The EMBO Journal vol.15 no.15 pp.3792–3805, 1996

# Molecular basis of two subfamilies of immunoglobulin-like chaperones

**Danielle L.Hung, Stefan D.Knight[1], Robert M.Woods[2], Jerome S.Pinkner and Scott J.Hultgren[3]**

Department of Molecular Microbiology, Washington University School of Medicine, St Louis, MO 63110, USA, [1]Department of Molecular Biology, Biomedical Centre, Uppsala, Sweden and [2]Medimmune, 35 W. Watkins Mill Road, Gaithersburg, MD 20878, USA

[3]Corresponding author

**The initial encounter of a microbial pathogen with the host often involves the recognition of host receptors by different kinds of bacterial adhesive organelles called pili, fimbriae, fibrillae or afimbrial adhesins. The development of over 26 of these architecturally diverse adhesive organelles in various Gram-negative pathogens depends on periplasmic chaperones that are comprised of two immunoglobulin-like domains. All of the chaperones possess a highly conserved sheet in domain 1 and a conserved interdomain hydrogen-bonding network. Chaperone–subunit complex formation depends on the anchoring of the carboxylate group of the subunit into the conserved crevice of the chaperone cleft and the subsequent positioning of the COOH terminus of subunits along the exposed edge of the conserved sheet of the chaperone. We discovered that the chaperones can be divided into two distinct subfamilies based upon conserved structural differences that occur in the conserved sheet. Interestingly, a subdivision of the chaperones based upon whether they assemble rod-like pili or non-pilus organelles that have an atypical morphology defines the same two subgroups. The molecular dissection of the two chaperone subfamilies and the adhesive fibers that they assemble has advanced our understanding of the development of virulence-associated organelles in pathogenic bacteria.**

*Keywords*: immunoglobulin-like chaperones/pathogenesis/virulence-associated pili/X-ray crystallography

## Introduction

Pathogenic bacteria assemble structures on their surfaces that allow them to colonize host tissues and cause disease. There are several molecular machineries used by Gram-negative bacteria to assemble adhesive surface organelles, including the chaperone/usher pathway, the general assembly pathway and the extracellular nucleation/precipitation pathway (Arnqvist *et al.*, 1992; Hultgren *et al.*, 1993a, 1996; Pugsley, 1993). The assembly of over 26 architecturally diverse adhesive organelles in Gram-negative bacteria proceeds via the chaperone/usher pathway and thus depends on one of the 26 proteins that are 25–56% identical to PapD, which is the prototypic periplasmic chaperone (Hultgren *et al.*, 1993a). PapD assembles P pili which are virulence factors produced by uropathogenic *Escherichia coli* (Hultgren *et al.*, 1991). The crystal structure of PapD has been solved and refined to 2.0 Å resolution (Holmgren and Brändén, 1989; A.Holmgren, in preparation). The structure of PapD consists of two globular domains oriented in the shape of a boomerang with a cleft between the two domains. Each domain has a β-barrel structure formed by two antiparallel β-pleated sheets with an overall topology similar to an immunoglobulin fold (Williams and Barclay, 1988; Holmgren and Brändén, 1989).

P pili are comprised of six distinct structural proteins that interact to form a composite fiber consisting of two distinct subassemblies: a 68 Å thick pilus rod comprised of PapA subunits arranged in a right-handed helical cylinder (Bullitt and Makowski, 1995) and a thin fibrillum (~20 Å) comprised mostly of repeating PapE subunits arranged in an open helical configuration (Kuehn *et al.*, 1992). The two subassemblies are joined together by the PapK adaptor protein. The PapG adhesin, a virulence factor in uropathogenic *E.coli*, is joined to the distal end of the tip fibrillum via the PapF adaptor protein (Jacob-Dubuisson *et al.*, 1993a; Roberts *et al.*, 1994). PapD binds to and caps interactive surfaces present on each subunit to prevent their premature aggregation during their secretion into the periplasmic space (Hultgren *et al.*, 1989; Kuehn *et al.*, 1991). The structural basis for part of the chaperone–subunit interaction was solved by co-crystallizing PapD with a peptide corresponding to the COOH terminus of PapG. The COOH terminus of pilus subunits assembled by immunoglobulin-like chaperones is part of a highly conserved motif that is recognized by the chaperone (Kuehn *et al.*, 1993). The PapG peptide was anchored in the chaperone cleft by hydrogen bonds between the terminal carboxylate group of the peptide and R8 and K112, which are positioned in the crevice of the cleft and invariant in the entire chaperone superfamily. The subsequent positioning of the peptide along the exposed edge of the G1 β strand of PapD was mostly the result of backbone hydrogen bonds forming an extended β sheet between PapD and the peptide that has been termed a β zipper (Kuehn *et al.*, 1993). The chaperone–subunit complexes are targeted to the outer membrane PapC usher where chaperone dissociation is coupled with incorporation of the subunits into the growing organelles (Dodson *et al.*, 1993).

Although the entire chaperone family can be grouped together as one PapD-like superfamily, we have discovered striking structural differences between them that are conserved in two subgroups. Remarkably, a division of the chaperones based on the molecular architecture of the organelles that they assemble defines the same two subgroups. Thus, 17 of the chaperones assemble pilus rod-like fibers defining the first subfamily. The second subfamily is comprised of chaperones that assemble either non-fimbrial

**Table I.** Superfamily of immunoglobulin-like chaperones

| Chaperone | Mass (kDa) | Organism | Surface structure assembled | % Identity to PapD[a] | Disease | Reference |
|---|---|---|---|---|---|---|
| Thick, rigid pili (rods) | | | | | | |
| PapD[b] | 28.5 | *E.coli* | P pili | – | pylonephritis/cystitis | Lindberg *et al.* (1989)<br>Kuehn *et al.* (1992) |
| SfaE[b] | 23.5 | *E.coli* | S fimbriae | 33 | UTI, NMB | Schmoll *et al.* (1990)<br>Hacker and Morschhäuser (1994) |
| FimC[b] | 26 | *E.coli* | Type 1 pili | 34 | cystitis | Jones *et al.* (1993)<br>Klemm and Krogfelt (1994) |
| HifB[b] | 25 | *H.influenzae* | *H.influenzae* fimbriae | 32 | otitis media, meningitis | Stull *et al.* (1984)<br>NIH sequence database |
| MrpD | 27.8 | *P.mirabilis* | MR/P fimbriae | 56 | nosocomial UTI | Bahrani *et al.* (1991)<br>Bahrani *et al.* (1994) |
| FocC | ? | *E.coli* | FIC fimbria | 33 | cystitis? | Klemm *et al.* (1994)<br>Klemm *et al.* (1995) |
| FimB | 24 | *B.pertussis* | Type 2 and 3 fimbriae | 31 | whooping cough | Steven *et al.* (1986)<br>Willems *et al.* (1992) |
| PefD | 23 | *S.typhimurium* | PEF | 32 | gastroenteritis/ salmonellosis | Friedrich *et al.* (1993) |
| PmfD | ? | *P.mirabilis* | PMF fimbriae | 47 | nosocomial UTI | Bahrani *et al.* (1993)<br>Massad and Mobley (1994) |
| LpfB | 23 | *S.typhimurium* | Long polar fimbriae | 33 | gastroenteritis?/ salmonellosis? | Bäumler and Heffron (1995) |
| Thin, flexible pili | | | | | | |
| FanE | 23 | *E.coli* | K99 fimbriae | 31 | neonatal diarrhea in pigs | Duchet-Suchaux *et al.* (1988)<br>Bakker *et al.* (1991) |
| FaeE | 26.5 | *E.coli* | K88 fimbriae | 28 | neonatal diarrhea in pigs, calves, lambs | Stirm *et al.* (1967)<br>Foged *et al.* (1986)<br>Bakker *et al.* (1991) |
| F17D | 28 | *E.coli* | F17 fimbriae | 30 | diarrhea | Lintermans *et al.* (1988)<br>Holmgren *et al.* (1992) |
| MrkB | 25 | *K.pneumoniae* | Type 3 fimbriae | 34 | pneumonia | Allen *et al.* (1991) |
| Unknown structure | | | | | | |
| EcpD | 27 | *E.coli* | ? | 35 | ? | Raina *et al.* (1993) |
| YehC | ? | ? | ? | 33 | ? | NIH sequence database |
| Atypical: very thin, fibrillae | | | | | | |
| ClpE | 25.3 | *E.coli* | CS31A capsule-like protein | 29 | diarrhea | Girardeau *et al.* (1988)<br>Bertin *et al.* (1993) |
| CssC | 23 | *E.coli* | Antigen CS6 | 27 | diarrhea | Knutton *et al.* (1989)<br>Wolf *et al.* (1989)<br>NIH sequence database |
| MyfB | 30.2 | *Y.enterocolitica* | Myf fimbriae | 31 | enterocolitis | Iriarte *et al.* (1993) |
| PsaB | 25 | *Y.pestis* | pH6 antigen | 29 | plague | Lindler and Tall (1993) |
| CS3–1 | 27 | *E.coli* | CS3 pili | 25 | travelers diarrhea | Levine *et al.* (1984)<br>Jalajakumari *et al.* (1989) |
| Caf1M | 26 | *Y.pestis* | Envelope antigen F1 | 25 | plague | Galyov *et al.* (1991)<br>Karlyshev *et al.* (1992) |
| NfaE | 30.5 | *E.coli* | non-fimbrial adhesins I | 32 | UTI, NBM | Goldhar *et al.* (1987)<br>Ahrens *et al.* (1993) |
| SefB | 27 | *S.enteritidis* | SEF14 fimbriae | 25 | gastroenteritis/ salmonellosis | Müller *et al.* (1991)<br>Clouthier *et al.* (1993) |
| AggD | 24.6 | *E.coli* | Aggregative adherence fimbria I | 31 | diarrhea | Savarino *et al.* (1994) |
| AfaB | 30 | *E.coli* | AFA-III | 32 | diarrhea | Le Bouguenec *et al.* (1993)<br>Garcia *et al.* (1994) |

[a]Each protein was aligned to PapD, the prototype member of the chaperone superfamily, from amino acid 22 to the end of the protein using the program GAP (version 7; Genetics Computer Group).
[b]Consists of two subassemblies: a thick pilus rod joined to a thin, flexible-tip fibrillum.

adhesins or very thin fibers that have an atypical morphology (Table I). In this paper, the structural differences that delineate the two chaperone subfamilies and the organelles that they assemble are defined. Structural similarities and differences between the two subfamilies are analyzed and discussed in terms of the β-zippering model of subunit–chaperone interactions.

