDBsum server<sup>67; 68; 69</sup>. Secondary structure assignments are from PROMOTIF v3.0<sup>70</sup>. (b) Side-by-side stereo-view of P7 $\Delta$ C. Key structural elements are labeled and the two 3<sub>10</sub> helices are colored cyan.

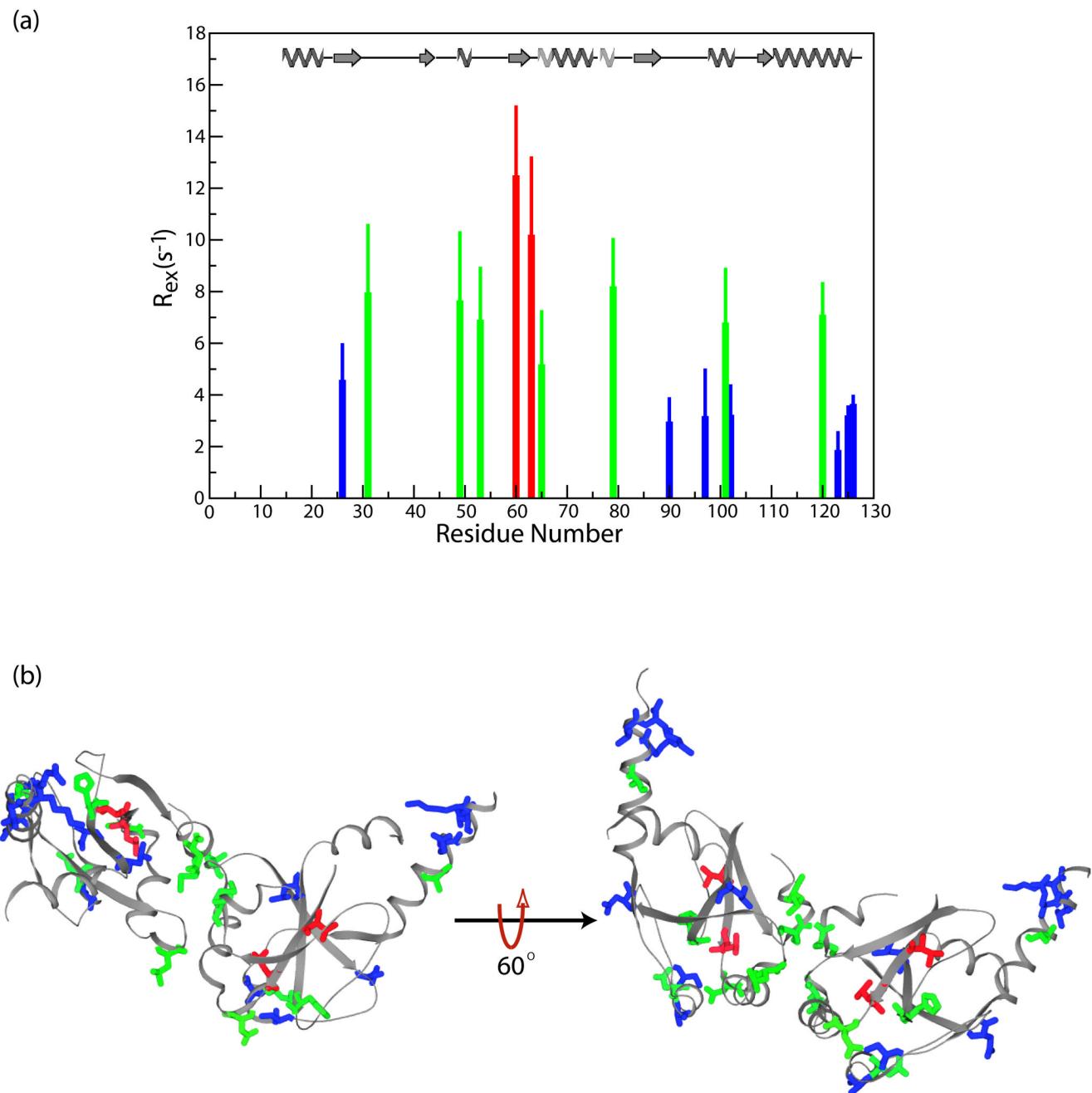
**Figure 3.**

Structure of the P7 $\Delta$ C biological unit. (a) P7 $\Delta$ C forms a symmetric homodimer in solution. The monomers are colored blue and red. (b) The dimerization surface of P7 $\Delta$ C is composed of mostly hydrophobic residues. Key residues forming the dimerization surface are labeled. Hydrophobic residues are colored green, the basic Arg77 and the acidic Asp69 are colored blue and red respectively. (c) The backbone amide H<sup>N</sup> and O atoms of Met49 and Ala80 are involved in intermolecular hydrogen bonds, as are the sidechains of Arg77 and Asp69. The two monomers are represented by ribbons colored red and blue with key residues shown in stick representation colored orange and cyan. Hydrogen bonds are shown by green dotted lines. (d) Gel filtration data (Pharmacia Superdex 200 HR 10/30) showing elution volumes of molecular weight markers (black) and P7fl (D69R). The Asp to Arg mutation disrupts dimer formation in P7.

