A Docking Model Based on Mass Spectrometric and Biochemical Data Describes Phage Packaging Motor Incorporation*[S]

Chi-yu Fu‡, Charlotte Uetrecht§¶, Sebyung Kang‡, Marc C. Morais‖, Albert J. R. Heck§¶, Mark R. Walter‡, and Peter E. Prevelige, Jr.‡**

The molecular mechanism of scaffolding protein-mediated incorporation of one and only one DNA packaging motor/connector dodecamer at a unique vertex during lambdoid phage assembly has remained elusive because of the lack of structural information on how the connector and scaffolding proteins interact. We assembled and characterized a ϕ29 connector-scaffolding complex, which can be incorporated into procapsids during *in vitro* assembly. Native mass spectrometry revealed that the connector binds at most 12 scaffolding molecules, likely organized as six dimers. A data-driven docking model, using input from chemical cross-linking and mutagenesis data, suggested an interaction between the scaffolding protein and the exterior of the wide domain of the connector dodecamer. The connector binding region of the scaffolding protein lies upstream of the capsid binding region located at the C terminus. This arrangement allows the C terminus of scaffolding protein within the complex to both recruit capsid subunits and mediate the incorporation of the single connector vertex. ***Molecular & Cellular Proteomics 9:1764–1773, 2010.***

The DNA packaging motor of double-stranded DNA bacteriophages translocates genomic DNA into a preformed procapsid to near crystalline density and is the strongest motor characterized to date. The packaging motor of the *Bacillus subtilis* phage ϕ29 can work against 57 piconewtons of internal force and translocate 2 bp of DNA per ATP hydrolyzed at a maximum velocity of 103 bp/s (1, 2). The motor complex is assembled on a dodecamer of the connector protein, which replaces a pentameric vertex in the procapsid and serves both as a portal for DNA passage and the docking site for the other packaging components (3).

From the ‡Deptartment of Microbiology, University of Alabama, Birmingham, Alabama 35294, §Biomolecular Mass Spectrometry and Proteomics Group, Utrecht Institute for Pharmaceutical Sciences and Bijvoet Center for Biomolecular Research, Utrecht University and Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands, and ‖Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, Texas 77555

Received, December 22, 2009, and in revised form, January 27, 2010

Published, MCP Papers in Press, February 2, 2010, DOI 10.1074/mcp.M900625-MCP200

To successfully package a full-length genome, incorporation of one and only one connector vertex is essential (4). *In vivo*, nearly every assembled procapsid has one and only one connector vertex and is able to package DNA and mature into an infectious phage (5). This narrow distribution in which 95% of particles have a single connector vertex cannot be explained by random statistical incorporation. The control mechanism is coupled to the procapsid assembly process. Procapsid assembly requires the copolymerization of hundreds of copies each of the capsid and scaffolding proteins as well as a dodecamer of the portal or connector protein. The scaffolding protein acts to both activate the coat protein for assembly and ensure proper form determination. In the absence of scaffolding protein, uncontrolled polymerization results in the assembly of aberrant structures. In a properly assembled procapsid, the portal protein is located at one vertex, whereas scaffolding protein occupies the bulk of the interior space and is subsequently removed during DNA packaging by either proteolysis or simple release. Mutational studies have indicated that scaffolding protein is involved either directly or indirectly in the incorporation of the connector vertex during procapsid assembly in a variety of phages (6–8).

In ϕ29, the connector vertex is specifically incorporated at one of the two 5-fold vertices lying on the long axis of a prolate procapsid composed of 235 copies of capsid protein and containing ~180 copies of scaffolding protein (9, 10). The structure of the 33-kDa connector protein subunit consists of three long central α-helices bridging wide and narrow domains that are rich in β-sheets and extended polypeptides (Fig. 1*A*) (10–12). The 12 subunits are arranged to form a 75-Å-long tapered grommet-shaped structure with an external diameter of 69 Å at the wide end and 33 Å at the narrow end. By fitting the crystal structure of the connector dodecamer into the cryo-EM[1] density of the procapsid, the orientation of connector at the unique vertex of the procapsid was revealed. The wide domain of connector protein lies inside the procapsid, and the narrow domain is exposed to the exterior and makes contacts with the other parts of the motor complex (11). The 11-kDa scaffolding protein subunits

[1] The abbreviations used are: EM, electron microscopy; DST, disuccinimidyl tartrate.

© 2010 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at http://www.mcponline.org

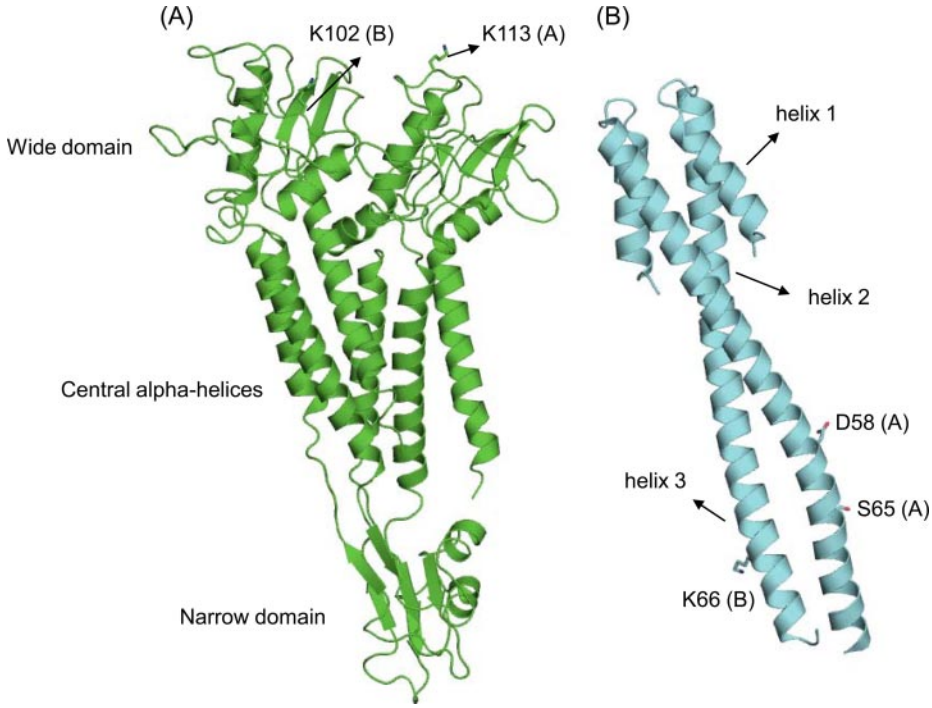

FIG. 1. **The x-ray crystal structures of connector protein (Protein Data Bank code 1FOU, chains A and B) (*A*) and scaffolding protein (Protein Data Bank code 1NO4, chains A and B) (*B*).**

form nanomolar affinity homodimers resembling arrows in solution. Each subunit contributes one side of the arrowhead and one-half of the long coiled coil shaft (Fig. 1*B*) (13). The subunit structure consists of three helical segments. A three-turn N-terminal helix (α1) followed by a five-residue loop, and an antiparallel five-turn helix (α2) makes up the arrowhead and part of the proximal part of the shaft. A three-residue loop and a seven-turn helix (α3) complete the shaft. The C-terminal 15 residues, which interact with capsid protein as determined in the *in vitro* assembly assay, are disordered in the crystal structure (14).

