STRUCTURAL AND DYNAMICAL FEATURES OF PROTEIN P7 FROM BACTERIOPHAGE ϕ12: INSIGHTS INTO A FUNCTIONAL ROLE IN THE CYSTOVIRAL POLYMERASE COMPLEX

by

ERTAN ERYILMAZ

A dissertation submitted to the Graduate Faculty in Physics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2010

This manuscript has been read and accepted for the
Graduate Faculty in Physics in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

	Ranajeet Ghose
Date	Chair of Examining Committee
	Steven Greenbaum
Date	Executive Officer

Ranajeet Ghose, PhD

Ruth Stark, PhD

Paul Gottlieb, PhD

Babis Kalodimos, PhD

Fabien Ferrage, PhD

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

STRUCTURAL AND DYNAMICAL FEATURES OF PROTEIN P7 FROM BACTERIOPHAGE ϕ12: INSIGHTS INTO A FUNCTIONAL ROLE IN THE CYSTOVIRAL POLYMERASE COMPLEX

by

Ertan Eryilmaz

Advisor: Professor Ranajeet Ghose

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. The cystoviral PX, that is amenable to *in vitro* and *in vivo* manipulation, comprises a unique model system for similar polymerase machinery. Containing three segmented double stranded RNA genome, the cystoviral PX is a simplified model for the polymerase machinery in more complex RNA viruses like reoviruses sharing structural and functional similarities at the level of the constituent proteins. Though the structures of the RNA dependent RNA polymerase (RdRp) and ATPase components of the cystoviral PX are known and their functional behaviors 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 αβ-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 through NMR-based 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 additional role of P7 in maintaining transcriptional fidelity, our data suggest a central biological role for P7/RNA interactions.

Acknowledgements

I would like to thank my wife Deniz Bilge Temel for her love and support for the entire time since we have met.

I thank my family for their unconditional love and support. They have raised me, guided me and encouraged me in all steps of my life.

I thank Dr. Ranajeet Ghose for his mentorship, support and understanding. He has always been there whenever I needed help.

This work would not be possible without the support of the members of the Ghose Laboratory, Dr. Andrea Piserchio, Aswin Natarajan, Deniz B. Temel, Ren Zhen, Dr. Hsin Wang and Dr. Padmanava Pradhan.

I acknowledge my committee members, Dr. Ruth Stark, Dr. Paul Gottlieb, Dr. Charalampos (Babis) Kalodimos and Dr. Fabien Ferrage, for their time and input through my advancement to my degree.

I thank my friends for making me a part of their life, for their love and support.

Table of Contents

List of Tables

List of Figures

1. Introduction

The cystoviruses (ϕ6-14) (Mindich, Qiao et al. 1999) form a unique group of enveloped bacteriophages that contain a three-segmented double-stranded RNA (dsRNA) genome. The RNA segments are named in accordance with their lengths (S, M and L segments for small, medium and large). The cystoviruses infect strains of the plant pathogen, *Pseudomonas phaseolicola* and share an overall organization and replicative mechanism as the pathogenic reoviruses (Mertens 2004). While the nine members of the cystovirus have very similar genetic organization and they express similar proteins, there is minimal similarity in the primary sequences for each dsRNA segment. This genetic variation allows variety in host selectivity (Mindich, Qiao et al. 1999) and infectivity (Hoogstraten 2000; Qiao 2000; Gottlieb, Wei et al. 2002). The cystovirus family members can be classified into two major groups depending on the similarity of each member to representative, first-discovered and most widely studied member, ϕ6 (Mindich, Qiao et al. 1999; Hoogstraten 2000; Qiao 2000; Gottlieb, Potgieter et al. 2002; Gottlieb, Wei et al. 2002); i.e. ϕ7, ϕ9, ϕ10, and ϕ11 are closely related to ϕ6, while ϕ8, ϕ12, and ϕ13 are distantly related to ϕ6.

There are four distinct proteins that exist in multiple stoichiometric copies that comprise the innermost layer of the cystoviral virion (Kainov 2004), the polymerase complex (PX): Protein P1 (major capsid protein) (Butcher, Dokland et al. 1997), P2 (RNA-directed RNA polymerase) (Makeyev 2000; Butcher, Grimes et al. 2001; Yang, Makeyev et al. 2001), P4 (packaging ATPase) (Mancini, Kainov et al. 2004) and P7 (putative assembly co-factor)(Juuti and Bamford 1997). These proteins are encoded by the L-segment of the dsRNA genome and comprise the molecular machinery responsible

for genome replication and transcription (Mindich, Nemhauser et al. 1988). The structures of the PX in the two cystoviruses most distant in genomic sequence, namely ϕ6 (Huiskonen, de Haas et al. 2006) and ϕ8 (Jaalinoja, Huiskonen et al. 2007), obtained from cryo-EM reconstructions, revealed an overall conservation of structure and organization.

Forming the skeleton of the viral capsid, P1 assembles on a T=13 lattice with overall icosahedral symmetry and determines the overall structural properties of the capsid. As indicated by the 7.5 and 8.5 Å cryo-EM reconstructions of the ϕ6 (Huiskonen, de Haas et al. 2006) and ϕ8 (Jaalinoja, Huiskonen et al. 2007) PX cores respectively, the P1 protein is highly α-helical and putatively exists as an asymmetric dimer. However, suggested by other studies, ϕ6 P1 is monomeric (Poranen, Paatero et al. 2001) and ϕ8 P1 is tetrameric in solution (Kainov, Butcher et al. 2003). Additionally, the electron densities from the EM reconstructions of ϕ6 and ϕ8 cores hint towards significant differences in the tertiary structures across the species at least at the level of domain organization (Jaalinoja, Huiskonen et al. 2007), though this is far from clear at the level of resolution available.

Recombinant P2 proteins from three different cystoviruses(ϕ6, ϕ8, and ϕ13) displayed RNA-dependent RNA polymerase activity catalyzing de novo initiation, elongation, replication and producing double-stranded RNA from single-stranded RNA templates (Makeyev and Bamford 2000; Yang, Makeyev et al. 2001). Multiple crystal structures are available for ϕ6 in apo-form, in complex with divalent metal cations, with substrate GTPs, ssDNA and ssRNA (Butcher, Grimes et al. 2001; Salgado, Makeyev et al. 2004; Poranen 2008). These structures reveal a conserved polymerase fold shaped like a cupped right-hand, containing “thumb”, “palm” and “fingers” domains and the

conserved sequence motifs A-F common to DNA and RNA polymerases (Bruenn 1993; Bruenn 2003). Structures of ϕ6 P2 in complex with ssRNA construct representing the 3'-end of the minus strands of the genomic RNA show that oligonucleotides bind within a tunnel lined with basic amino acids, leading to a pre-formed active site. Suggested model through observations is that the incoming RNA is unwound and directed through a template site to an active site (Butcher, Grimes et al. 2001). The mechanism of this process is unclear. However, recent NMR studies indicate that P2 forms a highly dynamic machine with several motional modes that are altered along the catalytic pathway (Ren, Wang et al. 2010).

P4, a hexameric ATPase, provides the energy for the virus to package its mRNA (genomic plus-strands). There are 12 copies of hexameric P4 each of which is assembled on each apex of the twelve 5-fold axis of the icosahedral PX creating a symmetry mismatch (Juuti and Thomas 1998; de Haas, Paatero et al. 1999). The hexameric organization of P4 creates a central pore through which the mRNA passes during the packaging process and is also translocated into the host cytosol during the process of semi-conservative replication (Kainov, Tuma et al. 2006). A crystal structure of the P4 protein from ϕ12 revealed similarities with other hexameric NTPase motors such as the AAA+ family (Kainov, Mancini et al. 2008). The structure revealed an NTP-binding loop on each monomer that is involved in RNA translocation coupled to NTP hydrolysis (Mancini, Kainov et al. 2004). Concerted changes in exchange kinetics, which were visualized by hydrogen-deuterium exchange (HDX) experiments, suggest that the translocation of RNA is cooperative through the action of a "swiveling helix" that links the RNA binding channel to NTP binding sites. HDX experiments also revealed a

transition state that likely proceeds with hexameric ring opening and is associated with RNA loading (Lisal, Lam et al. 2005).

The least characterized of all of the PX proteins is P7. Though it is suggested that P7 exists in dimeric form in solution (Juuti and Bamford 1997; Poranen, Paatero et al. 2001; Kainov 2004), very little was known about its structure, its exact location and stoichiometry within the PX with no obvious electron density corresponding to P7 found in either the ϕ6 or the ϕ8 EM reconstructions. In spite of the lack of structural data, significant biochemical data was available for P7. It was shown through *in vitro* assembly studies that the incorporation of the P7 protein into an incomplete ϕ6 PX stabilized the PX and significantly enhanced the assembly rate of the PX particles (Poranen, Paatero et al. 2001). This indicated that P7 was likely an assembly co-factor that stabilized the PX through interactions with P1 and possibly other PX proteins (Poranen, Paatero et al. 2001). However in ϕ8, due to the highly stable P1 tetramers formed as assembly intermediates, incorporation of P7 into incomplete particles did not alter the assembly kinetics significantly (Kainov, Butcher et al. 2003). Additionally, it was shown in ϕ6 that incomplete PX particles 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 PX particles, which may be a clue for another function that can be attributed to P7. Further, the specificity of the P147 particles towards the S and L segments was identical to that of the PX while minor differences that were noted with the M segment. Thus, in addition to stabilizing the assembly intermediates, P7 was also likely to be a packaging co-factor (Juuti and Bamford 1995; Juuti and Bamford 1997). In addition, it

was demonstrated that P7-null PX particles were producing RNA transcripts of the incorrect size during semi-conservative positive strand synthesis (Juuti and Bamford 1997) pointing towards a possible role for P7 in regulating the transcriptional fidelity of the PX through a process that likely requires its interaction with other PX proteins and/or genomic RNA segments. Further, it was noted that P7 possessed a flexible C-terminal tail (Juuti and Bamford 1997) that was essential for the formation of an infectious virion (Poranen and Tuma 2004; Poranen, Butcher et al. 2008).

Given this background, we embarked on comprehensive studies of P7 (from the cystovirus ϕ12) using a host of biophysical and biochemical tools including solution NMR spectroscopy and X-ray crystallography towards a complete structural and functional understanding of this protein in the context of the cystoviral PX. Towards this goal, we have obtained a crystal structure of the P7 protein (from the ϕ12 cystovirus) at a resolution of 1.8 Å. P7 forms a novel αβ-fold and exists as a symmetric homodimer in solution as revealed by the measurement of solution NMR relaxation rates combined with chemical cross-linking and light scattering studies. Through solution NMR measurements we have shown that the C-terminal tail of P7 though highly disordered possesses residual helical properties. We have also found evidence that P7 is capable of interacting with ssRNA 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 PX.

2. The Structural Characteristics and Possible Functions of the Protein P7 From Bacteriophage ϕ12

We have targeted the protein P7 from bacteriophage ϕ12 as representative member for our studies. As stated before, ϕ12 falls within the group which is classified as being distantly related to ϕ6 (Mindich, Qiao et al. 1999; Hoogstraten 2000; Qiao 2000; Gottlieb, Potgieter et al. 2002; Gottlieb, Wei et al. 2002).

2.1. Experimental Methods

2.1.1. Overexpression and Purification of ϕ12 P7

The ϕ12 P7 expression plasmid (pPG29) was prepared by PCR amplification of the P7 gene from a pP12L1 template with recombinant Pfu DNA polymerase (Stratagene) using the following oligonucleotides 5'- GGTAACCATGGAC-TTCATTACTGAC-3' and 5'-AGGATCCTTATTCGTCGGCATCATCGAT -3' as upstream and downstream primers, respectively. The PCR fragment was digested with NcoI and BamHI and inserted into pET-21d vector (Novagen) between the Ncol and BamHI sites.

Full-length (1-169) P7 (P7fl) was expressed in *E. coli* BL21 (DE3-RIL/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 necessary labeling supplements i.e. $^{15}NH_4Cl$ (or $^{15}NH_4Cl$ and $^{13}C_6$-glucose or $^{15}NH_4Cl$,) and grown at 37° C to an A_{600} of 0.6 followed by induction with 0.5 mM of isopropyl-β-D-1-thiogalactopyranoside (IPTG) and growth at 20° C for 16 hours. The cells were centrifuged and resuspended in 20 mL of TNM buffer (20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 7.5 mM $MgCl_2$) 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 16000 rpm for 60 min. at 4° C. Supernatant fractions were loaded onto 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 $MgCl_2$. 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 $MgCl_2$ at pH 6.5.

For the production of the ^{15}N, ^{13}C, REDPRO-^{2}H-labeled (Shekhtman, Ghose et al. 2002) protein, cells from the overnight culture were diluted 50 times and grown at 37° C in an M9 medium containing $^{15}NH_4Cl$ and $^{13}C_6$-glucose prepared in H_2O, until an A_{600} of 0.6. Cells were collected by centrifugation, washed with phosphate-buffered saline (PBS), re-suspended in an M9 medium (containing $^{15}NH_4Cl$ and $^{13}C_6$-glucose) prepared in > 99 % D_2O and induced with 0.5 mM of IPTG followed by overnight aeration at 20° C. Further purification was carried out as described above.

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).

Se-methionine labeled truncated P7 was expressed in the methionine-auxotroph *E. coli* strain B834 (DE3) or BL21-DE3-RIL-X 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 of each for a 2 L culture). The insoluble amino acids, tryptophan and tyrosine (0.5 g each), were dissolved in 10 mL of 2 M HCl, diluted to 100 mL and 20 mL 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 an M9 medium. 20 mL of this M9 culture was used to inoculate 2L of defined-M9 medium. The cells were grown at 37° C to an A_{600} of 0.6, 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 no gel-filtration column was used; instead fractions collected from a Q-Sepharose HP column were reloaded into 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 crystallization buffer containing 10 mM Tris, 100 mM NaCl at pH 7.5.

2.1.2. Site-directed Mutagenesis

The desired mutations were introduced using four primers and two PCR reactions (Figure 1). Primers 1 and 4 were essentially the same forward and reverse primers respectively that have been used to extract the P7 gene from pP12L1 template. On the other hand the primers 2 and 3 were the ones designed to introduce the single site mutation. For the D69R mutation:

Primer 2 : 5'-GACGACGCTCAACGCCTAGCCTATGT-3'

Primer 3 : 5'-ACATAGGCTAGGCGTTGAGCGTCGTC-3'.

For the R77D mutation:

Primer 2 : 5'-TATGTCTTCGGCGCCGATCCTCTGGCCGTCGGA-3'

Primer 3: 5'-TCCGACGGCAGGATCGGCGCCGAAGACATA-3'.

The underlined sequences were the introduced mutations for Arg and Asp respectively. In the 1st PCR two separate reactions were made to produce fragments A and B (Figure 1).

The first cycle of the 2nd PCR is crucial where T4 DNA ligase was not introduced and the fragments were extended after which the T4 DNA ligase was added to the reaction to generate the single site mutated full-length P7 gene. With this technique both D69R and R77D mutations could be achieved. Figure 1 illustrates the process for the D69R mutation.

