Structural Basis of Chaperone Function and Pilus Biogenesis

Frederic G. Sauer,[1]* Klaus Fütterer,[2]* Jerome S. Pinkner,[1] Karen W. Dodson,[1] Scott J. Hultgren,[1]† Gabriel Waksman[2]†

Many Gram-negative pathogens assemble architecturally and functionally diverse adhesive pili on their surfaces by the chaperone-usher pathway. Immunoglobulin-like periplasmic chaperones escort pilus subunits to the usher, a large protein complex that facilitates the translocation and assembly of subunits across the outer membrane. The crystal structure of the PapD-PapK chaperone-subunit complex, determined at 2.4 angstrom resolution, reveals that the chaperone functions by donating its G_1 β strand to complete the immunoglobulin-like fold of the subunit via a mechanism termed donor strand complementation. The structure of the PapD-PapK complex also suggests that during pilus biogenesis, every subunit completes the immunoglobulin-like fold of its neighboring subunit via a mechanism termed donor strand exchange.

Many pathogenic Gram-negative bacteria assemble hair-like adhesive pili on their surfaces that mediate microbial attachment by binding to receptors present in host tissues (*1*). P pili are adhesive organelles encoded by 11 genes (*papA* through *papK*) in the *pap* (pilus associated with pyelonephritis) gene cluster found on the chromosome of uropathogenic strains of *Escherichia coli* (*2, 3*). P pili are composite fibers consisting of a rod, 7 nm in thickness, joined to a thinner tip fibrillum (*4*). The rod consists of PapA subunits arranged in a right-handed helical cylinder with 3.28 subunits per turn (*5, 6*). The fibrillum primarily consists of repeating PapE subunits arranged in an open helical configuration. The PapG adhesin is joined to the distal end of the tip fibrillum via the PapF adaptor subunit, and the fibrillum is joined to the rod via the PapK adaptor subunit (*7*). PapG is a critical virulence factor in pyelonephritic strains of *E. coli* (*8*). It binds specifically to the Galα(1-4)Gal digalactoside receptor found in the globoseries of glycolipids of the human kidney (*9, 10*).

The biogenesis of P pili occurs via the highly conserved chaperone-usher pathway (*11, 12*). PapD is the prototypical periplasmic chaperone in a family that includes more than 30 members (*13*). It is a boomerang-shaped two-domain molecule with each domain having an immunoglobulin-like (Ig) fold (*14*). PapD has three primary functions: (i) It binds and caps interactive surfaces on pilus subunits to prevent them from participating in nonproductive interactions (*15–19*); (ii) it facilitates the import of subunits into the periplasm (*17*); and (iii) it may facilitate subunit folding (*19*). Chaperone-subunit complexes are targeted to the PapC outer membrane usher, which is an oligomeric pore-forming complex that facilitates chaperone uncapping and the subsequent translocation and assembly of subunits across the outer membrane (*20, 21*). To gain further insight into the processes of subunit capping and assembly in the chaperone-usher pathway of pilus biogenesis, we solved the structure of a complex of the subunit PapK bound to the chaperone PapD (Fig. 1A and Table 1) (*22, 23*).

PapK has the same overall variable-region Ig fold (*24*) as the NH_2-terminal domain (domain 1) of PapD (Figs. 1B, 2A, and 2B). However, the Ig fold of PapK is incomplete: It lacks the COOH-terminal seventh strand, G, which in canonical Ig folds runs antiparallel to strand F and contributes to the hydrophobic core of the protein. In the PapD-PapK complex, this missing strand is provided by PapD, which donates its G_1 β strand to complete the Ig fold of PapK (Figs. 1B and 2B), a mechanism we call donor strand complementation. However, an atypical Ig fold is produced because the donated strand runs parallel rather than antiparallel to strand F in PapK. Similar observations are made in the crystal structure of the FimC-FimH complex (*25*). The first eight NH_2-terminal residues of PapK are disordered. The Ig fold of PapK (Figs. 1B and 2B) begins with β strand A, which is divided into strands A1 and A2 by a 3_{10} helical turn. Strand A1 makes antiparallel β-sheet interactions with strand B while strand A2 makes antiparallel β-sheet interactions with the G_1 strand of the chaperone.