## Results

### *Adhesive organelles assembled by periplasmic chaperones have two basic architectures*

There are two major architectural classes of adhesive organelles assembled by the chaperone/usher pathway, all of which contain diverse adhesins with distinct receptor

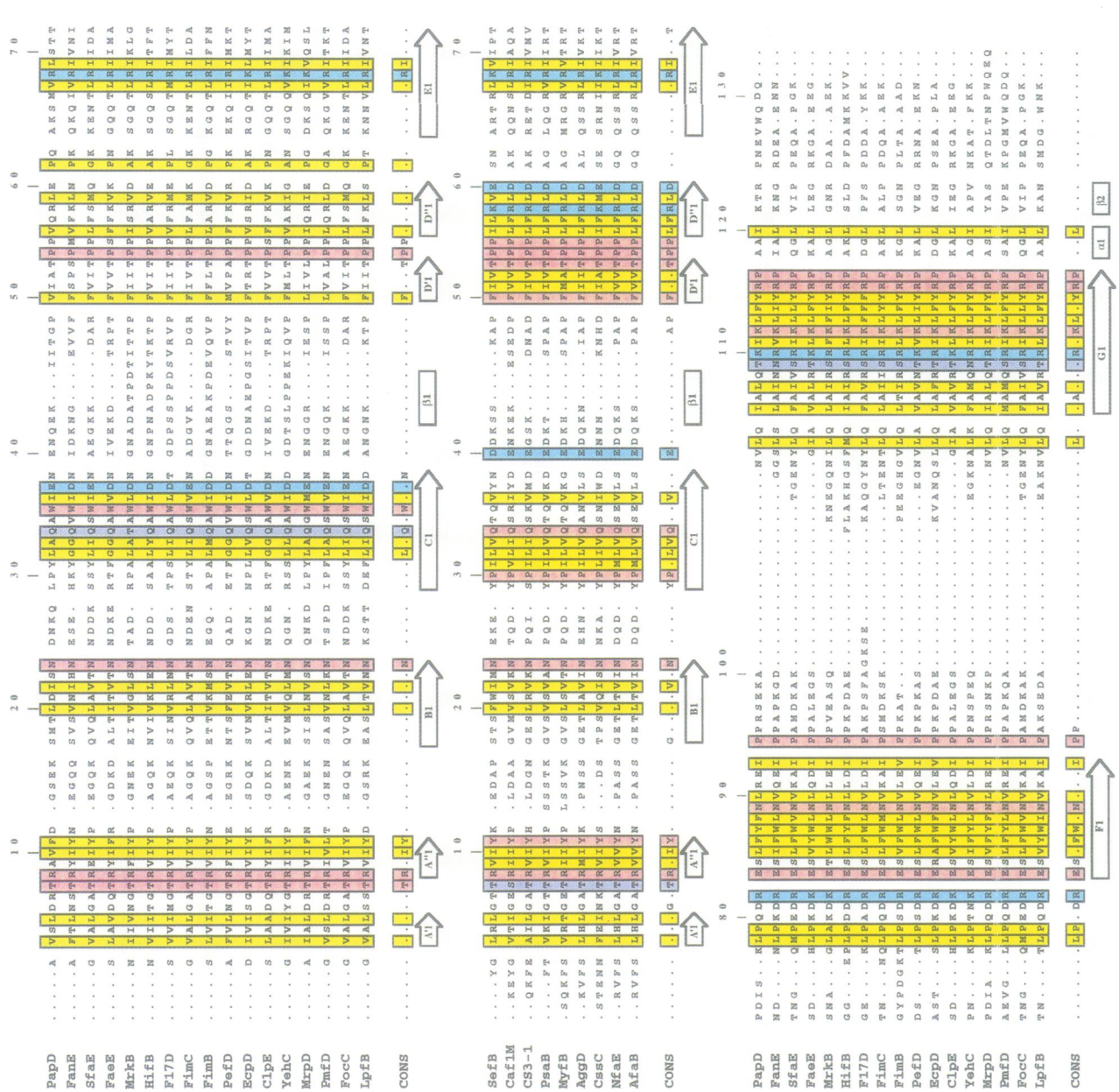
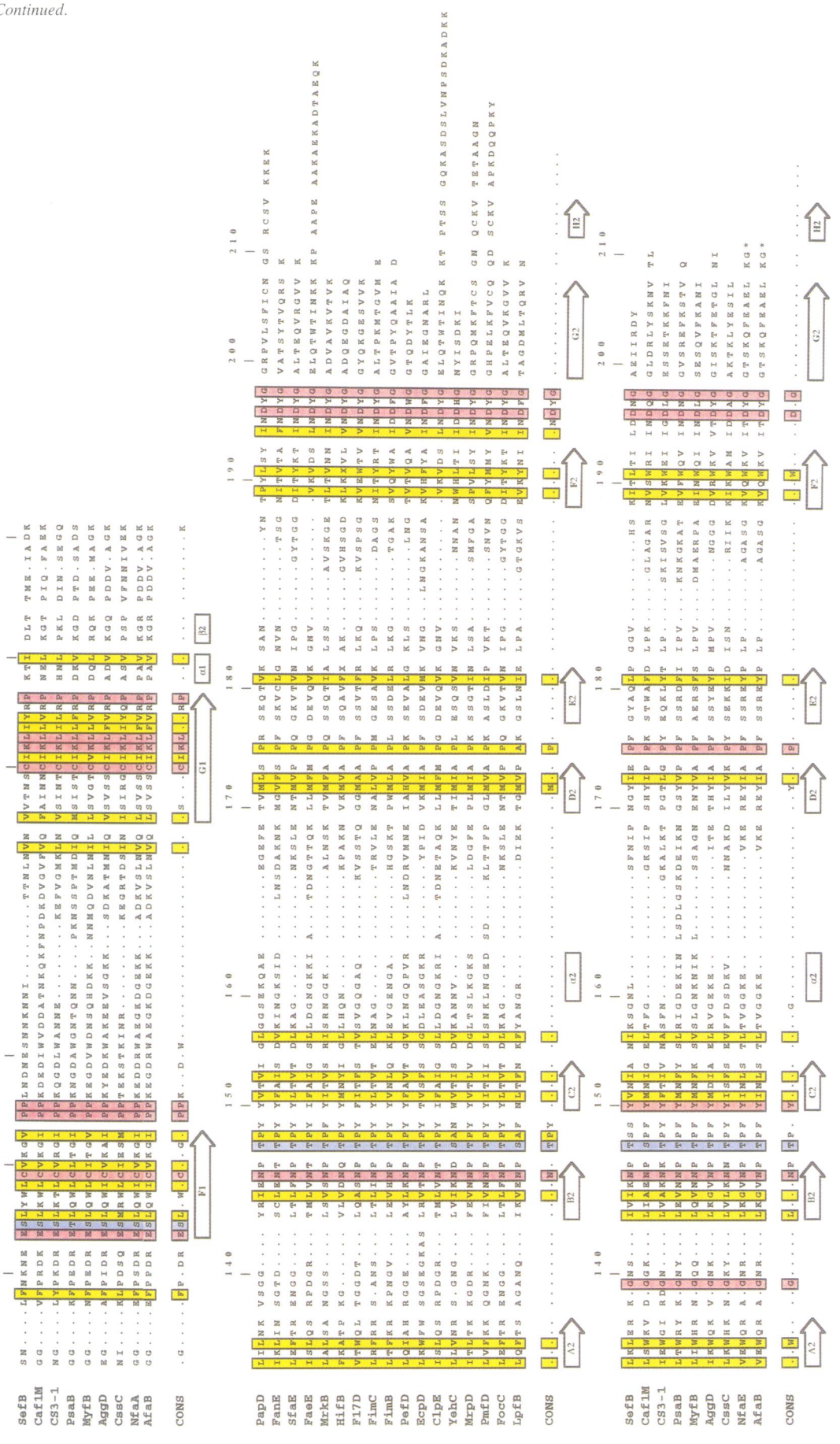


**Fig. 1.** Amino acid sequence alignment of the 26 known members of the chaperone superfamily. Sequences that belong to the FGS subfamily (PapD–LpfB) have been grouped at the top; FGL sequences (SefB–AfaB) are underneath. Secondary structure elements are indicated below each subgroup by arrows (β strands) and rectangles (α helices). Invariant residues and residues that are conserved in character are boxed and color coded using the following key: pink, invariant; yellow, conserved hydrophobic (A, L, V, I, P, M, W, F, C, Y, G); purple, conserved polar (N, Q, S, T, H); blue, conserved charged (D, E, K, R). Residues that are conserved in at least 70% of the sequences within each subfamily are given in the respective consensus sequences below each subfamily. Residues in the alignment are numbered according to PapD. *See Materials and methods for an explanation of the last 26 residues in these two chaperones.

binding specificities that contribute to the various host and tissue tropisms exhibited by the different pathogens (Table I). The most common type of adhesive organelles are the rod-like fibers called pili. Pilus rods have a wide range of diameters and helical symmetries. The adhesive component of the pilus is often present in a distinct subassembly, called a tip fibrillum, that is joined to the distal end of the thicker pilus rod. Seventeen of the 26 chaperones in the superfamily assemble rigid pilus-like rods ranging in diameter from 2 to 10 nm (Table I). All of the rods in this group that have been examined by high-resolution electron microscopy—P, S and type 1 pili of *E.coli*, and *Haemophilus influenzae* pili—have been shown to be composite structures consisting of thin, flexible tip fibrillae joined to the distal ends of thick rods (Kuehn *et al.*, 1994; Jones *et al.*, 1995; J.St.Geme, unpublished). The P pilus rod has a 15 Å helical cavity winding through the rod that communicates with the external environment by a set of radial channels (Bullitt and Makowski, 1995). The thinner pilus rods (2–5 nm in diameter) are more flexible and may be missing the axial hole (Kuehn *et al.*, 1992). The other nine chaperones in the superfamily assemble the second major class of adhesive organelles which are very thin fibers (<2 nm) and tend

**Fig. 1.** *Continued.*



to coil up into an amorphous mass on the surface of the bacterium (Table I). Very little is known about the fine molecular details of the architecture of this class of fibers. A subgroup of this class of adhesive organelle has been termed non-fimbrial (non-pilus-associated) adhesins (Goldhar *et al.*, 1987; Ahrens *et al.*, 1993; Le Bouguenec *et al.*, 1993; Garcia *et al.*, 1994). These adhesins are not thought to be a component of any oligomeric structure, although it is possible that these types of adhesins are also part of very thin fibers that have not yet been visualized by standard electron microscopy techniques.

### *Structural conservation in the chaperone superfamily*

The amino acid sequences of the 26 periplasmic chaperones required for the assembly of the adhesive organelles described above have been determined. All 26 chaperones have highly conserved structural features. The sequences of all the chaperones were aligned by superimposing them onto the known three-dimensional structure of PapD (Figure 1). Gaps were only introduced in loop regions and never in the β strands. Some conserved residues have been shown previously to play critical roles in chaperone function (Slonim *et al.*, 1992; Kuehn *et al.*, 1993). Ten residues are invariant in all 26 chaperones. Residues R8 and K112 at the crevice of the cleft form a critical part of the subunit binding site of PapD (Slonim *et al.*, 1992; Kuehn *et al.*, 1993). These residues are invariant in the entire chaperone superfamily and probably have a similar function in each chaperone. Mutations in these residues abolish the ability of PapD to bind subunits and mediate their assembly into pili (Slonim *et al.*, 1992; Kuehn *et al.*, 1993). In PapD, a buried interdomain charge–charge/hydrogen-bond network is formed by residues E83, R116 and D196. The positively charged side chain of R116 bridges the two negatively charged side chains, one from each domain. D196 and E83 are invariant, and R116 is conserved in 25 of the 26 members. This internal bridge is thought to orient the two domains towards one another, stabilizing the subunit binding cleft region between the domains (Holmgren *et al.*, 1992).