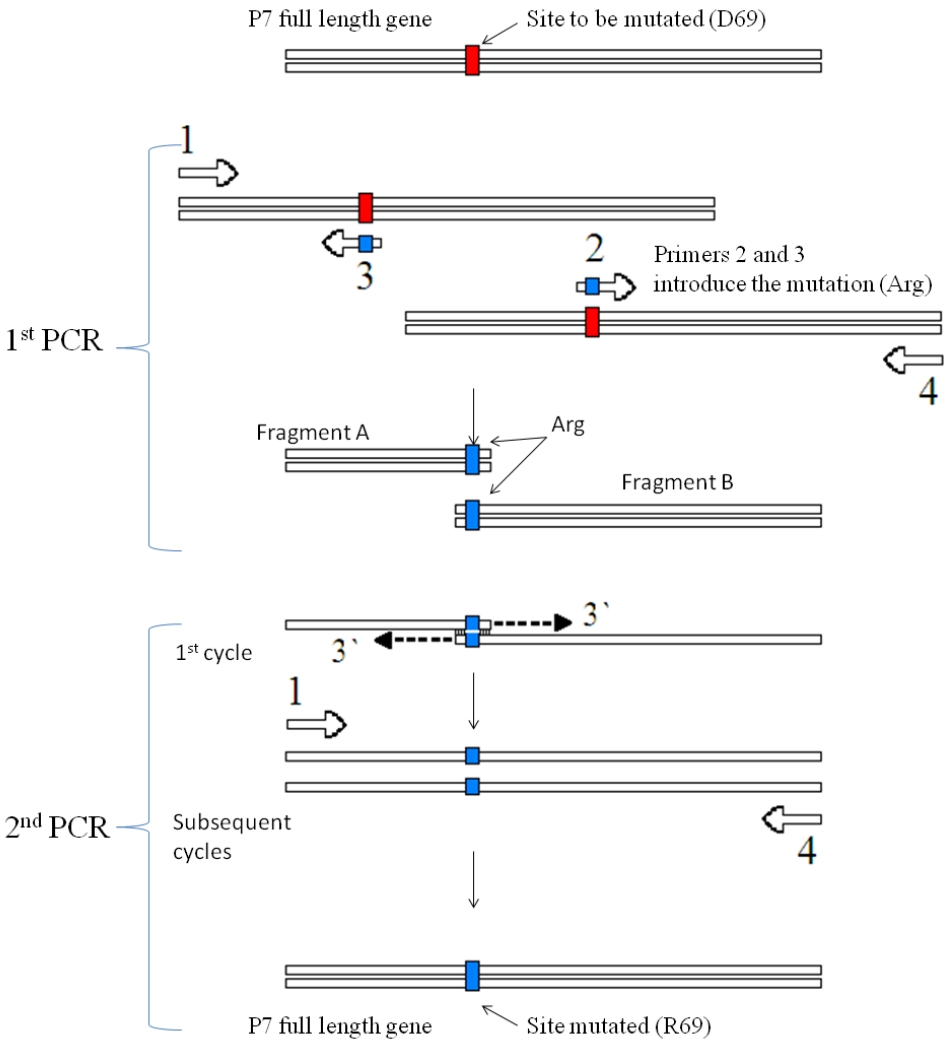

Figure 1: A schematic description of the method used to achieve single site mutations. Primers 2 and 3 introduce the desired single site mutation. Two separate PCRs were carried out, the first PCR, done in two separate vials, produced two separate mutated fragments. The second PCR was done to get a single site mutated full length P7 gene from fragments A and B produced in the 1st PCR.

2.1.3. Biochemical/Biophysical Assays

2.1.3.1. Oligomeric State of P7

It has been suggested that P7 exists as an elongated dimer in ϕ6 (Juuti and Bamford 1997), ϕ8 (Kainov, Butcher et al. 2003) and ϕ12 (Kainov, Simonov et al. 2004) cystoviruses. Chemical cross-linking studies (Figure 2) revealed the dimer as the dominant species (> 95 %), both in P7fl and P7ΔC indicating that C-terminal tail truncation does not interfere with the ability of P7 to dimerize, consistent with previous studies (Kainov, Simonov et al. 2004). Static light scattering studies on P7ΔC at protein concentrations ranging from 200 μM to 1.7 mM at pH values of 6.5 and 7.5 (corresponding to the NMR and crystallization conditions respectively) confirmed that P7ΔC was dimeric over a wide range of concentrations (Figure 3).

Chemical Cross-linking

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 samples were mixed with 2 μL of glutaraldehyde (up to a final glutaraldehyde concentration of 1.3 %). After 15 minutes of incubation, 20 μL of 2X sample buffer was added and the samples were boiled immediately and analyzed on an SDS-PAGE gel. The assay revealed that both P7fl (Figure 2) and P7ΔC (data not shown) exist in dimeric form in solution.

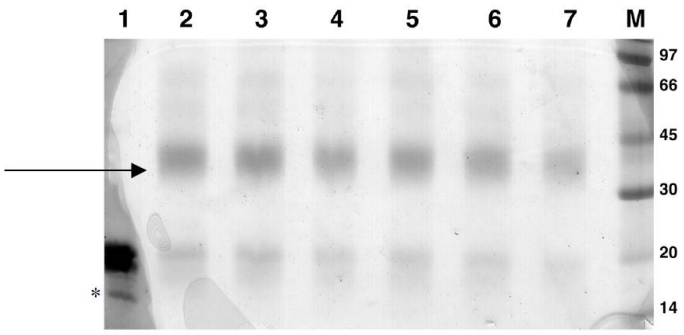

Figure 2: Chemical cross-linking experiments on P7fl shows that the dimer (indicated by the arrow) is the dominant species. The "*" in Lane 1 indicates a minor impurity.

Lane 1 – Control - 20 µL full-length P7 (MW: 18,517 Da),
Lane 2 – 20 µl full-length P7 + 4 µl, 2 % glutaraldehyde,
Lane 3 – 20 µl full-length P7 + 6 µl, 2 % glutaraldehyde,
Lane 4 – 20 µl full-length P7 + 10 µl, 2 % glutaraldehyde,
Lane 5 – 20 µl full-length P7 + 4 µl, 4 % glutaraldehyde,
Lane 6 – 20 µl full-length P7 + 6 µl, 4 % glutaraldehyde,
Lane 7 – 20 µl full-length P7 + 10 µl, 4 % glutaraldehyde,
Lane 8 – Molecular weight marker (molecular weights shown on the right)

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 x 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-3415α degasser (Showa Denko, Tokyo, Japan). 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 4º C. The SLS and RI detectors were located outside of the cold box. SLS measurements were conducted at ambient temperature (~25º C), but the sample chamber in the RI detector was maintained at 35º C to minimize instability due to thermal drift. 100 µl aliquots of sample containing the protein at concentrations of 200µM, 500µM and 2.5 mM (200 µM, 500 µM and 1.7 mM for the protein in the crystallography buffer) 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 Bis-Tris, 50 mM NaCl, 7.5 mM $MgCl_2$, pH 6.5 (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 based on a Debye plot of $R(\theta)/c$ *vs.* $\sin^2(\theta/2)$ (Figure 3).

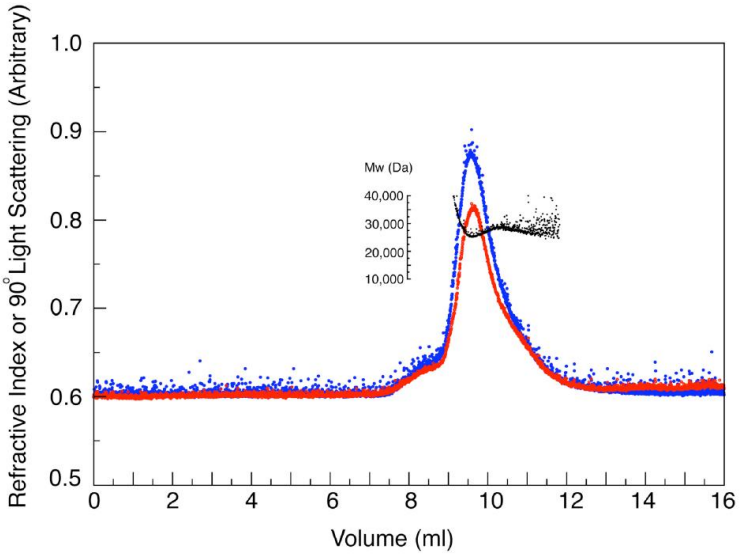

Figure 3: Analytical gel filtration and light scattering data on P7ΔC. P7ΔC in the crystallization or NMR buffers (see Materials and Methods) was injected into an analytical gel filtration column and the eluate was monitored using refractometric interferrometry (RI) and 90º static light scattering detectors (SLS). Representative RI (red, signal proportional to protein concentration) and SLS (blue, signal proportional to the protein concentration times its hydrodynamic volume) are shown. The inset represents the protein molecular weight as a function of elution volume determined from a fit of the individual data slices of the chromatogram to the Debye equation.

2.1.3.2. Differential Scanning Calorimetry

Differential scanning calorimetry (DSC) experiments were performed to evaluate if C-terminal truncation affects the protein stability. The experiments were performed using a VP-DSC microcalorimeter (MicroCal) with a scan rate of 1.5 º/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 P7fl and P7ΔC were 1 mg/ml (Figure 4). The melting temperatures of P7fl and P7ΔC were found to be 47.5º C and 48.4º C respectively.

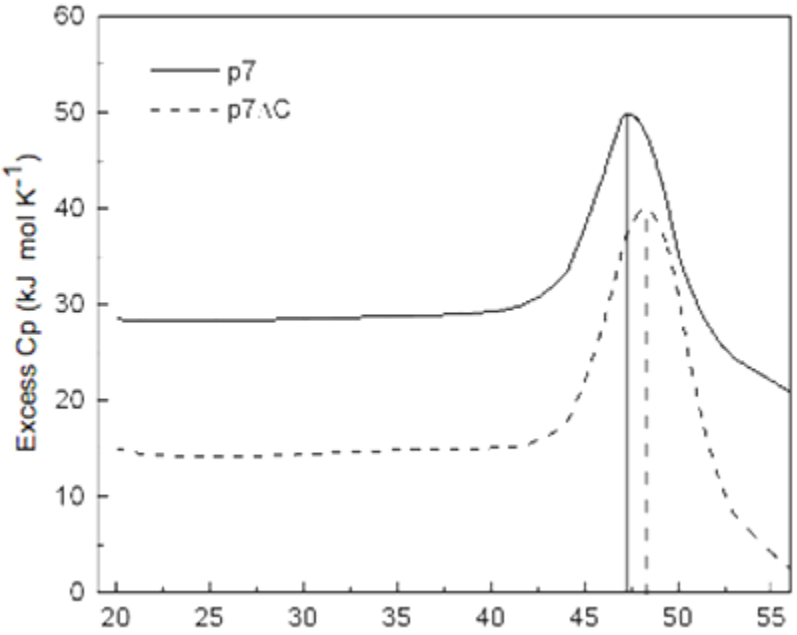

Figure 4: Comparison of DSC traces of the P7fl and P7ΔC indicates no significant difference in thermal stability. The T_m for full-length P7 and P7ΔC were found to be 47.5° C and 48.4° C respectively.

2.1.3.3. CD Spectroscopy

Far-UV CD spectroscopy was used to assess the secondary structure content of the protein constructs (Figure 5). 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 μM, 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 (Luo and Baldwin 1997).

$$f_H = \frac{\theta_{222} - 2220 + 53T}{\theta_H - 2220 + 53T}$$
$$\theta_H = (-44000 + 250T)\left(1 - \frac{3}{N}\right) \quad \textbf{(1)}$$

where f_H is the fractional helical content at temperature T (in °C), θ_{222} is the molar ellipticity measured at 222 nm, θ_H is molar ellipticity of a pure α-helix, and N is the number of residues in the protein. The CD spectra revealed a loss of approximately 6 helical residues upon C-terminal tail truncation.

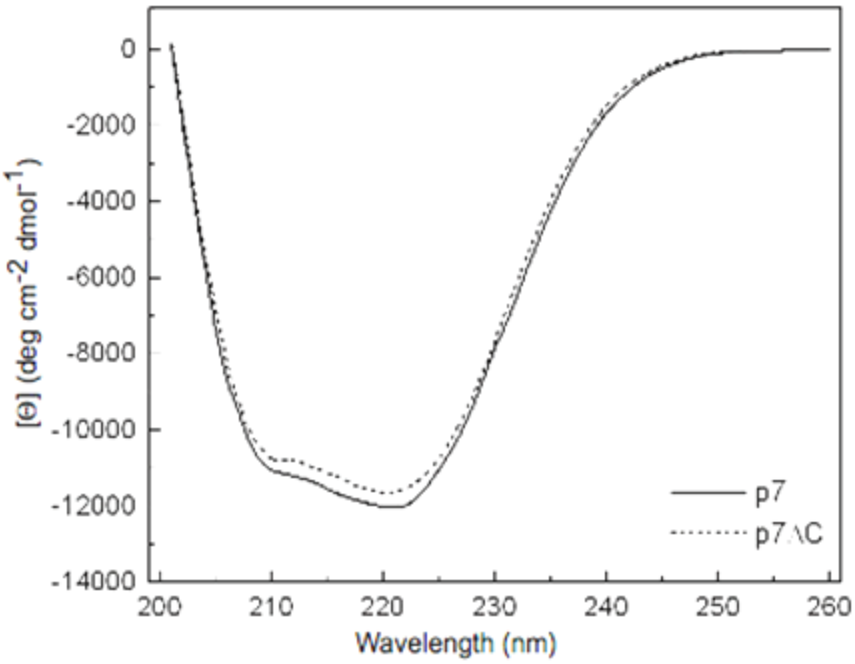

Figure 5: Comparison of far UV CD spectra of P7fl and P7ΔC reveals a slight change in the helical content (as determined by the ellipticity value at 222 nm) corresponding to a loss of approximately 6 helical residues upon tail truncation.

2.1.4. NMR Spectroscopy

NMR spectra were recorded using ^{15}N- or ^{15}N,^{13}C-labeled samples for P7fl and ^{15}N,REDPRO-^{2}H- or ^{15}N,^{13}C,REDPRO-^{2}H-labeled samples for P7ΔC on either a 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 operating 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 (Delaglio, Grzesiek et al. 1995) and analyzed using either NMRPipe (Delaglio, Grzesiek et al. 1995) or NMRViewJ (Johnson 2004) software suites.

2.1.4.1. Backbone Resonance Assignments

For sequential backbone assignment of the flexible (129-169) C-terminal tail of P7fl, standard gradient-enhanced backbone-directed triple-resonance experiments, HNCACB/CBCA(CO)NH, HN(CA)COHNCO and HNCA/HN(CO)CA (Sattler, Schleucher et al. 1999) were used on fully-protonated ^{15}N,^{13}C-labeled P7fl. Sidechain ^{13}C and ^{1}H assignments were obtained from (H)C(CO)NH and H(CCO)NH experiments respectively. $^{3}J(H^{N}H^{\alpha})$ coupling constants were obtained from an HNHA experiment (Vuister and Bax 1993) using a mixing time of 26.1 ms and an empirical correction factor of 0.9 to account for the differential relaxation of the anti-phase and multiple quantum coherences. In addition a ^{15}N-edited NOESY with a mixing time of 150 ms was used to locate regions of residual structure in the tail (129-169) of P7fl.

For P7ΔC, sequential backbone assignments were obtained for ^{15}N,^{13}C,REDPRO-^{2}H-labeled using TROSY-based HN(CA)CO/HNCO, HNCA/HN(CO)CA and HNCACB/HN(CO)CACB experiments (Salzmann, Pervushin et al. 1998).