[1]Department of Molecular Microbiology, [2]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110, USA.

*These authors contributed equally to this report.
†To whom correspondence should be addressed.

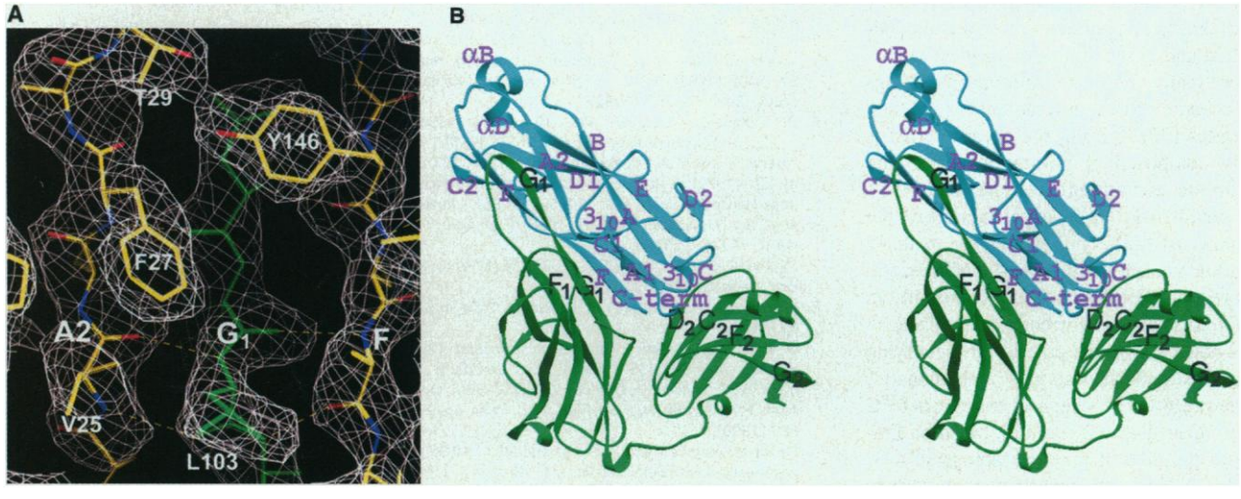

Fig. 1. (**A**) Density from a map calculated using MAD solvent-flattened phases. Atoms in PapK are yellow for carbon, red for oxygen, and blue for nitrogen; PapD is in green (*28*). (**B**) Stereo ribbon diagram of the PapD-PapK complex (*29*). PapK and PapD are in cyan and green, respectively. Secondary structures of PapK and PapD are labeled in magenta and black, respectively. Only secondary structures of PapD mentioned in the text are labeled. Subscripts 1 and 2 refer to domains 1 and 2 of PapD, respectively.

Strand B forms the edge of one of the two β sheets in the β sandwich of the Ig fold and runs antiparallel to strand E. Following strand B, the structure crosses over to the other side of the β sandwich through a short 3_{10} helix to form strand C1, which runs antiparallel to strand F. The COOH-terminus of strand C1 deviates from the β-sheet arrangement to form a β meander (strands C′ and C″), which eventually returns, as C2, to make main-chain hydrogen bonds with strand F. Strand D constitutes an edge of the D, E, B, A1 β sheet and runs antiparallel to strand E. However, strand D is divided in the middle by an insertion that meanders toward strands C′ and C″. Strand E is followed by a three-turn helix (αD) and a long loop structure that connects it to the COOH-terminal strand F. Finally, strand F forms a parallel β sheet with strand G_1 of PapD. Hence, strand G_1 of PapD is an integral part of the C, F, A2 β sheet of PapK.