The other six invariant residues occupy critical points in loops or are involved in intramolecular interactions which orient loops. The proline at position 94 is the first residue after the F1 β strand and is located on the surface of domain 1. Both N24 and N145 are involved in positioning loop regions, between β strands B1–C1 and B2–C2, respectively. Residue P54 is at a bend in domain 1 where β strand D shifts from one sheet to another. This residue is part of a conserved patch of hydrophobic residues as discussed below. Residue P117 is located at the end of the G1 β strand just before the hinge connecting the two domains. G198 is positioned at a reverse turn between the F2 and G2 β strands (Holmgren *et al.*, 1992).

There are 42 residues conserved in character, 38 of them are hydrophobic, three are polar and one is positively charged. In addition, 11 residue positions are consensus positions with the same side chain in >70% of the sequences in each subfamily, as defined below. As expected, most of the conserved hydrophobic residues are part of the hydrophobic core of the protein. The non-polar nature of core-forming residues is generally highly conserved within a protein family. In contrast, structural constraints on surface-exposed residues are weak and they are generally only conserved if they are involved in specific functions, such as binding of ligands or receptor molecules. To provide an operational definition of buried and exposed residues in the PapD structure, we calculated the solvent accessibility (Miller *et al.*, 1987) for each residue in the sequence. Residues with a side chain accessibility of <10% were classified as buried and residues with a side chain accessibility >10% were classified as exposed. Of the 42 residues conserved in character, 31 are buried. Thirty of these buried side chains are hydrophobic and one is polar. Eight of the surface-exposed conserved residues are hydrophobic, one is positively charged and two are polar (Table II). All except one of the conserved surface-exposed residues are in domain 1 where they cluster in three main regions: (i) along the G1 edge strand that defines part of the subunit binding cleft; (ii) as a diagonal patch extending from the tip of domain 1 across the β barrel and away from the cleft over the back of the molecule; (iii) at the beginning of the F1 strand, close to the domain interface and the hinge of the molecule (Figure 2). Many of the conserved surface-exposed residues described above are located on the β sheet formed by strands A1″, G1, F1, C1 and D1″ in domain 1. We will refer to this sheet as the conserved sheet. Adjacent residues on the solvent face of the conserved sheet form rows across the sheet that we will denote as rows 1, 2, 3 and 4, starting from the bottom of the subunit binding cleft (Figure 3).

The G1 strand is at the binding cleft proximal edge of the conserved sheet and the residues in the first half of the strand (positions 105–110) are all highly exposed on the surface of PapD. Positions 103 and 105, just before and at the start of the G1 strand, are occupied by conserved hydrophobic residues. Positions 107 and 109 are hydrophobic and polar, respectively, in all except one sequence each and have calculated accessibilities of 64.9 and 44.1%, respectively. In the PapD–PapG peptide crystal structure, the three hydrophobic residues at positions 103, 105 and 107 were in register with the alternating hydrophobic residues of the β-zipper motif of PapG's COOH terminus (Kuehn *et al.*, 1993). Although there were only weak van der Waals interactions between the hydrophobic residues of the peptide and PapD, all of these side chains point in the same direction and together define a hydrophobic surface that most likely would be buried in a chaperone–protein complex. The conservation of the three most highly exposed hydrophobic residues at positions 103, 105 and 107 in PapD suggests that the β-zippering interaction may represent a general binding paradigm used by all 26 chaperones.

Apart from the cleft, the most prominent conserved surface feature of the chaperones is the hydrophobic band, or elongated patch, that stretches from the tip of domain 1 and diagonally across the back of the domain (Figure 2). Five of the 11 conserved surface-exposed residues, together with one invariant and two consensus residues, cluster on the surface of PapD to form this patch (Figure 4A). Three conserved hydrophobic residues in rows 3 and 4 of the conserved sheet, at positions 32, 56 and 93, together with invariant P54 and consensus P55, define the hydrophobic part of the patch. The two proline residues could be important for positioning residue 56 next to

**Table II.** Invariant and conserved residues in the chaperone superfamily

| Residue | Acc. (%) | Type | Location |
|---|---|---|---|
| V2 | 1.4 | conserved hydrophobic | core |
| L4 | 1.2 | conserved hydrophobic | core |
| T7 | 79.1 | conserved polar (T) | floor of cleft |
| R8 | 12.0 | invariant | floor of cleft |
| V10 | 16.0 | conserved hydrophobic (I) | domain interface |
| F11 | 0.0 | conserved hydrophobic (Y) | core |
| L20 | 0.0 | conserved hydrophobic | core |
| I22 | 0.0 | conserved hydrophobic | core |
| N24 | 0.0 | invariant | end of strand B1; H-bonds to B1–C1 loop |
| L32 | 37.3 | conserved hydrophobic | back |
| A33 | 0.0 | conserved hydrophobic | core |
| Q34 | 29.8 | conserved polar (Q) | back |
| I37 | 1.5 | conserved hydrophobic | core |
| V50 | 0.0 | conserved hydrophobic (F) | core |
| T53 | 43.7 | consensus T | back |
| P54 | 53.2 | invariant | back; D1-strand switch |
| P55 | 79.8 | consensus P | back; D1-strand switch |
| V56 | 53.7 | conserved hydrophobic | back |
| L59 | 1.7 | conserved hydrophobic | core |
| V67 | 0.2 | conserved hydrophobic | core |
| R68 | 56.4 | conserved charge (R) | back |
| L69 | 0.0 | conserved hydrophobic (I) | core |
| L78 | 15.0 | conserved hydrophobic | edge of core |
| P79 | 49.6 | consensus P | E1–F1 (DRE) loop; elbow |
| D81 | 63.0 | consensus D | E1–F1 (DRE) loop; elbow |
| R82 | 51.4 | consensus R | E1–F1 (DRE) loop; elbow |
| E83 | 1.0 | invariant | domain interface; H-bond to R116 (2.9 Å) |
| S84 | 16.4 | consensus S | F1 strand; elbow |
| L85 | 5.8 | conserved hydrophobic | domain interface |
| Y87 | 13.3 | consensus W | floor of cleft |
| F88 | 0.0 | conserved hydrophobic | core |
| L90 | 0.1 | conserved hydrophobic | core |
| I93 | 29.1 | conserved hydrophobic | back |
| P94 | 7.8 | invariant | back; start of F1–G1 loop |
| P95 | 2.1 | consensus P | back; beginning of F1–G1 loop |
| L103 | 58.3 | conserved hydrophobic | cleft; end of F1–G1 loop |
| I105 | 89.8 | conserved hydrophobic | cleft |
| I111 | 6.6 | conserved hydrophobic | edge of core |
| K112 | 30.3 | invariant | floor of cleft |
| L113 | 0.0 | conserved hydrophobic (L) | core |
| F114 | 4.2 | conserved hydrophobic | domain interface |
| Y115 | 4.9 | conserved hydrophobic | packs between A1–B1 and E1–F1 (DRE) loop |
| R116 | 7.6 | consensus R | domain interface; H-bonds to E83 (2.9 Å) and D196 (3.0 Å) |
| P117 | 0.0 | invariant | hinge |
| I120 | 7.6 | conserved hydrophobic | domain interface; packs against E1–F1 (DRE) loop |
| L132 | 0.0 | conserved hydrophobic | core |
| L134 | 0.0 | conserved hydrophobic | core |
| I143 | 0.0 | conserved hydrophobic | core |
| N145 | 0.0 | invariant | end of strand B2; H-bonds to T147 OH, Y149 O |
| T147 | 3.4 | conserved polar (T) | elbow |
| P148 | 30.1 | consensus P | elbow; edge of domain interface |
| V151 | 0.4 | conserved hydrophobic | core |
| V153 | 0.0 | conserved hydrophobic | core |
| L156 | 0.6 | conserved hydrophobic | core |
| L173 | 0.1 | conserved hydrophobic | core |
| P175 | 4.6 | conserved hydrophobic | edge of domain interface; packs against E1–F1 (DRE) loop |
| V181 | 7.4 | conserved hydrophobic | core |
| P189 | 0.1 | conserved hydrophobic | core |
| L191 | 0.0 | conserved hydrophobic | core |
| I194 | 19.4 | conserved hydrophobic | floor of cleft |
| D196 | 24.6 | invariant | domain interface; H-bonds to R116 (3.0 Å), 2 waters |
| G198 | 32.0 | invariant | F2–G2 turn |

Residues are classified as invariant, conserved in character or consensus (present in >70% of the sequences in each subfamily). The consensus residue type for character conserved residues is given in parentheses. Acc. is the relative accessibility of the side chain compared with the same side chain X in a tripeptide Gly-X-Gly (Miller *et al.*, 1987). When X is a glycine, the listed value refers to the residue accessibility.

residues 32, 34 and 93, by maintaining the switch of strand D1 between sheets. Two polar and one positively charged residue are also associated with the conserved band (Figure 2C). The positive charge at position 68, at one end of the elongated patch, is conserved in all of the chaperones. Position 34 is almost invariantly a glutamine residue (T in one) and position 53 is a threonine residue in a majority of the chaperones. Two highly conserved

