doi:10.1006/jmbi.2000.3620 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 297, 1195–1202

JMB

Structure of the Coat Protein-binding Domain of the Scaffolding Protein from a Double-stranded DNA Virus

Yahong Sun[1], Matthew H. Parker[3], Peter Weigele[4], Sherwood Casjens[4] Peter E. Prevelige Jr[3] and N. Rama Krishna[1,2]*

[1]*Comprehensive Cancer Center University of Alabama at Birmingham, Birmingham AL 35294, USA*

[2]*Department of Biochemistry and Molecular Genetics University of Alabama at Birmingham, Birmingham AL 35294, USA*

[3]*Department of Microbiology University of Alabama at Birmingham, Birmingham AL 35294, USA*

[4]*Department of Oncological Sciences, University of Utah Health Science Center, Salt Lake City, UT, 84132, USA*

**Corresponding author*

Scaffolding proteins are required for high fidelity assembly of most high *T* number dsDNA viruses such as the large bacteriophages, and the herpesvirus family. They function by transiently binding and positioning the coat protein subunits during capsid assembly. In both bacteriophage P22 and the herpesviruses the extreme scaffold C terminus is highly charged, is predicted to be an amphipathic α-helix, and is sufficient to bind the coat protein, suggesting a common mode of action. NMR studies show that the coat protein-binding domain of P22 scaffolding protein exhibits a helix-loop-helix motif stabilized by a hydrophobic core. One face of the motif is characterized by a high density of positive charges that could interact with the coat protein through electrostatic interactions. Results from previous studies with a truncation fragment and the observed salt sensitivity of the assembly process are explained by the NMR structure.

© 2000 Academic Press

Keywords: scaffolding protein; NMR; structure; virus assembly

Introduction

The protein components of macromolecular structures are often thought to simply bind to one another spontaneously (by ''self assembly'') once they are properly folded, but in some cases still other proteins are required to catalyze such assembly reactions. The viral scaffolding proteins are the best understood catalysts of protein polymerization (Casjens & Hendrix, 1988; King & Chiu, 1997). These proteins co-assemble with viral coat proteins to form an intermediate structure, called a procapsid, in which the scaffolding protein is found within an external shell of icosahedrally arranged coat protein subunits. In a subsequent step the scaffolding protein molecules are released, leaving a coat protein shell that could not have assembled by itself. Such catalysts of macromolecular assembly may well be required more frequently than is currently suspected, since analysis of the components present in a mature structure gives no hint that it might not have been able to assemble by itself.

The bacteriophage P22 scaffolding protein was the first to be discovered and is the best studied scaffolding protein (King & Casjens, 1974; Prevelige *et al*., 1988). *In vivo*, the absence of scaffolding protein results in incorrect assembly of coat protein into improperly sized and asymmetric aberrant structures and in the failure of other minor capsid proteins to be incorporated into the structure (King & Casjens, 1974; King *et al*., 1973; Bazinet & King, 1988). *In vitro*, scaffolding protein nucleates co-assembly of scaffolding and coat proteins into procapsid-like structures at coat protein concentrations too low to assemble alone (Prevelige *et al*., 1988, 1993). The P22 procapsids contain 250-300 molecules of internal scaffold and a capsid shell built from about 420 molecules of coat protein in a $T = 7$ icosahedral arrangement (King & Casjens, 1974; Eppler *et al*., 1991; Thuman-Commike *et al*., 1996). All the scaffolding protein molecules are released intact at about the time DNA enters the capsid and can be reused in at

Abbreviations used: r.m.s.d., root-mean-square deviation; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; COSY, correlated spectroscopy; TOCSY, total COSY; HSQC, heteronuclear single quantum correlation; HMQC; heteronuclear multiple quantum correlation,.

E-mail address of the corresponding author: NRKRISHNA@bmg.bhs.uab.edu

0022-2836/00/051195–8 $35.00/0

© 2000 Academic Press

least five more rounds of procapsid assembly (King & Casjens, 1974; Casjens & King, 1974); thus the P22 scaffolding protein is a true catalyst of macromolecular assembly. This strategy of procapsid assembly and subsequent packaging of DNA into these structures is used by many dsDNA viruses including the herpesviruses, other tailed phages, and probably the adenoviruses (Casjens, 1997). In some cases, for example bacteriophages T4, lambda, and herpesvirus, the scaffold is proteolytically degraded as it is released from the procapsid and so causes proper coat assembly, but does not function catalytically.

The P22 scaffolding protein is 303 amino acid residues in length and is rich in hydrophilic amino acids (Eppler *et al*., 1991). A structural model in which mainly helical segments are connected by turns and random coil with no buried, hydrophobic core best explains the available biophysical analyses and secondary structure predictions (Tuma *et al*., 1996). In solution, at physiological concentrations, scaffolding protein exists in monomer-dimer-tetramer equilibrium, and the dimer is thought be rod-shaped with an axial ratio of 10:1 (Parker *et al.,* 1997a). The P22 scaffolding protein contains a number of functional domains: residues 1-74 include a post-transcriptional autoregulatory domain (Eppler *et al*., 1991; Casjens & Adams, 1985); residues 1-245 contain the dimerization and tetramerization domains (Parker *et al.,* 1998); residues 200-250 appear to be required for recruiting minor proteins into the procapsid (Greene & King, 1996). The C-terminal region of the scaffolding protein is responsible for binding to coat protein. Removal of the C-terminal 11 residues abolishes coat protein binding whereas a synthetic peptide corresponding to the last 29 amino acid residues shows coat protein binding activity (Parker *et al.,* 1998; P. Weigele, L. Sampson & S. C., unpublished results). Little is known concerning the functional domain structure of any other viral scaffolding proteins, except that herpesvirus scaffolding proteins' coat protein-binding sites are also about 25 amino acid α-helical regions which reside near the C termini of those proteins (Beaudet-Miller *et al*., 1996; Hong *et al*., 1996). The inherent flexibility and tendency to oligomerize has made structural analysis of scaffolding proteins challenging.