2.1.4.2. Backbone Relaxation Measurements

All relaxation experiments were performed at 600 MHz at 25° C. ^{15}N-{^{1}H} NOE measurements were made for fully-protonated ^{15}N-labeled P7fl and ^{15}N,^{13}REDPRO-^{2}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 lock stabilization. Additionally, a lower resolution ^{15}N-{^{1}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_1 and R_2 rates were

measured in ^{15}N,^{13}REDPRO-^{2}H-labeled P7ΔC using recycle delays of 1.5 s and the following relaxation delays – for R_1 : 7.0 (x2), 102, 202, 302, 502, 702, 1002, 1202 and 1502 ms; for R_2 – 10 (x2), 30, 50, 70, 90, 110 (x2) and 130 ms. All data were processed using the NMRpipe suite (Delaglio, Grzesiek et al. 1995) and relaxation rates were obtained using in-house software that utilized the ODRPACK (Boggs, Donaldson et al. 1989) library. Errors for the ^{15}N-{^{1}H} NOE were obtained by simple propagation of the uncertainties in peak intensities. Errors for the R_1 and R_2 (68.3 % confidence bounds) were obtained from the respective inverse covariance matrices of the fits.

2.1.4.3. Determination of the Hydrodynamic Properties

The hydrodynamic properties of P7ΔC were determined using the program DIFFTENS v2.0 utilizing established methods (Ghose, Fushman et al. 2001) using the measured relaxation rates and the crystal structure of P7ΔC (after addition of protons using the program MOLMOL (Koradi, Billeter et al. 1996) 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 a 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.

2.1.4.4. Determination of Backbone Microdynamic Parameters

An analysis of the micro-dynamic motional parameters using the Lipari-Szabo framework (Lipari and Szabo 1982; Clore, Szabo et al. 1990) was performed utilizing the DYNAMICS package (Fushman, Cahill et al. 1997) using the measured R_1, R_2 and ^{15}N-$\{^1H\}$ 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 (Fushman, Cahill et al. 1997).

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

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

where the symbols have their usual meaning.

2.1.4.5. P7/ RNA Interactions

5-nt single-stranded RNA constructs corresponding to the 5' ends of the plus strands ϕ12 dsRNA genome - 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 ^1H, ^{15}N HSQC spectra were recorded (at 600 MHz) using ^{15}N-labeled P7fl (conc. 300 μM) in the presence of Oligo1 or Oligo2 (final protein to RNA molar 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 Section 2.1.1) in the presence of commercial RNAse inhibitor (RNAasein from Promega, 10 μl of 40 units/μl in a 300 μl

NMR sample). The scaled chemical shift changes (Δδ) 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 ^1H and ^{15}N dimensions respectively in the presence of RNA (Shekhtman 2001). For Oligo3 HSQC spectra were recorded for final protein to RNA molar ratios of 1:0.25, 1:0.5 and 1:6.0.

In addition, {^1H}-^{15}N steady-state NOE values were measured (at 600 MHz) for ^{15}N-labeled P7 containing Oligo1 at a sub-stoichiometric ratio of 1:0.7 using the same acquisition parameters above (Section 2.1.4.2).

Chemical shift changes for P7ΔC in the presence of ssRNA were monitored using ^1H, ^{15}N HSQC spectra (at 600 MHz) using ^{15}N,^2H-REDPRO-labeled P7ΔC. Titrations were carried out using Oligo2 and Oligo3 to final P7ΔC:RNA molar ratios of 1:0.5, 1:1, 1:2, 1:4 and 1:0.25 and 1:4 respectively in the NMR buffer (see Section 2.1.1).

2.1.5. X-Ray Crystallography

2.1.5.1. Crystallization of P7ΔC

Monoclinic crystals of P7ΔC 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 a cryoprotectant.

2.1.5.2. X-Ray Data Collection

Single-wavelength anomalous diffraction (SAD) 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 (NSLS, Brookhaven National Laboratories, Upton, NY) in a single sweep of 400 x 1°, 8 s oscillations from a crystal maintained at ~100 K using radiation at a wavelength of 0.97947 Å.

2.1.5.3. Data Processing, Model Building and Refinement

Data were processed using HKL2000 and SCALEPACK (Otwinowski and Minor 1997). For each data set, the scaling thermal B factors of the final frames were within 3 $Å^2$ 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 $P3_221$ lattice, the packing density in the crystal was 2.7 $Å^3$/Da, that lies in the most probable range for proteins (Matthews 1968). 7 of the 8 Se atoms in the asymmetric unit were identified using BnP (Weeks, Blessing et al. 2002), yielding a map that was used for density modification (with 54% solvent content) and automated iterative model building using AUTO_RESOLVE (Terwilliger and Berendzen 1999). 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 (Emsley and Cowtan 2004). Completion of the structure required iterative cycles of refinement in CNS (Brunger, Adams et al. 1998) and manual rebuilding using standard stereochemical restraints (Engh and Huber 1991). The structure refinement was monitored using a randomly selected R_{free} set containing 10% of the reflections. B-factors were refined using standard vicinal restraints (1.5–2.0 $Å^2$ for main-chain atoms and 2.0–2.5 $Å^2$ for side-chain atoms). Water-

molecule sites were selected automatically using CNS and checked for consistency with $2F_o$-F_c electron-density and hydrogen-bonding criteria. The structure quality was assessed using PROCHECK (Laskowski, MacArthur et al. 1993). The biological unit, known to be a dimer from solution NMR, chemical cross-linking and SLS studies, was generated using the following symmetry operation; (1-x,1-x+y,2/3-z).

2.1.6. Fluorescence Anisotropy Measurements

3'- or 5'-fluorescein-tagged ssRNA constructs (custom synthesized by Integrated DNA Technologies) were used for the fluorescence anistropy measurements . For the fluorescence experiments, a Beacon 2000 (Invitrogen) spectrofluorimeter equipped with 488 nm excitation and 535 nm emission filters was used. The experiments were performed in NMR buffer (for consistency) with and without Mg^{2+} at room temperature (25º C). Fluorescein is in singly protonated form (pK_a=6.4) in NMR buffer (pH=6.5). Singly protonated form has reduced molar absorptivity (ε=29000$M^{-1}cm^{-1}$) compared to the corresponding dianionic form (ε=76900 $M^{-1}cm^{-1}$)(Nygren, Svanvik et al. 1998). Despite of reduced fluorescence in NMR buffer we were able to measure the change in polarization (anisotropy of 5 nM fluorescein dye (Fl)-tagged RNA segments) with high accuracy as a function of increasing protein (P7ΔC or P7fl) concentrations (150µl total volume).

The RNA constructs (either 5' or 3'-Fl tagged) used in the assays were as follows: 6rU, 12rU, 6rA, 12rA, 5'-rGrArArUrArA-3' (5'-end plus-strand S-segment: Oligo1), 5'-rGrArArUrU-3' (5'-end plus-strand M-segment: Oligo2), 5'-rArCrArArUrA-3' (5'-end plus strand L-segment: Oligo3), 5'-rGrArArUrUrArArUrUrArArA-3' (M-segment, 12 nt), 5'-rGrArArUrArArArCrUrArArA-3' (S-segment, 12 nt).

2.2. Results – Structure, Dynamics and Interactions of P7 from NMR, X-Ray Crystallography and Fluorescence Anisotropy Measurements

In this section we focus on the results obtained from NMR, X-Ray crystallography and fluorescence anisotropy measurements. Results from chemical cross-linking CD and DSC measurements have been summarized in the previous section.

2.2.1. Nature of the C-Terminal Tail in Full-length P7

The ^{15}N-^{1}H HSQC spectrum of the fully protonated P7fl revealed very sharp peaks on a background of broad resonances (Figure 6). These sharp peaks are indicative of highly unstructured regions of the protein. These intense resonances were not visible in a construct where the 40 residues from the C-terminal are truncated (P7ΔC, residues 1-129). This finding is consistent with the suggestions that P7 in both ϕ6 and ϕ12 possesses flexible C-terminal tails (Kainov, Simonov et al. 2004; Poranen and Tuma 2004).

Using fully protonated P7fl and standard hetoronuclear NMR experiments we obtained complete backbone and side-chain assignments for ^{1}H, ^{15}N and ^{13}C nuclei for the 41 (129-169) C-terminal residues (Figure 6b). Near complete backbone $^{1}H^{N}$, ^{15}N, $^{13}C'$, $^{13}C^{\alpha}$ and side-chain $^{13}C^{\beta}$ assignments of the P7 core domain (P7ΔC) were obtained using REDPRO-labeled protein (Shekhtman, Ghose et al. 2002) and TROSY based triple resonance experiments (Salzmann, Pervushin et al. 1998). For P7ΔC, 103 of the expected 123 resonances (excluding the N-terminal Met and 5 Pro residues) were observed. Out of these 103 resonances a total of 98 resonances ($^{1}H^{N}$, ^{15}N) could be assigned unambiguously.

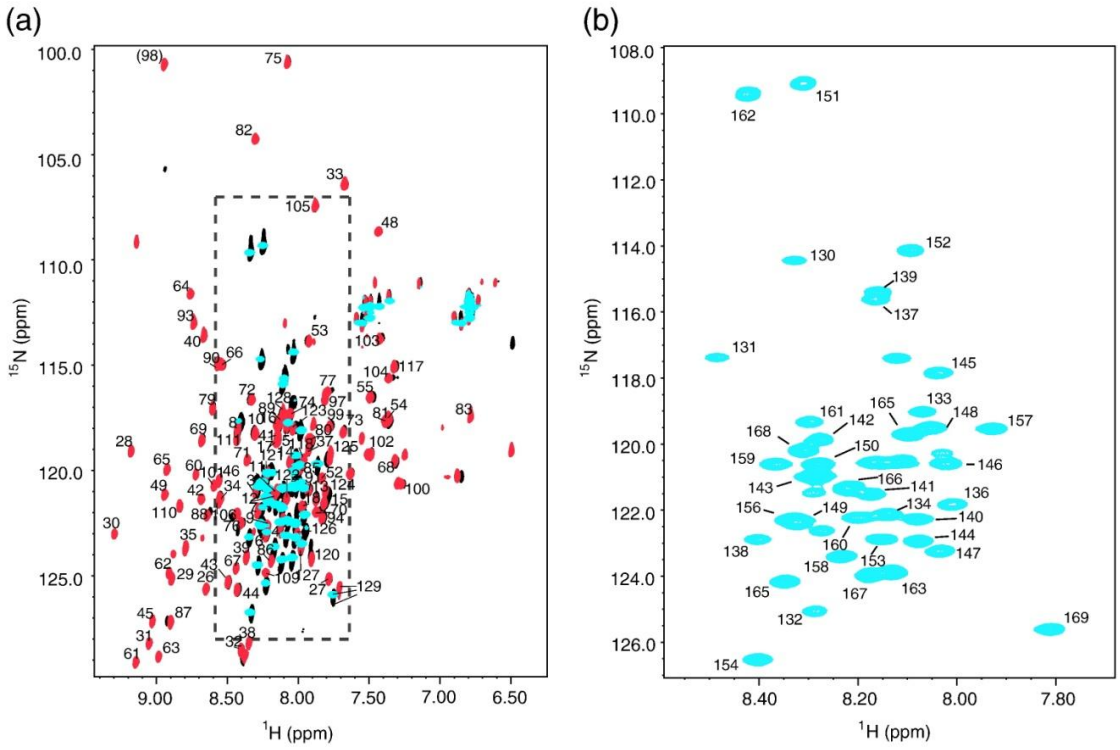

Figure 6: The C-terminal tail of P7 is flexible and interacts minimally with the protein core. (a) Overlay of ^{15}N,^{1}H TROSY spectra (800 MHz) of ^{15}N,REDPRO-^{2}H-labeled P7fl (black) and P7ΔC (red). The intense resonances corresponding to the C-terminal tail residues in a ^{15}N,^{1}H HSQC spectrum (600 MHz, cyan) of fully protonated P7fl are also shown. The C-terminal tail resonances were independently assigned in the context of P7fl. The boxed region is expanded in (b).

The $^{1}H^{N}$, ^{15}N resonances of P7ΔC were completely super-imposable on those of P7fl (Figure 6a) indicating that the C-terminal truncation has no effect on the core residues, and the C-terminal tail interacts minimally with the N-terminal core domain. In addition, the C-terminal truncation had no effect in the stability of the protein given by the differential scanning calorimetry experiment that the melting temperatures of P7ΔC and P7fl are 48.4 ºC and 47.5 ºC respectively (see Section 2.1.3.2).

2.2.2. Crystal Structure of the N-terminal Core Domain of P7

The N-terminal core of P7 (P7ΔC) crystallized in a $P3_221$ space group with similar unit cell parameters as was reported in preliminary crystallization studies (Kainov, Simonov et al. 2004). We solved the structure of P7ΔC to a resolution of 1.8 Å using single wavelength anomalous diffraction (SAD). The crystal parameters: data collection, refinement and structural statistics are given in Table 1.

P7ΔC forms a novel α-β sandwich (Figure 7) with two sets of α-helices flanking a central five-stranded parallel β-sheet (α-β-β-α-β-α-β-α-β-α 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}N-{^{1}H} NOE values varying from -0.47 to 0.37 (as discussed below). Additionally, no electron density corresponding to residues Glu56 and Leu57 was seen. Notably, ^{15}N-$^{1}H^{N}$ 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 μs-ms timescale.