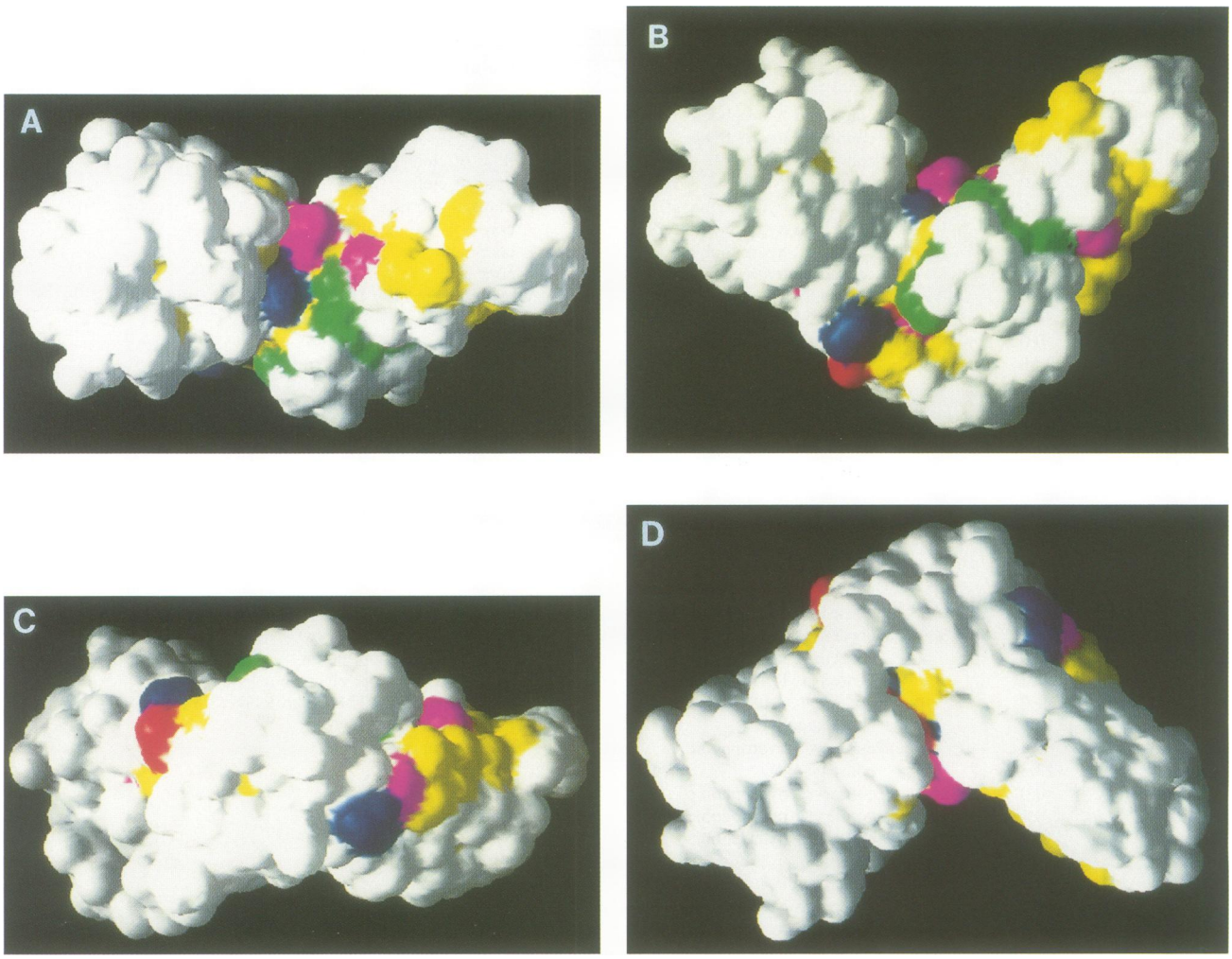



**Fig. 2.** Conserved surface features of PapD-like chaperones. The solvent-accessible surface of PapD has been color coded yellow for conserved hydrophobic residues, cyan for conserved polar residues, red for conserved negative and blue for conserved positively charged residues. Residues that are differently conserved within each subfamily are green. Four different views of the molecule at 90° intervals are shown with domain 1 to the right and domain 2 to the left in each figure. (**A**) View down the cleft. The crevice of the cleft is defined by polar and invariant charged (R8 and K112) side chains. Side chains that are different but conserved within each subfamily (green) cluster in the immediate vicinity of the invariant cleft residues. Apolar residues (yellow) in the G1 strand form a hydrophobic ridge along one side of the cleft. (**B**) The conserved side of the molecule. Part of the elongated hydrophobic patch across the side and around the back of domain 1 is visible to the right. The conserved feature near the elbow of the molecule is formed by residues in the DRE loop (see the text). (**C**) View from the back of the molecule. The elongated hydrophobic patch is clearly seen to the right in the picture. To the left of the elbow is a second view of the conserved DRE loop. (**D**) The non-conserved side of the molecule. This large V-shaped surface is almost entirely comprised of variable residues. The figure was prepared using GRASP (Nicholls, 1993).

proline residues, P94 and P95, are near the surface at the opposite end of the patch to R68. The side chain of L32 packs against the P95 ring. As for P54 and P55, these two proline residues could be important for proper positioning of residues in the patch.

The loop connecting strand E1 to F1, and the beginning of strand F1, near the hinge of the molecule, contains eight highly conserved residues: L78, P79, D81, R82, E83, S84, L85 and Y87. Three of these residues, P79, D81 and R82, are highly solvent exposed, whereas the rest are buried, or exposed only to a limited extent (Table II). Leucine at position 85 is buried in the domain interface where it packs against Y150 and M172 in domain 2. The invariant glutamic acid residue at position 83 is also buried in the domain interface where it forms a charge–charge/hydrogen-bond interaction with the side chain of R116. This almost invariant side chain in turn forms a second charge–charge/hydrogen-bond interaction to an invariant aspartic acid residue in domain 2, D196. The conserved charge–charge/hydrogen-bond network at the domain interface has been suggested to be important for orienting the domains with respect to each other (Holmgren *et al.*, 1992). The three residues P79, D81 and R82 are highly solvent exposed, and form a small patch near the hinge region of the molecule (Figure 2B). It is conceivable that interactions with other proteins (e.g. subunits or ushers) in this region could cause conformational changes that would be transmitted through the E1–F1 loop to the interdomain region, possibly disrupting the conserved charge–charge/hydrogen-bond network and triggering a reorientation of the domains.

### *FGS and FGL chaperones*

The chaperone sequences fall into two groups differentiated by conserved differences in rows 1 and 2, and by the

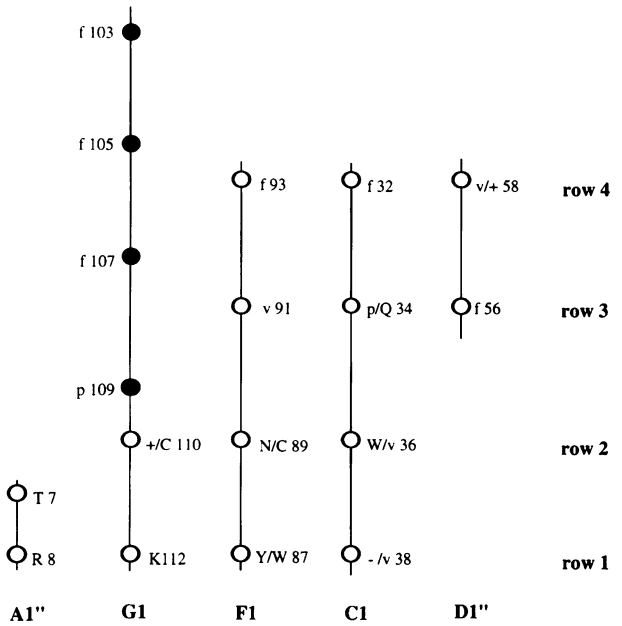



**Fig. 3.** Schematic diagram of the conserved sheet in domain 1 of PapD highlighting conserved rows in the two subfamilies of chaperones. The lines represent β strands in the 'conserved sheet' (A1", G1, F1, C1 and D1"), and each circle represents a residue in the β strand and is numbered according to the numbering for PapD. Only solvent-exposed positions are shown. Residue side chains that point up from the plane of the paper are represented by open circles, those that point into the plane of the paper by filled circles. Invariant residues are indicated by single-letter amino acid code in capitals. For other residues, the following key is used: f, conserved hydrophobic residue at that position; p, conserved polar residue at that position; v, variable residue at that position; +, conserved positively charged residue at that position; –, conserved negative residue at that position. Two symbols separated by a slash (/) are shown when different residues are conserved within each subfamily. In these cases, the letter on top represents the residue conserved in the FGS subfamily and the letter on the bottom represents the residue conserved in the FGL subfamily.

number of amino acids in the loop between β strands F1 and G1, termed the F1–G1 loop. Interestingly, this structural subdivision of the chaperone superfamily groups the chaperones into those that assemble pili versus those that assemble non-pilus organelles that have an atypical morphology. We have denoted the 17 chaperones that have a short F1–G1 loop ranging in size from 10 to 20 (only one having 20) residues as part of the FGS subfamily, deriving their name from the loop length (F1–G1 Short). The other nine chaperones have a long F1–G1 loop that ranges in size from 21 to 29 (only one has 21 residues). We have classified these chaperones as part of the FGL subfamily (F1–G1 Long).

### *Structural features unique to each subfamily*

Apart from the difference in the length of the F1–G1 loop, conserved structural differences between the two chaperone subfamilies are clustered in rows 1 and 2 across the conserved sheet (Figures 2, 3 and 4B). In PapD, row 1 residues are R8, K112, Y87 and E38, and row 2 residues are K110, N89 and W36. Residues N89 and W36 in row 2 are invariant, and position 110 is always a positively charged residue in the FGS subfamily. In contrast, in the FGL subfamily, row 2 positions 110 and 89 are occupied by invariant cysteine residues which by modeling are predicted to make a disulfide bond. In PapD, K110 participates in a salt link to E38 in row 1. Position 38 is occupied by a negatively charged residue in all of the FGS chaperones, but is variable in the FGL subfamily. No member of the FGL subfamily has a tryptophan at position 36 in row 2. Instead, the FGL chaperones always have a polar or charged residue at this position. Row 1 residues at positions 8 and 112 are occupied by invariant R8 and K112 throughout the entire superfamily, and position 87 is an aromatic Y or W residue throughout the whole superfamily. In PapD, both R8 and K112 have previously been shown to play critical roles in subunit binding and chaperone function (Slonim *et al.*, 1992; Kuehn *et al.*, 1993). Most of the side chains in rows 1 and 2 have quite low solvent accessibilities (Table II), although all of them are on the surface of the protein with their side chains pointing into solution. The long aliphatic parts of these side chains pack against each other, thus leaving only their polar or charged head groups solvent accessible, except for N89 which is completely buried beneath the side chains of W36, E38, Q108 and K110. The conserved residues in rows 1 and 2 may thus be expected to have rather fixed positions in comparison to side chains in loop regions, which often are highly flexible. The clustering of residues that are conserved differently in the two chaperone subfamilies (Figures 2 and 4B) in a region on the chaperone surface known to be critical in subunit binding probably reflects unique requirements within each subfamily for binding target subunits or for making interactions with the outer membrane usher.

### *A variation in the β-zipper motif recognized by FGS and FGL chaperones*

Subunits assembled by FGS chaperones have the following structural characteristics: (i) two cysteines spaced ~30 amino acids apart in the $NH_2$-terminal region that form disulfide bonds (Normark *et al.*, 1986; Simons *et al.*, 1990; Jacob-Dubuisson *et al.*, 1993b); (ii) a conserved pattern of alternating hydrophobic residues at positions 4, 6 and 8 from the COOH terminus; (iii) a penultimate tyrosine; (iv) a glycine at position 14 from the COOH terminus (Normark *et al.*, 1986; Kuehn *et al.*, 1993) (Figure 5A). The last three criteria are part of a conserved β-zipper motif recognized by the chaperone that was defined in detail by co-crystallizing a peptide corresponding to the COOH terminus of PapG with PapD (Kuehn *et al.*, 1993).

Subunits assembled by the FGL chaperones contain a variation of the β-zipper motif at their COOH terminus. All but four of the known subunits have a tyrosine at position 3 from the COOH terminus and one of the four has an aromatic residue (tryptophan) at this position. There are at least two alternating hydrophobic residues in each of the FGL subunits, most being at positions 6 and 8 from the COOH terminus. A majority of the FGL subunits have a tyrosine at position 12 from the COOH terminus. The glycine at position 14 from the COOH terminus is highly conserved in these subunits, only two have different residues (Figure 5B).