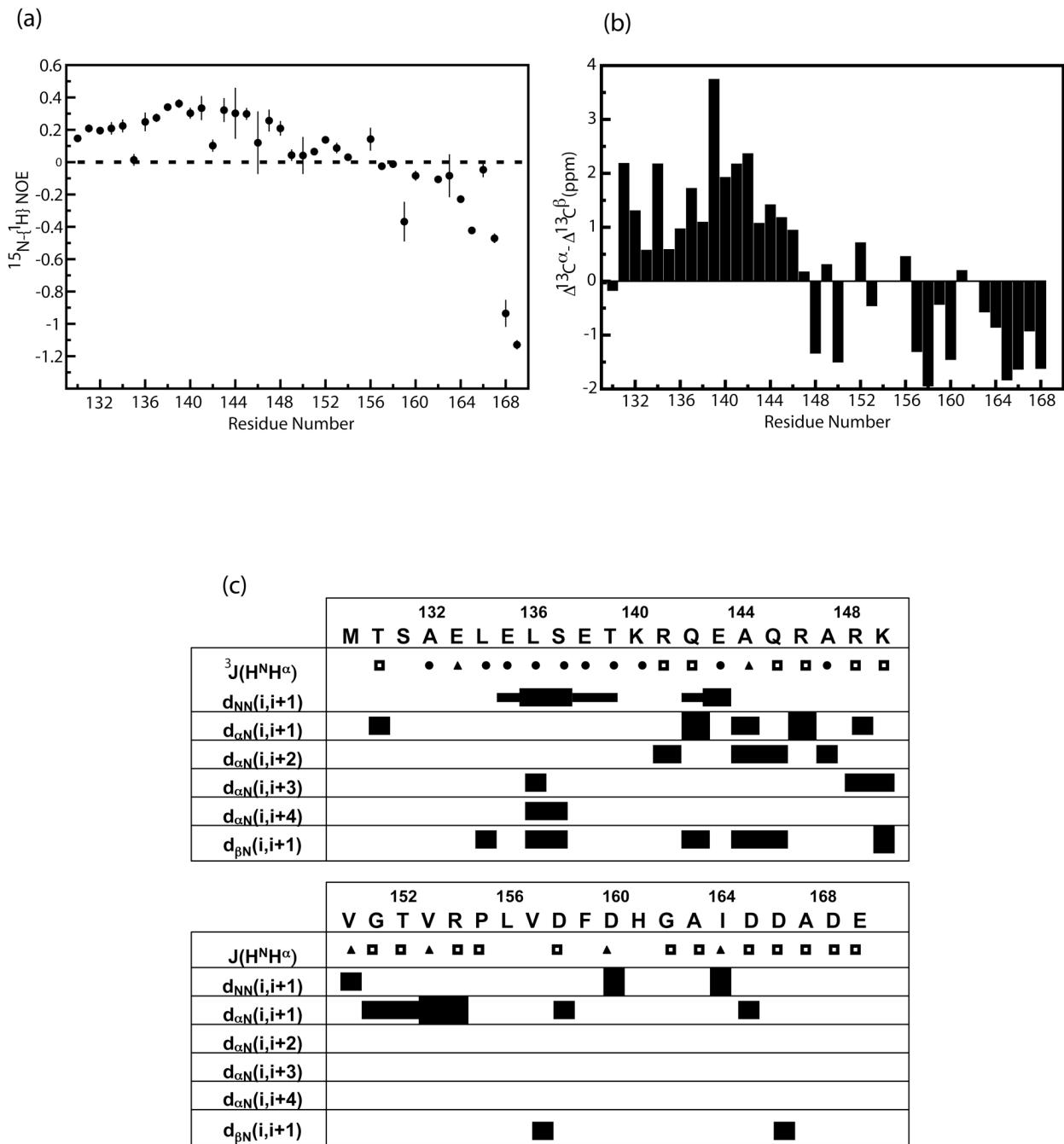


**Figure 4.**  
Fast dynamics of the P7 $\Delta$ C backbone. (a)  $S^2$  values were obtained by a model-free analysis utilizing backbone R<sub>1</sub>, R<sub>2</sub> and steady-state <sup>15</sup>N-<sup>1</sup>H NOE data at 600MHz. Data were fitted to an anisotropic model using the NH unit vectors generated utilizing the crystal coordinates. The microdynamic parameters of the flexible N-terminus of P7 $\Delta$ C, for which no crystal coordinates were available, were obtained by fitting to an isotropic model with the same effective correlation time as the anisotropic model.  $S^2$  values for the anisotropic and isotropic fits are represented by filled and open circles respectively. The regions of definite secondary structures are shown schematically – helices ( $\alpha$  – dark,  $\beta_{10}$  – light) by springs and  $\beta$ -strands by arrows. Phe74 and Val100 that have an unusually low  $S^2$  values are labeled. (b) Reduced

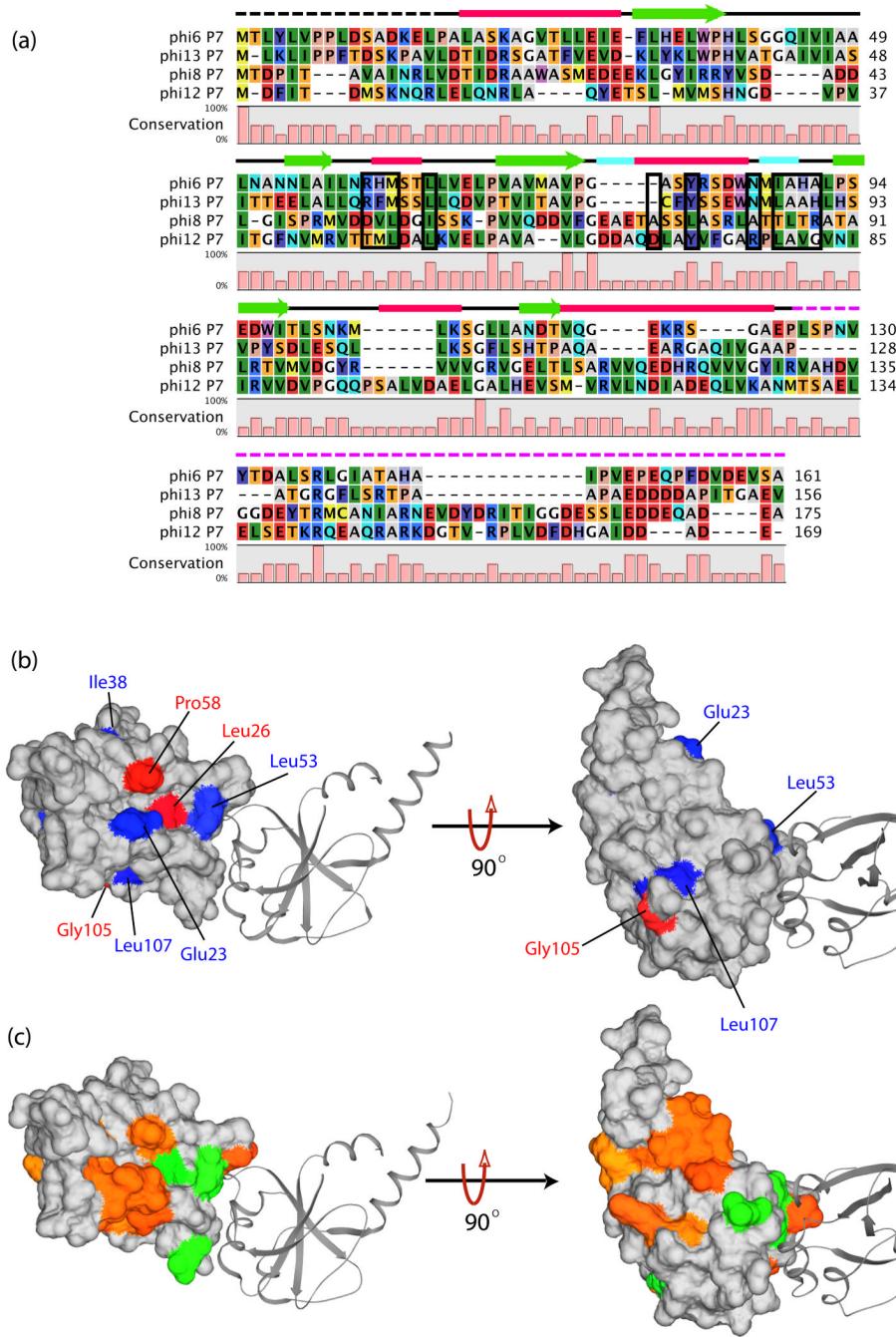
spectral density values near the  $^1\text{H}$  Larmor frequency –  $J(0.87\omega_{\text{H}})$ . Residues with increased  $J(0.87\omega_{\text{H}})$  values due to extensive sub-nanosecond timescale dynamics are labeled.