We have recently reported the development of an *in vitro* assembly system for phage ϕ29 in which purified connector protein complex can be successfully incorporated (15). The addition of connector protein dodecamers to coat and scaffolding subunits accelerated the rate of assembly and lowered the critical concentration, suggesting involvement in nucleation of assembly (15). Here we used native mass spectrometry, chemical cross-linking, and mutational analysis to characterize the interactions between the connector and the scaffolding proteins and develop a model of the scaffolding-connector complex, which provides a molecular model of how scaffolding protein might mediate stringent incorporation of one and only one connector dodecamer.

MATERIALS AND METHODS

Protein Preparations—The scaffolding and connector proteins were purified as described previously (14, 16). The scaffolding deletion mutants Δ79–98, Δ74–98, Δ70–98, and Δ66–97 were constructed by introducing a stop codon to replace residue Glu-79, Thr-74, Gln-70, and Lys-66, respectively, using the QuikChange site-directed mutagenesis kit (Stratagene) with primers 5′-gacagacaaacaggaataagatcacaagaaagctgatattagtg-3′ for Δ79–98, 5′-gcttttcagacagattggtttgtgagacaaacaggaagaagatc-3′ for Δ74–98, 5′-caaatagtaagcttttcagatagattggtttgacagacaaacagg-3′ for Δ70–98, and 5′-gatctgatcgtgtcaaatagttagcttttcagacagattgg-3′ for Δ66–97. The scaffolding mutant D58K was constructed using the QuikChange site-directed mutagenesis kit with primer 5′-gaaaagttagccgctgaaaaaaaggatctgatcgtgtcaaatagtaagc-3′. The nucleotide substitutions producing mutations are underlined.

In Vitro Solubilization Assay of Connector/Scaffolding Interactions—The *in vitro* assay was performed as described previously with some modifications (17). The connector protein at 1 mg/ml (29 μM) was mixed with scaffolding protein at either 2- or 5-fold molar excess. The mixture was dialyzed against buffer containing 89 mM Tris borate, pH 8.3 and 2.5 mM EDTA for 30 min at room temperature and subsequently against buffer containing 50 mM Tris, pH 7.8, 10 mM $MgCl_2$, and 100 mM NaCl for 1 h. The connector protein in the insoluble fraction was obtained by pelleting the sample at 13,000 rpm with a tabletop centrifugation for 15 min. The total amounts of protein in the input and in the pellet were analyzed by SDS-PAGE and quantified using a Bio-Rad gel documentation system.

Native Mass Spectrometry—The instrument used was a modified QTOF 1 (Waters) (18) operated in positive ion mode. To achieve optimal resolution the backing pressure was increased to 10 millibars. The capillary and cone voltages were 1500 and 150 V, respectively. The pressure in the collision cell was 1.5×10^{-2} millibar with xenon as the collision gas where the acceleration voltage was 50 V (19). No fragmentation was observed at this accelerating voltage.

Capillaries for electrospray ionization were prepared in house from borosilicate glass tubes of 1.2-mm outer diameter and 0.68-mm inner diameter with filament (World Precision Instruments, Sarasota, FL) using a P-97 micropipette puller (Sutter Instruments, Novato, CA) and gold-coated using an Edwards Scancoat Six Pirani 501 sputter coater (Edwards Laboratories, Milpitas, CA). Capillary tips were opened on the sample cone of the instrument.

For 2- and 10-fold molar excesses of scaffolding protein, 30 μM connector protein was mixed with 60 or 290 μM scaffolding protein. Samples were dialyzed against 500 mM ammonium acetate, pH 7.8

for 4 h at room temperature. Pure connector protein was prepared in the same way. For 25-fold molar excess, 10 μM connector was mixed with 250 μM scaffolding protein. Buffer exchange to 500 mM ammonium acetate, pH 7.8 was carried out with a centrifugal filter device at 4 °C (5-kDa cutoff; Millipore, Amsterdam, The Netherlands). Native mass spectrometry was performed at a final concentration of ~10 μM connector protein (based on the monomer).

Chemical Cross-linking and In-gel Digestion—29 μM connector protein and 145 μM scaffolding protein were mixed and dialyzed against buffer containing 20 mM sodium phosphate, 10 mM $MgCl_2$, and 100 mM NaCl, pH 7.8 for 3 h at room temperature. After spinning at 13,000 rpm for 15 min with a tabletop centrifuge, the supernatant containing connector-scaffolding protein complexes was chemically cross-linked with the lysine-reactive cross-linker disuccinimidyl tartrate (DST) (Pierce) at a final concentration of 500 μM. The remaining reactive groups were quenched with 50 mM Tris, pH 7.5 after 5 min of reaction.

The cross-linked species were separated by 15% SDS-PAGE. The bands of interest were excised, destained, and in-gel digested with sequencing grade trypsin (Roche Applied Science) at 37 °C for 16 h followed by quenching with 1% formic acid (20).

Detection of Cross-link Signature Peptides by Mass Spectrometry—The mass spectra were acquired with a hybrid ion trap FT-ICR mass spectrometer (LTQ-FT, Thermo Finnigan, San Jose, CA) equipped with a 7-tesla magnet. The tryptic peptides digested in gel were separated by a Magic C_{18} reverse phase column (0.2 × 50 mm, 200 Å) (Michrom Bioresources, Inc. Auburn, CA) using a linear gradient from 0 to 95% acetonitrile containing 0.1% formic acid for 1 h at a 5 μl/min flow rate. The automated data-dependent MS/MS was performed in the linear ion trap with a collision energy of 35 V. The five most intense precursor ions in each FT scan were selected and subjected to consecutive MS/MS with 3-min exclusion duration. The spectra were processed and analyzed with Qual Browser and Bioworks 3.2 (Thermo Finnigan).

Computational Docking—The crystal structures of connector protein (Protein Data Bank code 1FOU, chains A and B) and scaffolding protein (Protein Data Bank code 1NO4, chains A and B) were used to generate a structural model of connector-scaffolding complexes with the rigid body docking ZDOCK algorithm (21). The entire dimeric scaffolding molecule with a 15° rotational sampling interval was docked to the dimer of dodecameric connector protein with the interior surface blocked. The models were scored by the pairwise shape complementarity, desolvation, and electrostatic energies. The 2000 decoy models were first filtered with a maximal distance constraint of 8 Å between the NZ atoms of connector protein residue Lys-102 and scaffolding protein residue Lys-66. For the remaining complexes, it was required that the axis running from N to C terminus of the scaffolding protein align with the axis from wide to narrow ends of the connector protein to conform to the cross-linking constraints afforded by the cross-link of scaffolding Lys-83 to connector Lys-4 or Lys-19.