We describe here the solution structure of the C-terminal coat protein-binding domain of P22 scaffolding protein. A truncated scaffolding protein consisting of residues 238-303 was cloned, expressed and purified, and its three-dimensional structure was determined by multidimensional NMR methods.

Results

Assembly activity of the truncated scaffolding protein

The 238-303 fragment retains coat protein-binding activity both *in vivo* and *in vitro*. *In vivo*, procapsid-like particles containing 238-303 fragment are assembled when cells expressing this fragment are infected by phage carrying a nonsense mutation in the scaffolding protein gene (P.W., unpublished results). *In vitro*, purified scaffolding protein can enter and bind to empty procapsid shells from which the scaffolding proteins have been removed (Greene & King, 1994). The 238-303 fragment can also bind to procapsid shells (data not shown). *In vitro*, this scaffolding fragment induced rapid polymerization of the coat protein as indicated by an increase in turbidity. Samples of the *in vitro* assembly reaction were sedimented through sucrose gradients, fractionated, and analyzed by SDS-PAGE as described previously (Parker *et al*., 1997b, 1998; Prevelige *et al*., 1988). Some procapsid-like particles were assembled, although the vast majority of assembly products were faster-sedimenting aberrant structures which contained the fragment (data not shown). Taken together, these results indicate that the 238-303 fragment retains the ability to bind to the coat protein in a physiologically relevant manner, despite the removal of nearly 80 % of the protein from the amino-terminal end.

Multimeric state of the scaffolding protein

Oligomerization of the scaffolding protein appears to be part of the control mechanism for procapsid assembly (Parker *et al*., 1997a), and it has been suggested that decreased ability to oligomerize might cause the assembly of aberrant particles. Therefore, we determined whether the residue 238-303 fragment is capable of self-association. Sedimentation equilibrium analytical ultracentrifugation, conducted at pH 7.6, 20 °C, revealed that the 238-303 fragment only weakly self-associates to dimers. The dissociation constant of 4400 (±14) μM is in contrast to the 78 μM (Parker *et al*., 1997a) obtained for dimerization of the full-length scaffolding protein consistent with the suggestion that scaffolding protein oligomerization plays an important role in form determination.

Sequence-specific NMR assignments and secondary structure

The sequence-specific assignments for the uniformly ^{15}N-labeled scaffolding protein C-terminal functional domain (residues 238-303 plus an alanine at position 237 introduced during cloning) have been completed using established procedures that exploit through-space and through-bond correlations (Whthrich, 1986) in H_2O solutions. These correlations were obtained from a combination of 2D-NOESY, 2D-TOCSY, and ^{15}N edited 3D-NOESY-HSQC and 3D-TOCSY-HMQC measurements. The 3D NMR measurements were particularly useful in facilitating the sequential assignments for the N-terminal half of the molecule, since the signals from this region were more crowded. A representative strip plot from the ^{15}N

edited 3D-NOESY-HSQC, showing sequential connectivities from residues 283 to 290, is shown in Figure 1. The HSQC spectrum is supplied as a supplementary Figure, F1 (residues 237 and 238 did not exhibit peaks in the HSQC spectrum, presumably due to considerable disorder at the N terminus). The complete chemical-shift assignments of the scaffolding protein 239-303 fragment have been deposited with the BioMagResBank (http://www.bmrb.wisc.edu, BMRB accession number 4289) (Sun & Krishna, 1999).

The secondary structure elements of this scaffolding protein fragment were established by the unique NOESY contacts and $^3J_{NH}$ couplings (Figure 2). A helix-loop-helix motif is present at the C terminus of the protein and consists of 32 residues (S269-K300). The first α-helix has 4.2 turns (S269-A283) and the second α-helix has 3.3 turns (E289-K300). The loop consists of five residues from A284 to D288. The N-terminal segment from R240 to T267 has a relatively extended structure characterized mainly by long stretches of $d_{\alpha N}(i,i+1)$ contacts, with some weak d_{NN} contacts in the region 254 to 266. It does not show any long-range contacts to the C-terminal half of the molecule. Although this may be caused by the truncation, such a flexible extended region could account for the observation that while the coat protein-binding domain of the scaffolding protein is anchored to icosahedrally symmetrical positions in the procapsid shell, the bulk of the protein does not display icosahedral symmetry.