Table 1: Crystallographic Parameters

Space group	$P3_221$	
Unit-cell at 100 K	75.5, 75.5, 46.5 Å	90.0, 90.0, 120.0 °
Data quality		
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_{sym}	4.8% (41.0 %)	($I \geq -3\sigma_I$ for observations)
Mean redundancy	11.0 (9.0)	
Completeness	96.9% (80.3 %)	(All measured reflections)
	80.1% (40.3 %)	($I \geq 2\sigma_I$)
Mean I/σ_I	41.4 (20.4)	($I \geq \sigma_I$ after merging)
Refinement residuals ($F \geq 2\sigma_F$):		
R_{free}	22.6 %	
R_{work}	21.0 %	
Structure Quality		
RMSD bond lengths	0.019 Å	
RMSD bond angles	2.1°	
Ramachandran plot	92.0 %	Core
	8.0 %	Allowed
	0.0 %	Generously allowed
Average B factors (Å^2)		
All	35.4	
Main chain	32.6	
Side chain	40.6	
Waters	53.3	
Model contents:		
Protein molecules in the AU	1	
Protein residues	114 (A14-A55 A58-A129)	
Water molecules	190	
PDB accession code [b]	2Q82	

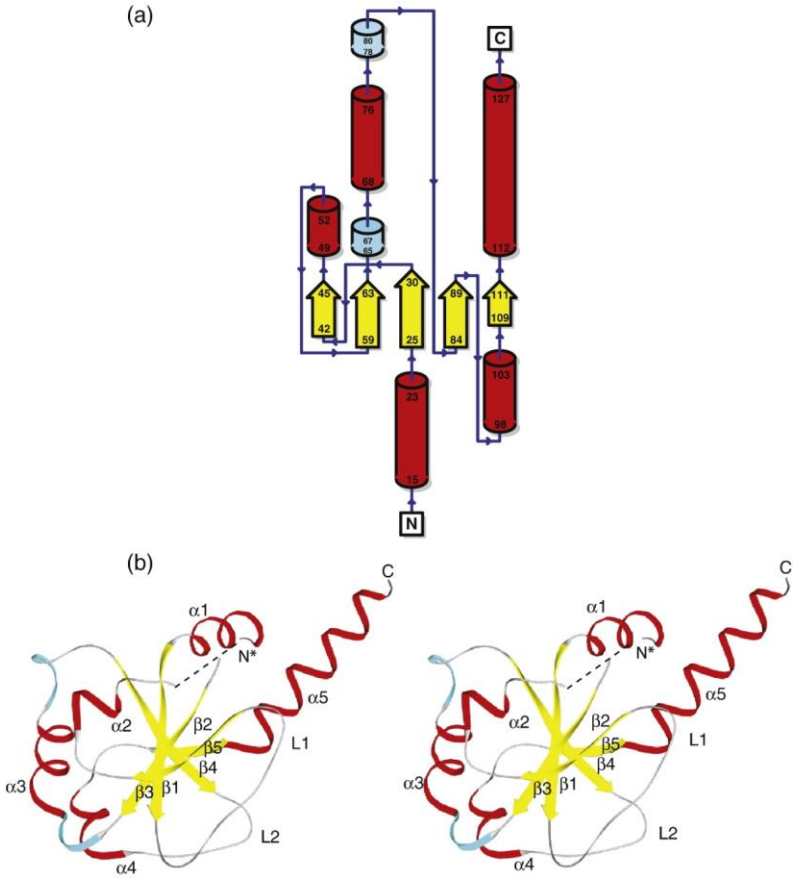

Figure 7: Structure of the P7 core domain (P7ΔC). (a) Topological diagram for P7ΔC generated using the PDBsum server (Laskowski 2007) (Laskowski 2001; Laskowski, Chistyakov et al. 2005; Laskowski 2007). Secondary structure assignments are from PROMOTIF v3.0 (Hutchinson and Thornton 1996). (b) Side-by-side stereo-view of P7ΔC. Key structural elements are labeled and the two 3_{10} helices are colored cyan.

P7ΔC contains 5 α-helices comprising residues 15-23 (α1), 49-52 (α2), 68-76 (α3), 98-103 (α4) and 112-127 (α5) and 5 β-strands comprising residues 25-30 (β1), 42-45 (β2), 59-63 (β3), 84-89 (β4) and 109-111 (β5). Additionally, two 3_{10} helices, between residues 65-67 and 78-80 connecting β3 to α3 and β3 to α4 respectively, and two large loops (L1, 32-41) connecting β1 to β2 and second (L2, 90-97) connecting α4 to β4, were seen. Multiple helix-helix contacts stabilize the core structure. Some of these key interactions involve Leu19 (α1) with Ile119 (α5), Tyr22 (α1) with Val115 (α5), Met49

(α2) with Leu70 (α3) and Phe74 (α3), Ala71 (α3) with Val100 (α4). The side-chain of Asp118 on the C-terminal helix (α5) forms a salt-bridge that of Arg18 on α1. Helix α5 is rather unique, being relatively long (16 residues) with few stabilizing contacts especially at its C-terminal end. It is stabilized by crystal contacts (intermolecular contacts with α1 and α4 helices on neighboring molecules) *in crystallo*, and 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 ϕ6 (Juuti and Bamford 1997), ϕ8 (Kainov, Butcher et al. 2003) and ϕ12 (Kainov, Simonov et al. 2004). Chemical cross-linking studies (see Section 2.1.3.1) revealed the dimer as the dominant species (> 95 %), both in P7fl and P7ΔC indicating that C-terminal tail truncation does not interfere with the ability of P7 to dimerize, consistent with previous studies (Kainov, Simonov et al. 2004). Static light scattering studies on P7ΔC at protein concentrations ranging from 200 μM to 1.7 mM at pH values of 6.5 and 7.5 (corresponding to the NMR and crystallization conditions respectively, see Section 2.1.3.1) confirmed that P7ΔC 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 P7ΔC (see Table 2).

Although crystals of P7ΔC contained a single molecule per asymmetric unit (AU), it is clear from the studies described above that it was dimeric. Analysis of the $^{15}N,^{1}H$ TROSY spectra (Figure 6a) revealed a single set of resonances implying that P7ΔC exists as a symmetric homodimer in solution. The dimer that produced the largest inter-monomer contact surface was generated using one of the 2-fold crystallographic

symmetry operations of the $P3_221$ space-group. This operation is represented by (1-x,1-x+y,2/3-z).

The putative dimerization surface consists of interactions involving helices α2 and α3 (and adjoining loops) on each monomer (Figure 8a) and is composed of the largely hydrophobic residues - Thr48, Met49, Leu50, Leu53, Tyr72, Leu79, Ala80, Val81 and Gly82 (Figure 8b). Inter-monomer hydrogen bonds involving the backbone amides (H^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 8c). Disruption of this salt bridge (see Section 2.1.2) results in dissociation of the dimer (data shown for an Asp69 to Arg mutation in Figure 8d). Dimer formation results in the burial of approximately 10.2 % (674.2 $Å^2$ of 6631.7 $Å^2$ per subunit) of the solvent accessible surface in the monomer, a value that is quite characteristic of proteins of this size.

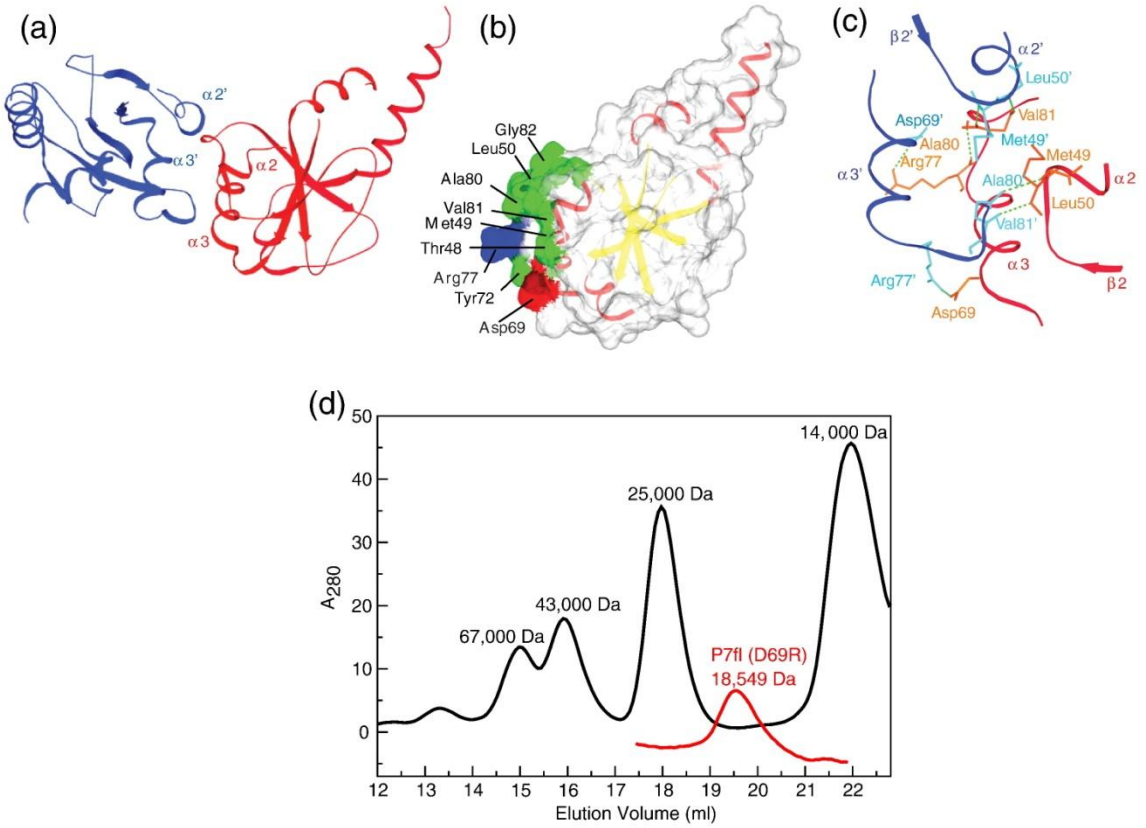

Figure 8: Structure of the P7ΔC biological unit. (a) P7ΔC forms a symmetric homodimer in solution. The monomers are colored blue and red. (b) The dimerization surface of P7ΔC is composed of mostly hydrophobic residues. Key residues forming the dimerization surface are labeled. Hydrophobic residues are colored green; the basic Arg77 and the acidic Asp69 are colored blue and red respectively. (c) The backbone amide H^N 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.

Structural Homologues of P7

A search for structural homologues in the PDB and SCOP archives using SSM (Secondary Structure Matching) (http://www.ebi.ac.uk/msd-srv/ssm) (Krissinel and Henrick 2004) yielded several structures with Z-scores > 3.0. The structures with the largest Z scores included the C-terminal BRCT domain from human BRCA1 (Gaiser,

Ball et al. 2004) (PDB code: 1OQA, Z-score = 4.1, RMSD = 2.78 Å, 52 residues aligned) (Figure 9) 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 (Figure 9). BRCT domains take part in key interactions involving proteins implicated in DNA damage repair and cell-cycle checkpoint control (Derbyshire, Basu et al. 2002; Joo, Jeffrey et al. 2002; Glover, Williams et al. 2004). 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. (Derbyshire, Basu et al. 2002)) 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.

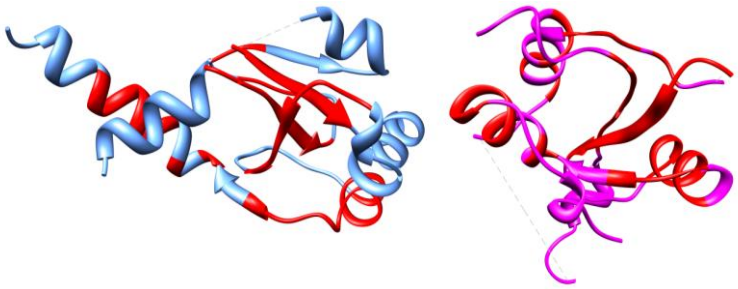

Figure 9: Ribbon representation of P7ΔC(left) and human BRCA1(PDB: 1OQA)(right) with Z-score = 4.1. 52 residues (red in both structures) could be aligned with C^{α} RMSD = 2.78 Å.

2.2.3. Dynamics of the P7 Core Domain

We measured a complete set of backbone amide ^{15}N relaxation rates–R_1 (0.78 ± 0.24 s^{-1}), R_2 (23.83 ± 6.83 s^{-1}) and ^{15}N-{^{1}H} NOE (0.58 ± 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 peak "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 (Ghose, Fushman et al. 2001) 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 (Koradi, Billeter et al. 1996). The analysis (see Table 2) yielded an axially-symmetric rotational diffusion tensor with a large axial ratio (D_{aniso}, see Table 2) of 1.66 and an effective isotropic rotational correlation time of 19.92 ns.

Table 2: Hydrodynamics of P7ΔC from NMR relaxation rates

Model	D_{xx}	D_{yy}	D_{zz}	α	β	γ	τ_c	D_{aniso}	D_{asym}	χ^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^{-7}

D_{ii} ($i = x, y, z$) are expressed in units of 10^7 s^{-1}; the angles α, β, and γ are in degrees. $\tau_c = 1/(6\text{Trace}\{D\})$ in nanoseconds; $D_{aniso} = 2D_{zz}/(D_{xx} + D_{yy})$; $D_{asym} = (D_{yy} - D_{xx})/D_{zz}$. *P*, the probability that the improvement in χ^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.

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 (Lipari and Szabo 1982; Lipari and Szabo 1982) model-free approach as implemented in the DYNAMICS package (Fushman, Cahill et al. 1997). 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^2 values (Figure 10a) for the first 9 residues were quite low, varying from 0.18 ± 0.01 to 0.42 ± 0.01, indicated substantial disorder in solution and a possible explanation for the lack of electron density for this region. For the core structure (14-129), S^2 values for the center of α1 were smaller (Arg18: 0.63 ± 0.04) than the average along the polypeptide chain (0.83 ± 0.16, 10 % trimmed – 0.87 ± 0.08) indicating some dynamics on the ps-ns timescale.

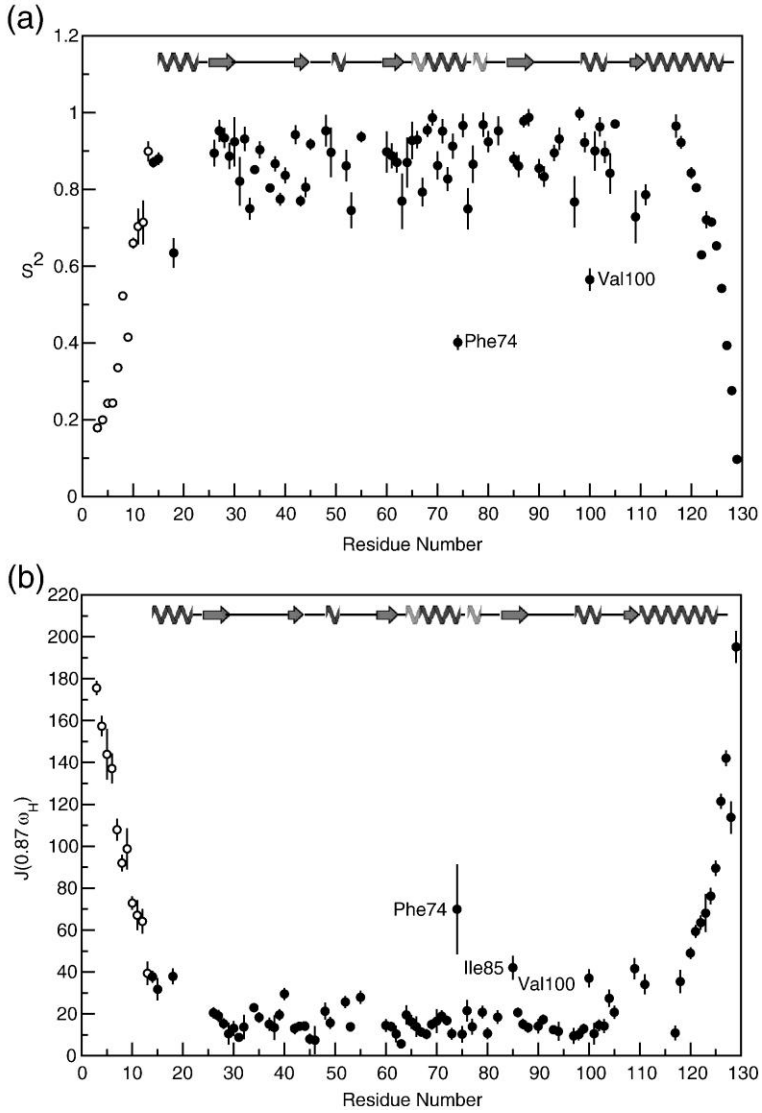

Figure 10: Fast dynamics of the P7ΔC backbone. (a) S^2 values were obtained from a Lipari-Szabo model-free analysis utilizing backbone R_1, R_2 and steady-state ^{15}N-{^1H} NOE data at 600MHz. Data were fitted to an anisotropic model using the NH unit vectors generated utilizing the crystal coordinates. The microdynamic parameters of the flexible N-terminus of P7ΔC, 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 (α – dark, 3_{10} – light) by springs and β-strands by arrows. Phe74 and Val100 that have an unusually low S^2 values are labeled. (b) Reduced spectral density values near the ^1H Larmor frequency – $J(0.87\omega_H)$. Residues with increased $J(0.87\omega_H)$ values due to extensive sub-nanosecond timescale dynamics are labeled.