The structure of apo-PapD has been described (*14*). The binding of PapK significantly alters the structure of PapD only in the F_1-G_1 loop [root mean square (rms) deviation in Cα atom positions, excluding the F_1-G_1 loop, of 0.65 Å]. The tip of this loop undergoes a flap motion of about 11 Å, such that residues 101 to 105 of PapD become part of the G_1 β strand that inserts into the Ig fold of PapK (Fig. 3A). Two distinct sites on PapK, K1 and K2, interact with two corresponding sites on PapD, D1 and D2 (Fig. 3B). The K2-D2 interface (910 Å^2) is less extensive than the K1-D1 interface (2524 Å^2). Site K2 is formed primarily by residues in helix 3_{10}C and the COOH-terminal Arg^{157} side chain of PapK (Fig. 1B). Residues in site K2 interact with residues in the C_2 and D_2 strands and with the F_2-G_2 loop of domain 2 of PapD (site D2). Site K1 contains a deep groove that runs the length of the subunit and whose base consists of the hydrophobic core of PapK (Fig. 3C). This groove is the result of the missing G β strand in the Ig fold of PapK. Site D1 includes residues 101 to 112 of the G_1 β strand of PapD, which insert into the K1 groove. The main-chain atoms of the G_1 strand make β-sheet interactions with strands F and A2 of PapK on either side of the K1 groove; the alternating hydrophobic side chains of the G_1 strand (Leu^{103}, Ile^{105}, and Leu^{107}) interact with the hydrophobic base of the groove (Figs. 2B and 3C). Pilus subunits,

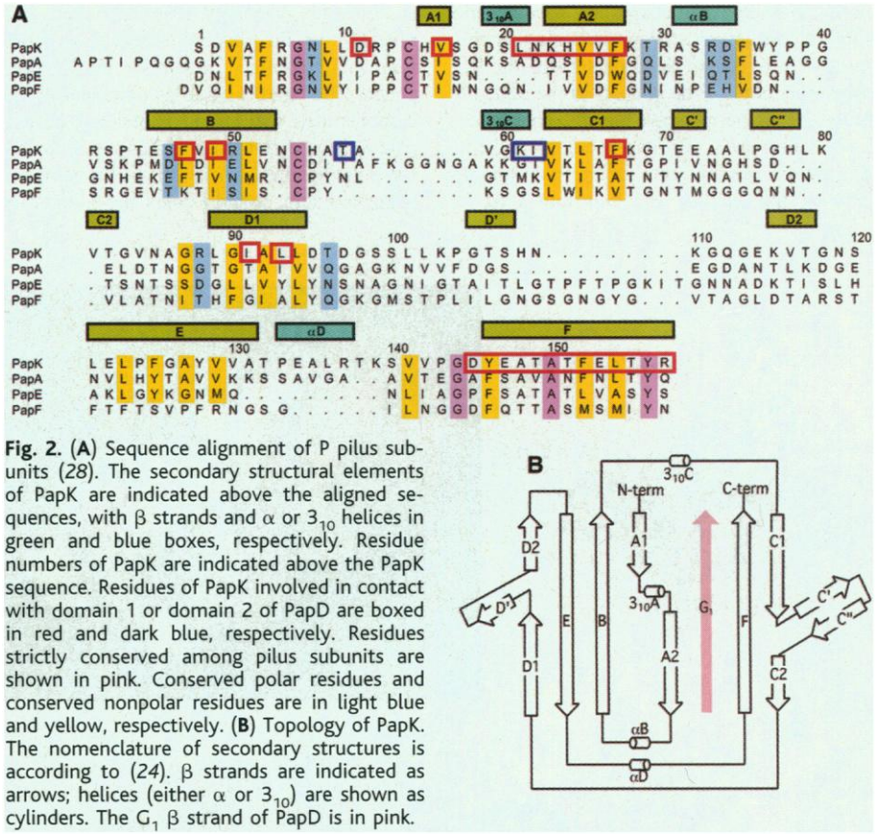

Fig. 2. (**A**) Sequence alignment of P pilus subunits (*28*). The secondary structural elements of PapK are indicated above the aligned sequences, with β strands and α or 3_{10} helices in green and blue boxes, respectively. Residue numbers of PapK are indicated above the PapK sequence. Residues of PapK involved in contact with domain 1 or domain 2 of PapD are boxed in red and dark blue, respectively. Residues strictly conserved among pilus subunits are shown in pink. Conserved polar residues and conserved nonpolar residues are in light blue and yellow, respectively. (**B**) Topology of PapK. The nomenclature of secondary structures is according to (*24*). β strands are indicated as arrows; helices (either α or 3_{10}) are shown as cylinders. The G_1 β strand of PapD is in pink.

