Structural Basis of the Interaction of the Pyelonephritic *E. coli* Adhesin to Its Human Kidney Receptor

Karen W. Dodson,[1] Jerome S. Pinkner,[1]
Thierry Rose,[2] Goran Magnusson,[3,5]
Scott J. Hultgren,[1,4] and Gabriel Waksman[2,4]
[1] Department of Molecular Microbiology
[2] Department of Biochemistry
  and Molecular Biophysics
Washington University School of Medicine
Saint Louis, Missouri 63110
[3] Lund University
Sweden

Summary

PapG is the adhesin at the tip of the P pilus that mediates attachment of uropathogenic *Escherichia coli* to the uroepithelium of the human kidney. The human specific allele of PapG binds to globoside (GbO4), which consists of the tetrasaccharide GalNAcβ1-3Galα1-4Galβ1-4Glc linked to ceramide. Here, we present the crystal structures of a binary complex of the PapG receptor binding domain bound to GbO4 as well as the unbound form of the adhesin. The biological importance of each of the residues involved in binding was investigated by site-directed mutagenesis. These studies provide a molecular snapshot of a host-pathogen interaction that determines the tropism of uropathogenic *E. coli* for the human kidney and is critical to the pathogenesis of pyelonephritis.

Introduction

Urinary tract infections (UTI) are one of the most prevalent infectious diseases, second only to infections of the respiratory tract, with 8 million physician visits per year and 1.5 million hospital discharge diagnoses per year (Hooton and Stamm, 1997). One-third of women in the United States will have contracted a UTI before the age of 65, and many women experience more than one such infection per year. Within six months after an initial UTI, approximately 25% of women will experience a second infection and about 3% will suffer a third (Foxman, 1990). A history of recurrent bladder infections is associated with an increased risk of developing kidney infections (pyelonephritis), which itself has a recurrence rate of about 40% (Patton et al., 1991; Ronald and Pattullo, 1991). The most frequent causative agent of pyelonephritis is *Escherichia coli* (Svanborg and Godaly, 1997; Warren, 1996), and the ability of pyelonephritic strains of *E. coli* to bind (or adhere) to the human kidney has been shown to be critical in the infection process (Roberts et al., 1994).

Bacterial attachment is a key event in the early stages of most infectious diseases. Without the ability to specifically adhere to host tissues, pathogens would be readily expelled from the host and disease would not occur. These interactions typically facilitate extracellular colonization or internalization and may initiate a complex cascade of molecular signaling at the host-pathogen interface. Recognition and attachment to host tissues is mediated by adhesins on the surface of the microbe interacting with receptors displayed on the surface of the host cell. In pyelonephritic *E. coli*, the adhesin PapG binds to a Galα1-4Gal containing glycolipid on the surface of human kidney epithelial cells (Leffler and Svanborg-Eden, 1980; Lund et al., 1987). This interaction allows the bacteria to gain a foothold on the tissue and resist being displaced by the mechanical and physical forces in the kidney.

The PapG adhesin is assembled into adhesive hair-like fibers called P pili via the chaperone/usher pathway (Sauer et al., 2000; Soto and Hultgren, 1999; Thanassi and Hultgren, 2000). This pathway is used by diverse pathogenic gram-negative bacteria for the assembly of over 30 different adhesins that are associated with host and tissue tropism (or selectivity) in a wide variety of diseases. P pili are composite fibers consisting of a thin tip structure called a tip fibrillum that is joined to the distal end of a thicker structure called the pilus rod (Kuehn et al., 1992). P pilus rods are 68 Å in diameter and are comprised of repeating PapA subunits arranged in a right-handed helical cylinder consisting of 3.28 subunits per turn (Bullitt and Makowski, 1995). The tip fibrillum of the P pilus is ~20 Å in diameter and is comprised mostly of repeating PapE subunits arranged in an open helical configuration. The PapG adhesin is joined to the distal end of the fibrillum via the PapF adaptor and the fibrillum is joined to the rod by the PapK adaptor (Jacob-Dubuisson et al., 1993). All of the pilin subunits share a great deal of homology. The recent crystal structures of chaperone-subunit complexes revealed that pilin domains have incomplete immunoglobulin-like (Ig) folds where the seventh strand of a typical Ig fold is missing (Choudhury et al., 1999; Sauer et al., 1999). The periplasmic chaperone, in a mechanism termed donor strand complementation, transiently completes the Ig-like fold of each pilin domain by providing the missing secondary structural element. The structure also suggested a mechanism of pilus assembly, termed donor strand exchange, in which each subunit contributes its N-terminal sequence to complete the Ig-like fold of its immediate neighbor in the pilus. The PapG adhesin is comprised of a C-terminal pilin domain and an N-terminal receptor binding domain (Hultgren et al., 1989). The pilin domain of the adhesin has the characteristic Ig-like fold and participates in a donor strand exchange reaction with PapF to link PapG to the tip fibrillum. Thus, the mature pilus consists of an array of over a thousand canonical Ig-like domains, each of which contributes a strand to the fold of the preceding subunit to produce the organelle. At the tip of this structure is the N-terminal receptor binding domain that we have studied here.

Adhesin-receptor interactions, in addition to anchoring the bacteria to the host tissue, can trigger signal transduction cascades that alter gene expression in

both bacterial and host cells (Finlay and Falkow, 1997; Martinez et al., 2000). Binding of host cell receptors by PapG, for instance, activates the transcription of a sensor-regulator protein, AirS, which regulates the bacterial iron acquisition system of uropathogenic *E. coli* (Pernestig et al., 2000; Zhang and Normark, 1996). On the host side, the binding of P pili to Galα1-4Gal-containing host receptors on uroepithelial cells induces the release of ceramides, important second messenger molecules that can influence a number of signal transduction processes (Hedlund et al., 1996; Svanborg et al., 1996). In addition, binding of P-piliated and type 1 piliated bacteria result in the up-regulation and eventual secretion of several immunoregulatory cytokines from host cells (Hedlund et al., 1999; Schilling et al., 2001). Thus, the interaction between PapG and its glycolipid receptor both initiates bacterial colonization of the kidney and modulates the host response.