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^{α} thermal factors for α1, which were, on an average, significantly higher (C^{α} B-factors = 46.9 ± 7.5) than other regions of the protein with definite secondary structure (C^{α} B-factors = 30.2 ± 8.7). Some low order parameters were also seen in the extreme C-terminus at the end of helix α5 though the C^{α} thermal factors for this region are relatively low (27.2 ± 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^2 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 ± 0.07 and 0.79 ± 0.03 for Glu109 and Ser111 respectively) were seen in β5. It is possible that the slight disorder in β5 could facilitate rigid body motions of α5. The Ser111 position is occupied by threonine residues in the ϕ6 and ϕ8 P7 proteins (ϕ13 also contains a threonine residue that is displaced by one towards the N-terminal end, see Figure 13a 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 (Huang and Nau 2003). The only other residues in the P7ΔC core that had low S^2 values were Phe74 (0.41 ± 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 ± 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 (Fischer, Zeng et al. 1997) and residue-by-residue variations in the ^{15}N chemical shift anisotropy (Fushman, Tjandra et al. 1998; Kroenke, Rance et al. 1999) 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 (Lipari and Szabo 1982; Lipari and Szabo 1982). 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 (Farrow, Zhang et al. 1995; Dayie, Wagner et al. 1996), that has a linear dependence on the measured relaxation rates, to evaluate the spectral density function around the ^{1}H Larmor frequency $J(0.87\omega_H)$ (Figure 10b). The spectral density function at this frequency is most sensitive to sub-nanosecond dynamics. Though the spectral density functions at zero and the ^{15}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 (Pfeiffer, Fushman et al. 2001). $J(0.87\omega_H)$, 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 ± 11.5 ps/rad for the core region (14-129) of P7ΔC. The same set of residues, as in the Lipari-Szabo model-free analysis described above, namely Phe74 (70.0 ± 21.5 ps/rad) and Val100 (37.0 ± 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 ± 5.7 ps/rad for $J(0.87\omega_H)$.

Several residues displayed medium to large R_{ex} values (Figure 11a) that varied from 1.87 ± 0.73 s^{-1} for Gln123 to 12.50 ± 2.71 s^{-1} for Val60, implying slow conformational dynamics on the μs-ms timescale. The largest R_{ex} values occurred in the hydrophobic core (Figure 11b) including two of the residues that showed the largest R_{ex} values – Val60 and Leu63 (12.11 ± 2.88 s^{-1}) (red in Figure 11) and in the inter-monomer interface (Leu79 8.21 ± 1.87 s^{-1}). Both these clusters of significant R_{ex} values occur near large internal cavities where slow, breathing motion is not unexpected. Smaller R_{ex} values were also seen at the C-terminal end of α5. As pointed out earlier, the C-terminal end of α5 is stabilized by intermolecular contacts in the P7ΔC crystal and disordered on both the fast, ps-ns timescale (low S^2 values, see above) and slow, μs-ms timescale. 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 P7ΔC could be dynamic on the μs-ms timescale. This region

contains a proline residue that is strictly conserved in cystoviruses (see Figure 13a), a slow sampling of alternate conformations by this residue cannot be ruled out.

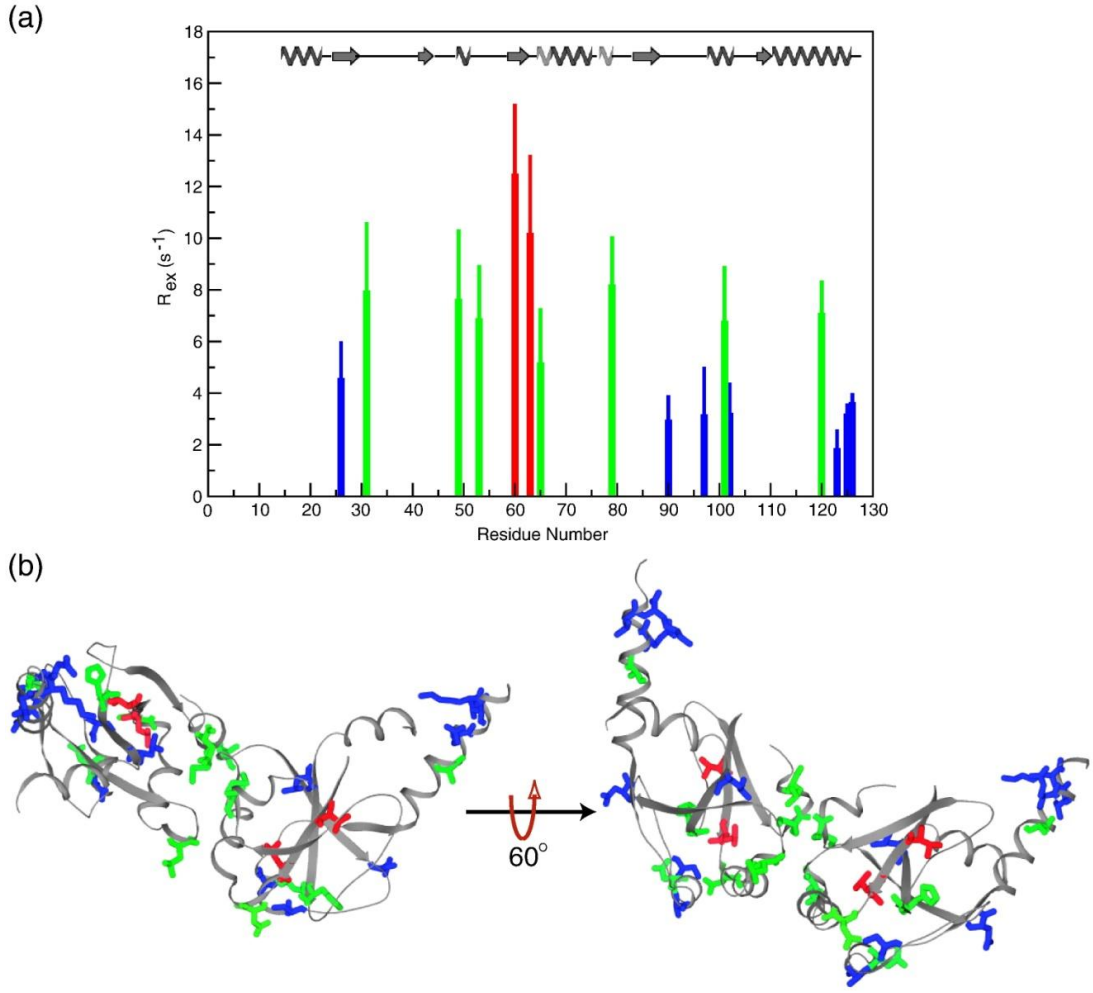

Figure 11: Slow dynamics of the P7ΔC backbone. (a) R_{ex} values obtained from the model-free analysis of the relaxation data at 600 MHz. R_{ex} > 10 s^{-1} (red), 5-10 s^{-1} (green) and < 5 s^{-1} (blue). (b) The residues that display R_{ex} values are mapped onto the structure of P7ΔC, the largest R_{ex} values were seen in the hydrophobic core and at the dimer interface.

2.2.4. Residual Structure in the C-terminal Tail of P7

As mentioned previously, the C-terminal tail residues in full-length P7 protein were highly dynamic. The ^{15}N-{^{1}H} NOE values for the 41 C-terminal tail resonances were very low indicative of extensive disorder on the ps-ns timescale (0.03 ± 0.32) (see Figure 12a). However, the residues comprising the stretch between Leu136 and Gln145

on the N-terminal end of the tail had significantly higher ^{15}N-{^{1}H} NOE values (0.31 ± 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 $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts from their random coil values (see Figure 12b) (Wishart and Sykes 1994). 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 H^{α}-H^{N} connectivities (Wüthrich 1986) corresponding to the region 136-140 (Figure 12c) confirming the helix-like nature of this region. Notably, $^{3}J(H^{N}H^{\alpha})$ values for this region were marginally smaller than the average (5.24 ± 0.38 Hz compared with 7.02 ± 1.67 Hz), these values are usually < 3-4 Hz for well-structured helices (Vuister and Bax 1993). 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 (see Section 2.1.3.3) of P7 and P7ΔC revealed a change of approximately 4 % in helical content (as determined by the ellipticity value at 222 nm). This may be interpreted qualitatively 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 (Poranen and Tuma 2004). This speculation is further supported by Raman difference spectra obtained by Benevides et. al. (Benevides, Juuti et al. 2002) which seem to indicate a higher α-helix content for full-length P7 in the fully assembled PX in ϕ6. It was suggested by the authors that this additional helicity was a result of P7/P1 interactions.

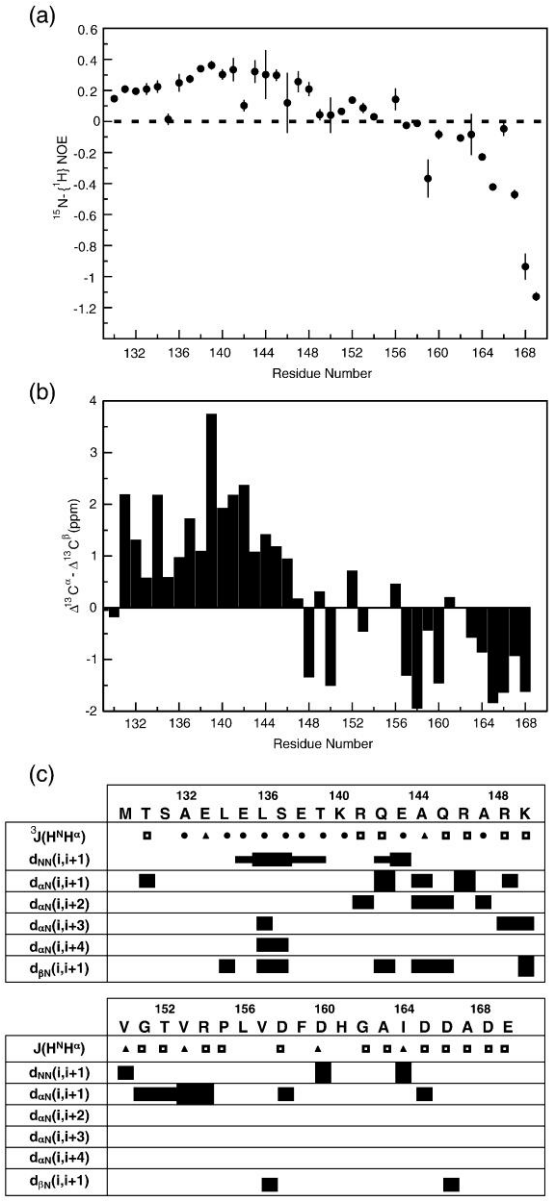

Figure 12: The C-terminal tail residues in full-length P7 are highly disordered. (a) Steady-state ^{15}N-{^{1}H} NOE data at 600 MHz for the 41 C-terminal tail residues in fully protonated full-length P7. (b) The difference in the deviation of the C^{α} and C^{β} chemical shifts from the corresponding 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 H^{N} are classified as strong, medium and weak represented by the thickness of the horizontal bars. Also shown are the $^{3}J(H^{N}H^{\alpha}$ values- < 6.5 Hz (filled circles), 6.5-8.5 Hz (open squares) and > 8.5 Hz (filled triangles).

2.2.5. Sequence Conservation Among Various Cystoviral P7 Proteins

An investigation of the P7 sequences from the ϕ6, ϕ8, ϕ12 and ϕ13 cystoviruses revealed that they could be broadly classified into two groups (see Figure 13a). As will be clarified in the discussion below, the hydrophobic residues at the dimer interface are largely conserved among all the cystoviral P7 proteins (Figure 13a). Additionally, as mentioned above, the dimer interface in ϕ12 P7 is stabilized by salt-bridges between residues Asp69 on one monomer and Arg77 on the other. Both these residues that lie on α3 form a D-$(X)_7$-R sequence with X denoting any (mostly hydrophobic) residue. A similar sequence – E-$(X)_7$-R can also be identified in ϕ8 P7 displaced by two residues towards the N-terminus of the protein. For ϕ6 and ϕ13 P7 proteins on the other hand, sequences given by R-$(X)_7$-E can be identified at the dimer interface. These sequences lie on the α2 helix (as opposed to the α3 helix as in ϕ8 and ϕ12) and the adjoining loop. Notably, the dimer interface is bounded by α2 on the N-terminal end and α3 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 ϕ6 and ϕ13 compared to ϕ8 and ϕ12. Nevertheless, based on the sequence analysis and expected structural features it is expected that the overall nature of the dimer interface is likely to be largely similar for P7 proteins in all the cystoviruses.

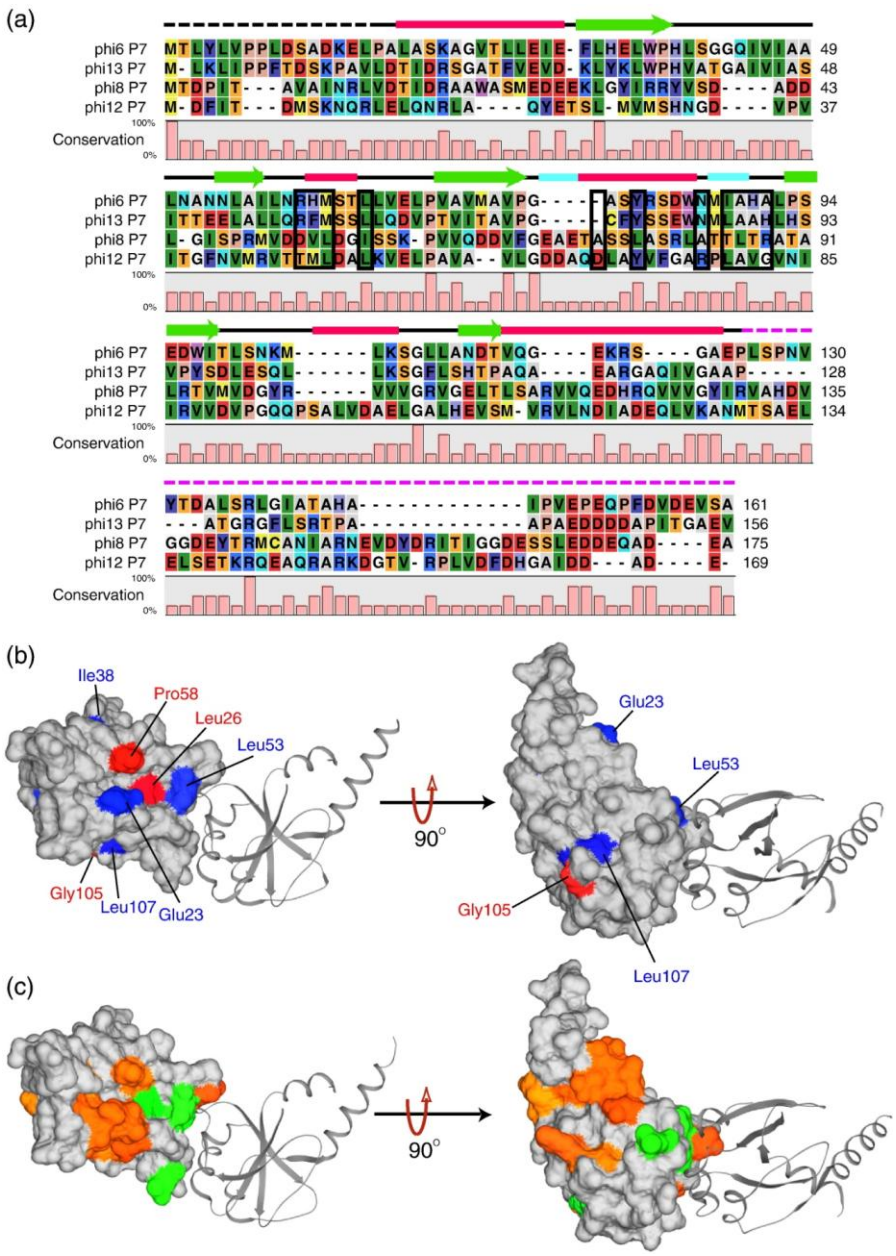

Figure 13: Sequence conservation among cystoviral P7 proteins. (a) Sequence alignment for the P7 proteins for the ϕ6, ϕ8, ϕ12 and ϕ13 cystoviruses are shown. The residues that comprise the dimer interface are boxed. The regions of definite secondary structure are indicated – α-helices (thick red line), 3_{10}-helices (thick cyan line), β-strand (green arrow). The dotted black line indicates the residues that have no electron density in the crystal structure of P7ΔC. 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 s^{-1} for P7ΔC are shaded green. Residues for which $^{1}H^{N}$,^{15}N resonance assignments could not be obtained in the TROSY spectra of P7ΔC are shaded gold.