Table 1. Data collection and refinement statistics (*22*, *23*).

Data collection

Data set	Radiation	Resolution	Total/unique reflections	Completeness (%)*	R_{sym} (%)†	R_{iso} (%)‡
Native	CuKα, Raxis	30 to 2.7 Å	22,046/8,960	84.5 (77.2)	6.4 (14.5)	
SeMet-single	CuKα, Raxis	30 to 2.5 Å	46,683/12,179	90.4 (75.2)	6.9 (19.2)	8.5
SeMet-1	0.9879 Å, X4A	30 to 2.4 Å	92,857/14,135	89.4 (79.1)	6.0 (13.9)	
SeMet-2	0.9792 Å, X4A	30 to 2.4 Å	105,506/14,343	92.9 (83.2)	6.1 (14.1)	
SeMet-3	0.9788 Å, X4A	30 to 2.4 Å	102,568/14,203	91.5 (79.4)	6.4 (14.6)	
SeMet-4	0.9667 Å, X4A	30 to 2.4 Å	102,187/14,314	92.5 (82.4)	6.3 (15.0)	
Figure of merit for MAD phasing SeMet-1–4 data (calculated for 30 to 2.4 Å)				0.49		

Refinement

Resolution (Å)	Number of reflections§	Total number of atoms	*R* factor/R_{free} (%)‖	rms deviations¶			
				Bonds (Å)	Angles (°)	*B* values (Å^2)	
30.0 to 2.4	12,678 (84.8%/77.3%)	2912#	23.8/27.4	0.011	1.45	1.62 (main chain)	2.20 (side chain)

*Completeness for $I/\sigma(I) > 1.0$, high-resolution shell in parentheses [2.80 to 2.70 Å (native), 2.59 to 2.50 Å (SeMet-single), 2.50 to 2.40 Å (MAD)]. †$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I = observed intensity, and $\langle I\rangle$ = average intensity from multiple observations of symmetry-related reflections; high-resolution shell in parentheses. ‡$R_{iso} = \Sigma||F_{PH}| - |F_P||/\Sigma|F_P|$, where F_P = native structure factor amplitude and F_{PH} = derivative (SeMet) structure factor amplitude. §Numbers reflect the "working set" of reflections at $F/\sigma(F) > 2.0$, overall/last shell (2.51 to 2.40 Å) completeness in parentheses. ‖R_{free} was calculated on the basis of 7.0% of the total number of reflections randomly omitted from the refinement. ¶Deviations from ideal bond lengths and angles and in *B* factors of bonded atoms. #Including 104 water molecules.

in the absence of the chaperone, are unstable (*26*); the structure of PapK reveals that this instability is due to the exposure of the hydrophobic core of the subunit to the aqueous milieu of the periplasm. Thus, the chaperone, by donating its G_1 β strand to complete the Ig fold of the subunit, shields the hydrophobic core of the subunit and contributes to its stabilization. The PapD-PapK structure also reveals how the chaperone prevents premature subunit-subunit assembly in the periplasm. Biochemical and genetic data have demonstrated that strand F residues participate in subunit-subunit interactions necessary for pilus assembly (*16, 19*). In the structure presented here, strand F lines one edge of the K1 groove of the subunit, but its residues interact with (are "capped" by) the G_1 strand residues of the chaperone. Thus, the insertion of the G_1 β strand of the chaperone into the K1 groove prevents strand F residues from participating in interactions with other subunits.