**Figure 5.**

Slow dynamics of the P7ΔC backbone. (a)  $R_{\text{ex}}$  values obtained from the model-free analysis of the relaxation data at 600 MHz.  $R_{\text{ex}} > 10 \text{ s}^{-1}$  (red),  $5-10 \text{ s}^{-1}$  (green) and  $< 5 \text{ s}^{-1}$  (blue). (b) The residues that display  $R_{\text{ex}}$  values are mapped onto the structure of P7ΔC, the largest  $R_{\text{ex}}$  values are seen in the hydrophobic core and at the dimer interface.

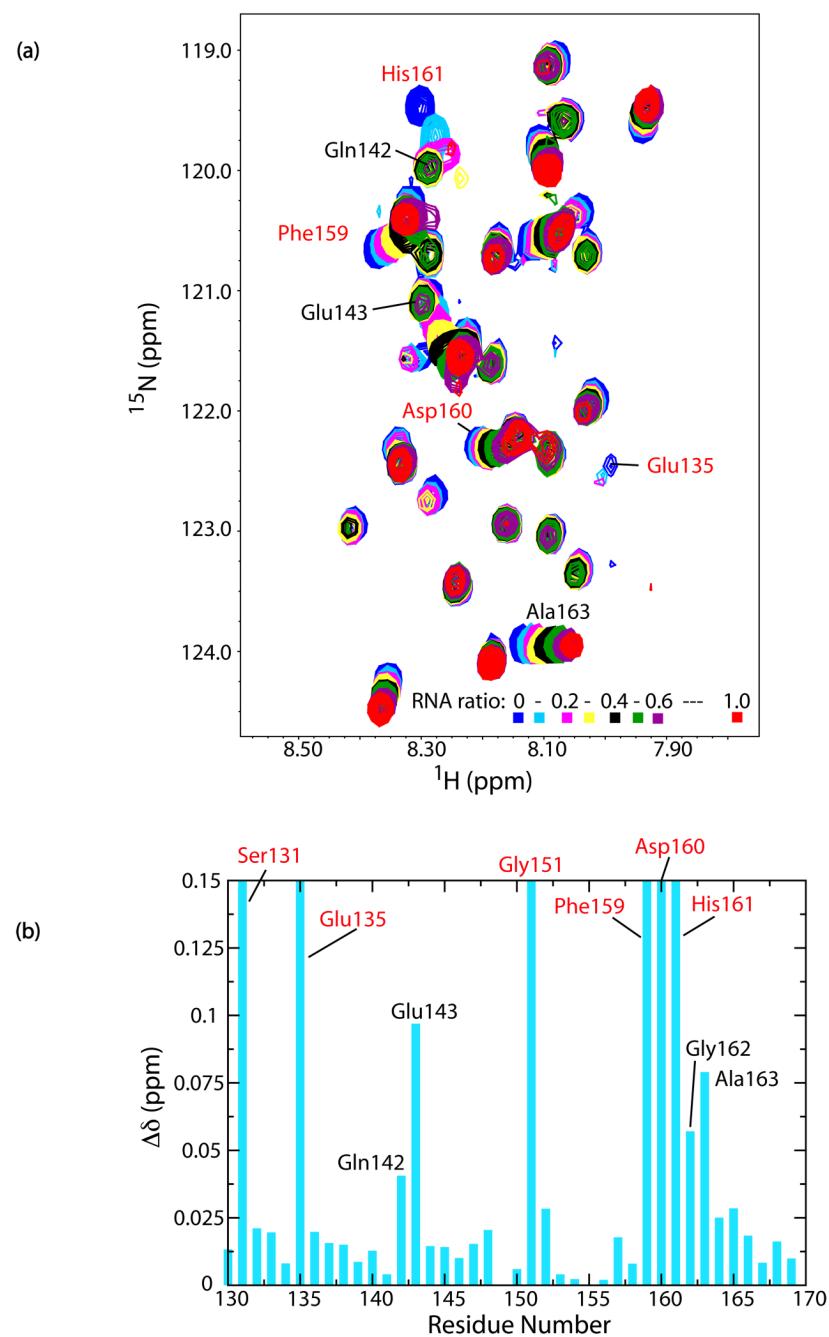
**Figure 6.**

The C-terminal residues in full-length P7 are highly disordered. (a) Steady-state  $^{15}\text{N}-\{\text{H}\}$  NOE data at 600MHz for the 41 C-terminal tail residues in fully protonated full-length P7. (b) The difference in the deviation of the  $\text{C}^\alpha$  and  $\text{C}^\beta$  chemical shifts from the corresponding random coil values for the 41 C-terminal residues in fully-protonated full-length P7 (P7fl). A string of positive values at the N-terminal end of this tail region indicates helical propensity. (c) NOE correlation patterns for the 41 C-terminal residues in full-length P7. NOEs to the backbone amide  $\text{H}^\text{N}$  are classified as strong, medium and weak represented by the thickness of the horizontal bars. Also shown are the  $^3\text{J}(\text{H}^\text{N}\text{H}^\alpha)$  values - < 6.5 Hz (filled circles), 6.5–8.5 Hz (open squares) and > 8.5 Hz (filled triangles).