Additionally, as shown earlier, disruption of the salt bridge at the inter-monomer interface results in dissociation of the P7 dimer (Figure 8d). It has been proposed that in ϕ6, dimeric P7 is required for assembly onto the nucleation complex formed by P4 hexamers and P1 dimers as an initial step in procapsid formation (Poranen, Paatero et al. 2001). 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 β3 and form a P-$(X)_{3/5}$-V-(X)-G motif. Of these residues only Pro58 (in ϕ12) 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 β1 (Leu26 in ϕ12), in the loop between α4 and β5 (Gly105) that facilitates formation of a tight β-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 13b). Interestingly, Leu26 ($4.6 \pm 1.4\ s^{-1}$) and Leu53 ($6.9 \pm 2.1\ s^{-1}$) show significant R_{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 13c). These observations considered together are indicative of the presence of slow dynamics in the μs-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 (Ma, Wolfson et al. 2001; DeLano 2002). 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 ϕ12). It is to be noted however, that the extreme C-terminal tail of P7 is highly acidic in all the sequenced cystoviral P7 proteins.

2.2.6. Monitoring P7/RNA Interactions Using NMR Chemical Shift Mapping

It has been shown that P7 negative particles in ϕ6 package the positive strands of the viral genome with greatly reduced efficiency (Juuti and Bamford 1995). 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 ϕ6 (Mindich 2004). On the basis of these biochemical results and noting the structural similarities of P7 with the BRCT domains which have shown to interact with DNA (Kobayashi, Figaroa et al. 2006), we considered the possibility whether P7 could play a role in RNA recognition i. e. whether P7 in full-length or tail-truncated form could interact with RNA. In order to test this hypothesis, we recorded a series of $^{15}N,^{1}H$ HSQC spectra of fully-protonated, full-length P7 containing varying amounts of a 5nt ssRNA constructs corresponding to the 5'-ends of the plus-strands of the S (Oligo1), M (Oligo2) or L (Oligo3) segments of the ϕ12 genome. Upon analysis of the dynamic C-terminal tail resonances (data for Oligo1 is shown in Figure 14), significant chemical shift changes were seen in the presence of ssRNA. 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 14b). 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 ssRNA since several other negatively charged residues located in this disordered region did not show chemical shift changes. Additionally, analysis of the ^{15}N-{^{1}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) (Figure 15) with the largest changes displayed by His161 (ΔNOE = 0.63 ± 0.18) and by Glu143 (ΔNOE = 0.22 ± 0.11). Small but significant changes were also seen for Asp160, Gly162 and Ala163. The resonance corresponding to Phe159 was completely broadened out even for this sub-stoichiometric protein/RNA ratio and could not be analyzed.

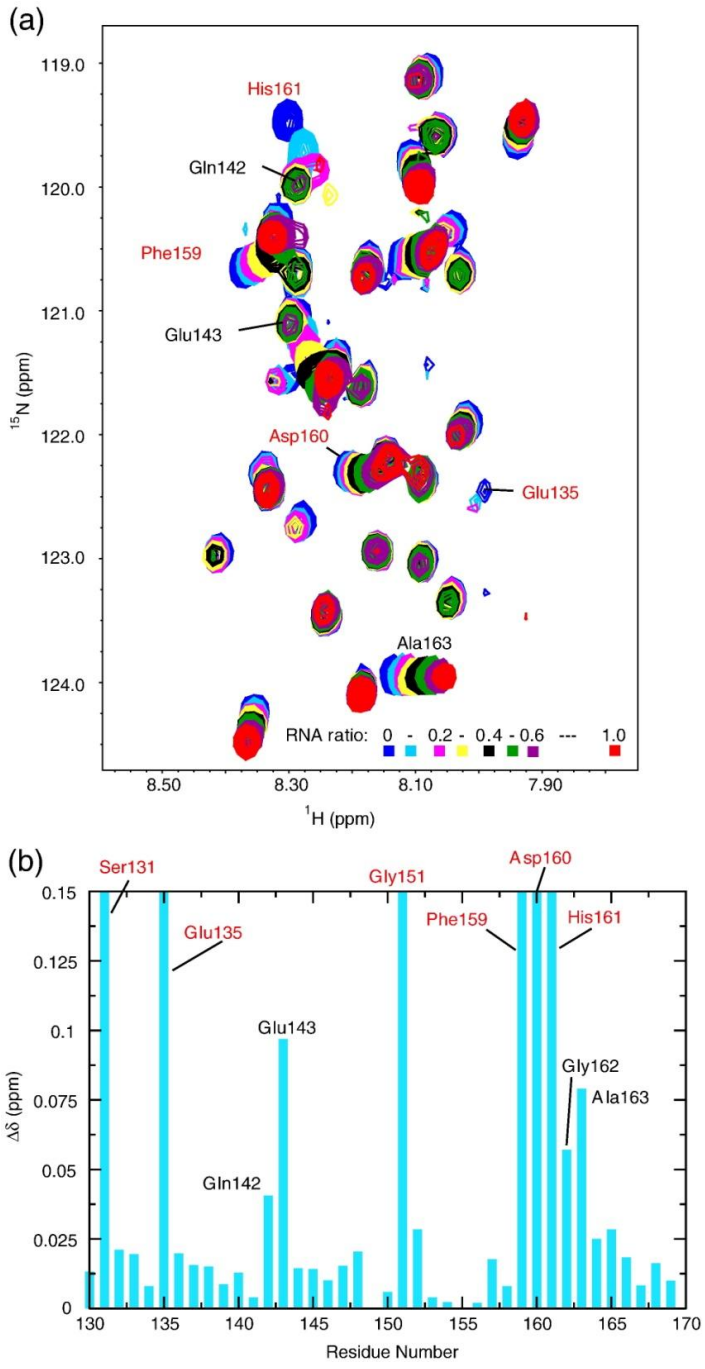

Figure 14: Interactions of the C-terminal tail in P7fl with ssRNA. (a) Expanded region of a ^{15}N,^{1}H HSQC spectrum (acquired at 600 MHz) of fully-protonated ^{15}N-labeled P7fl in the presence of varying amounts of Oligo1 (see Section 2.1.4.5). Substantial shift changes were are also seen for Gly151 and Gly162 (not shown in the expansion). The spectra are plotted at a suitable contour level to allow for optimal 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.

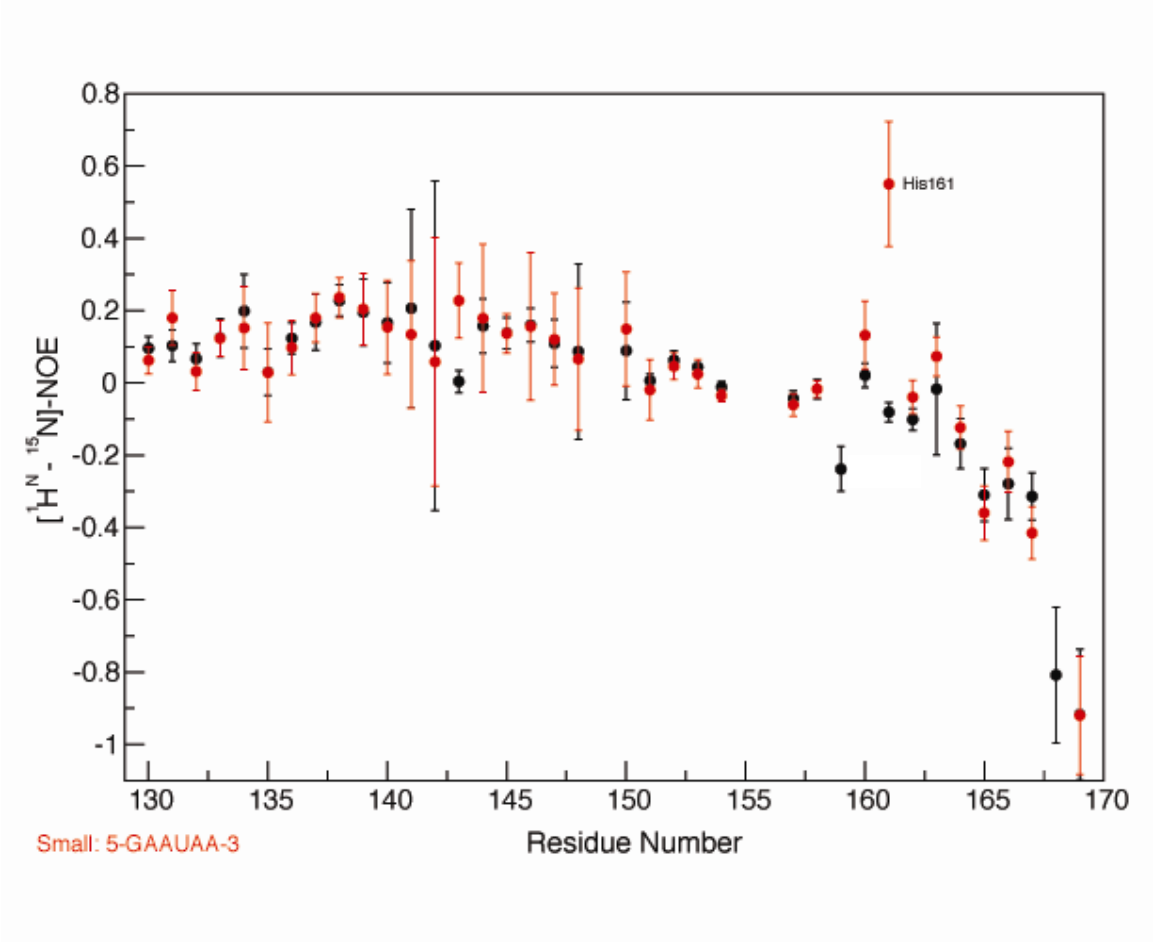

Figure 15: ^{15}N-{^{1}H} NOE values for the tail residues; apo (black) and in the presence of sub-stoichiometric (1:0.7) ssRNA (Oligo1) (red). Positive changes in ^{15}N-{^{1}H} NOE reveals ordering in the tail region. His161 displays the largest change, Phe159 is completely broadened and could not be analyzed.

In order to test whether P7 could interact with ssRNA in the absence of the dynamic C-terminal tail, experiments similar to those described in the previous paragraph were carried out using ^{15}N,^{2}H-REDPRO-labeled P7ΔC containing varying amounts of Oligo2 and Oligo3 (see Section 2.1.4.5). Substantial chemical shift changes were seen in the N-terminal core (results for Oligo2 are shown in Figure 16). 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 α5, 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 (c). Some changes were also seen at the extreme N-terminus (the N- and C-termini of P7ΔC are highly dynamic, note that electron densities 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 possible – (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 involving the C-terminal tail. These two scenarios are analyzed below (see Section 2.3). 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 P7ΔC to a simple quadratic equation (corresponding to a one-site binding model) (Lian and Roberts 1993) and obtained K_d values between 500-800 μM. However, it should be stressed here that this number likely does not provide a true measure of the affinity of P7/RNA interactions given the possibility of mutiple-site binding and/or allosteric effects.

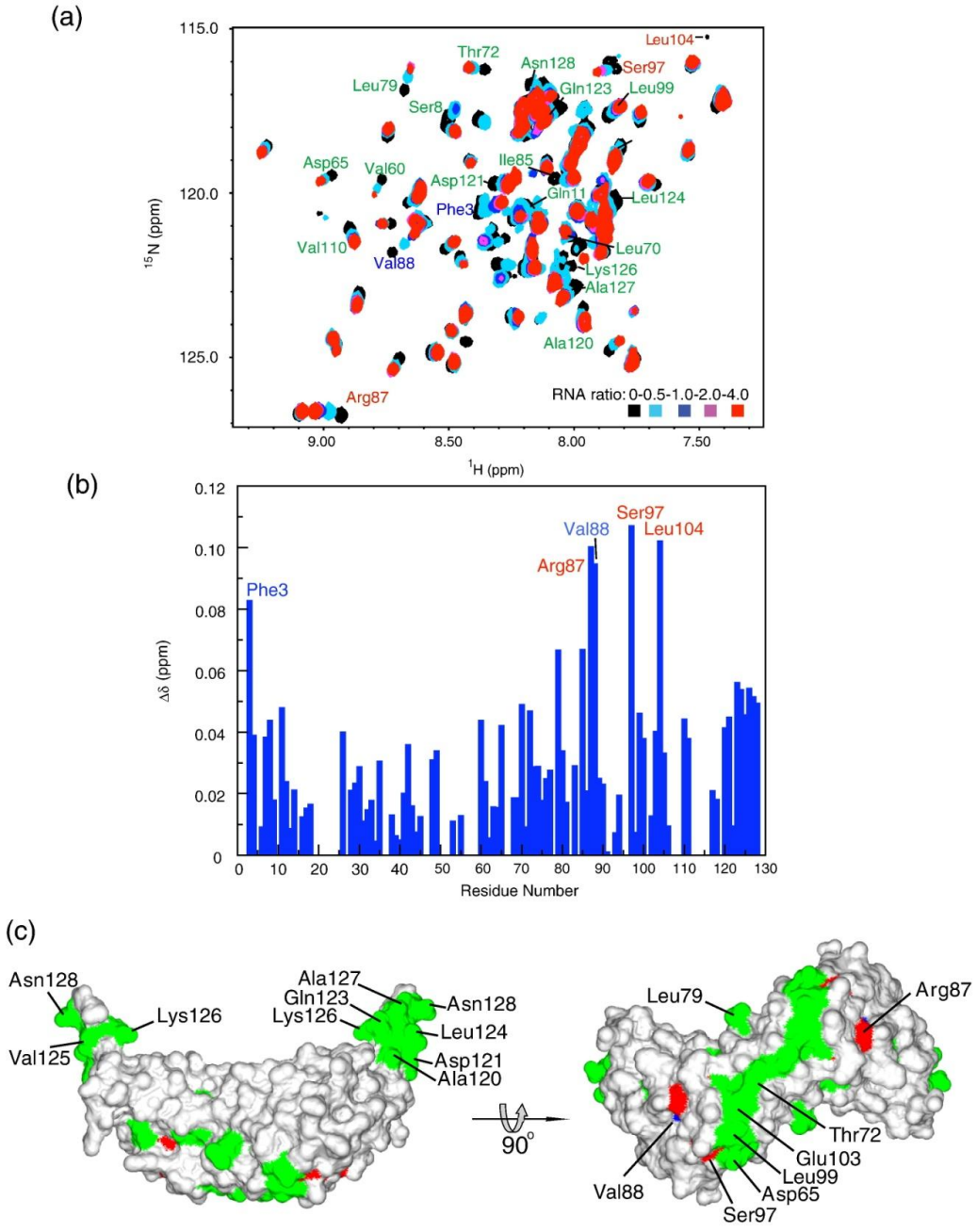

Figure 16: Interactions of P7ΔC with ssRNA. (a) Expanded region of a ^{15}N,^{1}H HSQC spectrum (acquired at 600 MHz) of ^{15}N,^{2}H-REDPRO-labeled P7ΔC with varying amounts of Oligo2 (see Section 2.1.4.5). Residues that undergo large chemical shift changes (scaled shift changes calculated using Equation 3 for a P7ΔC:RNA molar ratio of 1:2) are labeled - 0.04 ppm < Δδ < 0.07 ppm are labeled in green, 0.07 ppm ≤ Δδ < 0.10 ppm are labeled in blue and Δδ > 0.10 ppm are labeled in red. (b) Scaled chemical shift changes plotted against residue number for a P7ΔC: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).