The structural and functional relatedness amongst different members of the chaperone superfamily and the subunits that they assemble was borne out by showing that PapD was capable of functionally substituting for the FimC chaperone in binding type 1 pilus subunits and mediating

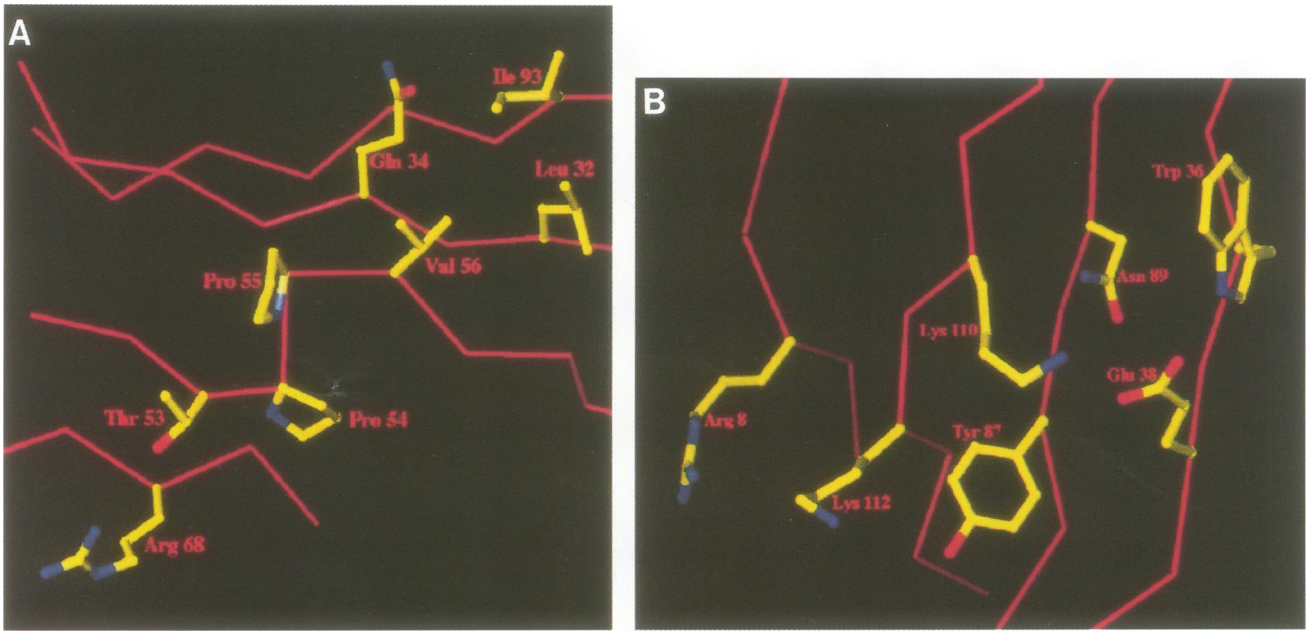



**Fig. 4.** (**A**) Close-up of residues in PapD forming the conserved diagonal patch over the side and back of the chaperone. Eight highly conserved residues cluster to form a mostly hydrophobic patch over the side and back of the molecule. The conserved hydrophobic character of this patch suggests that it is involved in protein–protein recognition and binding. (**B**) Close-up of residues in rows 1 (R8, K112, Y87, E38) and 2 (K110, N89, W36) of the conserved sheet in PapD. In the FGS family, N89 and W36 are invariant, and position 110 is a conserved positive charge. In contrast, in the FGL subfamily, row 2 positions 110 and 89 are occupied by invariant cysteine residues which by modeling are predicted to make a disulfide bond. In PapD, K110 participates in a salt link to E38 in row 1. Position 38 is occupied by a negatively charged residue in all of the FGS chaperones, but is variable in the FGL subfamily. No member of the FGL subfamily has a tryptophan at position 36 in row 2. Instead, the FGL chaperones always have a polar or charged residue at this position. Figures prepared with O (Jones *et al.*, 1991).

their assembly into type 1 pili (Jones *et al.*, 1993). Part of the PapD–subunit interaction involves the β-zippering interaction between the G1 β strand of PapD and the COOH terminus of the subunit which was elucidated by X-ray crystallography (Kuehn *et al.*, 1993). The β-zippering interaction can be measured by testing the ability of the chaperone to bind to COOH-terminal peptides (Kuehn *et al.*, 1993). We used the peptide binding assay to investigate whether the FGS chaperone, PapD, was capable of binding peptides corresponding to the COOH terminus of subunits assembled by FGL chaperones. PapD bound strongest to the PapG peptide, as shown previously (Kuehn *et al.*, 1993). Surprisingly, PapD also bound strongly or moderately to peptides corresponding to the COOH terminus of AggA, MyfA, C53-3 and F1 which are subunits that are assembled into very thin fibers with an atypical morphology by FGL chaperones. PapD bound only weakly to the P5A peptide (Figure 6A). The R8A mutant PapD had decreased ability to bind to all of the peptides (Figure 6B). These results demonstrate that the variation of the β-zipper motif present in the COOH terminus of FGL subunits is recognized by the FGS chaperone, PapD. The effect of the R8A mutation in PapD in reducing the binding to the FGL subunit peptides argues that the molecular anchoring interaction in the cleft is also critical in recognizing these peptides. The binding of PapD to the COOH-terminal peptides of Pap subunits also varies considerably (Kuehn *et al.*, 1993). PapD has recently been co-crystallized with the COOH-terminal PapK peptide (D.Ogg and S.J.Hultgren, unpublished data). Even though it is one of the weakest binding Pap peptides, it bound in the cleft of PapD in an analogous manner as the PapG COOH-terminal peptide (D.Ogg and S.J.Hultgren, unpublished data). The molecular basis for what determines the difference between strong and weak binding peptides is not known. The differences in affinity may reflect differences in the ability of the peptides to adopt a conformation that fits into the cleft of PapD and along the exposed edge of the conserved sheet.

## Discussion

The development of a diverse and widespread family of adhesive organelles in Gram-negative bacteria requires periplasmic chaperones that have immunoglobulin-like three-dimensional structures. The chaperones utilize their immunoglobulin-like structures in novel β-zippering recognition processes to cap associative surfaces on subunits to prevent premature assembly in the periplasm (Kuehn *et al.*, 1991). Dissociation of the chaperone from the subunit coincides with the assembly of the subunit into the adhesive fiber and requires outer membrane proteins called ushers (Dodson *et al.*, 1993). The chaperone/usher pathway is used to assemble two different families of adhesive fibers in Gram-negative bacteria that have dramatically different molecular architectures. One family of adhesive fibers are known as pili. These are rod-like structures that range in diameter from 2 to 10 nm and are comprised of subunits arranged in different helical symmetries (Table I). Rod-like fibers are typically part of composite organelles that contain thin fibrillae joined to their distal ends. These so-called tip fibrillae are composed of subunits arranged in a linear array and often contain the adhesive component of the pilus (Kuehn *et al.*, 1994; Jones *et al.*, 1995; J.St.Geme, unpublished). The second class of adhesive fibers are not associated with pilus rods,

but instead consist only of a thin linear fibrillum. In some pathogenic bacteria, these very thin fibers (<2 nm) are long and flexible, and often coil up into an amorphous mass on the bacterial surface (Levine *et al.*, 1984; Knutton *et al.*, 1989; Wolf *et al.*, 1989; Müller *et al.*, 1991; Karlyshev *et al.*, 1992; Clouthier *et al.*, 1993; Iriarte *et al.*, 1993; Lindler and Tall, 1993; Savarino *et al.*, 1994). Other adhesins that have been classified as afimbrial adhesins are either monomeric or part of very short oligomeric structures such as those found at the distal ends of pilus rods (Goldhar *et al.*, 1987; Ahrens *et al.*, 1993; Le Bouguenec *et al.*, 1993; Garcia *et al.*, 1994). Fine molecular details of the architecture of this second class of adhesive fibers are lacking. The immunoglobulin-like superfamily of chaperones can be divided into two subfamilies based on structural differences conserved within each subgroup. We discovered that one subfamily of chaperones assembles pilus-like rods, whereas the other subfamily assembles very thin fibers that have an atypical morphology or are non-fimbrial adhesins. An analysis of the fine molecular details that distinguish these two subgroups of chaperones and the subunits that they assemble has given insight into the molecular basis of the development of diverse adhesive organelles in pathogenic Gram-negative bacteria.

All of the periplasmic chaperones have a conserved hydrophobic core that maintains the overall immunoglobulin-like features of the two domains. The two immunoglobulin-like domains are oriented in a boomerang-shaped structure, creating a cleft between the domains. The chaperone cleft contains surface-exposed residues that are highly conserved or invariant throughout the entire superfamily. Surface-exposed residues are generally only conserved if they are involved in specific functions, such as binding of ligands or receptor molecules. Accordingly, the invariant residues R8 and K112 in the crevice of the cleft form a critical part of the subunit binding site (Slonim *et al.*, 1992; Kuehn *et al.*, 1993). In the PapD–$PapG_{peptide}$ crystal structure, the alternating hydrophobic residues in the peptide are in register with three of the conserved surface-exposed hydrophobic residues of PapD that are at positions 103, 105 and 107 extending from the F1–G1 loop into the G1 β strand (Kuehn *et al.*, 1993). The highly conserved nature of these interactions suggests that the anchoring and β-zippering mechanism is utilized as part of a subunit-recognition paradigm by all of the chaperones.

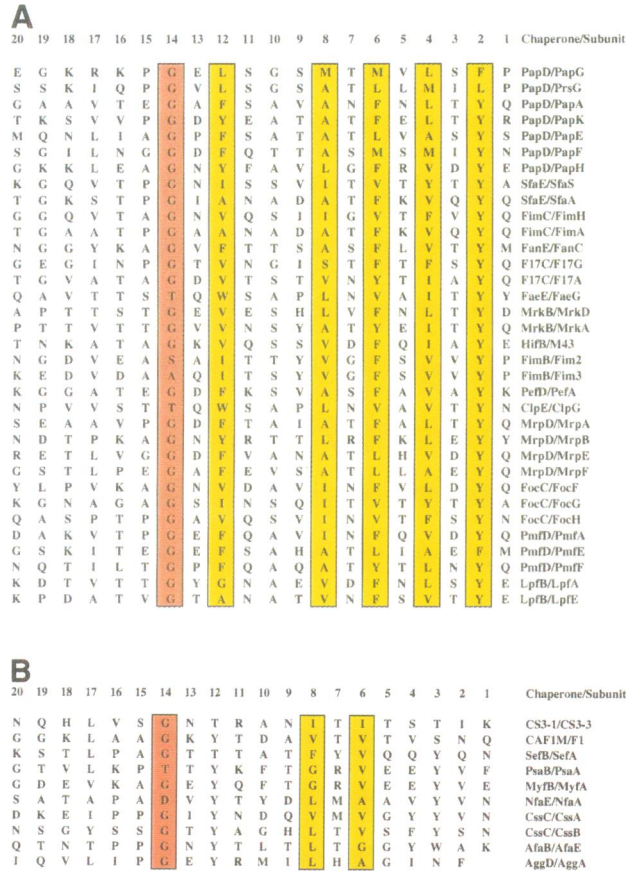



**Fig. 5.** (**A**) Conserved β-zipper motif present in subunits assembled by FGS chaperones. Sequences are numbered for each subunit with 1 being the COOH-terminal residue in the subunit. Key: yellow boxed residues, conserved alternating hydrophobic residues; orange boxed residues, conserved glycine at position 14 from the COOH terminus. (**B**) Variation of the β-zipper motif present in subunits assembled by FGL chaperones. Sequences are numbered as in (A). Notice the highly conserved tyrosine at position 3, and the hydrophobic residues at positions 6 and 8. The key is the same as in (A).