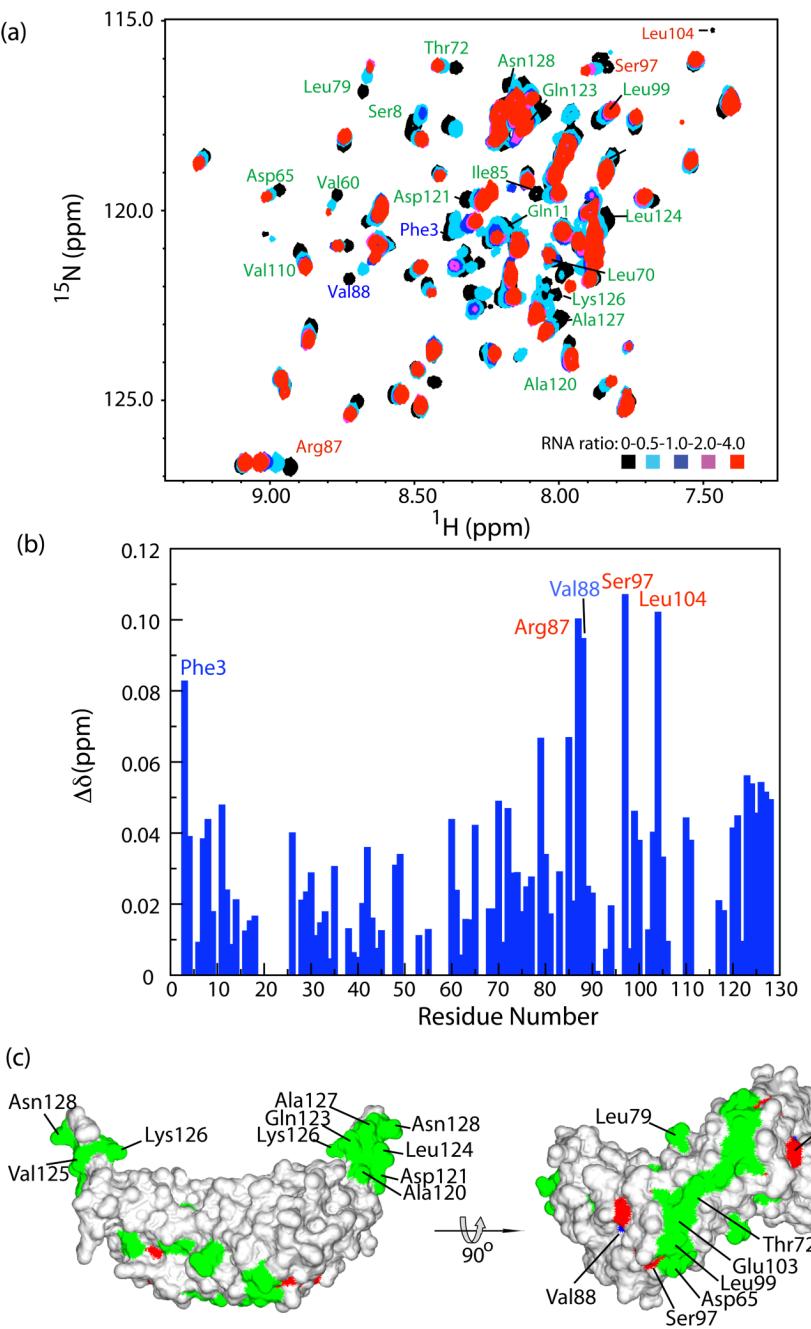
**Figure 7.**

Sequence conservation in cystoviral P7 proteins. (a) Sequence alignment for the P7 proteins for the  $\phi_6$ ,  $\phi_8$ ,  $\phi_{12}$  and  $\phi_{13}$  cystoviruses are shown. The residues that comprise the dimer interface are boxed. The regions of definite secondary structure are indicated –  $\alpha$ -helices (thick red line),  $\beta_10$ -helices (thick cyan line),  $\beta$ -strand (green arrow). The dotted black line indicates the residues that have no electron density in the crystal structure of P7 $\Delta$ C. The flexible C-terminal tail is indicated by the dotted magenta line. (b) Conserved surface residues in P7. Highly conserved residues are colored blue and strictly conserved residues are colored red. (c) Residues that show  $R_{ex}$  values  $> 4.5 \text{ s}^{-1}$  for P7 $\Delta$ C are shaded green. Residues for