RESULTS

Characterization of Connector-Scaffolding Complexes—The *in vitro* assembly experiments showed that premixing the connector and scaffolding proteins resulted in slightly more efficient connector incorporation than simultaneous addition, thereby implying the formation of a complex between them (data not shown).

With a mild procedure, using ESI from buffered aqueous solutions, non-covalent interactions in proteins and protein complexes can be preserved in the gas phase (22–24). Here this so-called native mass spectrometry approach was used to detect the formation of connector-scaffolding complexes and determine the stoichiometry of interactions. In ESI, the analytes become multicharged, leading to a series of ion signals for each species. From adjacent signals in a charge state distribution, the mass can be obtained; in practice, all ion signals are used for a more accurate determination. Recent developments in Q-TOF mass spectrometers (18, 25) have enabled the analyses of species with masses over 1 million Da (16, 23, 26, 27). The connector protein was analyzed previously using native mass spectrometry (16) and found to be dodecameric. Native mass spectrometry of the present connector preparation yielded equivalent results (data not shown). Next, connector-scaffolding complexes formed at various connector/scaffolding ratios were mass-analyzed. The ESI mass spectra of the connector-scaffolding complexes displayed several overlapping charge distributions, which could, however, all be assigned to the dodecameric connector with zero up to 12 scaffolding protein molecules bound (Fig. 2). Only dodecameric connector complexes with an even number of bound scaffolding protein molecules (0, 2, 4, etc.) were detected, suggesting that scaffolding protein binds to the connector as a dimer.

The relative abundance of each species is shown in Fig. 2*D* at various protein ratios. At a 2-fold excess of scaffolding to connector protein (based on the monomeric concentrations), the free connector and connector with two scaffolding molecules bound are the most abundant species. The broad distribution of stoichiometries observed at a 10-fold excess of scaffolding protein indicated that there is little or no cooperativity in the binding. With a 25-fold excess of scaffolding protein, connectors with five or six dimers bound became the dominant species. It is noteworthy that no complexes with more than six bound scaffolding dimers were observed.

To estimate the scaffolding binding affinity, the theoretical distribution of bound scaffolding dimers was calculated for K_D ranging from 25 to 500 μM and compared with the experimental data. Based on the native mass spectrometry data, it was assumed that dimeric scaffolding protein bound non-cooperatively (binomial distribution) and that there was an upper limit of six binding sites on the connector. The distribution calculated using a K_D of 25 μM agrees reasonably well with experimental data and suggests that the scaffolding/connector dissociation constant is in the low μM range (Fig. 2*E* and supplemental Fig. S1).

Identification of Connector Binding Region of Scaffolding Protein—Although the connector-scaffolding complexes were detected by native mass spectrometry, cryo-EM reconstructions of the connector in the mixture with a 25-fold excess of scaffolding protein did not reveal extra density corresponding to the scaffolding protein (data not shown). This may be attributable to the low molecular weight and dynamic nature of the scaffolding molecule. To determine the region of scaffolding protein that interacts with the connector protein, a series of increasingly large C-terminal scaffolding deletion

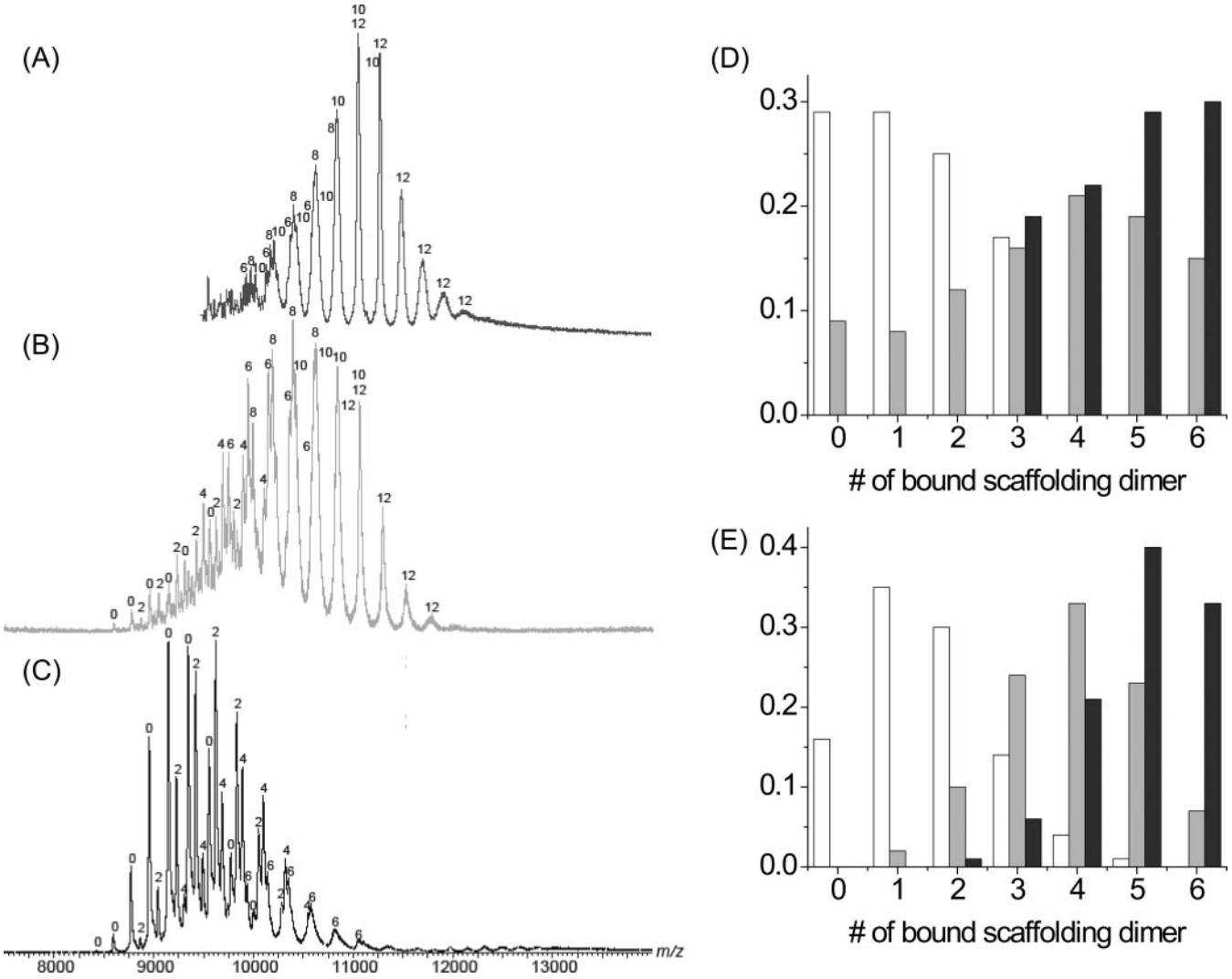

FIG. 2. **Native mass spectra of connector-scaffolding complexes.** The mass spectra of connector-scaffolding complexes formed at 25- (*A*), 10- (*B*), and 2-fold (*C*) molar excess of scaffolding protein are shown. The number of scaffolding protein molecules present in the assigned connector-scaffolding complexes is indicated *above* the peaks in the mass spectra. Bar graphs of the relative abundance of the connector-scaffolding species at 2-, 10-, and 25-fold molar excess of scaffolding protein (*white*, *light gray*, and *dark gray*, respectively) determined experimentally (*D*) and calculated assuming a K_D of 25 μM (*E*) are shown. Complexes with more than 12 scaffolding proteins bound to a connector dodecamer were not detected. Peaks around m/z = 12,000 and 13,000 in *C* correspond to dimers of the connector dodecamer.