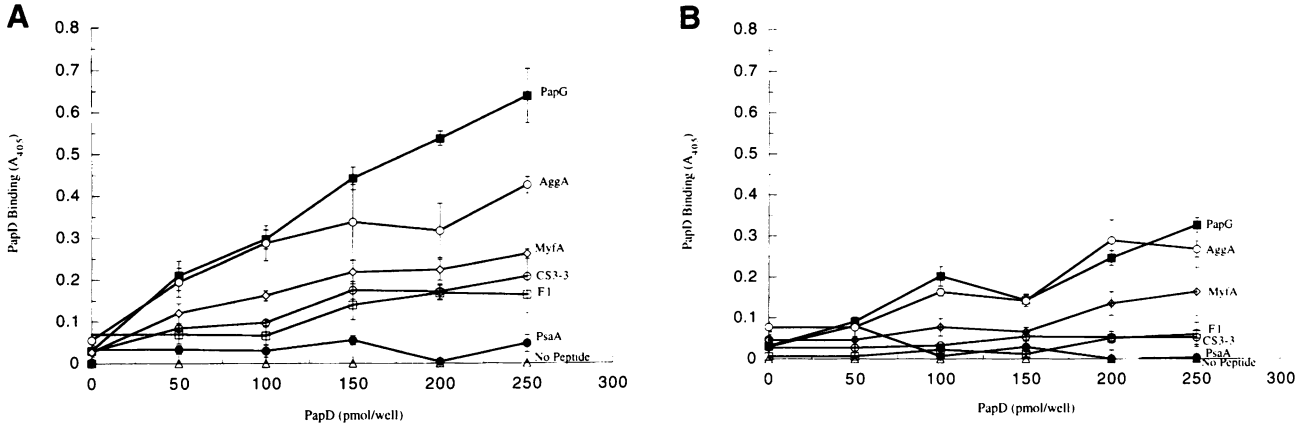



**Fig. 6.** (**A**) PapD binds to FGL subunit-related peptides. Eight synthetic peptides corresponding to COOH-terminal residues of the subunits indicated were coated to wells of microtiter plates. The concentration of each peptide is 2 nmol/well. Binding of 50–250 pmol/50 μl PapD to the immobilized peptides was determined by ELISA using anti-PapD antiserum. (**B**) PapD with R8A mutation has little affinity for FGL subunit-related peptides. The binding of 50–250 pmol/50 μl R8A PapD to the immobilized peptides was determined as in (A). Each graph represents the average of triplicate wells with the SD shown with bars.

In addition to the hydrophobic residues at positions 103, 105 and 107, there are eight other surface-exposed residues that are conserved in character in the chaperone superfamily: five more are hydrophobic, one is positively charged and two are polar (Table II). Many of these conserved surface-exposed residues are located on the β sheet formed by strands A1″, G1, F1, C1 and D1″ in domain 1 that we have defined as the conserved sheet. The surface-exposed residues on the conserved sheet form four rows across the sheet, extending from the crevice of the subunit binding cleft (Figure 3). Owing to the twist of the sheet, these residues are found both in the cleft (rows 1 and 2) and in the conserved diagonal patch across one side and around the back of domain 1 of the molecule (rows 3 and 4). Almost all of the conserved surface-exposed residues are in domain 1, either in the cleft (Figure 2A), at the side of the domain defined by the conserved sheet (Figure 2B) or at the back of the domain (Figure 2C). The side of domain 1 that is opposite to the conserved sheet is almost entirely comprised of variable residues, as is all of domain 2. This side of domain 1 together with the adjacent side of domain 2 form an extended V-shaped surface that defines the non-conserved side of the molecule (Figure 2D).

Pilus biogenesis depends on the targeting of chaperone–subunit complexes to the usher (Dodson *et al.*, 1993). Dissociation of the chaperone from the subunit exposes the surface on the subunit required for its assembly into the pilus structure. Nothing is known about the molecular basis of the interactions that ushers make with chaperone–subunit complexes. In contrast, it is known that the exposed edge of the G1 β strand of PapD and the crevice of the cleft form a critical part of the subunit binding site and are highly conserved in the entire chaperone superfamily. All of the chaperones probably utilize these conserved features to recognize subunits and form chaperone–subunit complexes. PapD, but not R8A or K112A mutant PapDs, will functionally substitute for FimC in the assembly of type 1 pili (Jones *et al.*, 1993), arguing that the subunit binding surfaces are conserved between the two chaperones and probably throughout the whole chaperone superfamily. Other chaperones are not able to assemble subunits from heterologous systems unless the chaperone and usher are molecular partners from the same assembly system (Klemm *et al.*, 1995), suggesting that an usher may interact in part with non-conserved surfaces of a chaperone presented to the usher in a chaperone–subunit complex.

In the absence of PapC, chaperone–subunit complexes accumulate in the periplasmic space, arguing that the usher is required to facilitate the dissociation of the chaperone from subunits, allowing their assembly into pili. Binding of PapD to the PapG COOH-terminal peptide was shown to induce a 13° jaw-closing or hinge-bending motion, making the angle between the two domains more acute (Kuehn *et al.*, 1993). Therefore, dissociation of the chaperone may be triggered by a rearrangement of chaperone domains that induces its release from the subunit. The binding of chaperone–subunit complexes to an usher, *in vitro*, did not trigger the dissociation of the chaperone (Dodson *et al.*, 1993). Jones *et al.* (1995) argued that dissociation of the chaperone is triggered upon an interaction of a chaperone–subunit–usher complex with an incoming chaperone–subunit complex. However, the mechanism of how this myriad of protein–protein interactions may trigger chaperone dissociation is unknown.

The charge–charge/hydrogen-bond network at the domain interface involving two carboxylate side chains (E83 in domain 1 and D196 in domain 2) bridged by an arginine side chain (R116) is invariant in all except one of the chaperones. This bridge is thought to be important for proper orientation of the domains with respect to each other (Holmgren *et al.*, 1992). The glutamic acid E83 is part of the consensus sequence DRES at the end of the E1–F1 loop which is located near the hinge of the molecule. The consensus residues P79, D81 and R82 in or near the DRE loop are all solvent exposed, their side chains forming a small patch on the conserved side of PapD (Figure 2B). The backbone of the DRE loop packs against residues 117–120 in the hinge region connecting the two domains. It is conceivable that interactions with other proteins (e.g. subunits or ushers) in this region could cause conformational changes that would be transmitted to the interdomain region, possibly disrupting the conserved charge–charge/hydrogen-bond network and triggering a reorientation of the domains. We are currently testing the hypothesis that incoming chaperone–subunit complexes may interact directly with the solvent-exposed residues in the DRES motif of the chaperone that is part of the chaperone–subunit–usher complex, which could trigger the release of the chaperone via a conformational change.

In this context, it is interesting to note that a similar motif (DREA) is present at the sequentially and topologically equivalent position in another class of immunoglobulin-like molecules called cadherins. Cadherins are a class of eukaryotic cell surface molecules that have tandemly repeated immunoglobulin-like domains in their extracellular regions (Shapiro *et al.*, 1995), and are important for controlling the development and maintenance of tissues. In cadherins, the DREA motif is involved in calcium ion binding which orients successive cadherin domains in a rigid conformation on the cell surface where they participate in cell–cell interactions. The glutamic acid side chain of the DREA motif, together with negatively charged side chains from other parts of the cadherin domain, forms a calcium binding site at the prospective interface between successive cadherin domains (Shapiro *et al.*, 1995). In chaperones, instead of binding calcium, E83 interacts with the positively charged side chain of R116. Thus, a similar motif, at the same location in both cadherins and periplasmic chaperones, may be involved in a similar function. In both cases, the glutamic acid side chain is part of a binding site for a positively charged group (a calcium ion in cadherins and an arginine side chain in the chaperones) that bridges successive domains. In both classes of molecules, these bridging interactions are thought to be important for proper orientation of successive immunoglobulin-like domains.

At the crevice of the cleft, row 1 contains the invariant R8 and K112 subunit binding residues. Conserved differences between the two subfamilies are concentrated to the immediate vicinity of these residues. In the FGS subfamily, row 2 residues W36 and N89 are invariant, and position 110 is always occupied by a positive charge. In the FGL subfamily, position 36 is never a tryptophan, but instead is occupied by either a polar or charged residue. In the

FGL subfamily, positions 89 and 110 are invariant cysteine residues that by modeling are predicted to form a disulfide bond. The chaperones containing C89 and C110 possess a long F1–G1 loop, and thus were named FGL chaperones. The other subfamily of chaperones contain a short F1–G1 loop and were named FGS chaperones. Given the increased number of residues between the F1 and G1 β strands in the FGL chaperones, the two cysteines may play a role in folding and/or in stabilizing the structure. The F1–G1 loop occurs at the same relative position in a chaperone molecule as the hypervariable region three (CDR3) in an immunoglobulin molecule which forms part of the antibody combining site (Amit *et al.*, 1986; Holmgren *et al.*, 1992; Hultgren *et al.*, 1993b). CDR3 has the highest degree of variability in its amino acid composition and thus provides for much of the specificity in an antibody–antigen interaction (Amit *et al.*, 1986). The F1–G1 loop is also variable in length and sequence, and may provide part of the specificity in chaperone–subunit interactions (Holmgren *et al.*, 1992; Hultgren *et al.*, 1993b).

All subunits assembled by the FGS chaperones contain the β-zippering motif in their COOH terminus that is defined as a penultimate tyrosine, alternating hydrophobic residues at positions 4, 6 and 8, and a glycine at position 14 from the COOH terminus (Kuehn *et al.*, 1993). The subunits assembled by the FGL chaperones contain a variation of this β-zipper motif. The majority of these subunits contain a tyrosine at position 3 from the COOH terminus, two alternating hydrophobic residues at positions 6 and 8, a tyrosine at position 12 and a glycine at position 14 from the COOH terminus. FGL chaperones contain the invariant R8 and K112 residues, as well as the conserved alternating hydrophobic residues in the G1 β strand that are known to form a critical part of the subunit binding site of FGS chaperones. This conservation argues strongly that FGL chaperones utilize a similar subunit binding paradigm. The ability of PapD, and the reduced ability of R8A PapD, to bind to the AggA, MyfA, C53-3 and F1 COOH-terminal peptides demonstrates that the COOH termini of FGL subunits bind in the PapD cleft, even though their sequences represent a variation of the β-zipper motif. The affinity of PapD for different peptides may in each case be determined by the specific packing interactions made between side chains in the peptide and in PapD, and by the capability of each peptide to adopt a conformation that fits into the PapD cleft and along the exposed edge of the G1 β strand. Whether PapD is capable of binding FGL subunits and forming assembly-competent complexes is not yet known.

Residues in the β-zipper motif of the PapA subunit of P pili have been shown to participate in subunit–subunit interactions after the dissociation of the chaperone (E.Bullitt, C.H.Jones, R.Striker, G.Soto, F.Jacob-Dubuisson, J.Pinkner, M.J.Wick, L.Makowski and S.J.Hultgren, in preparation). If the COOH-terminal residues of subunits assembled by FGL chaperones participate in subunit–subunit interactions, the structural alteration of the β-zipper motif may be transmitted throughout the entire fiber, resulting in the subunits being assembled into linear fibers as opposed to helical rods. Thus, the variation of the β-zipper motif in the FGL subunits may evoke a requirement for a different type of subunit–subunit interaction that has important ramifications on the molecular architecture of the fiber that is formed. Alternatively, it may preclude subunit–subunit interactions, resulting in afimbrial adhesins.

