Published in final edited form as:

Biomol NMR Assign. 2017 April ; 11(1): 35–38. doi:10.1007/s12104-016-9716-5.

NMR assignments for the insertion domain of bacteriophage Sf6 coat protein

Therese N. Tripler¹, Carolyn M. Teschke^{1,2}, and Andrei T. Alexandrescu¹

¹Department of Molecular and Cell Biology, University of Connecticut, 91 N. Eagleville Road, Storrs, CT 06269-3125, USA

²Department of Chemistry, University of Connecticut, 55 N. Eagleville Rd., Storrs, CT 06269-3060, USA

Abstract

The P22 bacteriophage group is a subgroup of the λ phage supercluster, comprised of the three major sequence types Sf6, P22, and CUS-3, based on their capsid proteins. Our goal is to investigate the extent to which structure–function relationships are conserved for the viral coat proteins and I-domains in this subgroup. Sf6 is a phage that infects the human pathogen *Shigella flexneri*. The coat protein of Sf6 assembles into a procapsid, which further undergoes maturation during DNA packaging into an infectious virion. The Sf6 coat protein contains a genetically inserted domain, termed the I-domain, similar to the ones present in the P22 and CUS-3 coat proteins. Based on the P22 example, I-domains play important functional roles in capsid assembly, stability, viability, and size-determination. Here we report the ¹H, ¹⁵N, and ¹³C chemical shift assignments for the I-domain of the Sf6 phage coat protein. Chemical shift-based secondary structure prediction and hydrogen-bond patterns from a long-range HNCO experiment indicate that the Sf6 I-domain adopts a 6-stranded β -barrel fold like those of P22 and CUS-3 but with important differences, including the absence of the D-loop that is critical for capsid assembly and the addition of a novel disordered loop region.

Keywords

Structure homology; Protein dynamics; Intrinsically disordered proteins; Icosahedral particles; Conserved folding motifs

Biological context

A recent bioinformatics analysis of more than 150 P22-like phages, showed that these cluster into three branches based on coat protein sequence identity: P22, CUS-3 and Sf6-like (Casjens and Grose 2016). The Sf6 phage infects the human pathogen *Shigella flexneri* and is suggested to increase the pathogenicity of the bacterium (Parent et al. 2014). The Sf6 coat

Correspondence to: Carolyn M. Teschke; Andrei T. Alexandrescu.

Compliance with ethical standards

Ethical standards All experiments complied with all laws of the United States of America.

Conflict of interest The authors declare no conflict of interest.

protein has low sequence identity (<20 %) to the P22 and CUS-3 coat proteins but nevertheless adopts an HK97 folding motif, and contains a distinct genetically inserted I-domain (Parent et al. 2014, Rizzo et al. 2014). To date the only structurally characterized I-domain is that of the P22 phage, which has a 6-stranded β -barrel fold, a smaller accessory subdomain consisting of three β -strands, and a one and a half-turn α -helix (Rizzo et al. 2014). The P22 I-domain plays critical roles in capsid assembly, stability, viability, and size determination (Rizzo et al. 2014, D'Lima and Teschke 2015). Cryo-electron reconstructions show that the Sf6, P22, and CUS-3 I-domains protrude from the surfaces of T = 7 icosahedral particles (Parent et al. 2014). Interestingly, the morphology of the protrusions in Sf6 are different than in P22 and CUS-3, prompting us to examine if these changes in cryoEM data are due to differences in the Sf6 I-domain structure (Parent et al. 2014). Here we present NMR assignments for the Sf6 I-domain to complement those previously reported for the I-domains of P22 (Rizzo et al. 2014) and CUS-3 (Tripler et al. 2015).