The glycolipid receptor for PapG consists of a digalactoside (Galα1-4Gal) core linked by a β-glucose (Glc) residue to a ceramide group that anchors the receptor in the membrane (Stromberg et al., 1990, 1991). This minimum receptor isotype is called globotriasylceramide (GbO3). The various members of this receptor family differ by the addition of sugar residues distal to the Galα1-4Gal core of GbO3. The addition of a single N-acetyl-galactosamine (GalNAc) sugar to GbO3 creates GbO4 (globoside), whereas the addition of two GalNAc sugars to GbO3 creates the Forssman antigen (GbO5). Sialic acids can also be added to form more complex receptor structures (Stapleton et al., 1998; Stroud et al., 1998). Three different PapG alleles exist—class I, II, and III—which bind with different specificity to different receptor isotypes, with class II being the allele predominantly associated with human pyelonephritis, and class III being correlated with human cystitis. The differential distribution of the receptor isotypes in different hosts and tissues and the binding specificity of the various PapG adhesins account for the host and tissue tropisms of uropathogenic *E. coli*.

In this study, we describe the crystal structures of the apo and GbO4-bound forms of the receptor binding domain of the class II PapG adhesin (PapGII). The adhesin-receptor interaction described here is a critical event in the ability of pyelonephritic *E. coli* to cause disease (Roberts et al., 1994). The adhesin-receptor interactions were further investigated by a structure-based site-directed mutagenesis analysis of the binding pocket. This analysis elucidates how *E. coli* recognizes the globoside receptor in the kidney and provides fundamental insights into a key event in bacterial pathogenesis.

Results and Discussion

Structure of the PapGII Receptor Binding Domain
The structure of the receptor binding domain of PapGII bound to GbO4 was determined using the multiwavelength anomalous dispersion (MAD) phasing method to a resolution of 1.8 Å. The quality of the map was excellent and all protein and sugar residues could be built unam-

biguously into electron density (Figure 1A). The structure of the PapGII receptor binding domain is a mostly β sheet structure that can be subdivided into two regions (Figure 1B). The first region (region 1), composed of strands b, p, o, c, n, i, and h, forms a β barrel. The topology of this region is similar to that of the equivalent region of the sugar-free form of the receptor binding domain of the mannose-specific FimH adhesin (no bound form for this adhesin has been characterized yet structurally) (Choudhury et al., 1999). The second region (region 2) is composed of a central antiparallel β sheet formed by strands e, a, o, g, k, and j, which is flanked on one side by two double-stranded β sheets (strands f and l, and d and m, respectively), and, on the other side, by an α helix (A) and a large loop connecting this helix to strand o. This region contains the receptor binding site and its structure has no equivalent in the database.

The GbO4 Binding Site: Structural Features and Mutagenesis
The receptor binding site of PapGII is located in a pocket formed by strands j, g, and k, helix A, and the loop connecting strand o to helix A (Figures 1 and 2). The solvent-accessible area of PapGII buried upon binding of the receptor amounts to 716 Å^2. The tetrasaccharide binds in a V shape with the reducing end Glc residue (residue D) and the Gal residue next to it (C) forming one branch of the V, and the following Gal (B) and GalNAc (A) residues forming the other branch.

The Galβ1-4Glc (residues C and D) moiety lies over Trp107 (strand k) and runs parallel to the backbone structure of strand k between residues 104 and 107. Hydrophobic and aromatic contacts are made between Trp107 and the nonpolar face of Galβ1-4Glc composed of the C1, C3, and C5 atoms of residue C and the C4 atom of residue D. One polar contact is observed between the Nε$_1$ of Trp107 and the O3 hydroxyl of residue D. These interactions are critical in receptor binding: when Trp107 was changed to an alanine (W107A), receptor binding was abolished (Figure 3). Interactions of the Galβ1-4Glc moiety with Trp107 occur in the vicinity of a bulge in strand k. This bulge is created by the displacement of Lys106 and Asp108 out of a classical alternating β strand arrangement. Lys106 and Tyr105 are on one side of strand k while Trp107 and Asp108 are on the other. A water molecule (W1 in Figure 2A) interacts with the backbone nitrogens of Lys106 and Tyr105 and the carbonyl oxygens of Gly104 and Tyr60. Further stabilization of the bulge is achieved by a salt bridge between Asp108 and Arg170. W1 is also involved in receptor binding since it is within H bond distance of the O6 hydroxyl of residue C. Asp108 is not in direct or water-mediated contact with the receptor. However, Arg170 (in the loop between strand o and helix A [the oA loop]) makes interactions with the O2 and O3 hydroxyls of residue D (Figure 2). Mutation of Lys106 to alanine (K106A) resulted in a slight reduction in hemaglutination titer, suggesting that this side chain plays only a minor role in stabilizing the binding site (Figure 3). In contrast, an R170A mutation abolishes binding (Figure 3). This result is consistent with the observed roles of Arg170 in the structure. Glu59 in strand g interacts directly with the O6 hydroxyl of