Discussion

Description of structures

Since the N-terminal 28 residues of the 66 amino acid scaffolding protein fragment have a relatively extended structure with no long-range NOE

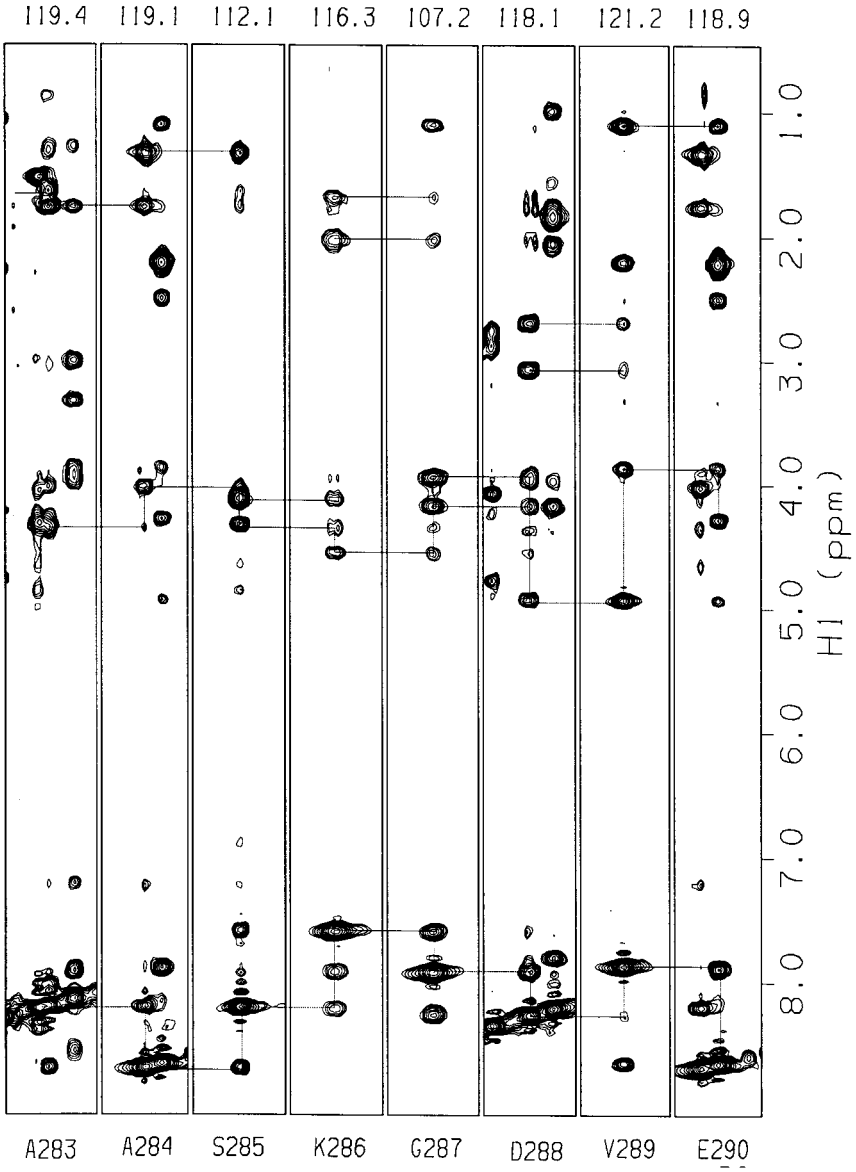

Figure 1. Typical example of strips from the 3D ^{15}N-edited NOESY-HSQC spectrum of the scaffolding protein 238-303 fragment. Sequential connectivities for residues 283-290 are shown. The ^{15}N frequency (ppm) for each plane is shown at the top.

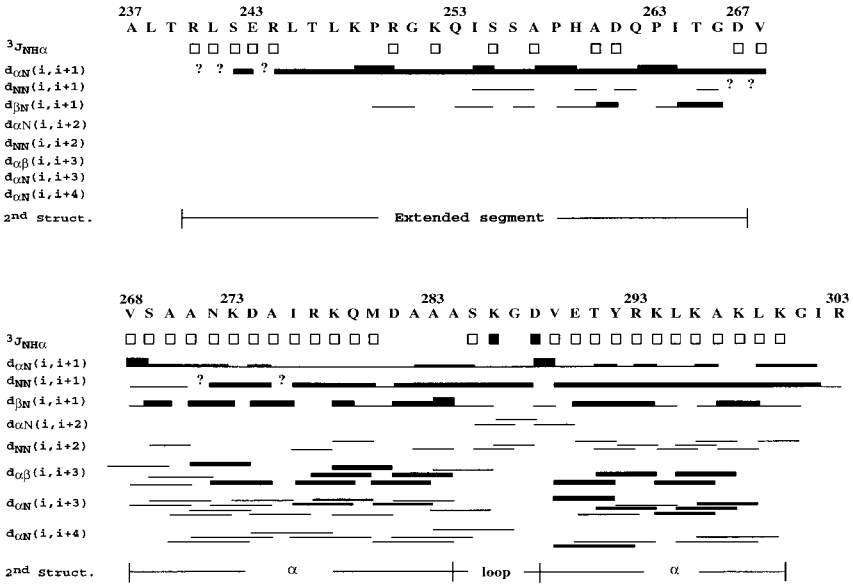

Figure 2. Summary of NMR data for the scaffolding protein 238-303 fragment. NOESY contacts are shown as filled bars. The relative strength of the sequential NOESY contacts is indicated by the thickness of the bars. In the case of proline residues, the δ protons are substituted for the amide protons. Vicinal coupling constants are indicated by (□) for $^3J_{HNa} < 7$ Hz and (■) for $^3J_{HNa} > 8$ Hz. Regions where sequential connectivities could not be confirmed due to resonance overlap are indicated by a ? mark.

contacts to the C-terminal half, only the C-terminal 40 residues were subjected to the structural statistics calculations (Table 1). Figure 3 shows a superposition of the 15 lowest energy structures from the final ensemble of 35 simulated annealing structures. Some selected hydrophobic side-chains are also shown in the Figure. The structures consist of two α-helices that are connected by a short loop of five amino acid residues, with the two helices oriented nearly antiparallel to each other. The five-residue loop causes a near 180° turn in the polypeptide chain. The r.m.s.d. for the 35 backbone structures with respect to the average coordinates is 0.63 Å (S269-K300), 0.70 Å for the first helix (S269-A283), 0.56 Å for the second helix (V289-K300), and 0.55 Å for the loop (S284-D288). A summary of the structural statistics is provided in Table 2.