mutants were prepared and tested in an *in vitro* solubilization assay in which scaffolding protein prevents connector protein aggregation. In the absence of scaffolding protein, the connector protein could be completely pelleted from a 100 mM NaCl solution with a tabletop centrifuge (Fig. 3*A*). However, in the presence of a 2-fold excess of scaffolding protein, most of the connector protein remained soluble. Mutants lacking residues 74 to the C terminus (Δ74–98) still solubilized connector protein as efficiently as the wild type scaffolding protein did. However, deleting residues 70 to the C terminus (Δ70–98) resulted in loss of solubilization activity even at a 5-fold molar excess. The data suggested that residues between 74 and the C terminus are dispensable for interaction but that deletion of an additional four residues (70–74) blocks complex formation.

A temperature-sensitive scaffolding mutation, S65N, has been isolated *in vivo*. Phage carrying this mutation produce normal prolate capsids at permissive temperature (30 °C) but produce isometric capsids with a reduced amount of incorporated connector protein at a non-permissive temperature (42 °C) (9, 10, 28). To investigate the possibility that this phenotype arises directly from an altered interaction with the connector protein, recombinant S65N scaffolding protein was tested in the *in vitro* solubilization assay. Relative to wild type, the S65N scaffolding protein displayed a 60% reduction in solubilization activity at the permissive temperatures of 20 and 30 °C (Fig. 3*B*). When the temperature was raised to 42 °C, S65N lost solubilization activity, whereas the wild type protein did not. The data indicated that connector binding

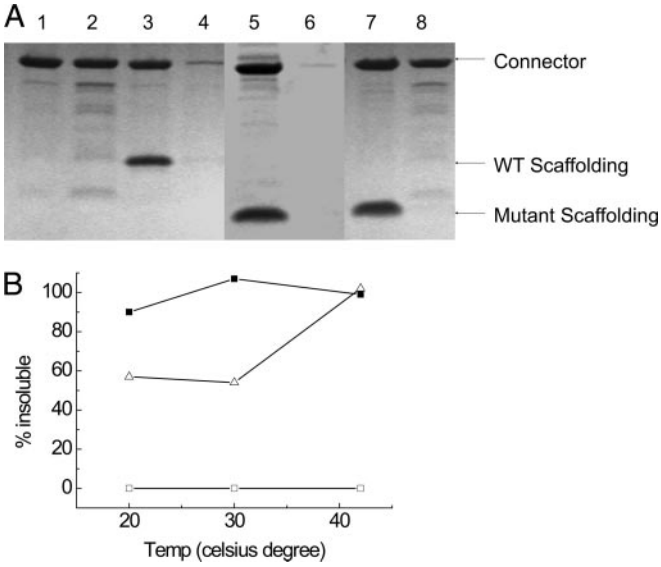

FIG. 3. **Identification of connector binding region of scaffolding protein.** *A*, the ability of scaffolding deletion mutants to solubilize connector protein at low ionic strength buffer was assayed by SDS-PAGE following centrifugation. The amounts of input and insoluble protein are shown in the *odd* and *even lanes*, respectively. *Lanes 1* and *2*, connector protein only; *lanes 3* and *4*, connector plus wild type (*WT*) scaffolding protein; *lanes 5* and *6*, connector plus Δ74–98 scaffolding protein; *lanes 7* and *8*, connector plus Δ70–98 scaffolding protein. *B*, the effect of temperature on the ability of wild type and temperature-sensitive mutant scaffolding protein to solubilize connector protein. The fraction of connector solubilized following incubation at 20, 30, and 42 °C was quantified by centrifugation and SDS-PAGE for connector protein alone (*closed squares*) or connector in the presence of wild type (*open squares*) or S65N scaffolding protein (*open triangles*).

regions likely span at least 10 residues of the scaffolding molecule (residues 65–74) and are located upstream of the capsid binding region, which was previously localized to the crystallographically disordered C terminus (14). The correlation between the *in vivo* and *in vitro* phenotypes of the S65N scaffolding protein suggested that the connector/scaffolding interactions formed in the complexes *in vitro* recapitulate those involved in recruiting connector protein *in vivo*.

Identification of Connector/Scaffolding Interfaces by Chemical Cross-linking Mass Spectrometry—Next, to locate the region of connector protein in contact with scaffolding protein, the complexes were chemically cross-linked with the 6.4-Å lysine-reactive cross-linker DST. To minimize the degree of cross-linking and avoid perturbing the structure, mild cross-linking conditions were used under which ~80% of connector protein and 90% of scaffolding protein remained monomeric. Species containing intramolecular cross-links and homo- or heterointermolecular cross-links were separated by SDS-PAGE. By comparing the banding pattern of the cross-linked complexes with that of the connector proteins alone, species uniquely present in the complex could be identified (Fig. 4*A*).

A strong band, unique to the cross-linked connector-scaffolding complex, that migrated between monomeric and dimeric connector proteins (35 and 70 kDa) was observed. This species was considered likely to be a covalent 1:1 complex of connector and scaffolding proteins (calculated molecular mass of 46 kDa) and appeared to be a doublet, a faster migrating cross-linked band (CS1) and a slightly more slowly migrating cross-linked band of about 5-fold less intensity (CS2). As this migration difference could reflect branching differences arising from different cross-linking locations, both bands were excised and digested in gel with trypsin. The extracted peptides were separated by reverse phase liquid chromatography and analyzed using a hybrid ion trap FT-ICR mass spectrometer. Peptides covering almost the entire sequence space of scaffolding and connector proteins were seen in the digestion of both complex-specific bands, confirming that they represent heteromolecular cross-links. The ion profiles of the digested cross-linked complex were compared with those of uncross-linked connector and scaffolding proteins to identify unique peptides generated by cross-linking. Modification of the lysine side chain by the cross-linker prevents tryptic cleavage at that residue. Therefore, the masses of the combinations of two peptides (each one with at least one missed cleavage site) plus the cross-linker were calculated and compared with the observed masses of peptides unique to the cross-linked samples.