A

B
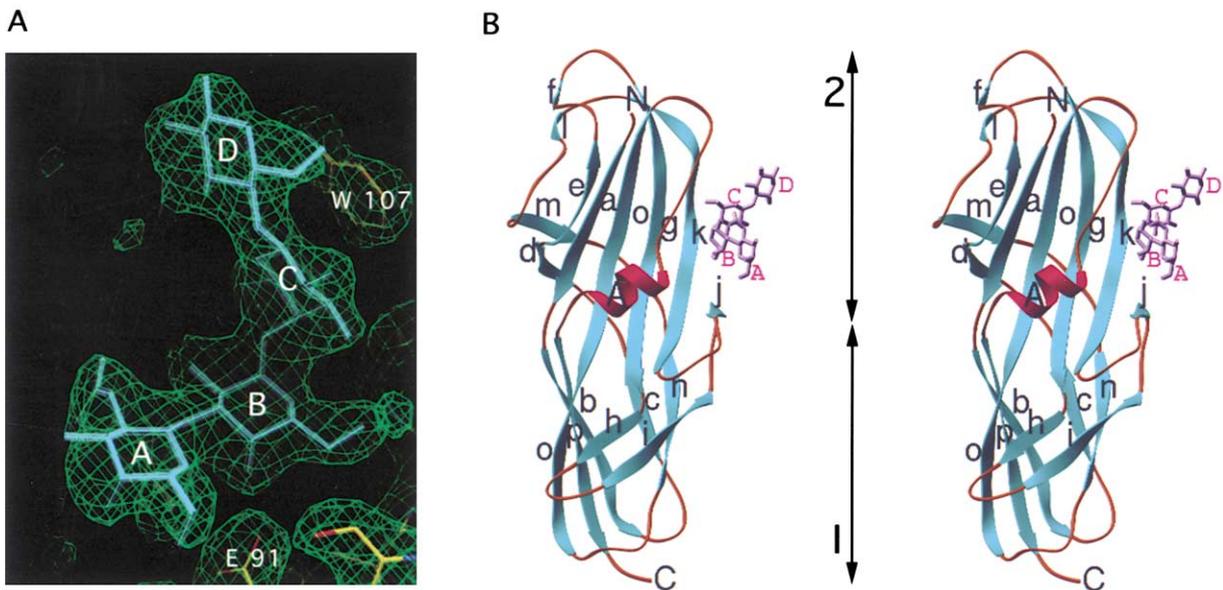
Figure 1. Structure of the Binary Complex of the Receptor Binding Domain of the PapGII Adhesin Bound to GbO4 Glycolipid

(A) Experimental solvent-flattened electron density map. The represented region is that of the tetrasaccharide binding site. The final model of the adhesin domain–tetrasaccharide complex is shown in ball-and-stick representation. The tetrasaccharide is color-coded in blue. Each sugar residue is labeled, A to D, with A indicating the GalNAc, B and C indicating the Galα1-4Gal galactobiose, and D indicating Glc. The protein is color-coded by atom types with oxygens in red, nitrogens in blue, and carbons in yellow.

(B) Structure of the PapGII receptor binding domain bound to GbO4. The protein is in ribbon-representation with β strands as cyan arrows, α helices in red, and loop and coil structures in orange. The tetrasaccharide is in ball-and-stick representation color-coded in magenta. The Glc, Gal, Gal, and GalNAc residues are labeled D to A, respectively. Strands are labeled from a to p, while the α helix is labeled A. Regions 1 and 2 of the structure are indicated by the vertical black arrows and labeled 1 and 2, respectively.

residue C and makes water-mediated contacts with the O2 hydroxyl of residue B. These interactions are shown to be important by two separate experiments. We demonstrated that an E59A mutation reduced the ability of PapG to mediate hemagglutination of human erythrocytes (Figure 3) and previous studies have shown that deoxy analogs of galabiose, with hydrogen or methyl groups at the O6 position of residue C, have reduced affinity for the PapG adhesin (Kihlberg et al., 1989; Striker et al., 1995). No contact is observed between PapG and the O2 hydroxyl of residue C (Figure 2). This is consistent with previous findings showing that deoxy analogs of O2 of residue C had no effect on PapG-galabiose interactions (Kihlberg et al., 1989; Striker et al., 1995).

Binding of residue B is mediated in large part by water molecules (Figure 2). Its O2 hydroxyl is located in the proximity of W2 and W3. W2 contacts Glu59 while W3 interacts with Arg170 through W4 (Figure 2A). Another water molecule (W7) on the other side of the binding site mediates contact between the O6 hydroxyl and the Nζ group of Lys103 (in strand k). However, this interaction must be weak since a K103A mutation results only in a slight reduction in hemagglutination titer (Figure 3). Residue B also makes direct contacts with the PapG binding pocket via H bond interactions between its O4 hydroxyl and the charged tips of Glu91 (in strand j) and Lys172 (in the oA loop). In addition, residue B participates in hydrogen bonding interactions with the main chain nitrogen of Gly104 through its O6 hydroxyl. Although previous studies reported that a double E90A-

E91A mutation did not affect PapGII function (Klann et al., 1994), the E91A and K172A mutations reported here reduced or completely abolished receptor binding, respectively (Figure 3), confirming the involvement of both these residues in receptor binding. Furthermore, deoxy analogs in the O2, O4, and O6 hydroxyls of residue B dramatically reduce PapG-receptor interactions (Kihlberg et al., 1989; Striker et al., 1995), further supporting the significance of the interactions described in the present study. The C1 and C2 nonpolar region of residue B lies on top of a hydrophobic platform comprised of the aliphatic portion of Lys172, Ile61 (g strand), and Leu102 (strand k).