The loop consists of residues 284 to 288, of which the first four residues form a type I β-turn. The fifth residue in the loop, D288, extends the β-turn allowing the necessary conformational freedom for the helices to associate along their lengths. It is interesting to note that the first residue in the second helix, V289, which is located near the end of the loop, has solvent-exposed methyl groups which could participate in a site-specific hydrophobic interaction between this domain and the coat protein (Figure 3).

The two helices are amphipathic. The hydrophobic residues I276, M280 and A283 on the N-terminal helix and Y292, L295 and L299 on the C-terminal helix are positioned against each other to form a partially buried hydrophobic core between the helices (Figure 3). A rather large number of inter-helical NOE contacts have been observed (supplementary Figure F2). These hydrophobic interactions stabilize the helix-loop-helix motif and orient the two helices. This solution structure is in good agreement with results from structure predictions, Raman spectroscopy, and circular dichroism (Tuma *et al.*, 1998), and explains our earlier observation that the deletion of the 11 amino acids at the carboxyl terminus results in the complete elimination of coat protein-binding activity, and the disordering of approximately 20 additional residues (Tuma *et al.*, 1998).

Table 1. Structural statistics for the family of 35 structures, $\langle SA \rangle_k$

r.m.s. deviations from experimental distance restraints (Å)	
All (371)	0.020
Sequential ($\|i-j\|=1$) (114)	0.007
Short-range ($1<\|i-j\|\leqslant 5$) (108)	0.006
Long-range ($\|i-j\|>5$) (40)	0.004
Intraresidues (81)	0.040
Hydrogen bond (28)	0.009
r.m.s. deviations from experimental dihedral restraints	
Dihedral restraints (31)	0.085
Mean energies	
E_{all} (all) (kcal mol^{-1})	52.00
E_{repel} (kcal mol^{-1})	0.705
E_{cdih} (kcal mol^{-1})	0.021
$E_{L\text{-}J}$ (kcal mol^{-1})[a]	−174.5
Deviations from idealized covalent geometry	
Bonds (Å)	0.001
Angles (deg.)	0.459
Impropers (deg.)	0.360
% Residues (non-Gly and non-Pro) in most favored and additional allowed regions of the Ramachandran plot[b]	100

$\langle SA \rangle_k$ are the final 35 simulated annealing structures. The structural statistics were calculated for residues 264 to 303 forming the helix-loop-helix structure. None of the structures exhibited inter-proton distance violations greater than 0.3 Å and dihedral angle violations greater than 3°. The torsion angle restraints consist of 26 ϕ and five χ_1 angles. No restraints were employed for the angle ψ.

[a] $E_{L\text{-}J}$ is the Lennard-Jones/van der Waals potential calculated using CHARMm empirical energy function.

[b] Ramachandran plot statistics calculated using Procheck.

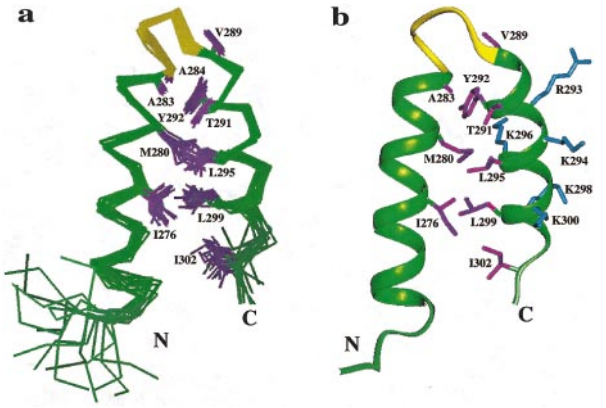

Figure 3. (a) Superposition of the 15 lowest energy structures from the family of 35 distance geometry/simulated annealing structures of the region 264-303 from the C-terminal functional domain of the scaffolding protein. The side-chains for residues constituting the hydrophobic core, and for Val289 at the beginning of the C-terminal helix are shown in purple. The N and C stand, respectively, for the N and C-terminals of the region. (b) Ribbon diagram of the energy-minimized average structure, identifying the location of some of the hydrophobic (purple) and basic (blue) residues.

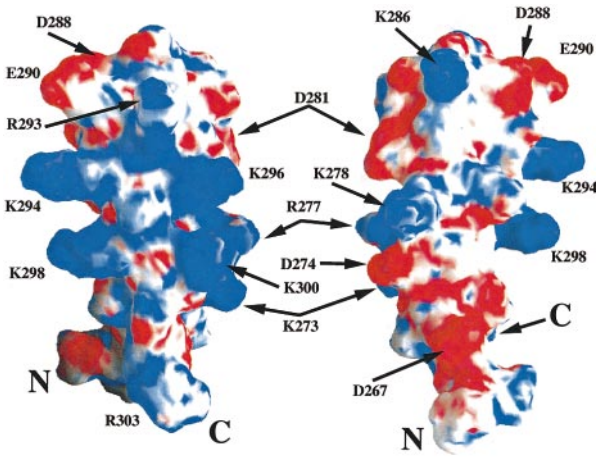

Figure 4. Electrostatic potential surface for residues 264-303 from the C-terminal functional domain of the scaffolding protein. The potentials are shown at ±5 *kT* with positive surfaces in blue and negative surfaces in red. The coat protein-binding domain is viewed from the C-terminal end (a), and from the N-terminal end (b). The charged residues are identified.