Methods and experiments

Expression and purification

The gene encoding the Sf6 I-domain, corresponding to amino acids 222–345 of the coat protein, was cloned into a *pET21a* plasmid (Novagen, Madison, WI). The presence of the cloned sequence was confirmed with DNA sequencing through Genewiz (South Plainfield, New Jersey). The vector was transformed into *Escherichia coli* BL21 (DE3) cells for protein expression. To make isotopically enriched samples for NMR, cells were grown at 37 °C in minimal medium supplemented with 3 g/L of ^{13}C D-glucose and/or 1 g/L $^{15}\text{NH}_4\text{Cl}$. Samples labeled with ^2H , ^{13}C , and ^{15}N were obtained similarly, except that cells were grown in 99.8 % D_2O at 30 °C, and 3 g/L ^2H , ^{13}C D-glucose was used as the carbon source. Protein expression was induced with a final concentration of 1 mM IPTG at mid-log phase, and cells were grown for a further 16 h at 30 °C. Cells were harvested by centrifugation at 3300g using a Sorvall SH-3000 swinging bucket rotor and resuspended in 20 mM sodium phosphate buffer pH 7.6, containing a 1:100 dilution of Sigma P8849 protease inhibitor, 0.1 % triton, 200 $\mu\text{g/mL}$ lysozyme, 5.8 mM MgSO_4 , 0.58 mM CaCl_2 , and 115 $\mu\text{g/mL}$ deoxyribonuclease and ribonuclease. Cells were lysed using a Misonix sonicator operating at an amplitude of 37, with 15 s pulse and 30 s rest times for a total process time of three minutes. Following sonication, cell debris was removed by centrifugation for 30 min at 38,621g using a F18-12 $\times$ 50 rotor. The supernatant was further centrifuged at 162,635g for 90 min in a Sorvall T-865 rotor to remove cell membranes. The supernatant was then run through a TALON metal affinity column (Clontech Laboratories, Mountain view, CA) at a flow rate of 1 mL/min for I-domain purification via an engineered C-terminal His₆-tag. Fractions containing the Sf6 I-domain were visualized using 15 % SDS-PAGE gel electrophoresis. The fractions were pooled, followed by precipitation of the protein with 0.4 g/mL of ammonium sulfate for 30 min. The precipitated protein was sedimented by centrifugation at 38,621g using a F18-12 $\times$ 50 rotor. The pellet was resuspended in 20 mM sodium phosphate buffer, pH 6.0 and dialyzed three times against the same buffer, followed by concentration using an Amicon Ultra 3 kDa molecular weight cutoff filter (Millipore, Billerica, MA). Final sample concentrations were 2.1 mM for $^{13}\text{C}/^{15}\text{N}$ -labeled protein and 1.5 mM for ^{15}N labeled protein.

In addition to the I-domain of the Sf6 coat protein (amino acids 222–345), our construct contains the C-terminal 8-residue extension LEHHHHHH, an artifact of the affinity tag used for purification.

NMR Spectroscopy

Experiments from the Varian Protein Pack were collected on 600 and 800 MHz NMR spectrometers operating with cryogenic probes. Backbone assignments were made using $3D^1H-^{15}N$ -HSQC, HNCACB, CBCA(CO)NH, HN(CA)CO, and HNCO experiments (Cavanagh et al. 2006). Aliphatic side-chain assignments were made with 3DHCCH-TOCSY, CCH-TOCSY, HBHA(CO)NH, and ^{15}N -TOCSY-HSQC experiments (Cavanagh et al. 2006). To assign aromatic side-chain protons, we collected additional 2D NOESY, DQF-COSY, and TOCSY spectra. Data were processed with the programs FELIX-NMR (San Diego, CA) and NMRpipe. The data were analyzed with CcpN (Vranken et al. 2005) using the platform NMRbox (<http://nmrbox.org/>).

Assignments and data deposition

Based on the quality reports from CcpN Analysis (Vranken et al. 2005), backbone chemical shift assignments were completed to greater than 98 % (Fig. 1). Residues P290 and I291 could not be assigned. Side-chain assignment coverage was 82 % for carbons, 92 % for hydrogens, and 89 % for nitrogens. Chemical shifts are deposited in the Biological Magnetic Resonance Bank (<http://www.bmrb.wisc.edu>) under accession number 26844.

Information on the secondary structure and dynamics of the three I-domains from phages Sf6, CUS-3, and P22 is summarized in Fig. 2. The three I-domains show similar secondary structure elements and we have confirmed that the Sf6 I-domain adopts a 6-stranded β -barrel core fold, like that of P22, based on hydrogen-bonding patterns identified from long-range HNCO data (Cordier and Grzesiek 1999). An important difference for Sf6, is that the D-loop (residues V239 to N254) found in the P22 and CUS-3 I-domains is absent in the Sf6 I-domain. This D-loop forms critical interactions between coat proteins to form proper capsid structures in P22 (D'Lima and Teschke 2015). Its absence in the Sf6 I-domain may account for the differences in morphology observed in cryoEM reconstructions of Sf6 capsids compared to those from P22 and CUS-3 (Parent et al. 2014). In lieu of the D-loop, the Sf6 I-domain appears to have a new disordered loop between strands 5 and 6 of the β -barrel that is not observed in P22 or CUS-3 (Fig. 2). Another difference is that the Sf6 I-domain appears to be missing the accessory sub-domain (consisting of three β -strands and a short α -helix) seen in P22. This conclusion is tentative, however, and must await the complete structure determination of the Sf6 I-domain, since short β -strands could be hard to discern at the current level of structural resolution. The NMR structure of the Sf6 I-domain, the determination of which is under way, will allow us to examine the degree to which structure–function relationships are conserved between the Sf6, CUS-3 and P22 coat protein sequence types within the P22 phage subgroup. We will also ascertain if differences in the cryoEM morphologies of the capsids are linked to differences in the structures of the I-domain coat protein components (Parent et al. 2014).