Finally, the structure of the PapD-PapK complex suggests a mechanism for pilus biogenesis at the outer membrane usher. In addition to strand F, a highly conserved motif at the NH_2-terminus of the subunits has also been shown to participate in subunit-subunit interactions (*19*). This motif, located upstream of strand A1 and disordered in the structure, is highly homologous to the G_1 strand of the PapD chaperone. Both regions consist of an alternating arrangement of hydrophobic residues: Leu^{103}, Ile^{105}, and Leu^{107} in the G_1 strand of PapD, and Val^{11}, Phe^{13}, Gly^{15}, and Val^{17} in the NH_2-terminal region of PapA, for example (Fig. 2A). Thus, this region could fit into the groove of a neighboring subunit during pilus biogenesis. Because this NH_2-terminal motif apparently protrudes away from the main body of the chaperone-subunit complex, it would be free to exchange with the G_1 strand of the chaperone by a mechanism we term donor strand exchange. Thus, the Ig fold of every subunit would be complemented by the NH_2-terminal motif of its neighboring subunit. Donor strand exchange differs from domain swapping, a well-characterized mechanism of protein assembly. Domain swapping implies that the swapped domains make identical interactions in both the monomeric and the oligomeric forms and that the monomeric form of the protein is stable (*27*). However, pilus subunits are unstable as monomers and can only exist either bound to a chaperone or as part of the pilus (*26*).

We used the PapK structure to model a PapA pilus rod. The disordered NH_2-terminal motif of each pilus subunit was built as a β

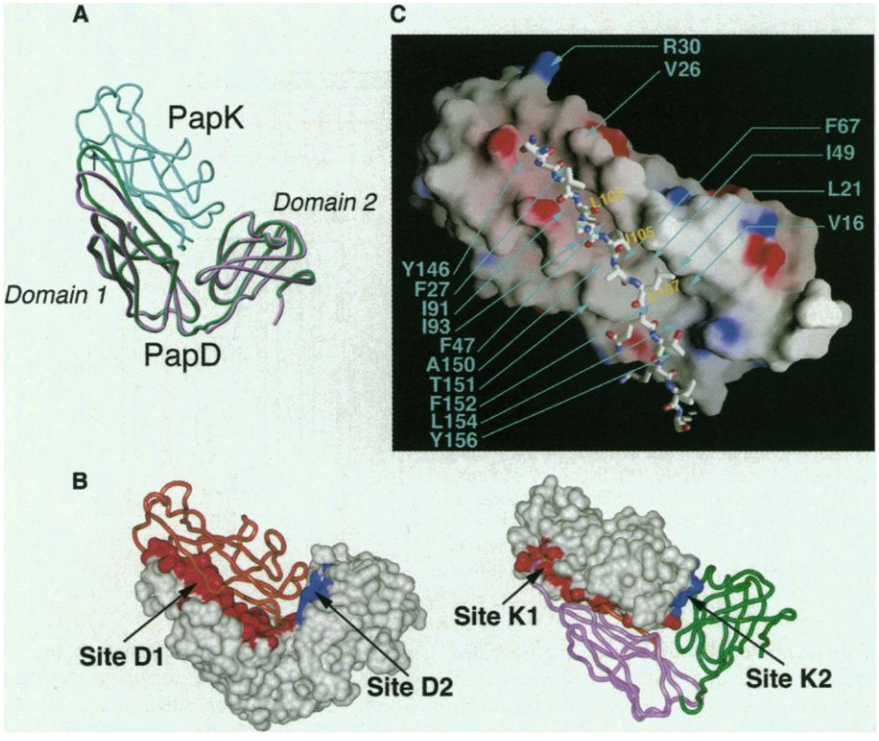

Fig. 3. (**A**) Superposition of the structures of apo-PapD (purple) and PapD complexed to PapK (green) (*29*). PapK is in cyan. The arrow indicates the conformational change in the F_1-G_1 loop upon subunit binding. (**B**) Definition of the binding sites in PapD and PapK. On the left, PapD is shown as a space-filling model and PapK as a ribbon. On the right, PapK is shown as a space-filling model and PapD as a ribbon. The binding sites defined in the text are labeled. (**C**) The G_1 β strand of PapD as it inserts into the groove of PapK. The PapD G_1 strand is represented as a stick model and PapK as a molecular surface [program GRASP (*30*) with color coding according to charge, blue for positive and red for negative]. PapD and PapK residues are labeled in yellow and cyan, respectively (*28*).