Electrostatic potential surfaces

A striking feature of this structure is the high density of charged residues on the outer surface of the C-terminal helix. Figure 4 shows the electrostatic potential surface of the protein obtained from a DELPHI calculation (Nicholls *et al.*, 1991; Honig & Nicholls, 1995). Five basic residues (R293, K294, K296, K298 and K300) are located on the outer face of the 12-residue C-terminal helix, and R303 is also located on this face. Such a positively charged surface might facilitate binding to the coat protein through electrostatic interactions, and may explain the observation that high concentrations of NaCl inhibit scaffolding/coat protein binding (Parker & Prevelige, 1998). The deletion of the C-terminal 11 residues from the scaffolding protein eliminates the positively charged surface, and may also explain the concomitant loss of coat protein binding activity. Even though the N-terminal helix has three positively charged residues (K273, R277 and K278), it also has three negatively charged residues (D267, D274 and D281). The loop consists of two charged residues (K286, D288), and is followed by a solvent-exposed hydrophobic residue, V289, which is the first residue in the second helix.

Electrostatic interactions are somewhat non-specific, especially when they do not involve a single pair of oppositely charged residues. However, this type of interaction, active over long distances, would serve well for the initial recognition and binding events. In order to lock the scaffolding and coat proteins into a specific interaction (or a small set of appropriate interactions), it might be necessary to provide additional, more specific binding surfaces, perhaps through hydrophobic interactions. The solvent-exposed valine at position 289 might function in this way.

Table 2. r.m.s.d to the averaged coordinates (Å)

Residues	Backbone	All heavy atoms
269-283	0.695	1.158
284-288	0.545	1.026
289-300	0.561	1.197
269-300	0.633	1.165

Defined as average r.m.s. difference between the final 35 simulated annealing structures and the average structure. The average structure was computed from the individual structures after best-fit superposition of residues 269-300.

Structural homology

The helix-loop-helix motif found in the C-terminal functional domain is distinct from the helix-turn-helix motif that is found in many DNA- and calcium-binding proteins, in which the two helices are perpendicular. A DALI (Holm & Sander, 1993, 1995) search was performed to identify helix-loop-helix motifs in known proteins that are structurally similar to that observed here in the coat protein-binding domain of the scaffolding protein. Because the coat protein-binding domain may be an isolated structure connected to the rest of the protein by a flexible region, we considered only matching structures significant if they were composed of a single stretch of contiguous amino acid residues. Using these criteria, the scaffolding coat-binding domain displays structural homology to the α-helical zig-zag structures of TPR1 (tetratricopeptide

repeat) domain of the human protein phosphotase PP5 ($Z = 2.8$, r.m.s.d.= 2.1 Å for G266-I302 w.r.t. A21-L57 in PP5) (Das *et al*., 1998) and a similar domain of the clathrin heavy chain ($Z = 2.5$, r.m.s.d.= 2.0 Å for A271-I302 w.r.t. G359-G390 in clathrin) (ter Harar *et al*., 1998). Superposition of the coat protein-binding domain on these structures revealed very good agreement in the disposition of the helices and the loop. For example, residues S269 to K300 of the scaffolding protein show alignment with residues A24 through I55 of the TPR1 domain of PP5 protein with an average r.m.s.d. of 1.40 Å. It has been proposed that both of these elements are involved in modulating transient protein/protein interactions. TPR domains are generally presented in tandem arrays of 3-16 helix-turn-helix motifs which form a scaffold for protein/protein interactions. The α-zig-zag is a repeated array which forms a flexible linker region of the clathrin heavy chain connecting the C-terminal domain to the distal leg and may also serve to bind interacting proteins. Unlike these structures in which the motif is repeated in tandem, the region immediately adjacent to this motif in the coat protein-binding domain of the scaffolding protein appears to be unstructured, and based on previous spectroscopic studies is likely to remain so upon binding. Perhaps the use of a single binding domain allows for the rapid release of the scaffolding protein during DNA packaging.

Conclusions

Here, we have determined the detailed structure of the C-terminal functional domain (coat protein-binding portion) of the phage P22 scaffolding protein by multidimensional NMR spectroscopy. The functional domain structure consists of a highly charged, helix-loop-helix structure that has structural similarity to motifs in several other proteins that participate in transient macromolecular interactions. Although scaffolding protein has only one of these motifs and these other proteins have multiple tandem motifs, it is of interest to note that the scaffolding protein dimer only appears to interact with coat protein when both subunits of the dimer carry this domain. Thus, it is possible that as has been suggested in the other cases, partner proteins bind through interactions with two adjacent helix-loop-helix motifs. Or, perhaps the use of a single binding domain allows for the rapid release of the scaffolding protein during DNA packaging. The fact that this domain may be attached to the rest of the dimeric scaffolding protein by an unstructured ''linker'' suggests that part of the role of scaffolding protein may be simply to bring two coat protein molecules into close proximity, without dictating an exact spatial juxtaposition, but more detailed information on how the scaffolding function operates may have to await additional structural information about the participating proteins.