Based on our analysis of the RNA induced perturbations in both P7fl and P7ΔC, the residues on P7 that undergo significant alterations in their spectral properties (undergo chemical shift changes and/or are broadened can be represented schematically in Figure 17 below.

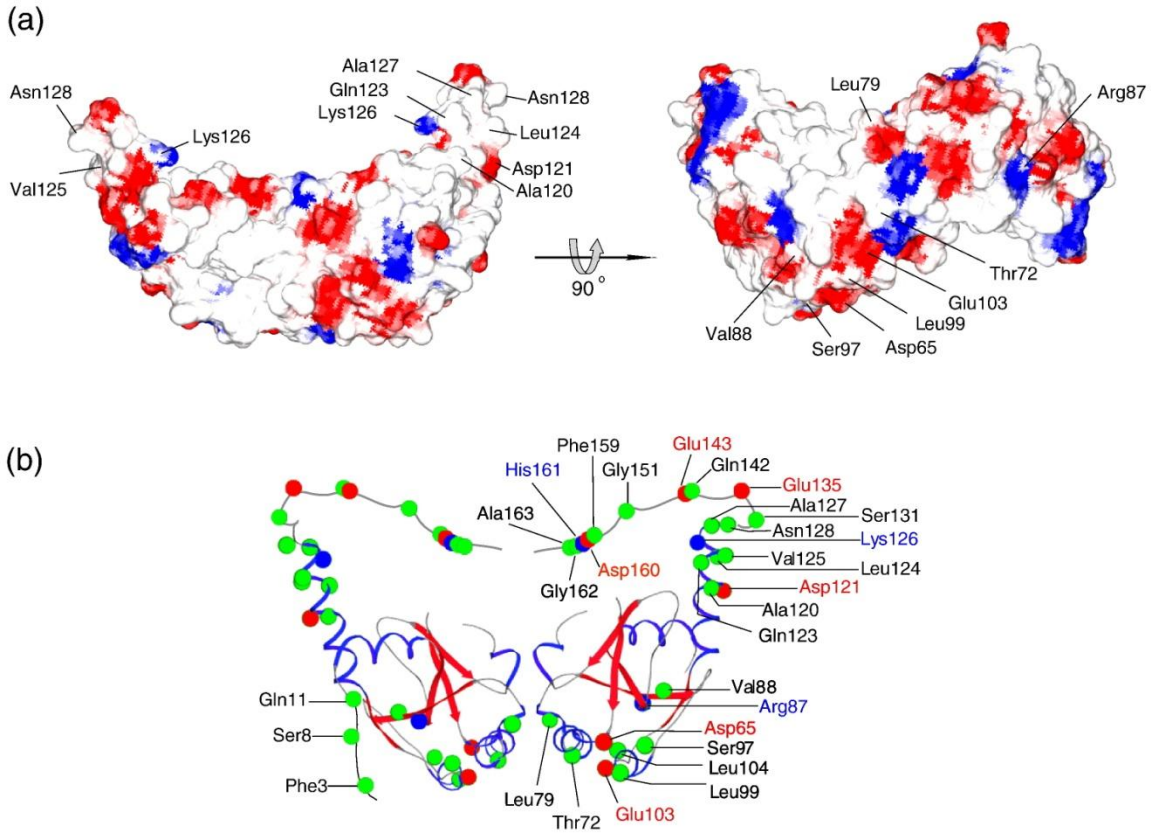

Figure 17: P7/RNA interactions. (a) Surface plot of P7ΔC colored by electrostatic potential. Residues on P7ΔC that show chemical shift changes > 0.04 ppm in the presence of RNA (P7ΔC:Oligo2 ratio 1:2) are labeled. (b) Residues that undergo large chemical shift changes in P7 in the presence of RNA are shown. Red, blue and green spheres represent acidic, basic and hydrophobic residues respectively. The dynamic N- and C-terminal tail conformations are represented schematically.

2.2.7. Fluorescence Anisotropy Measurements of P7/RNA Interactions

As mentioned earlier, two sets of binding sites for ssRNA were found on P7. It was therefore necessary to test whether a longer piece of ssRNA that could engage both sites simultaneously would possibly bind with higher affinity. In addition, experiments were required to ascertain the dependence, if any, of P7/RNA interactions on RNA sequence. These experiments would require copious amounts of ssRNAs of various lengths and sequences to be synthesized commercially, making these experiments prohibitively expensive if solution NMR were to be used to determine these interactions. Therefore, we switched to using fluorescence anisotropy which, while being an extremely sensitive probe of binding, requires smaller amounts of material.

For the fluorescence anisotropy measurements we utilized 5'- or 3'-Fluorescein (Fl) tagged Oligo1, Oligo2 Oligo3, 6nt PolyU, 12nt PolyU and a 12nt construct that corresponds to the extreme 5'-end of the genomic M strand (see Section 2.1.6) (Figure 18). The affinities determined using 5'- and 3' polyU constructs were comparable. Therefore we show data only for the 5'-Fl-tagged constructs. These experiments revealed the following conclusions:

(1) Full length P7 has a far higher affinity for ssRNA compared with P7ΔC in agreement the with NMR chemical shift perturbation analysis.

(2) Higher affinity binding is seen in the presence of Mg^{2+}. This observation is also in agreement with NMR experiments where large chemical shift perturbations are seen in several negatively charged residues (Figure 14 and Figure 16). This effect, while less pronounced for the shorter 6nt polyU construct, is quite substantial for the 12nt polyU construct. Because as shown by NMR experiments two specific surfaces bind ssRNA,

longer ssRNA (12nt-polyU) can bind these surfaces simultaneously but 6nt-polyU can bind one surface at a time, hence the binding of longer ssRNA is stronger.

(3) On the whole, 12nt long RNA oligos seem to have marginally higher affinity for P7 compared to 6nt RNA oligos, however, the dramatic difference in affinity that was expected was not seen at least for 6nt polyU oligos. (Figure 19).

However, dramatically different results were obtained upon employing the 12nt construct that corresponded to the 5'-end of the plus strand of the genomic M-segment. The affinity of this construct towards P7 (in the presence of Mg^{+2}) was found to be several-fold higher (K_d=8 μM) than the polyU construct of the same length.

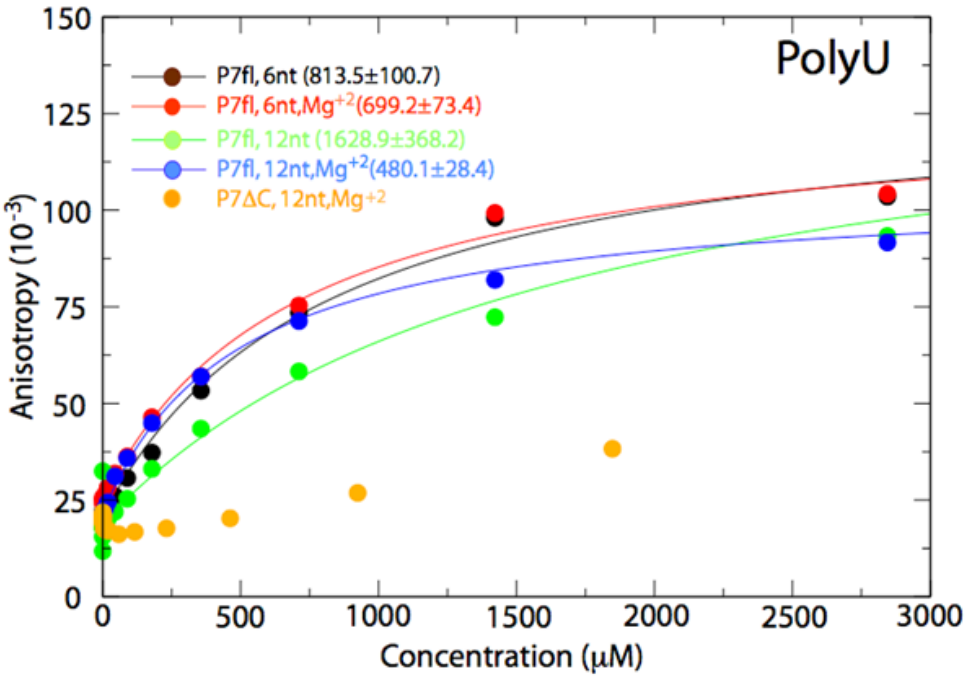

Figure 18: Fluorescence anisotropy experiments on P7fl and P7ΔC with 6nt and 12nt RNA oligos (polyU) with and without Mg^{2+}.

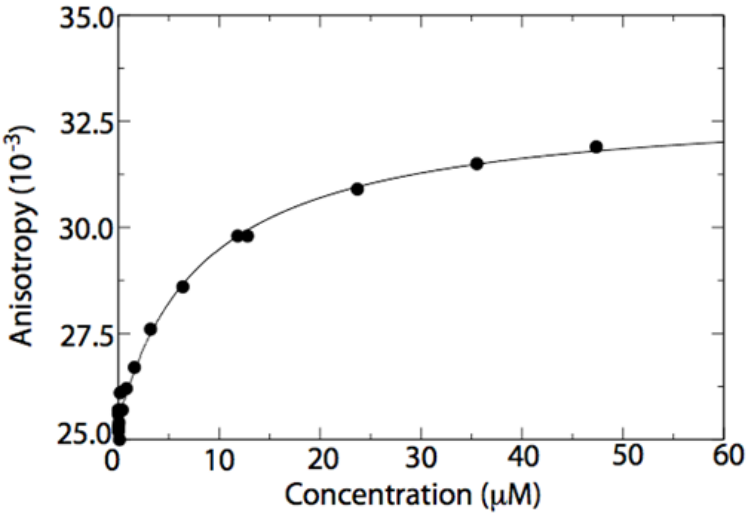

Figure 19: Fluorescence anisotropy measurement on a 12nt construct corresponding to the 5'-end of the plus strand of the genomic medium (M)-segment tagged with fluorescein at the 5'-end (5'-Fl-rGrArArUrUrArArUrUrArArA-3') with P7fl in the presence Mg^{2+}.

2.3. Possible Role of P7 in the Cystoviral PX

The similarities between the α/β-fold of the homodimeric P7 core (P7ΔC) with tandem BRCT domains hints towards a structural role for P7 in the cystoviral polymerase complex similar to that seen for BRCT domains (Glover, Williams et al. 2004) in DNA repair proteins. This is supported by *in vitro* assembly studies in ϕ6 which reveal that incorporation of P7 greatly accelerates formation of the P1-P4 nucleation and complete procapsids (Poranen, Paatero et al. 2001). It is likely that this stabilization occurs through interactions between P1 and P7 (Kainov, Butcher et al. 2003). 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 (Huiskonen, de Haas et al. 2006). 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 (Benevides, Juuti et al. 2002). Additionally, given the sensing/recognition of RNA by P7, seemingly in a manner that depends on the RNA sequence, it is possible that 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 (Mindich 2004). This role for P7 would be consistent with the fact that that ϕ6 procapsids lacking P7 package the positive strands of the viral genome with greatly reduced efficiency (Juuti and Bamford 1995). Given that chemical shift changes are seen for P7ΔC, P7 does not require the C-terminal tail to interact with RNA. However, it is clear from the widespread observed chemical shift changes that, 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. 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 14 and Figure 16). Experiments monitoring the change of fluorescence anisotropy of 5'-fluorescein-tagged RNA in the presence of P7 indicate that P7 in full-length form binds RNA with a higher affinity than P7ΔC (Figure 18). Further, the affinity of RNA towards both P7fl and P7ΔC increases in the presence of Mg^{+2} (Figure 18). 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^{+2} ions is also possible (Lee and Richardson 2005). 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.

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) (Juuti and Bamford 1997). 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 (Kainov 2004), 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 (Juuti and Bamford 1997). P4 hexamers are also known to be located on the 5-fold axes, forming a symmetry mismatch (Huiskonen, de Haas et al. 2006; Jaalinoja, Huiskonen et al. 2007). This would place P4 in close proximity to P7. A 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 (Wei, Cheng et al. 2009). 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 (Sen, Heymann et al. 2008).

3. Open Questions

While we have completely characterized the structure and the dynamics of P7 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? 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 (Juuti and Bamford 1995) and regulating the fidelity of transcription (Juuti and Bamford 1997) as evidenced by the available biochemical data?

Some of these answers can only be obtained upon determining the nature of interactions between P7 and the other components of the cystoviral PX, namely P1, P2 and P4. However, obtaining information on interactions between P7 and P1 in atomic detail is extremely challenging due to the unavailability of a high-resolution structure of P1. Obtaining P1 (from several cystoviral species) in soluble form at high concentrations has proved to be highly elusive. While crystal structures of P2 and P4 exist, their interactions with P7 may be weak or transient to allow crystallization. Given the size of the proteins involved (210 kDa hexamer for P4 and 75 kDa for P2), NMR experiments would be extremely challenging though not beyond the realm of possibility given some advances in NMR studies of large proteins and protein complexes (Gelis, Bonvin et al. 2007; Sprangers and Kay 2007). In the short term, however, a more achievable goal is

likely to understand the nature of RNA binding and RNA induced conformational changes in P7 through the determination of the structure of the P7/RNA complex. Given that the affinity of P7 towards genomic RNA seems greatly enhanced, attempts to crystallize P7 in complex with optimal constructs genomic 5'-ends (or possibly 3'-end) may be an excellent place to start.