Acknowledgments

We thank Dr. Mark Maciejewski for help in setting up NMR experiments, Anne R. Kaplan for help in data processing and analysis, and Drs. A.D. Schuyler, M.R. Gryk, and J.C. Hoch for use of the computational platform NMRbox (<http://nmrbox.org/>). This work was supported by NIH Grant R01 GM076661.

References

- Casjens SR, Grose JH. Contributions of P2- and P22-like prophages to understanding the enormous diversity and abundance of tailed bacteriophages. *Virology*. 2016; 496:255–276. [PubMed: 27372181]
- Cavanagh, J., Fairbrother, WJ., Palmer, AG., III, Rance, M., Skelton, NJ. Protein NMR spectroscopy. 2. Elsevier Inc; Amsterdam: 2006. Protein NMR spectroscopy principles and practice.
- Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. *Nucl Acids Res*. 2008; 36:W197–W201. [PubMed: 18463136]
- Cordier F, Grzesiek S. Direct observation of hydrogen bonds in proteins by interresidue 3HJNC' scalar couplings. *J Am Chem Soc*. 1999; 121:1601–1602.
- D'Lima NG, Teschke CM. A molecular staple: d-loops in the I-domain of bacteriophage P22 coat protein make important intercapsomer contacts required for procapsid assembly. *Virology*. 2015; 89:10569–10579.
- Parent KN, Tang J, Cardone G, Gilcrease EB, Janssen ME, Olson NH, Casjens SR, Baker TS. Three-dimensional reconstructions of the bacteriophage CUS-3 virion reveal a conserved coat protein I-domain but a distinct tailspike receptor-binding domain. *Virology*. 2014; 464–465:55–66.
- Rizzo AA, Suhanovsky MM, Baker ML, Fraser LC, Jones LM, Rempel DL, Gross ML, Chiu W, Alexandrescu AT, Teschke CM. Multiple functional roles of the accessory I-domain of bacteriophage P22 coat protein revealed by NMR structure and CryoEM modeling. *Structure*. 2014; 22:830–841. [PubMed: 24836025]
- Shen Y, Bax A. Protein structural information derived from NMR chemical shift with the neural network program TALOS-N. *Methods Mol Biol*. 2015; 1260:17–32. [PubMed: 25502373]
- Tripler TN, Maciejewski MW, Teschke CM, Alexandrescu AT. NMR assignments for the insertion domain of bacteriophage CUS-3 coat protein. *Biomol NMR Assign*. 2015; 9:333–336. [PubMed: 25694158]
- Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*. 2005; 4:687–696.

Fig. 1.
 ^1H - ^{15}N HSQC of the Sf6 I-domain at pH 6.0, 25 °C. Backbone chemical shifts are labeled green and side-chain protons in magenta. Residues are numbered according to their positions in the amino acid sequence of the full-length Sf6 coat protein

Comparison of secondary structure and dynamics in the I-domains of phages **a** Sf6, **b** P22 (Rizzo et al. 2014) and **c** CUS-3 (Tripler et al. 2015). For each protein the following are shown: amino acid sequence, residues protected from hydrogen exchange (indicated by *filled circles*), secondary structure prediction from sequence calculated with the JPred program (Cole et al. 2008), chemical-shift based secondary structure prediction calculated with TALOS-N (except for P22 where the schematic is based on the NMR structure) and prediction of backbone dynamics from TALOS-N (Shen and Bax 2015) where the *green symbols* “x” indicate flexible regions with small predicted S^2 order parameters (except for P22 where the dynamics information is based on ^{15}N -relaxation measurements). The β -strands that make up the core 6-stranded β -barrel fold of the I-domain are shown in magenta while those from the smaller accessory subdomain are shown in orange. For the Sf6 I-domain, the β -strand numbering is based on a long-range HNCO spectrum that identifies residues participating in H-bonds (Cordier and Grzesiek 1999)