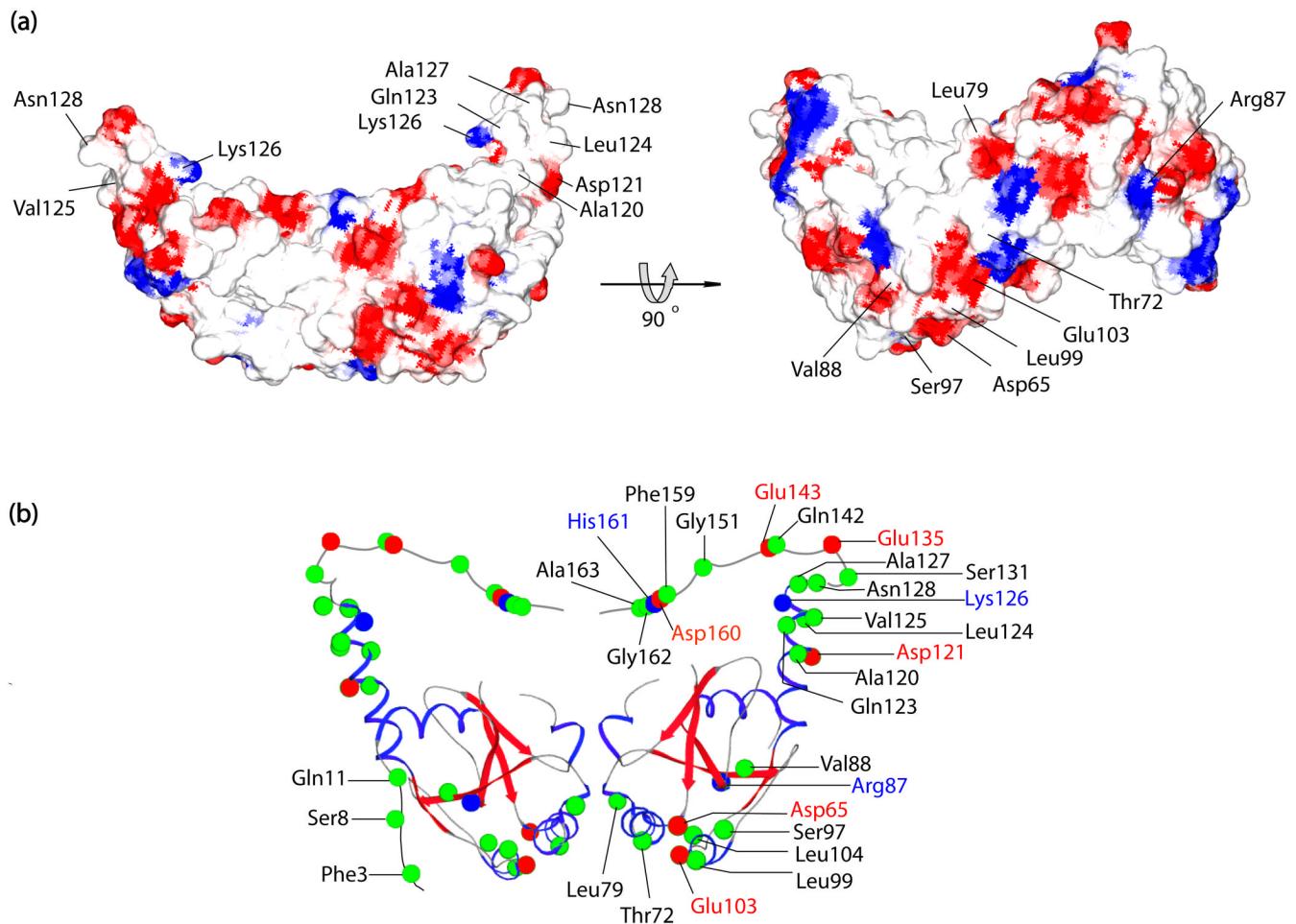
which  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$  assignments could not be obtained in the TROSY spectra of P7 $\Delta$ C are shaded gold.



**Figure 8.**  
Interactions of the C-terminal tail in full-length P7 with RNA. (a) Expanded region of a  $^{15}\text{N},^1\text{H}$  HSQC spectrum (acquired at 600 MHz) of fully-protonated  $^{15}\text{N}$ -labeled P7fl in the presence of varying amounts of Oligo1 (see text). Substantial shift changes are also seen for Gly151 and Gly162 (not shown in the expansion). The spectra are plotted at a suitable contour level to allow clear visualization of the sharp peaks corresponding to the C-terminal tail. (b) Scaled chemical shift changes occurring for tail-resonances of P7fl in the presence of RNA at a 1:1 molar ratio. Residues that are broadened beyond the observation threshold are labeled in red. Other key residues that show substantial chemical shift changes are also labeled.

**Figure 9.**

Interactions of P7 $\Delta$ C with RNA. (a) Expanded region of a  $^{15}\text{N},^1\text{H}$  HSQC spectrum (acquired at 600 MHz) of  $^{15}\text{N},^2\text{H}$ -REDPRO-labeled P7 $\Delta$ C with varying amounts of Oligo2 (see text). Residues that undergo large chemical shift changes (scaled shift changes calculated using Equation 3 for a P7 $\Delta$ C:RNA ratio of 1:2) are labeled. -0.04 ppm <  $\Delta\delta$  < 0.07 ppm are labeled in green, 0.07 ppm  $\leq \Delta\delta < 0.10$  ppm are labeled in blue and  $\Delta\delta > 0.10$  ppm are labeled in red. (b) Scaled chemical shift changes plotted against residue number for a P7 $\Delta$ C:RNA ratio of 1:2. Key residues with large shift changes ( $> 0.04$  ppm) are labeled using the same color scheme as in (a). (c) Two specific surfaces display large chemical shift changes in the presence of RNA – key residues are labeled and the coloring scheme used is as in (a).