A unique quintuply charged ion (m/z = 660.7504) was identified in the CS1 band. This ion had a mass close to that predicted for scaffolding residues 53–68 cross-linked to connector residues 94–105. The simulated mass spectrum of the predicted match (based on the chemical composition) agreed with the observed mass spectrum within the mass accuracy of the instrument (~2 ppm). The parent mass-based assignment of this ion and the location of the cross-linked residues were confirmed by MS/MS analysis. Following CID fragmentation, six b-ions from scaffolding residues 53–68, one b-ion from connector residue 94–105, and three ions with DST cross-linker were assigned (29) in the fragmentation spectra (Fig. 4*B*). The MS/MS data confirmed that the scaffolding Lys-66 was cross-linked to the connector Lys-102. This cross-linked peptide was repeatedly observed in different sample preparations, suggesting that the interaction between connector and scaffolding proteins in this region is specific and most favorable.

A unique triply charged ion (m/z = 731.3822) was identified in the CS2 band. The calculated mass of residues 83–98 of scaffolding protein cross-linked to either residues 4 and 5 or residues 19 and 20 of connector protein was m/z = 731.3813 and in excellent agreement (1.2 ppm) with the observed monoisotopic mass. The MS/MS spectra of this ion produced a series of b-ions containing cross-linker and y-ions that confirmed that scaffolding residue Lys-83 was cross-linked to the N-terminal region of the connector protein at either residue Lys-4 or residue Lys-19 (Fig. 4*C*). Because of the short length

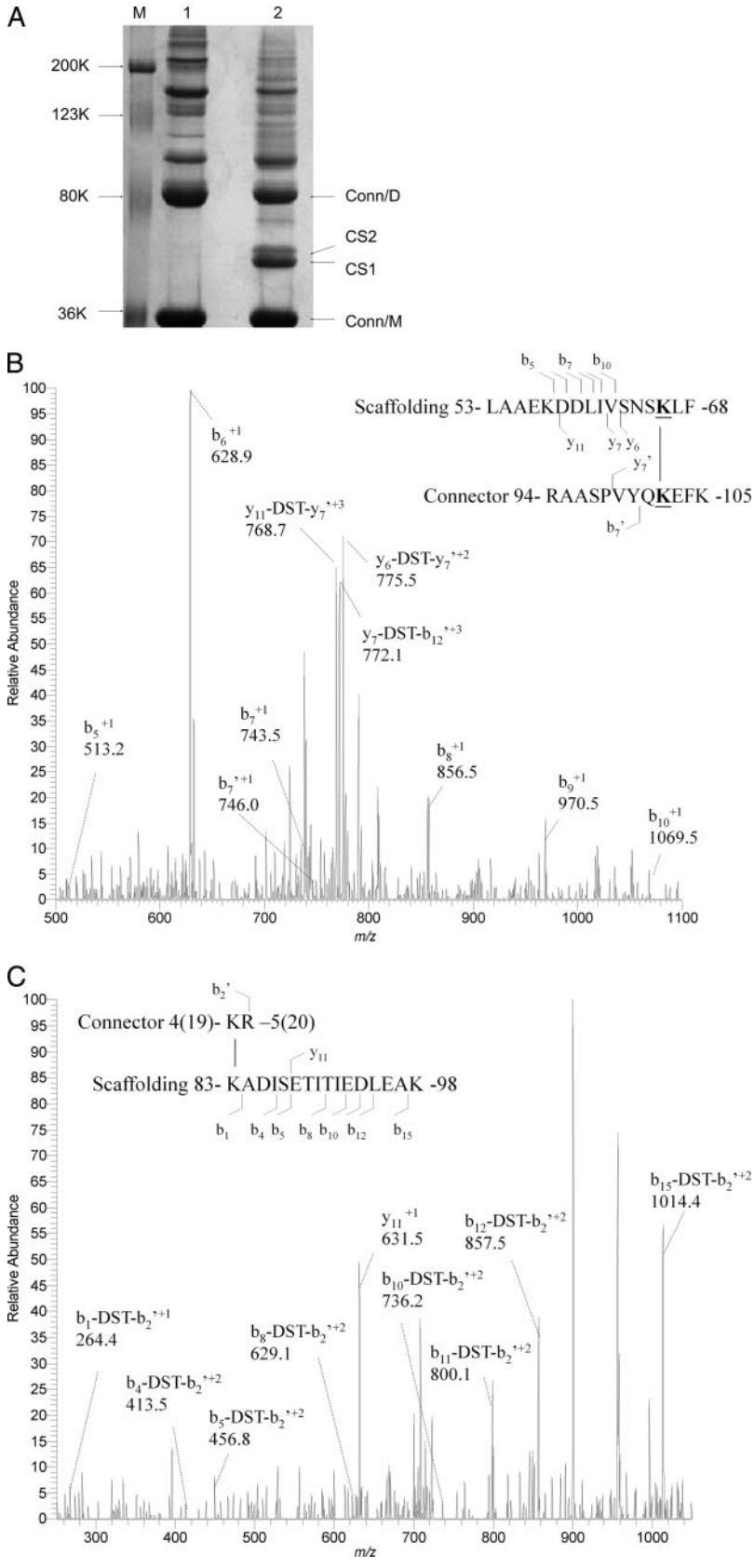

FIG. 4. **Identification of cross-linked peptides derived from connector-scaffolding complex.** *A*, SDS-PAGE separation of cross-linked connector protein (*lane 1*) and connector-scaffolding complexes (*lane 2*). Connector monomer (*Conn/M*), connector dimer (*Conn/D*), and the connector-scaffolding species (*CS1* and *CS2*) resulting from cross-linking are indicated with *arrows*. Molecular weight markers are in *lane M*. *B*, the MS/MS spectrum of the cross-linked peptide $m/z = 660.7504$ (5+ charge). The assigned b-ions and y-ions generated by fragmentation of the $m/z = 660.8$ parent ion and their corresponding masses are labeled. The *inset* shows the cross-linked fragments with the identified b- and y-ions labeled. *C*, the MS/MS spectrum of the cross-linked peptide $m/z = 731.3822$ (3+ charge). The assigned b-ions and y-ions generated by fragmentation of the $m/z = 731.4$ parent ion and their corresponding masses are labeled. The *inset* shows the cross-linked fragments with the identified b- and y-ions labeled.