In conclusion, the division of the PapD-like chaperone superfamily into the two subfamilies is supported by both structural and functional data. The implications of this division are far and widespread. All chaperones in the superfamily have the same invariant residues involved in anchoring subunits in the cleft of PapD. However, the ability of the FGL chaperones to assemble subunits, which have variations of the β-zipper motif in the COOH terminus, into atypical fibers or afimbrial adhesins suggests an underlying difference that separates the chaperones into two subfamilies. The long-term goal of being able to design novel therapeutics to inhibit the assembly of virulence-associated organelles requires knowledge of the molecular mechanisms used by pathogenic organisms to assemble virulence-associated adhesive surface structures. The division of this superfamily is an important step in elucidating the molecular details of the assembly of surface organelles.

## Materials and methods

### *Amino acid sequence alignment*

The sequences of all the chaperones were aligned by superimposing them onto the known three-dimensional structure of PapD. Gaps were only introduced in loop regions and never in the β strands. Conserved residues were grouped in three classes: invariant, consensus or conserved in character. Residue positions where the same residue is present in >70% of the sequences in each subfamily were classified as consensus residues. Positions were classified as conserved in character if all side chains at that position belonged to one of three character groups: hydrophobic (A, C, F, G, I, L, M, P, V, W, Y), polar (H, S, T, N, Q) or charged (D, E, K, R). The published sequences of AfaB (Garcia *et al.*, 1994) and NfaE (Ahrens *et al.*, 1993) are shorter than any of the other chaperones, ending after the fifth strand of domain 2. In the sequence alignment we have used the AfaB and NfaE sequences that would result if, for the last 26 residues, the reading frame is shifted by one additional nucleotide position (S.D.Knight, unpublished). Residues in the alignment are numbered according to PapD so that residues in β strands have the same numbering as PapD even if there are extra residues in between. Continued use of this numbering system is suggested for future analysis of this superfamily of proteins.

### *Peptide synthesis*

Synthetic peptides were constructed on a Perkin Elmer–Applied Biosystems (Foster City, CA) 430A peptide synthesizer, using standard t-Boc chemistry and the stepwise solid-phase approach of Merrifield (1963). Following their synthesis, peptides were deprotected and cleaved from the resin with anhydrous liquid HF containing a 'cocktail' of appropriate scavengers. The cleaved peptide materials were analyzed and HPLC purified to >90% homogeneity using Vydac (Hesperia, CA) C8 and C18 reversed-phase columns developed with a TFA-$H_2O$/$CH_3CN$ gradient system. The amino acid content of each purified peptide was found to be consistent with theoretical values. The peptide sequences are as follows: AggA, $H_2N$-IPGEYRMILHAGINF-$CO_2H$; CS3-3, $H_2N$-SGNTRANITITSTIK-$CO_2H$; F1, $H_2N$-AGKYTDAVTVTVSNQ-$CO_2H$; MyfA, $H_2N$-KAGEYQFTGRVEEYVE-$CO_2H$; PapG, $H_2N$-KRKPG-ELSGCMTMVLSFP-$CO_2H$; PsaA, $H_2N$-KPTTYKFTGRVEEYVF-$CO_2H$. The PapG peptide contains a non-native C in place of native S residue at position nine from the COOH terminus. This change will not effect the binding since in the crystal structure the hydrogen-bonding pattern between PapD and the PapG COOH terminus peptide broke at position nine (Kuehn *et al.*, 1993).

### *Peptide ELISA*

Stock solutions of peptides (0.2 mg) were dissolved in water or 1% acetic acid to a concentration of 2 nmol/μl. Peptides were diluted in phosphate-buffered saline (PBS) (120 mM NaCl, 2.7 mM KCl, 10 mM PBS, pH 7.4) to 2 nmol/50 μl and coated overnight onto microtiter wells

with 50 µl/well at 4°C. The ELISA assay was carried out as described previously (Kuehn *et al.*, 1993). Briefly, the wells were then washed with PBS and blocked with 3% bovine serum albumin (BSA) in PBS (BSA–PBS) for 2 h at 25°C. The wells were washed three times with PBS and incubated with 50 µl of PapD diluted at 50–250 pmol PapD in 50 µl of 3% BSA–PBS for 45 min at 25°C. The wells were washed three times with PBS and incubated with a 1:500 dilution of mouse anti-PapD antiserum in 3% BSA–PBS for 45 min at 25°C. The wells were washed three times with PBS and then incubated with a 1:1000 dilution of goat antiserum to mouse IgG (immunoglobulin G) coupled to alkaline phosphatase (Sigma) in 3% BSA–PBS for 45 min at 25°C. The wells were washed three times with PBS and three times with developing buffer (10 mM diethanolamine, 0.5 mM $MgCl_2$). For developing, 50 µl of substrate (50 µl of filtered 1 mg/ml *p*-nitrophenyl phosphate; Sigma) in developing buffer were added. The reaction was incubated in the dark at 25°C for 1 h and the absorbance at 405 nm was read.

## Acknowledgements

We thank G.Soto for computer assistance. This work was supported by NIH training grant AI07172-16 (D.L.H.), a grant from the Swedish research council NFR (S.D.K.) and National Institutes of Health grant RO1AI29549 (S.J.H.).

## References

Ahrens,R., Ott,M., Ritter,A., Hoschützky,H., Bühler,T., Lottspeich,F., Boulnois,G.J., Jann,K. and Hacker,J. (1993) Genetic analysis of the gene cluster encoding nonfimbrial adhesin I from an *Escherichia coli* uropathogen. *Infect. Immun.*, **61**, 2505–2512.

Allen,B.L., Gerlach,G.F. and Clegg,S. (1991) Nucleotide sequence and functions of *mrk* determinants necessary for expression of type 3 fimbriae in *Klebsiella pneumoniae*. *J. Bacteriol.*, **173**, 916–920.

Amit,A.G., Marrizzua,R.A., Phillips,S.E. and Poljak,R.J. (1986) Three dimensional structure of an antibody-antigen complex at 2.8A resolution. *Science*, **233**, 747–753.

Arnqvist,A., Olsén,A., Pfieffer,J., Russel,D.G. and Normark,S. (1992) The Crl protein activates cryptic genes for curli formation and fibronectin binding in *Escherichia coli*. *Mol. Microbiol.*, **6**, 2443–2453.

Bahrani,F.K. and Mobley,H.L.T. (1994) *Proteus mirabilis* MR/P fimbrial operon: genetic organization, nucleotide sequence, and conditions for expression. *J. Bacteriol.*, **176**, 3412–3419.

Bahrani,F.K., Johnson,D.E., Robbins,D. and Mobley,H.L.T. (1991) *Proteus mirabilis* flagella and MR/P fimbriae: isolation, purification, N-terminal analysis and serum antibody response following experimental urinary tract infection. *Infect. Immun.*, **59**, 3574–3580.

Bahrani,F.K., Cook,S., Hull,R.A., Massad,G. and Mobley,H.L.T. (1993) *Proteus mirabilis* fimbriae: N-terminal amino acid sequence of a major fimbrial subunit and nucleotide sequences of the genes from two strains. *Infect. Immun.*, **61**, 884–891.

Bakker,D., Vader,C.E.M., Mooi,F.R., Oudega,B. and de Graaf,F.K. (1991) Structure and function of periplasmic chaperone-like proteins involved in the biosynthesis of K88 and K99 fimbriae in enterotoxigenic *Escherichia coli*. *Mol. Microbiol.*, **5**, 875–886.

Bäumler,A.J. and Heffron,F. (1995) Identification and sequence analysis of *lpf*ABCDE, a putative fimbrial operon of *Salmonella typhimurium*. *J. Bacteriol.*, **177**, 2087–2097.

Bertin,Y., Girardeau,J.P., Vartanian,M.D. and Martin,C. (1993) The ClpE protein involved in biogenesis of the CS31A capsule-like antigen is a member of a periplasmic chaperone family in gram-negative bacteria. *FEMS Microbiol. Lett.*, **108**, 59–68.

Bullitt,E. and Makowski,L. (1995) Structural polymorphism of bacterial adhesion pili. *Nature*, **373**, 164–167.

Clouthier,S.C., Müller,K.H., Doran,J.L., Collinson,S.K. and Kay,W.W. (1993) Characterization of three fimbrial genes, *sefABC*, of *Salmonella enteritidis*. *J. Bacteriol.*, **175**, 2523–2533.

Dodson,K.W., Jacob-Dubuisson,F., Striker,R.T. and Hultgren,S.J. (1993) Outer membrane PapC usher discriminately recognizes periplasmic chaperone-pilus subunit complexes. *Proc. Natl Acad. Sci. USA*, **90**, 3670–3674.

Duchet-Suchaux,M., Bertin,A. and Dubray,G. (1988) Morphological description of surface structures on strain B41 of bovine enterotoxigenic *Escherichia coli* bearing both K99 and F41 antigens. *J. Gen. Microbiol.*, **134**, 983–995.

Foged,N.T., Klemm,P., Elling,F., Jorsal,S.E. and Zeuthen,J. (1986) Monoclonal antibodies to K88ab, K88ac and K88ad fimbriae from enterotoxigenic *Escherichia coli*. *Microb. Pathogen.*, **1**, 57–69.

Friedrich,M.J., Kinsey,N.E., Vila,J. and Kadner,R.J. (1993) Nucleotide sequence of a 13.9 kb segment of the 90 kb virulence plasmid of *Salmonella typhimurium:* the presence of a fimbrial biosynthetic gene. *Mol. Microbiol.*, **8**, 543–558.

Galyov,E.E., Karlishev,A.V., Chernovskaya,T.V., Dolgikh,D.A., Smirnov, O.Yu., Volkovoy,K.I., Abramov,V.M. and Zav'yalov,V.P. (1991) Expression of the envelope antigen F1 of *Yersinia pestis* is mediated by the product of *caf1M* gene having homology with the chaperone protein PapD of *Escherichia coli*. *FEBS Lett.*, **286**, 79–82.

Garcia,M.I., Labigne,A. and LeBouguenec,C. (1994) Nucleotide sequence of the afimbrial-adhesin-encoding *afa-3* gene cluster and its translocation via flanking IS*1* insertion sequences. *J. Bacteriol.*, **176**, 7601–7613.

Girardeau,J.P., Vartanian,M.D., Ollier,J.L. and Contrepois,M. (1988) CS31A, a new K88-related fimbrial antigen on bovine enteropathogenic and septicemic *Escherichia coli* strains. *Infect. Immun.*, **56**, 2180–2188.

Goldhar,J., Perry,R., Golecki,J.R., Hoschutzky,H., Jann,B. and Jann,K. (1987) Nonfimbrial, mannose-resistant adhesins from uropathogenic *Escherichia coli* O83:K1:H4 and O14:K?:H11. *Infect. Immun.*, **55**, 1837–1842.

Hacker,J. and Morschhäuser,J. (1994) S and F1C fimbriae. In Klemm,P. (ed.), *Fimbriae: Adhesion, Genetics, Biogenesis, and Vaccines*. CRC Press, Boca Raton, FL, pp. 27–36.

Holmgren,A. and Brändén,C.-I. (1989) Crystal structure of chaperone protein PapD reveals an immunoglobulin fold. *Nature*, **342**, 248–251.