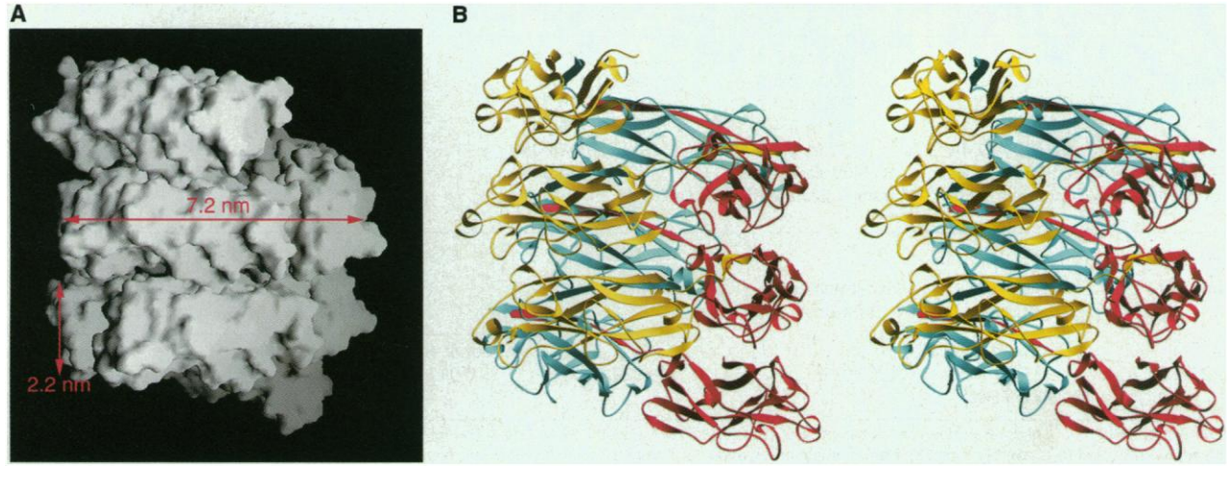

Fig. 4. (**A**) Molecular surface of a pilus rod [program GRASP (*30*)]. The disordered residues at the NH_2-terminus of the subunit were modeled as a β strand that inserts into the groove of the preceding subunit antiparallel to strand F. (**B**) Stereo ribbon diagram of the rod model.

strand and inserted into the groove of each preceding subunit. Insertion parallel to strand F yielded a rod with a star-shaped cross section, inconsistent with electron microscopy data. Insertion antiparallel to strand F produced a pilus with a helical symmetry having dimensions similar to those experimentally observed (*5, 6*) (Fig. 4). Thus, donor strand complementation with the chaperone results in an atypical Ig fold, whereas donor strand exchange between subunits produces a canonical variable-region Ig fold in the mature pilus (*24*). Stereochemical complementarity between the NH_2-terminal motifs and grooves of the various subunits most likely restricts the order of subunit assembly. Thus, the molecular basis for the adaptor function of PapK may in part be a consequence of its NH_2-terminal motif fitting the groove of PapE and its groove accommodating the NH_2-terminal motif of PapA with stereochemical specificity.