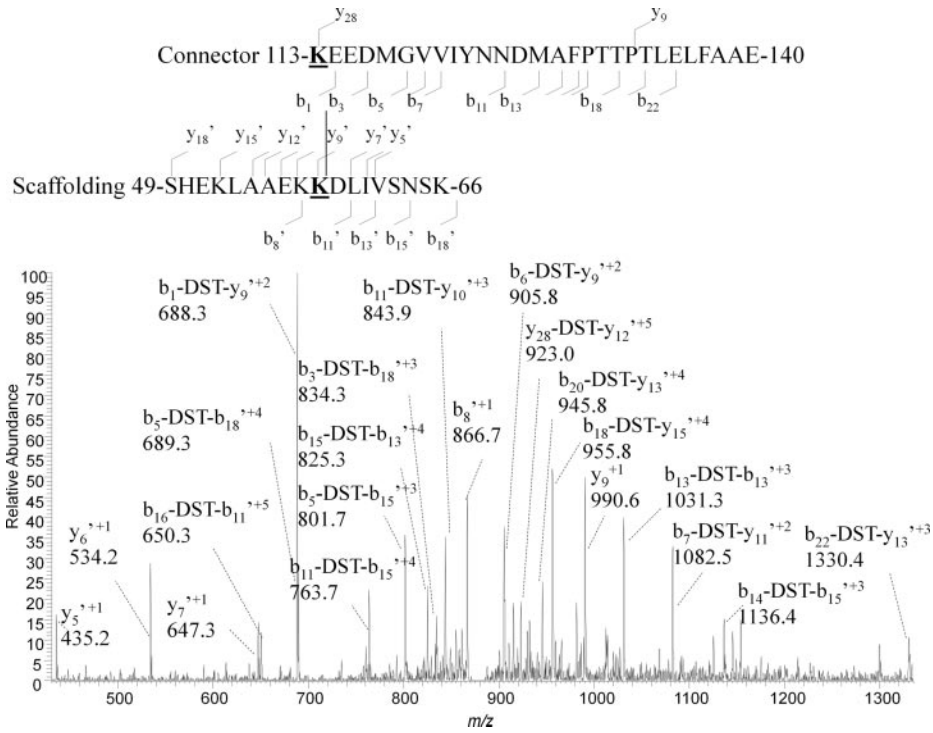

FIG. 5. **Identification of cross-link peptide in mutant complex.** The complex formed between mutant scaffolding protein D58K and wild type connector was cross-linked. A unique band in the SDS-PAGE was in-gel digested and yielded a unique parent ion of m/z = 876.7647 (6+ charge) that was subjected to MS/MS. The series of b- and y-ions generated by fragmentation and their assignments are indicated and mapped onto the sequence of the cross-linked peptide.

of the connector-derived peptide, it was not possible to discriminate between these two possibilities. Although connector residue 4 is disordered in the structure, extrapolation from the last ordered residue (Ser-11) suggested that it is located in the central helical region of the connector as is residue 19.

Generation and Verification of Docking Model of Connector-Scaffolding Complexes—A structural model of the connector/scaffolding interaction was generated using the rigid body docking algorithm ZDOCK (21). Dimeric scaffolding protein (Protein Data Bank code 1NO4, chains A and B, residues 2–76 (13)) was docked onto adjacent chains A and B of the 12-fold symmetric connector protein (Protein Data Bank code 1FOU (11)). Dimeric scaffolding protein was used as the building block because native mass spectrometry indicated that it binds as a dimer. Because the biochemical data suggested that the missing density in the C terminus of scaffolding crystal structure was not required for connector binding, the available crystal structure was considered sufficient to model the complex structure. Because the crystal structure of the connector is 12-fold symmetric, a dimer of the connector should capture all the essential protein surfaces and was used to reduce the computational demands.

Computational docking using the ZDOCK algorithm provided 2000 potential connector-scaffolding complexes that were ranked by optimized pairwise shape complementarity, desolvation, and electrostatic energies. Based on the cross-linking data, a distance constraint of 8 Å or less between the NZ atoms of connector Lys-102 and scaffold Lys-66 was used to remove unlikely connector-scaffolding complexes from the list of 2000 models. The 26 models that complied with this constraint were further winnowed based on the criteria that the scaffolding orientation had to allow cross-linking of scaffolding residue 83 to either connector residue 4 or 19. This cross-link does not constitute a stringent constraint because the scaffolding Lys-83 and connector Lys-4 are disordered in the respective crystal structures. However, their approximate positions can be traced from the ordered backbone (scaffolding Lys-76 and connector Ser-11); therefore, it was used to define the scaffolding orientation and select the preferred model.

As a test of the predictive value of the model, it was used to design and engineer a potential new cross-link into the connector/scaffolding interface. Scaffolding residue Asp-58, which was predicted to be adjacent to connector residue Lys-113, was mutated to a lysine. DST treatment resulted in the appearance of a new cross-linked band on the SDS-PAGE. The new cross-linked species was analyzed by mass spectrometry. A unique ion carrying six positive charges (m/z = 876.7647) was identified and assignable to scaffolding residues 49–66 cross-linked to connector residues 113–140 at 3-ppm mass accuracy from the predicted mass. The MS/MS data confirmed that the mutated scaffolding Lys-58 was cross-linked to the connector Lys-113 as the model predicted (Fig. 5).

In the model, the C terminus of the coiled coil dimer of helix α3 (residues 62–74) is docked to the exterior of the wide domain of connector protein at the interface between two adjacent subunits (Fig. 6). The A subunit of the scaffolding coiled coil predominately contacts the A subunit of the connector, and the B subunit predominately contacts the B sub-

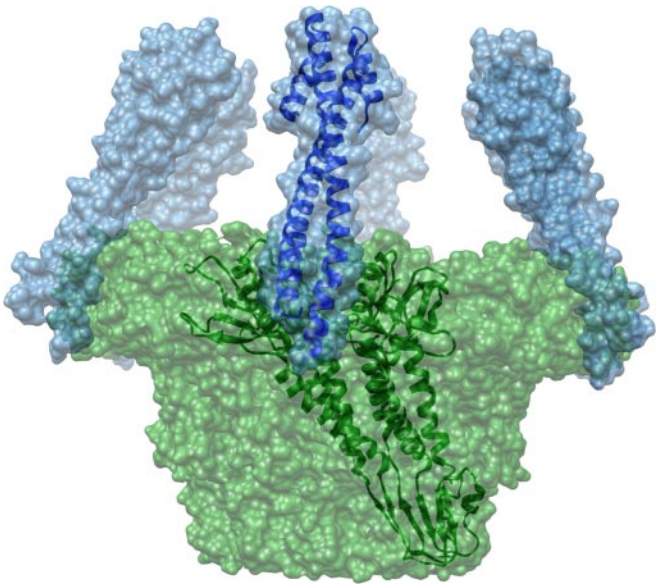

FIG. 6. **Docking model of connector-scaffolding complexes.** The docking model showing six scaffolding dimers (*blue*) docked to the wide domain interface between two adjacent connector subunits (*green*) in the dodecamer. The N-terminal arrowhead motif of the scaffolding points upward.

unit of the connector, but because of the twist in the coiled coil, the subunits do not make equivalent contacts. For example, scaffolding residue 65, the locus of the temperature-sensitive mutation, points toward the interface in the A chain and outward in the B chain, and the cross-link between Lys-66 and Lys-102 is favorable only in the B chain. Although the detailed molecular interactions of connector/scaffolding still remain unknown, the biochemical and chemical cross-linking data indicated that the current model closely locates the connector/scaffolding binding interfaces.