Publications

Most of he work presented in this thesis has been published in

Eryilmaz, E., J. Benach, et al. (2008). "Structure and Dynamics of the P7 Protein from the Bacteriophage ϕ12." J. Mol. Biol. **382**(2): 402-422

Other Publications

Nicholas, M. P., E. Eryilmaz, et al. (2010). "Nuclear spin relaxation in isotropic and anisotropic media." Prog. Nucl. Magn. Reson. Spectrosc. **57**(2): 111-158

Goktas, H. Ayhan, U.B., Gunduz, G., Disbudak , H., Eryilmaz, E., Oke, G., Cicek, B., and Somer, M. (2006). "Synthesis of carbon nanotubes by a plasma based pulsed electron beam generator." Physica Scripta **2006**(T123): 145

References

Benevides, J. M., J. T. Juuti, et al. (2002). "Characterization of subunit-specific interactions in a double-stranded RNA virus: Raman difference spectroscopy of the phi6 procapsid." Biochemistry **41**(40): 11946-11953.

Boggs, P. T., J. R. Donaldson, et al. (1989). "ODRPACK software for weighted orthogonal distance regression." ACM Trans. Math. Software **15**: 348-364.

Bruenn, J. A. (1993). "A closely related group of RNA-dependent RNA polymerases from double-stranded RNA viruses." Nucleic Acids Res.(21): 5667-5669.

Bruenn, J. A. (2003). "A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases." Nucleic Acids Res.(31): 1821-1829.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr. D Biol. Crystallogr. **54**(Pt 5): 905-921.

Butcher, S. J., T. Dokland, et al. (1997). "Intermediates in the assembly pathway of the double-stranded RNA virus phi6." EMBO J. **16**(14): 4477-4487.

Butcher, S. J., J. M. Grimes, et al. (2001). "A mechanism for initiating RNA-dependent RNA polymerization." Nature **410**(6825): 235-240.

Clore, G. M., A. Szabo, et al. (1990). "Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins." J.Am.Chem.Soc. **112**: 4989-4936.

Dayie, K. T., G. Wagner, et al. (1996). "Theory and practice of nuclear spin relaxation in proteins." Annu. Rev. Phys. Chem. **47**: 243-282.

de Haas, F., A. O. Paatero, et al. (1999). "A symmetry mismatch at the site of RNA packaging in the polymerase complex of dsRNA bacteriophage phi6." J. Mol. Biol. **294**(2): 357-372.

Delaglio, F., S. Grzesiek, et al. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J. Biomol. NMR **6**(3): 277-293.

DeLano, W. L. (2002). "Unraveling hot spots in binding interfaces: progress and challenges." Curr. Opin. Struct. Biol. **12**(1): 14-20.

Derbyshire, D. J., B. P. Basu, et al. (2002). "Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor." EMBO J. **21**(14): 3863-3872.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr. D Biol. Crystallogr. **60**(Pt 12 Pt 1): 2126-2132.

Engh, R. and R. Huber (1991). "Accurate bond and angle parameters." Acta Crystallogr. A **47**: 392-400.

Farrow, N., O. Zhang, et al. (1995). "Spectral density function mapping using ^{15}N relaxation data." J. Biomol. NMR **6**: 153-162.

Fischer, M. W. F., L. Zeng, et al. (1997). "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. **119**: 12629-12642.

Fushman, D., S. Cahill, et al. (1997). "The main chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: Analysis of ^{15}N relaxation with monomer/dimer equilibration." J. Mol. Biol. **266**: 173-194.

Fushman, D., N. Tjandra, et al. (1998). "Direct measurement of ^{15}N chemical shift anisotropy in solution." J. Am. Chem. Soc. **120**(42): 10947-10952.

Gaiser, O. J., L. J. Ball, et al. (2004). "Solution structure, backbone dynamics, and association behavior of the C-terminal BRCT domain from the breast cancer-associated protein BRCA1." Biochemistry **43**(51): 15983-15995.

Gelis, I., A. M. J. J. Bonvin, et al. (2007). "Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR." Cell **131**(4): 756-769.

Ghose, R., D. Fushman, et al. (2001). "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. **149**(2): 204-217.

Glover, J. N., R. S. Williams, et al. (2004). "Interactions between BRCT repeats and phosphoproteins: tangled up in two." Trends Biochem. Sci. **29**(11): 579-585.

Gottlieb, P., C. Potgieter, et al. (2002). "Characterization of phi12, a bacteriophage related to phi6: nucleotide sequence of the large double-stranded RNA." Virology **295**(2): 266-271.

Gottlieb, P., H. Wei, et al. (2002). "Characterization of phi 12, a bacteriophage related to phi 6: nucleotide sequence of the small and middle double-stranded RNA." Virology **293**(1): 118-124.

Hoogstraten, D., Qiao, X., Sun, Y., Hu, A., Onodera, S., and Mindich, L. (2000). "Characterization of phi8, a bacteriophage containing three double-stranded RNA genomic segments and distantly related to phi6." Virology **272**: 218-224.

Huang, F. and W. M. Nau (2003). "A conformational flexibility scale for amino acids in peptides." Angew. Chem. Int. Ed. Engl. **42**(20): 2269-2272.

Huiskonen, J. T., F. de Haas, et al. (2006). "Structure of the bacteriophage phi6 nucleocapsid suggests a mechanism for sequential RNA packaging." Structure **14**(6): 1039-1048.

Hutchinson, E. G. and J. M. Thornton (1996). "PROMOTIF--a program to identify and analyze structural motifs in proteins." Protein Sci. **5**(2): 212-220.

Jaalinoja, H. T., J. T. Huiskonen, et al. (2007). "Electron cryomicroscopy comparison of the architectures of the enveloped bacteriophages phi6 and phi8." Structure **15**(2): 157-167.

Johnson, B. A. (2004). "Using NMRView to visualize and analyze the NMR spectra of macromolecules." Methods Mol. Biol. **278**: 313-352.

Joo, W. S., P. D. Jeffrey, et al. (2002). "Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brca1 BRCT structure." Genes Dev. **16**(5): 583-593.

Juuti, J. T. and D. H. Bamford (1995). "RNA binding, packaging and polymerase activities of the different incomplete polymerase complex particles of dsRNA bacteriophage phi 6." J. Mol. Biol. **249**(3): 545-554.

Juuti, J. T. and D. H. Bamford (1997). "Protein P7 of phage phi6 RNA polymerase complex, acquiring of RNA packaging activity by in vitro assembly of the purified protein onto deficient particles." J. Mol. Biol. **266**(5): 891-900.

Juuti, J. T. B. D. H. T. R. and G. J. Thomas (1998). "Structure and NTPase activity of the RNA-translocating protein (P4) of bacteriophage phi6." J. Mol. Biol. **279**: 347-359.

Kainov, D., R. Tuma, et al. (2006). "Hexameric molecular motors: P4 packaging ATPase unravels the mechanism." Cell. Mol. Life Sci. **63**(10): 1095-1105.

Kainov, D. E., S. J. Butcher, et al. (2003). "Conserved Intermediates on the Assembly Pathway of Double-stranded RNA Bacteriophages." J. Mol. Biol. **328**: 791-804.

Kainov, D. E., Lisal, J., Bamford, D. H., and Tuma, R. (2004). "Packaging motor from double-stranded RNA bacteriophage phi12 acts as an obligatory passive conduit during transcription. ." Nucleic Acids Res. **32**(12): 3515-3521.

Kainov, D. E., E. J. Mancini, et al. (2008). "Structural basis of mechanochemical coupling in a hexameric molecular motor." J. Biol. Chem. **283**(6): 3607-3617.

Kainov, D. E., V. Simonov, et al. (2004). "Crystallization and preliminary X-ray diffraction analysis of bacteriophage varphi12 packaging factor P7." Acta Crystallogr. D Biol. Crystallogr. **60**(Pt 12 Pt 2): 2368-2370.

Kobayashi, M., F. Figaroa, et al. (2006). "Characterization of the DNA binding and structural properties of the BRCT region of human replication factor C p140 subunit." J. Biol. Chem. **281**(7): 4308-4317.

Koradi, R., M. Billeter, et al. (1996). "MOLMOL: a program for display and analysis of macromolecular structures." J. Mol. Graphics **14**(1): 51-55, 29-32.

Krissinel, E. and K. Henrick (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Crystallogr. D Biol. Crystallogr. **60**(Pt 12 Pt 1): 2256-2268.

Kroenke, C. D., M. Rance, et al. (1999). "Variability of the 15N chemical shift anisotropy in Escherichia coli ribonuclease H in solution." J. Am. Chem. Soc. **121**(43): 10119-10125.

Laskowski, R. A. (2001). "PDBsum: summaries and analyses of PDB structures." Nucleic Acids Res. **29**(1): 221-222.

Laskowski, R. A. (2007). "Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature." Bioinformatics **23**(14): 1824-1827.

Laskowski, R. A., V. V. Chistyakov, et al. (2005). "PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids." Nucleic Acids Res. **33**(Database issue): D266-268.

Laskowski, R. A., M. W. MacArthur, et al. (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J. Appl. Crystallogr. **26**: 283-291.

Lee, S. J. and C. C. Richardson (2005). "Acidic residues in the nucleotide-binding site of the bacteriophage T7 DNA primase." J. Biol. Chem. **280**(29): 26984-26991.

Lian, L. Y. and G. C. K. Roberts (1993). NMR of Macromolecules: A Practical Approach. New York, Oxford Univ. Press.

Lipari, G. and A. Szabo (1982). "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity." J.Am.Chem.Soc. **104**: 4546-4559.

Lipari, G. and A. Szabo (1982). "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2." J.Am.Chem.Soc. **104**: 4559-4570.

Lisal, J., T. T. Lam, et al. (2005). "Functional visualization of viral molecular motor by hydrogen-deuterium exchange reveals transient states." Nat. Struct. Mol. Biol. **12**(5): 460-466.

Luo, P. and R. L. Baldwin (1997). "Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water." Biochemistry **36**(27): 8413-8421.

Ma, B., H. J. Wolfson, et al. (2001). "Protein functional epitopes: hot spots, dynamics and combinatorial libraries." Curr. Opin. Struct. Biol. **11**(3): 364-369.

Makeyev, E. V., and Bamford, D. H. (2000). "Replicase activity of purified recombinant protein P2 of double-stranded RNA bacteriophage phi6." EMBO J. **19**: 124-133.

Makeyev, E. V. and D. H. Bamford (2000). "Replicase activity of purified recombinant protein P2 of double-stranded RNA bacteriophage phi6." EMBO J. **19**: 124-133.

Mancini, E. J., D. E. Kainov, et al. (2004). "Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation." Cell **118**(6): 743-755.

Matthews, B. W. (1968). "Solvent content of protein crystals." J. Mol. Biol. **33**(2): 491-497.

Mertens, P. (2004). "The dsRNA viruses." Virus Res. **101**: 3-13.

Mindich, L. (2004). "Packaging, replication and recombination of the segmented genome of bacteriophage Phi6 and its relatives." Virus Res. **101**(1): 83-92.

Mindich, L., I. Nemhauser, et al. (1988). "Nucleotide sequence of the large double-stranded RNA segment of bacteriophage phi 6: genes specifying the viral replicase and transcriptase." J. Virol. **62**(4): 1180-1185.

Mindich, L., X. Qiao, et al. (1999). "Isolation of additional bacteriophages with genomes of segmented double-stranded RNA." J. Bacteriol. **181**(15): 4505-4508.

Nygren, J., N. Svanvik, et al. (1998). "The interactions between the fluorescent dye thiazole orange and DNA." Biopolymers **46**(1): 39-51.

Otwinowski, Z. and W. Minor (1997). "Processing of X-ray Diffraction Data Collected in Oscillation Mode." Methods Enzymol. **276**: 307-326.

Pfeiffer, S., D. Fushman, et al. (2001). "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. **123**(13): 3021-3036.

Poranen, M. M. (2008). "Structural explanation for the role of Mn2+ in the activity of {phi}6 RNA-dependent RNA polymerase." Nucleic Acids Res. (36): 6633-6644.

Poranen, M. M., S. J. Butcher, et al. (2008). "Roles of the Minor Capsid Protein P7 in the Assembly and Replication of Double-Stranded RNA Bacteriophage [phi]6." J. Mol. Biol. **383**(3): 529-538.

Poranen, M. M., A. O. Paatero, et al. (2001). "Self-assembly of a viral molecular machine from purified protein and RNA constituents." Mol. Cell **7**(4): 845-854.

Poranen, M. M. and R. Tuma (2004). "Self-assembly of double-stranded RNA bacteriophages." Virus Res. **101**(1): 93-100.

Qiao, X., Qiao, J., Onodera, S., and Mindich, L. (2000). " Characterization of phi13, a bacteriophage related to phi6 and containing three dsRNA genomic segments." Virology **275**: 218-224

Ren, Z., H. Wang, et al. (2010). "Dynamics on multiple timescales in the RNA-directed RNA polymerase from the cystovirus {phi}6." Nucleic Acids Res.: gkq210.

Salgado, P. S., E. V. Makeyev, et al. (2004). "The structural basis for RNA specificity and Ca2+ inhibition of an RNA-dependent RNA polymerase." Structure **12**(2): 307-316.

Salzmann, M., K. Pervushin, et al. (1998). "TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins." Proc. Natl. Acad. Sci. U S A **95**(23): 13585-13590.

Sattler, M., J. Schleucher, et al. (1999). "Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients." Prog. Nucl. Magn. Reson. Spectrosc. **34**(2): 93-158.

Sen, A., J. B. Heymann, et al. (2008). "Initial location of the RNA-dependent RNA polymerase in the bacteriophage Phi6 procapsid determined by cryo-electron microscopy." J. Biol. Chem. **283**(18): 12227-12231.

Shekhtman, A. (2001). "Novel mechanism of regulation of the non-receptor protein tyrosine kinase Csk: insights from NMR mapping studies and site-directed mutagenesis." J. Mol. Biol. **314**(1): 129-138.

Shekhtman, A., R. Ghose, et al. (2002). "NMR structure determination and investigation using a reduced proton (REDPRO) labelling strategy for proteins." FEBS Lett. **524**: 177-182.

Sprangers, R. and L. E. Kay (2007). "Quantitative dynamics and binding studies of the 20S proteasome by NMR." Nature **445**(7128): 618-622.

Terwilliger, T. C. and J. Berendzen (1999). "Automated MAD and MIR structure solution." Acta Crystallogr. D Biol. Crystallogr. **55**(Pt 4): 849-861.

Vuister, G. W. and A. Bax (1993). "Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins." J. Am. Chem. Soc. **115**(17): 7772-7777.

Weeks, C. M., R. H. Blessing, et al. (2002). "Towards automated protein structure determination: BnP, the SnB-PHASES interface." Z. Krystallogr. **217**: 686-693.

Wei, H., R. H. Cheng, et al. (2009). "Three-Dimensional Structure of the Enveloped Bacteriophage Φ12: An Incomplete T=13 Lattice Is Superposed on an Enclosed T=1 Shell." PLoS ONE **4**(9): e6850.

Wishart, D. S. and B. D. Sykes (1994). "Chemical shifts as a tool for structure determination." Methods Enzymol. **239**: 363-392.

Wüthrich, K. (1986). NMR of proteins and nucleic acids. New York, John Wiley and Sons.

Yang, H., E. V. Makeyev, et al. (2001). "Comparison of polymerase subunits from double-stranded RNA bacteriophages." J. Virol. **75**(22): 11088-11095.