Finally, residue A of the receptor lies on a platform comprised of Lys172 (in the oA loop) and Glu91. This platform is extended by two water molecules (W5 and W6 in Figure 2) that also interact with residues Tyr175 in the helix A and Asn92 in strand j. The importance of the W5/W6-mediated interaction to Tyr175 is confirmed by the Y175A mutation which results in a reduction of the hemagglutination titer (Figure 3). Lys172 directly interacts with the O5 and the C8 methyl of the N-acetyl group and makes water-mediated contacts (through W6) with the O4 hydroxyl. Glu91 makes direct H bond interactions with the O6 hydroxyl. A few direct intramolecular contacts between receptor residues are observed. O3 of residue C contacts O6 and O5 of residue B, and O2 of residue B contacts O7 of residue A. No electron density was observed for the trimethylsilylethyl (TMSEt) group O-linked to the Glc residue of GbO4. TMSEt represents a substitute for the ceramide in the
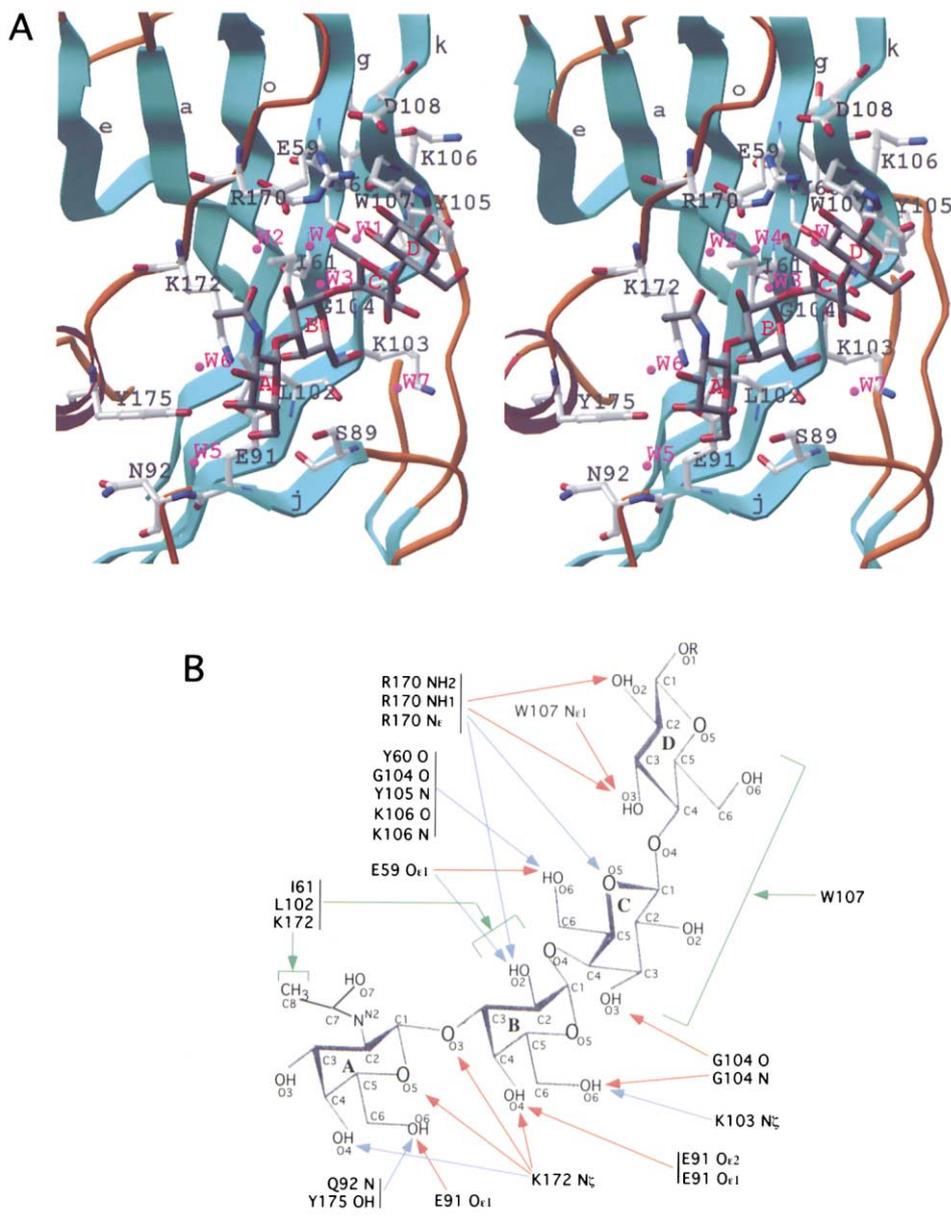
Figure 2. The GbO4 Binding Site

(A) Stereo-diagram of binding site. The protein backbone is in ribbon representation as in Figure 1B and secondary structures are labeled according to Figure 1B. Residues in the protein interacting with the GbO4 receptor are in ball-and-stick representation with carbon atoms in gray, oxygen atoms in red, and nitrogen atoms in blue. Receptor residues are in ball-and-stick representation with carbon atoms in silver, and oxygen and nitrogen atoms as in the protein. Water molecules involved in interactions between the protein and receptor residues are in ball representation color-coded in magenta and labeled W1 to 7. Receptor residues are labeled A to D as in Figure 1B.

(B) Schematic representation of interactions between protein and receptor. Direct polar interactions are indicated by red arrows. Water-mediated interactions are indicated by blue arrows. Brackets and arrows in green indicate contacts with aromatic/hydrophobic platforms.

glycolipid. Hence, the ceramide group is likely not involved in binding in vivo.

Apo PapG

In order to investigate whether the interaction between PapG and its globoside receptor requires conformational changes in PapG or a rearrangement of the water molecules around the binding site, the unbound form of the PapGII receptor binding domain was crystallized. Crystallization conditions were very different for the un-

bound form compared to the bound form (see Experimental Procedures). Nevertheless, the crystals belonged to the same space group and the unit cell dimensions were very similar. The structure of the unbound form of the protein is strikingly similar to that of the complex form (root-mean-square deviation of 0.16 Å in backbone atom positions; Figure 4). Hence, the interactions between the protein and its receptor appear to be of a rigid body type which does not involve conformational changes in the protein. Residues involved in GbO4
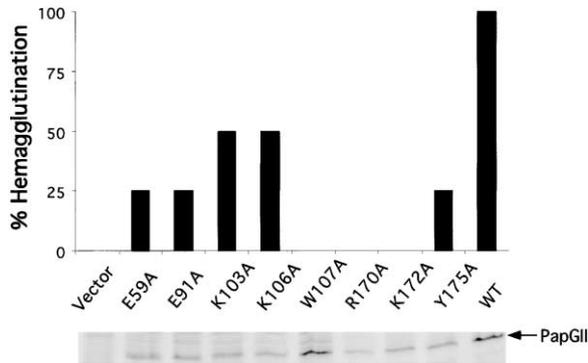
Figure 3. Mutational Studies of the GbO4 Binding Site