DISCUSSION

Biological processes generally involve controlled cooperation between multiple protein subunits in both time and space (30). Identification of the components involved and how they interact to carry out their biological functions is the key to understanding biological processes. Although structural data on individual proteins provide valuable insights into their mechanisms at the molecular level, obtaining structures of protein complexes, particularly in multiple regulatory states, is often very challenging. Intermediate resolution structural information obtained by cryo-EM and tomography has been integrated with high resolution data obtained by x-ray crystallography or NMR to provide multiscale pictures of dynamic interactions of protein molecules and complexes in the context of cellular space (31, 32). In addition, mass spectrometry has emerged as a powerful tool to identify protein interaction networks (33–36). Chemical cross-linking and mass spectrometry analysis are valuable approaches to identify protein/protein interfaces and obtain distance constraints of interacting residues (37–39). Here we applied experimentally constrained computational modeling to provide structural information on a connector-scaffolding complex involved in assembly of the viral packaging motor.

The ability of native mass spectrometry to resolve the distribution of species within an ensemble of complexes made it possible to demonstrate that scaffolding dimers are the dominant building block of the complex, and this constraint was also incorporated in the docking model. The model suggested that scaffolding dimer binds to the interfaces of two adjacent connector subunits. Binding of scaffolding protein around the wide domain of connector dodecamer will cause steric hindrance and prevent the head to tail stacking of connector dodecamers that has been proposed as the basis of its aggregation at low ionic strength. The model positioned the naturally occurring temperature-sensitive scaffolding S65N mutation and the C-terminal α3 of scaffolding protein in contact with connector protein in agreement with the biochemical data.

An interesting finding was that at most six rather than 12 scaffolding dimers could bind to the connector. The sharp cutoff of binding at six dimers suggests that every other connector location is occupied (binding sites 1, 3, 5, 7, 9, and 11). If adjacent sites could be occupied (*i.e.* sites 1, 2, and 3) there would be no reason for a sharp cutoff at six. The docking model provides an explanation for this experimental finding as it predicted that the inward canted docking of a dimer at one binding site sterically precludes docking of a dimer at an adjacent binding site.

There are two cryo-EM-based reconstructions of ϕ29 procapsids available. In one in which symmetry was not imposed a core of scaffolding protein was observed (13). However, in this reconstruction, density corresponding to the connector and capsid was subtracted from individual particle images, and therefore, scaffold density that is intimately associated with either the capsid or the connector would not be visible. In a second reconstruction, the density corresponding to the connector was preserved, but the procapsid was 5-fold-averaged (10). Because the imposition of 5-fold symmetry blurred the molecular boundaries of the 6-fold connector protein as well as any associated scaffolding protein, it was not possible to directly observe the scaffolding/connector interaction in this reconstruction. Nevertheless, the connector-scaffolding complex model was fit into the pseudoatomic structure of the 5-fold-averaged ϕ29 procapsid (Fig. 7). The results of the fitting demonstrated that the procapsid structure can easily accommodate the model of the complex and that the scaffolding protein does not have any significant clashes with surrounding capsid subunits. In this model, the disordered C termini of the scaffolding protein subunits, which have been demonstrated to host the capsid protein site, are proximal to the inner surface of the capsid shell.

How might the connector-scaffolding complex mediate portal incorporation? The model of the complex suggested

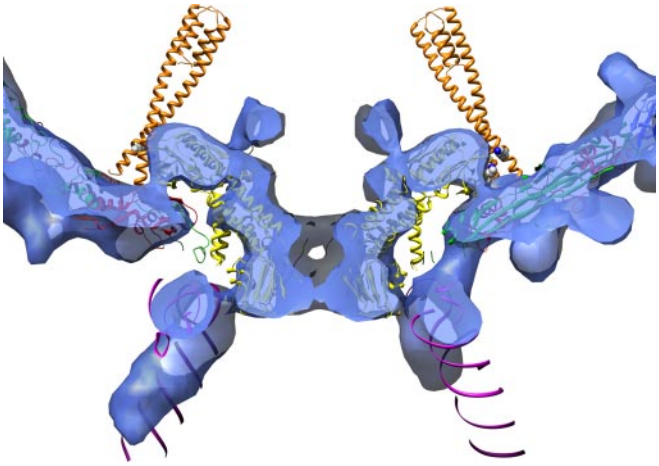

FIG. 7. **Model of connector-scaffolding complex docked in procapsid.** The docking model of the connector-scaffolding complex was fit into the connector vertex of a cryo-EM reconstruction of the procapsid (*blue*). The crystal structures of scaffolding (*orange*), connector (*yellow*), fitted HK97 capsid proteins (*red*, *green*, and *blue*), and packaging RNA (*pink*) are shown as ribbon diagrams. Interacting residues from the connector and the scaffolding are shown as van der Waals spheres colored by element (*gray* for carbon, *blue* for nitrogen, and *red* for oxygen). This figure was generated using CHIMERA (42).

that it consists of a dodecamer of connector protein onto which six dimers of scaffolding protein are docked with their C-terminal capsid binding region protruding from the narrow end. The simplest way to ensure the high fidelity incorporation of a single connector would be to couple it to nucleation of assembly as proposed for T4 (40, 41). The connector-scaffolding complex would make an ideal nucleation complex. Consistent with this suggestion is the observation that the complex lowers the critical concentration for assembly. Coat protein binding to the bottom of the connector-scaffolding complex would result in its inclusion in the growing shell in the correct orientation with the wide domain of connector dodecamer inside the procapsid (11). The flexible tail of scaffolding protein could provide the necessary soft coupling to allow five coat protein subunits to cluster around the dodecameric connector protein. We therefore propose that the connector-scaffolding complex acts as a platform to display the ~12 disordered C-terminal scaffolding residues, which in turn bind capsid protein subunits and nucleate assembly, thereby ensuring specific incorporation of one and only one connector vertex.

Acknowledgments—We thank Dr. Paul Jardine for ϕ29 clones.

* This work was supported, in whole or in part, by National Institutes of Health Grant GM47980.

S This article contains supplemental Fig. S1.

¶ Supported by the Netherlands Proteomics Centre.

** To whom correspondence should be addressed: Dept. of Microbiology, University of Alabama, BBRB 416, 845 19th St. South, Birmingham, AL 35294-2170. Tel.: 205-975-5327; Fax: 205-975-5479; E-mail: prevelig@uab.edu.