Materials and Methods

Protein expression and purification

A plasmid encoding residues 238-303 of the P22 scaffolding protein was created by polymerase chain reaction amplification using genomic DNA from P22 strain c1-7, 13⁻amH101 as a template. Two oligonucleotide primers, GACGATGCATATGGCTCTCACTCGACTATCCGAACGC and GTAGAGAGGATCCTTGGAGTGATTGCGGAGATG (*Nde*I and *Bam*HI sites underlined) were used to amplify a fragment that encodes residues 238 through 303 and 165 base-pairs of P22 sequence 3′ of the scaffolding protein termination codon. In addition, these primers created an *Nde*I site immediately 5′ to codon 238 (Leu CTC), as well as a *Bam*HI site 3′ of the scaffolding protein gene. This amplified DNA was cleaved with *Nde*I and *Bam*HI, ligated into similarly cleaved plasmid pET3a (Studier *et al*., 1990) and used to transform $CaCl_2$-treated competent *Escherichia coli* strain NF1829 (Shultz *et al*., 1982). Minilysates of ampicillin-resistant transformants were screened for plasmids with correctly sized DNA inserts.

In addition to encoding residues 238-303, this gene contains a methionine start codon followed by an alanine codon. Electrospray ionization mass spectroscopy of the purified (non-isotopically labeled) protein gave a molecular weight of 7271.5 Da, in good agreement with the predicted value of 7272.4 for the protein with the N-terminal methionine removed (Flinta *et al*., 1986). This was confirmed by six cycles of N-terminal protein sequencing. The uniformly ^{15}N labeled 67-residue protein was expressed in *E. coli* grown in M9 minimal media (Miller, 1992) containing 1 g/l $^{15}NH_4Cl$ as the sole nitrogen source; unlabeled protein was produced using LB medium. Cells were lysed by repeated freeze-thaw cycles, treated with DNase I, and centrifuged as described previously (Prevelige *et al*., 1988). A solution of EDTA to a final concentration of 10 mM was added to the supernatant and the lysate was loaded onto two 5 ml Hi-Trap SP columns (Pharmacia) connected in series. The column was washed with several volumes of buffer B (50 mM Tris, 25 mM NaCl, 2 mM EDTA, pH 7.6), and the protein was eluted with a linear gradient of 100-250 mM NaCl in buffer B. Fractions were dialyzed into buffer C (50 mM sodium acetate, 2 mM EDTA, pH 5.5) and then loaded onto 5 ml High-Trap Q and SP columns connected in series. After washing with buffer C, the Q column was disconnected and the protein was eluted from the SP column with buffer C containing 250 mM NaCl. The protein was concentrated to approximately 2 ml, dialyzed into degassed water, and lyophilized with buffer as described previously (Greene & King, 1994). One liter of culture yielded 8-9 mg of protein of >95 % purity.

Analytical ultracentrifugation

Analytical ultracentrifugation was performed essentially as described previously (Parker *et al*., 1998; Tuma *et al*., 1998), using loading concentrations of 1.3, 2.7, and 5.4 mg/ml, rotor speeds of 40,000, 44,000, and 48,000 revs/min, and detection at 277 nm. Extinction coefficients, measured using the method of Gill & von Hippel (1989) were 0.19 (±0.003) and 0.21 (±0.003) ml mg^{-1} cm^{-1} at 280 and 277 nm, respectively.

NMR spectroscopy

Both the labeled and unlabeled proteins were reconstituted at a concentration of approximately 1.1 mM at pH 4.4 in either 90% H_2O/10% ^2H_2O or 99.9% ^2H_2O solutions. Samples contained the following protease inhibitors (Boehringer Mannheim): antipain (73 μM), aprotinin (0.3 μM), bestatin (13 μM), chymostatin (10 μM), E-64 (28 μM), leupeptin (1 μM), Pefabloc (40 μM), pepstain (1 μM), and phosphoramidon (7 μM), as well as 0.02% (w/v) sodium azide.

NMR measurements were carried out at 20 °C on a Bruker AM600 (600 MHz), a Bruker AVANCE 600 (600 MHz) and a Varian Inova 600 (600 MHz) spectrometer. The NMR experiments performed include 2D-COSY, NOESY, TOCSY, ^{15}N-^1H HSQC, 3D ^{15}N-^1H HMQC-TOCSY and 3D ^{15}N-^1H HSQC-NOESY. In order to resolve some of the overlapping peaks, 2D NOESY experiments were also performed at 10 °C and 30 °C on the AM600 spectrometer.

Structure calculations

The φ angle restraints were obtained from the pH-COSY with $\phi = -55(\pm 45)^\circ$ for $6.0\ \text{Hz} \leqslant {}^3J_{NH} < 7.0$ Hz, and $\phi = -55(\pm 30)^\circ$ for ${}^3J_{NH} < 6.0$ Hz. They were supplemented with ${}^3J_{NH}$ calculated from the linewidth of amide protons in the NOESY and TOCSY spectra (Wang *et al.*, 1997). Four additional φ angle restraints were obtained by this method. In this case, restraints of $\phi = -55(\pm 45)^\circ$ for ${}^3J_{NH} < 6.0$ Hz, and $\phi = -120(\pm 40)^\circ$ for ${}^3J_{NH} > 9.0$ Hz were used. The χ_1 angle restraints, $-60(\pm 60)^\circ$, $60(\pm 60)^\circ$, or $-180(\pm 60)^\circ$, were determined from the ${}^3J_{\alpha\beta}$ and ${}^3J_{\alpha\beta'}$ coupling constants in the DQF-COSY in ^2H_2O, together with NH-H^β, NH-$H^{\beta'}$, H^α-H^β and H^α-$H^{\beta'}$ NOE patterns. No angle restraints were used for the glycine residues.