References and Notes

1. S. J. Hultgren, C. H. Jones, S. N. Normark, in Escherichia coli *and* Salmonella *Cellular and Molecular Biology*, F. C. Neidhardt, Ed. (ASM Press, Washington, DC, ed. 2, 1996), pp. 2730–2756.
2. R. A. Hull, R. E. Gill, P. Hsu, B. H. Minshew, S. Falkow, *Infect. Immun.* **33**, 933 (1981).
3. S. J. Hultgren *et al.*, *Cell* **73**, 887 (1993).
4. M. J. Kuehn, J. Heuser, S. Normark, S. J. Hultgren, *Nature* **356**, 252 (1992).
5. E. Bullitt and L. Makowski, *ibid.* **373**, 164 (1995).
6. M. Gong and L. Makowski, *J. Mol. Biol.* **228**, 735 (1992).
7. F. Jacob-Dubuisson, J. Heuser, K. Dodson, S. Normark, S. Hultgren, *EMBO J.* **12**, 837 (1993).
8. J. A. Roberts *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11889 (1994).
9. B. Lund, F. Lindberg, B.-I. Marklund, S. Normark, *ibid.* **84**, 5898 (1987).
10. R. Striker, U. Nilsson, A. Stonecipher, G. Magnusson, S. J. Hultgren, *Mol. Microbiol.* **16**, 1021 (1995).
11. D. G. Thanassi, E. T. Saulino, S. J. Hultgren, *Curr. Opin. Microbiol.* **1**, 223 (1998).
12. D. L. Hung, S. D. Knight, R. M. Woods, J. S. Pinkner, S. J. Hultgren, *EMBO J.* **15**, 3792 (1996).
13. F. Lindberg, J. M. Tennent, S. J. Hultgren, B. Lund, S. Normark, *J. Bacteriol.* **171**, 6052 (1989).
14. A. Holmgren and C.-I. Branden, *Nature* **342**, 248 (1989).
15. M. J. Kuehn, S. Normark, S. J. Hultgren, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10586 (1991).
16. E. Bullitt *et al.*, *ibid.* **93**, 12890 (1996).
17. C. H. Jones, P. N. Danese, J. S. Pinkner, T. J. Silhavy, S. J. Hultgren, *EMBO J.* **16**, 6394 (1997).
18. M. J. Kuehn *et al.*, *Science* **262**, 1234 (1993).
19. G. E. Soto *et al.*, *EMBO J.* **17**, 6155 (1998).
20. K. W. Dodson, F. Jacob-Dubuisson, R. T. Striker, S. J. Hultgren, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3670 (1993).
21. D. G. Thanassi *et al.*, *ibid.* **95**, 3146 (1998).
22. Crystals of wild-type or selenomethionine (SeMet)–containing PapD-PapK complexes (12 mg/ml) were grown by vapor diffusion using the hanging drop method against a reservoir containing 10 to 15% (w/v) polyethylene glycol (molecular weight 6000), 100 mM potassium acetate, and 200 to 400 mM sodium acetate at pH 4.6 [A. McPherson, *Eur. J. Biochem.* **189**, 23 (1990)]. Crystals were in space group $P2_12_12_1$, with cell dimensions a = 62.1 Å, b = 63.6 Å, and c = 92.7 Å, and with one complex in the asymmetric unit. Native and SeMet data sets to 2.7 Å and 2.5 Å resolution, respectively, were collected on single flash-cooled crystals in the laboratory setting (native and SeMet-single data sets, Table 1). A SeMet PapD-PapK crystal was also used to collect multiwavelength anomalous dispersion (MAD) data at four wavelengths to a resolution of 2.4 Å at the National Synchrotron Light Source (NSLS) (SeMet-1 through -4, Table 1). All data were reduced and processed using the programs DENZO and SCALEPACK [Z. Otwinoski, in *Proceedings of the CCP4 Study Weekend*, L. Sawyers, N. Isaacs, S. Bailey, Eds. (SERC Daresbury Laboratory, Warrington, UK, 1993), pp. 56–62].
23. The structure of the PapD-PapK complex was solved using MAD phasing. The native and SeMet-single data sets were used to generate a difference Patterson map with the program HEAVY [T. C. Terwilliger and D. Eisenberg, *Acta Crystallogr.* **A39**, 813 (1983)]. The positions of the three selenium atoms in SeMet PapD (residues 18, 66, and 172) were determined with the program HASSP [T. C. Terwilliger, S.-H. Kim, D. Eisenberg, *ibid.* **A43**, 1 (1987)] and used to calculate phases based on the MAD data with the program SHARP [E. De La Fortelle and G. Bricogne, *Methods Enzymol.* **276**, 472 (1997)]. An interpretable electron density map was obtained after solvent flipping with the program SOLOMON [J. P. Abrahams and A. G. W. Leslie, *Acta Crystallogr.* **D52**, 32 (1996)]. PapD and PapK were built into the electron density with the program O [T. A. Jones and S. Thirup, *EMBO J.* **5**, 819 (1986); T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991)]; the resulting density map was of sufficient quality to unequivocally assign the sequence (Fig. 1A). The model was refined using CNS version 0.5 [A. T. Brünger *et al.*, *ibid.* **D54**, 905 (1998)] against the SeMet-3 structure factor amplitudes, using both positional and simulated annealing refinement. The final model containing 104 well-ordered water molecules has R and R_{free} values of 23.8% and 27.4%, respectively (Table 1). The model of PapD does not include residues 216 to 218 of PapD, and residues Arg^{96} and Glu^{98} in PapD were built as alanines. The model of PapK is complete but for eight residues located at the NH_2-terminus. All residues in PapK and PapD are located in either the most favored or the allowed regions of the Ramachandran plot [G. N. Ramachandran and V. Sasisekharan, *Adv. Protein Chem.* **23**, 283 (1968)]. Coordinates have been deposited at the Protein Data Bank (entry code 1PDK).
24. E. Y. Jones, *Curr. Opin. Struct. Biol.* **3**, 846 (1993).
25. D. Choudhury *et al.*, *Science* **285**, 1061 (1999).
26. L. N. Slonim, J. S. Pinkner, C.-I. Branden, S. J. Hultgren, *EMBO J.* **11**, 4747 (1992).
27. M. P. Schlunegger, M. J. Bennett, D. Eisenberg, *Adv. Protein Chem.* **50**, 61 (1997).
28. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
29. M. Carson, *Methods Enzymol.* **277**, 493 (1997).
30. A. Nicholls, K. Sharp, B. Honig, *Proteins Struct. Funct. Genet.* **11**, 281 (1991).