Holmgren,A., Kuehn,M.J., Brändén,C.-I. and Hultgren,S.J. (1992) Conserved imunoglobulin-like features in a family of periplasmic pilus chaperones in bacteria. *EMBO J.*, **11**, 1617–1622.

Hultgren,S.J., Lindberg,F., Magnusson,G., Kihlberg,J., Tennent,J.M. and Normark,S. (1989) The PapG adhesin of uropathogenic *Escherichia coli* contains separate regions for receptor binding and for the incorporation into the pilus. *Proc. Natl Acad. Sci. USA*, **86**, 4357–4361.

Hultgren,S.J., Normark,S. and Abraham,S.N. (1991) Chaperone-assisted assembly and molecular architecture of adhesive pili. *Annu. Rev. Microbiol.*, **45**, 383–415.

Hultgren,S.J., Abraham,S.N., Caparon,M.G., Falk,P., St. Geme,J.W.,III and Normark,S. (1993a) Pilus and non-pilus bacterial adhesins: assembly and function in cell recognition. *Cell*, **73**, 887–901.

Hultgren,S.J., Jacob-Dubuisson,F., Jones,C.H. and Brändén,C.-I. (1993b) PapD and superfamily of periplasmic immunoglobulin-like pilus chaperones. *Adv. Protein Chem.*, **44**, 99–123.

Iriarte,M., Vanooteghem,J.C., Delor,I., Díaz,R., Knutton,S. and Cornelis,G.R. (1993) The Myf fibrillae of *Yersinia enterocolitica*. *Mol. Microbiol.*, **9**, 507–520.

Jacob-Dubuisson,F., Heuser,J., Dodson,K., Normark,S. and Hultgren,S.J. (1993a) Initiation of assembly and association of the structural elements of a bacterial pilus depend on two specialized tip proteins. *EMBO J.*, **12**, 837–847.

Jacob-Dubuisson,F., Kuehn,M. and Hultgren,S. (1993b) A novel secretion apparatus for the assembly of adhesive bacterial pili. *Trends Microbiol.*, **1**, 50–55.

Jalajakumari,M.B., Thomas,C.J., Halter,R. and Manning,P.A. (1989) Genes for biosynthesis and assembly of CS3 pili of CFA/II enterotoxigenic *Escherichia coli:* novel regulation of pilus production by bypassing an amber codon. *Mol. Microbiol.*, **3**, 1685–1695.

Jones,C.H., Pinkner,J.S., Nicholes,A.V., Slonim,L.N., Abraham,S.N. and Hultgren,S.J. (1993) FimC is a periplasmic PapD-like chaperone that directs assembly of type 1 pili in bacteria. *Proc. Natl Acad. Sci. USA*, **90**, 8397–8401.

Jones,C.H., Pinkner,J.S., Roth,R., Heuser,J., Nicholes,A.V., Abraham, S.N. and Hultgren,S.J. (1995) FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the *Enterobacteriaceae*. *Proc. Natl Acad. Sci. USA*, **92**, 2081–2085.

Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Karlyshev,A.V., Galyov,E.E., Smirnov,O.Y., Guzayev,A.P., Abramov, V.M. and Zav'yalov,V.P. (1992) A new gene of the *f1* operon of *Y. pestis* involved in the capsule biogenesis. *FEBS Lett.*, **297**, 77–80.

Klemm,P. and Krogfelt,K.A. (1994) Type 1 fimbriae of *Escherichia coli*. In Klemm,P. (ed.), *Fimbriae: Adhesion, Genetics, Biogenesis, and Vaccines*. CRC Press, Boca Raton, FL, pp. 9–26.

Klemm,P., Christiansen,G., Kreft,B., Marre,R. and Bergmans,H. (1994) Reciprocal exchange of minor components of Type 1 and F1C fimbriae results in hybrid organelles with changed receptor specificities. *J. Bacteriol.*, **176**, 2227–2234.

Klemm,P., Jørgensen,B.J., Kreft,B. and Christiansen,G. (1995) The export systems of type 1 and F1C fimbriae are interchangeable but work in parental pairs. *J. Bacteriol.*, **177**, 621–627.

Knutton,S., McConnell,M.M., Rowe,B. and McNeish,A.S. (1989) Adhesion and ultrastructural properties of human enterotoxigenic *Escherichia coli* producing colonization factor antigens III and IV. *Infect. Immun.*, **57**, 3364–3371.

Kuehn,M.J., Normark,S. and Hultgren,S.J. (1991) Immunoglobulin-like PapD chaperone caps and uncaps interactive surfaces of nascently translocated pilus subunits. *Proc. Natl Acad. Sci. USA*, **88**, 10586–10590.

Kuehn,M.J., Heuser,J., Normark,S. and Hultgren,S.J. (1992) P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips. *Nature*, **356**, 252–255.

Kuehn,M.J., Ogg,D.J., Kihlberg,J., Slonim,L.N., Flemmer,K., Bergfors,T. and Hultgren,S.J. (1993) Structural basis of pilus subunit recognition by the PapD chaperone. *Science*, **262**, 1234–1241.

Kuehn,M.J., Haslam,D., Normark,S. and Hultgren,S.J. (1994) Structure, function, and biogenesis of *Escherichia coli* P pili. In Klemm,P. (ed.), *Fimbriae: Adhesion, Genetics, Biogenesis, and Vaccines*. CRC Press, Boca Raton, FL, pp. 37–51.

Le Bouguenec,C., Garcia,M.I., Ouin,V., Desperrier,J.-M., Gounon,P. and Labigne,A. (1993) Characterization of plasmid bourne afa-3 gene clusters encoding afimbrial adhesins expressed by *Escherichia coli* strains associated with intestinal or urinary tract infections. *Infect. Immun.*, **61**, 5106–5114.

Levine,M.M. *et al.* (1984) Coli surface antigens 1 and 3 of colonization factor antigen II-positive enterotoxigenic *Escherichia coli:* morphology, purification, and immune responses in humans. *Infect. Immun.*, **44**, 409–420.

Lindberg,F., Tennent,J.M., Hultgren,S.J., Lund,B. and Normark,S. (1989) PapD, a periplasmic transport protein in P-pilus biogenesis. *J. Bacteriol.*, **171**, 6052–6058.

Lindler,L.E. and Tall,B.D. (1993) *Yersinia pestis* pH 6 antigen forms fimbriae and is induced by intracellular association with macrophages. *Mol. Microbiol.*, **8**, 311–324.

Lintermans,P.F. *et al.* (1988) Characterization and purification of the F17 adhesin on the surface of bovine enteropathogenic and septicemic *Escherichia coli*. *Am. J. Vet. Res.*, **49**, 1794–1799.

Massad,G. and Mobley,H.L.T. (1994) Genetic organization and complete sequence of the *Proteus mirabilis pmf* fimbrial operon. *Gene*, **150**, 101–104.

Merrifield,R.B. (1963) Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.*, **85**, 2149–2154.

Miller,S., Janin,J., Lesk,A.M. and Chothia,C. (1987) Interior and surface of monomeric proteins. *J. Mol. Biol.*, **196**, 641–656.

Müller,K.H., Collinson,S.K., Trust,T.J. and Kay,W.W. (1991) Type 1 fimbriae of *Salmonella enteritidis*. *J. Bacteriol.*, **173**, 4765–4772.

Nicholls,A. (1993) *GRASP: Graphical Representation and Analysis of Surface Properties*. Columbia University Press, New York.

Normark,S., Båga,M., Goransson,M., Lindberg,F.P., Lund,B., Norgren,M. and Uhlin,B.E. (1986) Genetics and biogenesis of *Escherichia coli* adhesins. In Mirelman,D. (ed.), *Microbial Lectins and Agglutinins: Properties and Biological Activity*. Wiley Interscience, New York, pp. 113–143.

Pugsley,A. (1993) The complete general secretory pathway in Gram negative bacteria. *Microbiol. Rev.*, **57**, 50–108.

Raina,S., Missiakas,D., Baird,L., Kumar,S. and Georgopoulos,C. (1993) Identification and transcriptional analysis of the *Escherichia coli htrE* operon which is homologous to *pap* related pilin operons. *J. Bacteriol.*, **75**, 5009–5021.

Roberts,J.A. *et al.* (1994) The Gal α(1-4) Gal-specific tip adhesin of *Escherichia coli* P-fimbriae is needed for pyelonephritis to occur in the normal urinary tract. *Proc. Natl Acad. Sci. USA*, **91**, 11889–11893.

Savarino,S.J., Fox,P., Yikang,D. and Nataro,J.P. (1994) Identification and characterization of a gene cluster mediating enteroaggregative *Escherichia coli* aggregative adherence fimbriae I biogenesis. *J. Bacteriol.*, **176**, 4949–4957.

Schmoll,T., Morschhäuser,J., Ott,M., Ludwig,B., van Die,I. and Hacker,J. (1990) Complete genetic organization and functional aspects of the *Escherichia coli* S fimbrial adhesin determinant: nucleotide sequence of the genes *sfa* B, C, D, E, F. *Microb. Pathogen.*, **9**, 331–343.

Shapiro,L. *et al.* (1995) Structural basis of cell–cell adhesion by cadherins. *Nature*, **374**, 327–337.

Simons,B., Rathman,P., Malij,C., Oudega,B. and de Graf,F.K. (1990) The penultimate tyrosine residue of the K99 fibrillar subunit is essential for the stability of the protein and its interaction with the periplasmic carrier protein. *FEMS Microbiol. Lett.*, **67**, 107–112.

Slonim,L.N., Pinkner,J.S., Brändén,C.I. and Hultgren,S.J. (1992) Interactive surface in the PapD chaperone cleft is conserved in pilus chaperone superfamily and essential in subunit recognition and assembly. *EMBO J.*, **11**, 4747–4756.

Steven,A.C., Bisher,M.E., Trus,B.L., Thomas,D., Zhang,J.M. and Cowell,J.L. (1986) Helical structure of *Bordetella pertussis* fimbriae. *J. Bacteriol.*, **167**, 968–974.

Stirm,S., Ørskov,F., Ørskov,E. and Birch-Andersen,A. (1967) Episome-carried surface antigen K88 of *Escherichia coli*. *J. Bacteriol.*, **93**, 740–748.

Stull,T.L., Mendelman,P.M., Haas,J.E., Schoenborn,M.A., Mack,K.D. and Smith,A.L. (1984) Characterization of *Haemophilus influenzae* type b fimbriae. *Infect. Immun.*, **46**, 787–796.

Willems,R.J.L., van der Heide,H.G.J. and Mooi,F.R. (1992) Characterization of a *Bordetella pertussis* fimbrial gene cluster which is located directly downstream of the filamentous haemagglutinin gene. *Mol. Microbiol.*, **6**, 2661–2671.

Williams,A.F. and Barclay,A.N. (1988) The immunoglobulin superfamily-domains for cell surface recognition. *Annu. Rev. Immunol.*, **6**, 381–405.

Wolf,M.K., Andrews,G.P., Tall,B.D., McConnell,M.M., Levine,M.M. and Boedeker,E.C. (1989) Characterization of CS4 and CS6 antigenic components of PCF8775, a putative colonization factor complex from enterotoxigenic *Escherichia coli* E8775. *Infect. Immun.*, **57**, 164–173.

*Received on November 24, 1995*