The 100 ms and 200 ms NOESY experiments were used to determine the NOE restraints. NOE restraints were grouped into four categories, strong (1.8-2.7 Å), medium (1.8B3.3 Å), medium to weak (1.8-4.0 Å), and weak (1.8-5.0 Å), based on the cross-peak intensities at short mixing time (Wüthrich, 1986; Driscoll *et al.*, 1989; Montelione *et al.*, 1992). Pseudo atom corrections were used for the non-stereospecifically assigned atoms (Wüthrich, 1986) and tight restraints were used for methyl groups (Knoning *et al.*, 1990). The structure refinement calculations were performed with the XPLOR/QUANTA/CHARMm package (Molecular Simulations, Inc.). The structures were first calculated without any hydrogen bond restraints. By inspecting the resulting structures, the hydrogen bond restraints were introduced in the final refinement only for those secondary structural region backbone amide protons with obvious acceptor oxygen atoms. The solution NMR structures were generated using a hybrid method, consisting of distance geometry combined with dynamical simulated annealing (Driscoll *et al.*, 1989) with some minor modifications (Lee *et al.*, 1994). A total of 14 hydrogen bond pairs (two restraints per pair) were introduced when supported by NOESY evidence. A total of 100 structures were generated of which 35 final structures with NOE distance violations of less than 0.3 Å and dihedral angle violations of less than 3°, designated as $\langle SA \rangle_k$, were selected. The average structure ($\underline{SA}$) was obtained by taking the averages of the atomic coordinates of these accepted structures. This structure was subjected to further energy minimization to remove bond length and bond angle distortions and to obtain the final energy minimized average structure, $\langle \underline{SA} \rangle_{kr}$.

Data Bank accession numbers

The coordinates for the 35 simulated annealing structures and the restrained energy minimized average structure have been deposited in the RCSB Protein Data Bank (PDB ID codes 1gp8 and 2gp8). The chemical shifts have been deposited with the BioMagResBank (http://www.bmrb.wisc.edu, BMRB accession number 4289).

Acknowledgments

The authors thank Kenneth W. French for excellent technical assistance, and Dr Mike Jablonsky at the NMR Facility for valuable suggestions. This work was supported in part by NIH grant GM-47980, NSF grants MCB-9630775 and MCB-9600574, NCI grant CA-13148 (NMR Facility), and by an NIH training grant T32-AI07150.

References

Bazinet, C. & King, J. (1988). Initiation of P22 procapsid assembly *in vivo*. *J. Mol. Biol.* **202**, 77-86.

Beaudet-Miller, M., Zhang, R., Durkin, J., Gibson, W., Kwong, A. & Hong, Z. (1996). Virus-specific interaction between the human cytomegalovirus major capsid protein and the C terminus of the assembly protein precursor. *J. Virol.* **70**, 8081-8088.

Casjens, S. (1997). Principles of virion structure. In *Structural Biology of Viruses* (Chiu, W., Burnett, R. & Garcea, R. L., eds), pp. 3-37, Oxford Press, New York.

Casjens, S. & Adams, M. (1985). Posttranscriptional modulation of bacteriophage P22 scaffolding protein gene expression. *J Virol.* **53**, 185-191.

Casjens, S. & Hendrix, R. (1988). Control mechanisms in dsDNA bacteriophage assembly. In *The Bacteriophages* (Calendar, R., ed.), vol. 1, pp. 15-91, Plenum Press, New York.

Casjens, S. & King, J. (1974). P22 morphogenesis I: catalytic scaffolding protein in capsid assembly. *J. Supramol. Struct.* **2**, 202-224.

Das, A. K., Cohen, P. W. A. & Burford, D. (1998). The structure of tetratricopeptide repeats of protein phosphatase: implications for TPR-mediated protein-protein interactions. *EMBO J.* **17**, 1192-1199.

Driscoll, P. C., Gronenborn, A. M., Beress, L. & Clore, G. M. (1989). Determination of the three-dimensional solution structure of the antihypertensive and antiviral protein BDS-I from the sea anemone *Anemonia sulcata*: a study using nuclear magnetic resonance and hybrid distance geometry-dynamic simulated annealing. *Biochemistry,* **28**, 2188-2198.

Eppler, K., Wyckoff, E., Goates, J., Parr, R. & Casjens, S. (1991). Nucleotide sequence of the bacteriophage P22 genes required for DNA package. *Virology,* **183**, 519-538.

Flinta, C., Persson, B., Jornvall, H. & von Heijne, G. (1986). Sequence determinants of cytosolic N-terminal protein processing. *Eur. J. Biochem.* **154**, 193-196.

Gill, S. & von Hippel, P. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319-326.

Greene, B. & King, J. (1994). Binding of scaffolding subunits within the P22 procapsid lattice. *Virology,* **205**, 188-197.