31. We thank A. B. Herr for help with MAD data collection, G. Soto and D. Hung for help in preparing figures, and C. Ogata and the staff of beamline X4A at NSLS. Supported by NIH grants RO1DK51406 and RO1AI29549 (S.J.H.) and RO1GM54033 (G.W.).

25 March 1999; accepted 9 July 1999

X-ray Structure of the FimC-FimH Chaperone-Adhesin Complex from Uropathogenic *Escherichia coli*

Devapriya Choudhury,[1] Andrew Thompson,[2] Vivian Stojanoff,[3] Solomon Langermann,[4] Jerome Pinkner,[5] Scott J. Hultgren,[5]* Stefan D. Knight[1]*

Type 1 pili—adhesive fibers expressed in most members of the Enterobacteriaceae family—mediate binding to mannose receptors on host cells through the FimH adhesin. Pilus biogenesis proceeds by way of the chaperone/usher pathway. The x-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli* at 2.5 angstrom resolution reveals the basis for carbohydrate recognition and for pilus assembly. The carboxyl-terminal pilin domain of FimH has an immunoglobulin-like fold, except that the seventh strand is missing, leaving part of the hydrophobic core exposed. A donor strand complementation mechanism in which the chaperone donates a strand to complete the pilin domain explains the basis for both chaperone function and pilus biogenesis.

Type 1 pili are adhesive fibers expressed in *E. coli* as well as in most members of the Enterobacteriaceae family (*1*). They are composite structures in which a short-tip fibrillar structure containing FimG and the FimH adhesin (and possibly the minor component FimF as well) are joined to a rod composed predominantly of FimA subunits (*1*). The FimH adhesin mediates binding to mannose oligosaccharides (*2*, *3*). In uropathogenic *E. coli*, this binding event has been shown to play a critical role in bladder colonization and disease (*4*). Type 1 pilus biogenesis proceeds by way of a highly conserved chaperone/usher pathway that is involved in the assembly of over 25 adhesive organelles in Gram-negative bacteria (*5*). The usher forms an oligomeric channel in the outer membrane with a pore size of ~2.5 nm (*6*) and mediates subunit translocation across the outer membrane. Periplasmic chaperones consist of two immunoglobulin-like domains with a deep cleft between the two domains (*7–9*). Chaperones stabilize pilus subunits and prevent them from participating in premature interactions in the