The hemagglutination relative to wild-type PapGII is indicated in the bar graph for each of the mutants. Expression and display of the adhesin mutants in the tip fibrillum of purified P pili is shown on the gel underneath the bar graph.

binding undergo only slight changes in side chain conformation or backbone structure (Figure 4). The primary difference between the apo and bound forms of PapG is the structure of water molecules in and around the receptor binding site. There exist seven water molecules mediating interactions between the receptor and the protein in the bound complex (described above). In the

apo form, only five waters were visible in the binding site, and only one of those (W2) was found bound in the same location (Figure 4). The four other water molecules were found in positions equivalent to where atoms in the GbO4 receptor were located when bound to PapG. Three of the water molecules were most closely aligned with three of the Galα1-4Gal core hydroxyls known to be important in receptor binding: the O6 of the residue C, and the O4 and O6 of the residue B. In the bound form, these hydroxyls make contacts with the side chains of residues 59 and 91 and with the main chain nitrogen of residue 104. Thus, binding of PapG to its receptor must involve displacement of these water molecules.

Model for Pilus-Uroepithelium Interaction

From the crystal structure of the PapGII-GbO4 complex, a model can be derived as to how the tip of the P pilus approximates its receptor on the cell surface of the kidney (Figure 5). One intriguing feature of the receptor binding site of the PapG adhesin is its location on the side of the molecule. The PapG adhesin consists of two domains, one (N-terminal) is the receptor binding domain and the other (C-terminal) termed "pilin" domain has an Ig-like fold structure and is connected to the tip fibrillum via donor-strand exchange with the PapF adaptor. The pilin and receptor binding domains of PapG are arranged head-to-tail with a short 3 amino acid linker
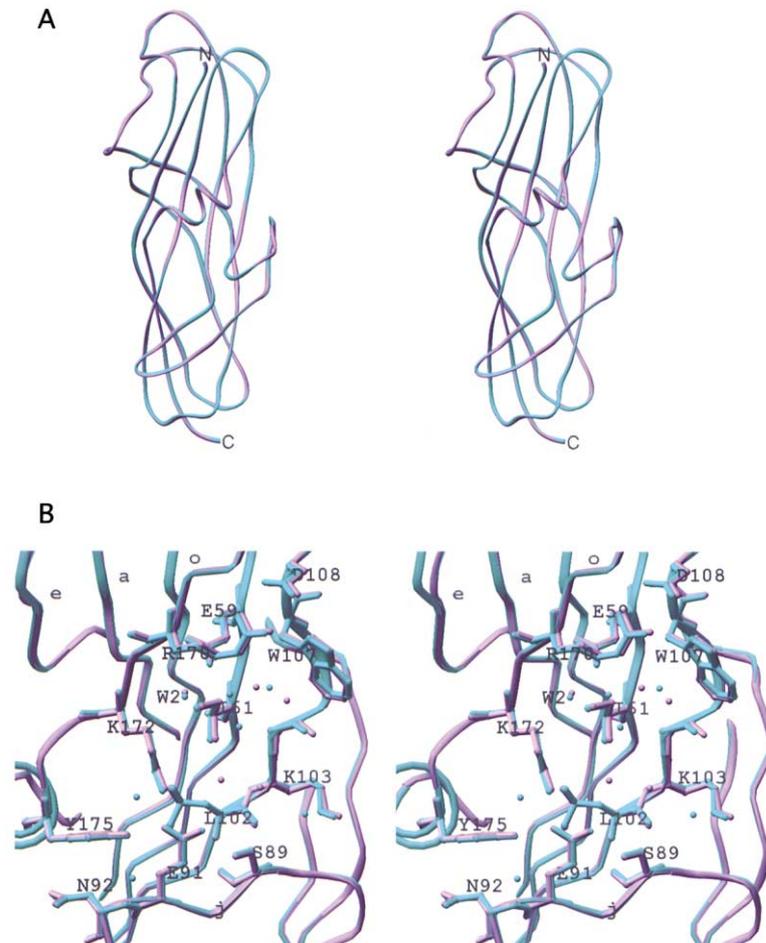
Figure 4. The Structure of the Unbound Form of the PapGII Receptor Binding Domain

(A) Superimposition of the structures of the apo (in lavender) and bound (in cyan) forms of the PapGII receptor binding domain. The two forms are in stereo-ribbon representation; the N- and C-terminal ends of the protein are indicated.

(B) Superimposition of the GbO4 binding site in the apo (in lavender) and bound (in cyan) forms. Side chains of residues involved in GbO4 binding are indicated in ball-and-stick representation. The 7 waters mediating protein-receptor interactions in the bound state are shown with cyan balls. The 5 waters bound in the GbO4-free state of the protein are indicated with lavender balls. The receptor-bound water, W2, is indicated.
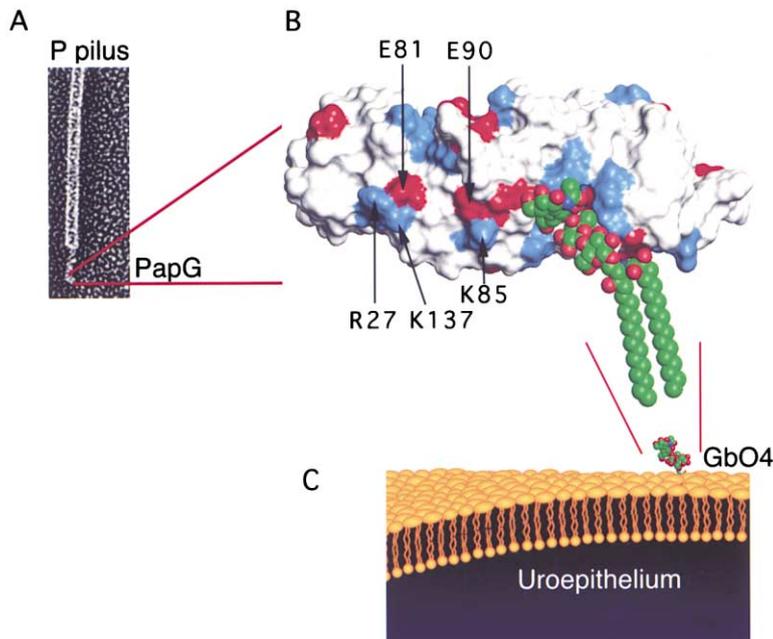
Figure 5. Model of the Interactions of the PapGII Adhesin at the Membrane