REFERENCES

1. Smith, D. E., Tans, S. J., Smith, S. B., Grimes, S., Anderson, D. L., and Bustamante, C. (2001) The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature* **413,** 748–752
2. Rickgauer, J. P., Fuller, D. N., Grimes, S., Jardine, P. J., Anderson, D. L., and Smith, D. E. (2008) Portal motor velocity and internal force resisting viral DNA packaging in bacteriophage phi29. *Biophys. J.* **94,** 159–167
3. Bazinet, C., and King, J. (1985) The DNA translocating vertex of dsDNA bacteriophage. *Annu. Rev. Microbiol.* **39,** 109–129
4. Moore, S. D., and Prevelige, P. E., Jr. (2002) Bacteriophage p22 portal vertex formation in vivo. *J. Mol. Biol.* **315,** 975–994
5. Casjens, S., and King, J. (1974) P22 morphogenesis. I: Catalytic scaffolding protein in capsid assembly. *J. Supramol. Struct.* **2,** 202–224
6. Bazinet, C., and King, J. (1988) Initiation of P22 procapsid assembly in vivo. *J. Mol. Biol.* **202,** 77–86
7. Greene, B., and King, J. (1996) Scaffolding mutants identifying domains required for P22 procapsid assembly and maturation. *Virology* **225,** 82–96
8. Earnshaw, W., and King, J. (1978) Structure of phage P22 coat protein aggregates formed in the absence of the scaffolding protein. *J. Mol. Biol.* **126,** 721–747
9. Tao, Y., Olson, N. H., Xu, W., Anderson, D. L., Rossmann, M. G., and Baker, T. S. (1998) Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell* **95,** 431–437
10. Morais, M. C., Choi, K. H., Koti, J. S., Chipman, P. R., Anderson, D. L., and Rossmann, M. G. (2005) Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29. *Mol. Cell* **18,** 149–159
11. Simpson, A. A., Tao, Y., Leiman, P. G., Badasso, M. O., He, Y., Jardine, P. J., Olson, N. H., Morais, M. C., Grimes, S., Anderson, D. L., Baker, T. S., and Rossmann, M. G. (2000) Structure of the bacteriophage phi29 DNA packaging motor. *Nature* **408,** 745–750
12. Guasch, A., Pous, J., Ibarra, B., Gomis-Rüth, F. X., Valpuesta, J. M., Sousa, N., Carrascosa, J. L., and Coll, M. (2002) Detailed architecture of a DNA translocating machine: the high-resolution structure of the bacteriophage phi29 connector particle. *J. Mol. Biol.* **315,** 663–676
13. Morais, M. C., Kanamaru, S., Badasso, M. O., Koti, J. S., Owen, B. A., McMurray, C. T., Anderson, D. L., and Rossmann, M. G. (2003) Bacteriophage phi29 scaffolding protein gp7 before and after prohead assembly. *Nat. Struct. Biol.* **10,** 572–576
14. Fu, C. Y., Morais, M. C., Battisti, A. J., Rossmann, M. G., and Prevelige, P. E., Jr. (2007) Molecular dissection of o29 scaffolding protein function in an in vitro assembly system. *J. Mol. Biol.* **366,** 1161–1173
15. Fu, C. Y., and Prevelige, P. E., Jr. (2009) In vitro incorporation of the phage Phi29 connector complex. *Virology* **394,** 149–153
16. Poliakov, A., van Duijn, E., Lander, G., Fu, C. Y., Johnson, J. E., Prevelige, P. E., Jr., and Heck, A. J. (2007) Macromolecular mass spectrometry and electron microscopy as complementary tools for investigation of the heterogeneity of bacteriophage portal assemblies. *J. Struct. Biol.* **157,** 371–383
17. Guo, P. X., Erickson, S., Xu, W., Olson, N., Baker, T. S., and Anderson, D. (1991) Regulation of the phage phi 29 prohead shape and size by the portal vertex. *Virology* **183,** 366–373
18. van den Heuvel, R. H., van Duijn, E., Mazon, H., Synowsky, S. A., Lorenzen, K., Versluis, C., Brouns, S. J., Langridge, D., van der Oost, J., Hoyes, J., and Heck, A. J. (2006) Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* **78,** 7473–7483
19. Lorenzen, K., Versluis, C., van Duijn, E., van den Heuvel, R. H., and Heck, A. J. (2007) Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *Int. J. Mass Spectrom.* **268,** 198–206
20. Rosenfeld, J., Capdevielle, J., Guillemot, J. C., and Ferrara, P. (1992) In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal. Biochem.* **203,** 173–179
21. Chen, R., Li, L., and Weng, Z. (2003) ZDOCK: an initial-stage protein-docking algorithm. *Proteins* **52,** 80–87
22. van den Heuvel, R. H., and Heck, A. J. (2004) Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* **8,** 519–526
23. Benesch, J. L., Ruotolo, B. T., Simmons, D. A., and Robinson, C. V. (2007)

Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem. Rev.* **107,** 3544–3567
24. Heck, A. J. (2008) Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* **5,** 927–933
25. Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A., and Robinson, C. V. (2002) A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74,** 1402–1407
26. Lorenzen, K., Vannini, A., Cramer, P., and Heck, A. J. (2007) Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure* **15,** 1237–1245
27. Uetrecht, C., Versluis, C., Watts, N. R., Roos, W. H., Wuite, G. J., Wingfield, P. T., Steven, A. C., and Heck, A. J. (2008) High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 9216–9220
28. Choi, K. H., Morais, M. C., Anderson, D. L., and Rossmann, M. G. (2006) Determinants of bacteriophage varphi29 head morphology. *Structure* **14,** 1723–1727
29. Schilling, B., Row, R. H., Gibson, B. W., Guo, X., and Young, M. M. (2003) MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J. Am. Soc. Mass Spectrom.* **14,** 834–850
30. Alberts, B. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92,** 291–294
31. Rossmann, M. G. (2000) Fitting atomic models into electron-microscopy maps. *Acta Crystallogr. D Biol. Crystallogr.* **56,** 1341–1349
32. Baumeister, W., and Steven, A. C. (2000) Macromolecular electron microscopy in the era of structural genomics. *Trends Biochem. Sci.* **25,** 624–631
33. Yates, J. R., 3rd, Gilchrist, A., Howell, K. E., and Bergeron, J. J. (2005) Proteomics of organelles and large cellular structures. *Nat. Rev. Mol. Cell Biol.* **6,** 702–714
34. Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., and Emili, A. (2005) Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature* **433,** 531–537
35. Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dümpelfeld, B., Edelmann, A., Heurtier, M. A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B., and Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440,** 631–636
36. Gavin, A. C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415,** 141–147
37. Back, J. W., de Jong, L., Muijsers, A. O., and de Koster, C. G. (2003) Chemical cross-linking and mass spectrometry for protein structural modeling. *J. Mol. Biol.* **331,** 303–313
38. van Dijk, A. D., Boelens, R., and Bonvin, A. M. (2005) Data-driven docking for the study of biomolecular complexes. *FEBS J.* **272,** 293–312
39. Sinz, A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom. Rev.* **25,** 663–682
40. Black, L. W., and Silverman, D. J. (1978) Model for DNA packaging into bacteriophage T4 heads. *J. Virol.* **28,** 643–655
41. Traub, F., and Maeder, M. (1984) Formation of the prohead core of bacteriophage T4 in vivo. *J. Virol.* **49,** 892–901
42. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612