Greene, B. & King, J. (1996). Scaffolding mutants identifying domains required for P22 procapsid assembly and maturation. *Virology,* **224**, 82-96.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.

Holm, L. & Sander, C. (1995). DALI: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478-480.

Hong, Z., Beaudet-Miller, M., Durkin, J., Zhang, R. & Kwong, A. D. (1996). Identification of a minimal hydrophobic domain in the herpes simplex virus type 1 scaffolding protein which is required for interaction with the major capsid protein. *J. Virol.* **70**, 533-540.

Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science,* **268**, 1144-1149.

King, J. & Casjens, S. (1974). Catalytic head assembling protein in virus morphogenesis. *Nature,* **251**, 112-119.

King, J. & Chiu, W. (1997). The procapsid to capsid transition in double-stranded DNA bacteriophages. In *Structural Biology of Virus* (Chiu, W., Burnett, R. M. & Garcea, R. L., eds), pp. 288-311, Oxford University Press, New York.

King, J., Lenk, E. V. & Botstein, D. (1973). Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. II. Morphogenetic pathway. *J. Mol. Biol.* **80**, 697-731.

Knoning, T., Boelens, R. & Kaptein, R. (1990). Calculation of the nuclear Overhauser effect and the determination of proton-proton distances in the presence of internal motions. *J. Magn. Reson.* **90**, 111-123.

Lee, W., Moore, C. H., Watt, D. D. & Krishna, N. R. (1994). Solution structure of the variant-3 neurotoxin from *Centruroides sculpturatus*. Ewing. *Eur. J. Biochem.* **218**, 89-95.

Miller, J. (1992). *A Short Course in Bacterial Genetics*, Cold Spring Harbor Laboratory Press, Plainview, NY.

Montelione, G. T., Wüthrich, K., Burgess, A. W., Nice, E. C., Wagner, G., Gibson, K. D. & Scheraga, H. (1992). Solution structure of murine epidermal growth factor determined by NMR spectroscopy and refined by energy minimization with restrains. *Biochemistry,* **31**, 236-249.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.

Parker, M. H. & Prevelige, P. E., Jr (1998). Electrostatic interactions drive scaffolding/coat protein binding and procapsid maturation in bacteriophage P22. *Virology,* **250**, 337-349.

Parker, M. H., Stafford, W. F., III & Prevelige, P., Jr (1997a). Bacteriophage P22 scaffolding protein forms oligomers in solution. *J. Mol. Biol.* **268**, 655-665.

Parker, M. H., Jablonsky, M., Casjens, S., Sampson, L., Krishna, N. R. & Prevelige, P. E., Jr (1997b). Cloning, purification, and preliminary characterization by circular dichroism and NMR of a carboxyl-terminal domain of the bacteriophage P22 scaffolding protein. *Protein Sci.* **6**, 1583-86.

Parker, M. H., Casjens, S. & Prevelige, P. E., Jr (1998). Functional domains of bacteriophage P22 scaffolding protein. *J. Mol. Biol.* **281**, 69-79.

Prevelige, P. E., Jr, Thomas, D. & King, J. (1988). Scaffolding protein regulates the polymerization of P22 coat subunit into icosahedral shells *in vitro*. *J. Mol. Biol.* **202**, 743-757.

Prevelige, P. E., Jr, Thomas, D. & King, J. (1993). Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys. J.* **64**, 824-835.

Shultz, J., Silhavy, T. J., Berman, M. L., Fiil, N. & Emr, S. D. (1982). A previously unidentified gene in the spc operon of *Escherichia coli* K12 specifies a component of the protein export machinery. *Cell,* **31**, 227-235.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60-89.

Sun, Y. & Krishna, N. R. (1999). ^{1}H and ^{15}N chemical shift assignments of a carboxyl-terminal function domain of the bacteriophage P22 scaffolding protein. *Magon. Reson. Chem.* **37**, 602-604.

ter Haar, E., Mustachio, A., Harrison, S. C. & Kirchhausen, T. (1998). Atomic structure of clathrin: a β propeller terminal domain joins an a zigzag linker. *Cell,* 95563-59573.

Thuman-Commike, P. A., Greene, B., Jakana, J., Prasad, B. V., King, J., Prevelige, P. E., Jr & Chiu, W. (1996). Three-dimensional structure of scaffolding-containing phage P22 procapsids by electron cryomicroscopy. *J. Mol. Biol.* **260**, 85-98.

Tuma, R., Prevelige, P. E., Jr & Thomas, G. J., Jr (1996). Structural transitions in the scaffolding and coat proteins of P22 virus during assembly and disassembly. *Biochemistry,* **35**, 4619-4627.

Tuma, R., Parker, M. H., Weigele, P., Sampson, L., Sun, Y., Krishna, N. R., Casjens, S., Thomas, G. J., Jr & Prevelige, P. E., Jr (1998). A helical coat protein recognition domain of the bacteriophage P22 scaffolding protein. *J. Mol. Biol.* **281**, 81-94.

Wang, Y., Nip, A. M. & Wishart, D. S. (1997). A simple method to quantitatively measure polypeptide $J_{HNH\alpha}$ coupling constants from TOCSY or NOESY spectra. *J. Biomol. NMR,* **10**, 373-382.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, J. Wiley, New York.

Edited by M. Summers

(Received 15 December 1999; accepted 15 February 2000)

http://www.academicpress.com/jmb

Supplementary material for this paper, comprising two figures, is available from JMB Online.