(A) A P pilus. The PapG adhesin is located at the thin tip of the pilus. Note the bend at the tip of the pilus, which places the PapG adhesin in a proper orientation for side-on binding to the membrane.

(B) Model of PapGII receptor binding domain interacting with GbO4-ceramide. This model was generated by attaching a ceramide group to the Galβ1-4Glc moiety of GbO4 such that the resulting Galβ1-4Glcβ-ceramide would adopt a configuration similar to that crystallographically observed for a digalacto-side-ceramide (Pascher et al., 1992). The molecular surface is that of the protein, color-coded according to charge, blue for positive (Arg and Lys) and red for negative (Glu and Asp). Charged residues which could potentially interact with the head groups of the eukaryotic membrane are labeled. The GbO4-ceramide is in CPK representation with carbon atoms in green, oxygen atoms in red, and nitrogen atoms in blue.

(C) Model of uroepithelium. The GbO4-ceramide is represented as in (B) and is labeled "GbO4".

between the two domains. Such a configuration thus places the receptor binding surfaces which we have characterized here on the side of the pilus tip. Interestingly, the structure of a modeled galactosylceramide indicates that the globoside moiety of the receptor approaches a perpendicular angle in relation to the ceramide group (Pascher et al., 1992). Thus, in order for the PapG's receptor binding pocket to interact with the GbO4 glycolipid, the receptor binding domain must be oriented with its N- to C-terminal axis approaching a parallel orientation to the membrane so that it can dock to the receptor. This orientation may be facilitated by the flexibility inherent in the tip fibrillum of the P pilus as evidenced by high-resolution EM (Figure 5). In fact, a flexible tip fibrillum may be required to present the receptor binding surface "side-on" to the membrane-bound receptor. The side-on orientation appears also to be faciliated by the preponderance of charged residues on the surface adjacent to the binding site that may facilitate docking to the polar head groups of the eukaryotic lipid membrane (Figure 5).

Binding Specificity of the PapG Adhesins

The P pilus architecture serves as a platform to present three known classes of adhesin variants: PapGI, PapGII, and PapGIII (Stromberg et al., 1990). The binding properties of the PapG adhesins differ depending on whether binding is assessed with glycolipid substrates inserted in natural membranes or immobilized on artificial surfaces (Stromberg et al., 1990, 1991). PapGI, II, and III adhesins preferentially bind to membrane-inserted GbO3, GbO4, and GbO5, respectively (Stromberg et al., 1990). On artificial surfaces, PapGIII still clearly exhibits specificity toward GbO5; however, PapGI or PapGII appear to bind equally well to GbO3 and GbO4 (Stromberg et al., 1991). Recent studies have further elucidated the specificities that differentiate receptor recognition between the PapGII and PapGI adhesins (Striker et al., 1995). Using a large set of receptor analogs, it was

shown that PapGII, but not PapGI, specifically recognized the reducing Glc residue (Kihlberg et al., 1989; Striker et al., 1995).

The structural basis of this binding specificity is suggested by the structure presented here. The amino acid sequence of PapGI (Figure 6A) was threaded onto the PapGII structure and amino acid differences on the PapGII surface were identified (Figure 6B, top panel). The region involved in contacting residue D (Glc) was found to be comprised of amino acid residues that vary between PapGI and PapGII. The most notable substitution is Arg (in PapGII) to His (in PapGI) at position 170 (Figure 6A). Arg170 makes interactions with Glc. These interactions would presumably not occur when a His occupies this position, thus explaining the lack of interactions between PapGI and the Glc at the reducing end of globoside (Striker et al., 1995). The regions of PapGII interacting with the galabiose core (residues B and C) and GalNAc (residue A) of GbO4 are well conserved between PapGI and PapGII (Figure 6B). This explains why these adhesins cannot discriminate between GbO3 and GbO4 in vitro as both would presumably make the same interactions with the galabiose core. However, in vivo, PapGII exhibits a distinct preference for membrane-inserted GbO4 over GbO3 (Stromberg et al., 1991). This binding behavior was investigated by Stromberg et al. (1991), who showed that GbO3 and GbO4 are likely to adopt different conformations when inserted in the membrane. The Glc residue in membrane-anchored GbO3 was predicted to be partially buried in the membrane, while the same residue in membrane-anchored GbO4 was predicted to be more exposed due to the effect of the GalNac residue on the spatial orientation of the tetrasaccharide with respect to the membrane surface (Stromberg et al., 1991). Since Glc is required for optimal PapGII binding, the masking of Glc in membrane-inserted GbO3 would negatively affect PapGII binding to this glycolipid (Striker et al., 1995).