**Figure 10.**  
P7/RNA interactions. (a) Surface plot of P7 $\Delta$ C colored by electrostatic potential. Residues on P7 $\Delta$ C that show chemical shift changes  $> 0.04$  ppm in the presence of RNA (P7 $\Delta$ C:Oligo2 ratio 1:2) are labeled. (b) Residues that undergo large chemical shift changes in P7 in the presence of RNA are shown. Red, blue and green spheres represent acidic, basic and hydrophobic residues respectively. The dynamic N- and C-terminal tail conformations are represented schematically.

**Table 1**  
Data Collection, Refinement Statistics and Structure Characteristics

<b>Crystal Parameters</b>		
Space group	P3 <sub>2</sub> 21	
Unit-cell at 100 K	75.5, 75.5, 46.5 Å	90.0, 90.0, 120.0 °
<b>Data quality <sup>a</sup></b>		
Resolution (last shell)	20-1.83 Å (1.90-1.83 Å)	
No. of measured reflections	280625 (18974)	
No. of unique reflections	25437 (2107)	
R <sub>sym</sub>	4.8% (41.0 %)	(I ≥ -3σ <sub>I</sub> for observations)
Mean redundancy	11.0 (9.0)	
Completeness	96.9% (80.3 %)	(All measured reflections)
	80.1% (40.3 %)	(I ≥ 2σ <sub>I</sub> )
Mean I/σI	41.4 (20.4)	(I ≥ σ <sub>I</sub> after merging)
Refinement residuals (F≥2σ <sub>F</sub> ):		
R <sub>free</sub>	22.6 %	
R <sub>work</sub>	21.0 %	
<b>Structure Quality</b>		
RMSD bond lengths	0.019 Å	
RMSD bond angles	2.1°	
Ramachandran plot	92.0 %	Core
	8.0 %	Allowed
	0.0 %	Generously allowed
<b>Average B factors (Å<sup>2</sup>)</b>		
All	35.4	
Main chain	32.6	
Side chain	40.6	
Waters	53.3	
<b>Model contents:</b>		
Protein molecules in the AU	1	
Protein residues	114 (A14-A55 A58-A129)	
Water molecules	190	
PDB accession code <sup>b</sup>	2Q82	

<sup>a</sup>Standard definitions were used for all parameters <sup>71</sup>. Data reduction and refinement statistics come from SCALEPACK <sup>58</sup> and CNS <sup>63</sup> respectively.

<sup>b</sup>Both co-ordinates and structure factors have been deposited in the PDB

**Table 2**Hydrodynamics of P7ΔC from NMR Relaxation Rates<sup>a</sup>

Model	D <sub>xx</sub>	D <sub>yy</sub>	D <sub>zz</sub>	$\alpha$	$\beta$	$\gamma$	$\tau_c$	D <sub>aniso</sub>	$\chi^2$	F	P
Asymmetric	0.64 ± 0.07	0.76 ± 0.11	1.10 ± 0.12	22.6 ± 0.8	81.9 ± 1.3	96.9 ± 3.8	20.07	1.57	0.11	199.7	–
Axially-symmetric	0.68 ± 0.05	0.68 ± 0.05	1.14 ± 0.09	21.4 ± 1.2	82.7 ± 1.4	–	19.92	1.66	–	218.6	2.13 0.15
Isotropic	0.83 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	–	–	–	20.04	–	–	497.9	20.01 5.57 × 10 <sup>-7</sup>

<sup>a</sup>D<sub>ii</sub> (i = x, y, z) are expressed in units of 10<sup>7</sup> s<sup>-1</sup>, the angles  $\alpha$ ,  $\beta$  and  $\gamma$  in degrees.

$\tau_c = 1/(6\text{Trace}\{\mathbf{D}\})$  in ns; D<sub>aniso</sub> = 2D<sub>zz</sub>/(D<sub>xx</sub>+D<sub>yy</sub>); D<sub>asym</sub> = (D<sub>yy</sub>-D<sub>xx</sub>)/D<sub>zz</sub>.

P, the probability that the improvement in  $\chi^2$  values obtained with the increase in the complexity of the model can be obtained by random noise alone, is expressed as a fraction. Values of P < 0.01 (1 %) are considered statistically significant.