Differences between the receptor binding sites of
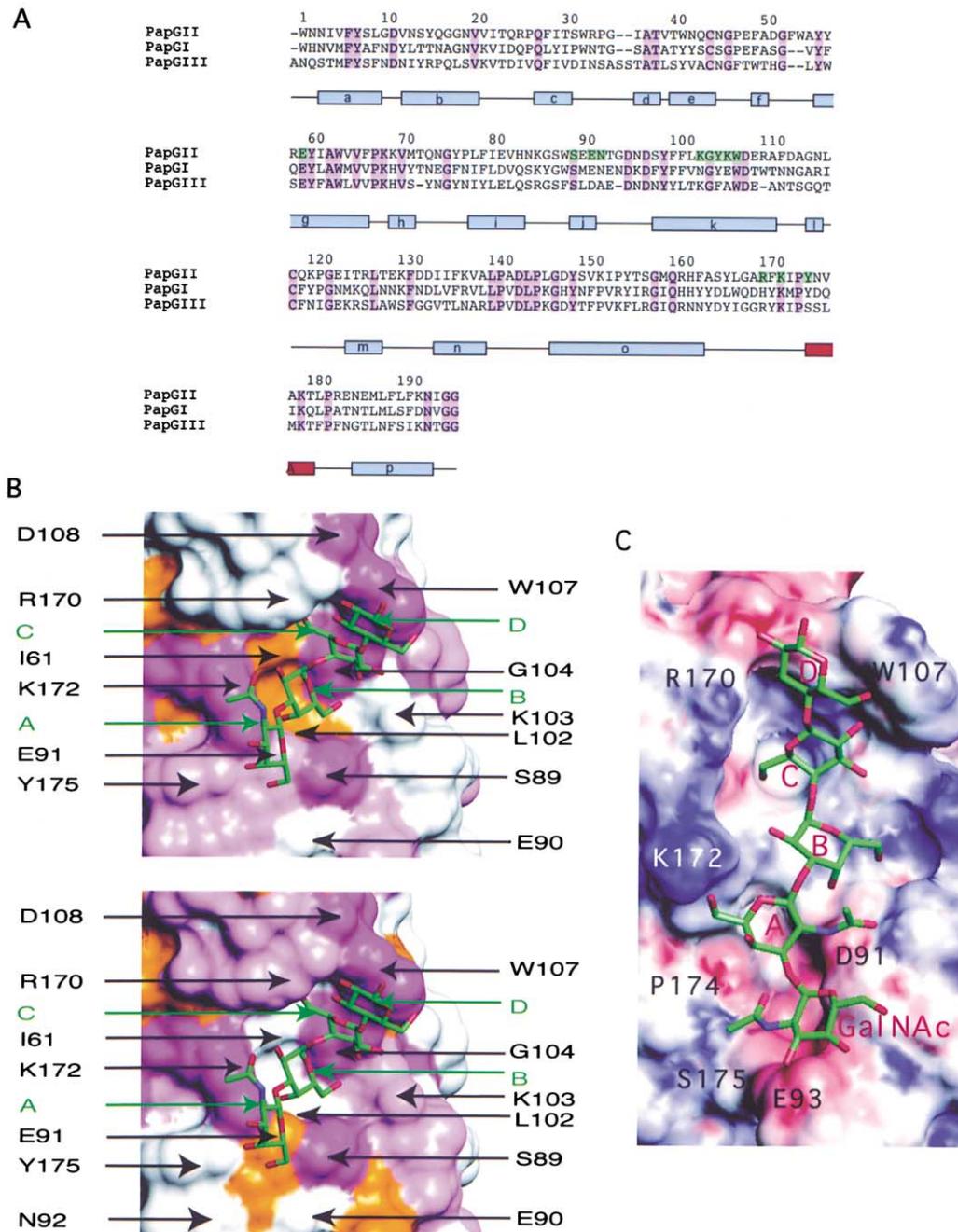
Figure 6. Conservation of Residues among PapG Adhesins and Model of the PapGIII Adhesin

(A) Sequence alignment of the PapGI, PapGII, and PapGIII sequences. The secondary structural elements are indicated below the aligned sequences, with β strands and α helices in light blue and red boxes, respectively. Strictly conserved residues between the three adhesin classes are in purple boxes. Residues involved in receptor binding in PapGII are indicated in green.

(B) Surface mapping of conserved (purple), similar (gold), and nonconserved (white) residues between the PapGII and PapGI adhesins (top panel) and the PapGII and PapGIII adhesins (bottom panel). The surface of the represented binding site is that of PapGII. The receptor is in stick representation color-coded as in Figure 4B. Similarity (gold) between residues is defined by the following groups (E:D), (R:K), (A:L:V:I:M), (F:Y:W), (N:Q), and (S:T:C) where residues within parentheses are defined as similar. Residues strictly conserved among the three adhesins are in deep purple while residues only conserved among the two adhesins compared in either the top or bottom panel are in light purple.

(C) Theoretical model of the PapGIII receptor binding site bound to GbO5 (sequence GalNAcα1-3GalNAcβ1-3Galα1-4Galβ1-4Glc). The receptor (GbO5) is in stick representation color-coded as in Figure 5B and labeled as in Figure 1. The nonreducing end GalNAcα is labeled GalNAc. The solvent-accessible surface of the modeled PapGIII binding site is represented and color-coded according to surface electrostatic potential (calculated using DELPHI [Nicholls and Honig, 1991] and displayed [blue for positive and red for negative] using INSIGHTII [MSI, San Diego]). PapGIII's residues in contact with GbO5 are labeled. The sequence of PapGIII was threaded onto the PapGII structure using the program SwissModel (Guex et al., 1999) and optimized using the program Discover (MSI, San Diego). A model of GbO5 was initially obtained by adding a GalNAcα residue to the O3 of the GalNAcβ residue of GbO4. The resulting GbO5 was then docked onto the PapGIII structure using a Monte-Carlo-based algorithm implemented by the program Discover. The GbO5/PapGIII complex with lowest energy was minimized using conjugate gradient method (program Discover).

Table 1. Data Collection and Refinement Statistics

Data collection*

Data set	Resolution	Reflections (Total/Unique)	Completeness (%)* ($I/\sigma I > 0$)	R_{sym} (%)[#],*	$I/\sigma(I)$*
SeMet-inflection	30–1.8 Å	395,851 / 32,839	93.7 (74.8)	4.7 (28.5)	17.1 (1.4)
Se-Met peak	30–1.8 Å	395,514 / 32,770	94.4 (80.9)	4.6 (27.9)	18.4 (1.6)
Se-Met remote	30–1.8 Å	395,870 / 32,819	94.1 (79.2)	4.3 (30.5)	17.7 (1.4)
Apo form	30–2.1 Å	155,672 / 20,308	93.5 (82.4)	5.6 (23.2)	24.1 (4.3)

Refinement

	Bound	Unbound
Resolution	30–1.8 Å	30–2.1 Å
$|F|/\sigma|F|$	>1	>1
Number of reflections (working / test)	28,277 / 1,489	17,220 / 1,867
Completeness (overall/last shell)*	91.0 (75.8)	93.3 (81.8)
Total number of atoms	1,838	1,716
Protein atoms	1,605 (196 residues)	1,605
GbO4	48 (1 tetrasaccharide)	
Water molecules	185	111
R factor	0.22	0.23
R free factor	0.24	0.25
Averaged B factors ($Å^2$)		
Main chain	30.2	30.1
Side chain	31.0	30.9
Rms deviations		
Bonds (Å)	0.010	0.006
Angles (°)	1.587	1.317
B values ($Å^2$)	0.8 / 1.8	0.8 / 1.7

*Numbers in parentheses indicate values in the highest resolution shell (1.86–1.80 Å for bound state data and 2.18–2.1 Å for apo state data).
[#]$R_{sym} = \Sigma|I - <I>|/\Sigma I$, where I = observed intensity, and $<I>$ = average intensity for symmetry-related reflections.

PapGII and PapGIII mostly lie in the region where the second GalNAcα residue of GbO5 (the PapGIII substrate) would locate (see the Glu90, Tyr175, and Asn92 region in Figure 6B, bottom panel). Hence, the difference in selectivity between PapGII and PapGIII may be explained by the fact that substitutions to PapGIII residues in that region may result in the formation of an additional GalNAc binding subsite capable of accommodating the second nonreducing end GalNAcα of GbO5. To explore this possibility, a model of PapGIII bound to GbO5 was generated (Figure 6C). In PapGIII, Glu91 is substituted to Asp. This allows space for a groove in which the GalNAcα1-3GalNAc moiety of GbO5 could lie. This groove is completed on the Lys172 side (i.e., the oA loop side) by Pro174 and Ser175 (substituting to Tyr175 in PapGII) and on the other side by residues in strand j and in the jk loop (Asp91 and Glu93). Glu93 in PapGIII is predicted to project toward the GalNAcα of GbO5 and make H bond interactions with its O3 and O4 hydroxyls as well as with the N2 of its N-acetyl group. GbO4 is the dominating galabiose-containing isoreceptor in the human kidney in contrast to the canine kidney where GbO5 predominates. PapGII and PapGIII favor colonization of the human and dog kidney, respectively (Stromberg et al., 1990). Thus, the structure presented here not only reveals the molecular basis for the tropism conferred by the PapGII adhesin but also suggests a plausible model that accounts for the tropism conferred by PapGIII. This model also predicts that GbO5-containing receptors may predominate in the adult human bladder over GbO3 or GbO4, since PapGIII containing strains are more prevalent in human cystitis than PapGII containing strains (Johnson et al., 1998).

Conclusion

The ability of pathogens to recognize host receptors is a critical event in most infectious diseases. The recognition event is often mediated by a bacterial adhesin located at the tip of a pilus structure. The P pilus is one of the most extensively studied adhesive fibers produced by bacteria and has served as a prototype for understanding host pathogen interactions for over a decade (Sauer et al., 2000; Soto and Hultgren, 1999; Thanassi and Hultgren, 2000). A vast amount of work has been carried out in an attempt to characterize the molecular basis of the tropism that the PapG adhesin determines and the function of the PapG-globoside interaction in disease. The three-dimensional structure presented here provides a view of these interactions, thereby revealing the molecular basis of a critical event in pyelonephritis: the colonization of the kidney. The structure of the adhesin-receptor complex reveals an intricate network of interactions between the adhesin and the glycolipid receptor that is likely to reflect a general mechanism for the function of all bacterial adhesins. The mutational analysis presented here, in concert with earlier studies using specific receptor analogs, pinpoint the importance of the interactions required for a stable adhesin-receptor interaction. In addition, the structure predicts that the adhesin domain binds "side-on" to the eukaryotic membrane, which suggests a requirement for a tip fibrillum flexible enough to orientate the receptor binding surfaces parallel to the membrane. Understanding the fine molecular details of host pathogen interactions will lead to the development of adhesin-based vaccines that are rationally designed to inhibit the patho-

genic process at an early stage and will provide a framework that will be used in the rational design of novel antibiotic compounds. Thus, continued research into the structure, function, and biogenesis of bacterial adhesins promises not only to enhance our knowledge of pathogenic processes, but may also help augment our current arsenal of antimicrobial agents.

Experimental Procedures

Cloning, Protein Expression, and Purification

Standard PCR and recombinant techniques were used to amplify and subclone the signal sequence and the N-terminal 196 amino acids of PapG class II which contain the receptor binding domain of PapGII to create the plasmid pTRCGII196. C600/pTRCGII196 was then grown and expression of GII196 was induced by addition of IPTG to 0.1 mM. Cells were harvested and processed for periplasmic extracts by resuspension in 20 mM Tris HCl (pH 8.0), 20% sucrose, 5 mM EDTA, and 150 μg/ml lysozyme. MgCl$_2$ was added to 10 mM, followed by centrifugation at 12 K for 40 min. The PapGII receptor binding domain was selectively precipitated using 30% ammonium sulfate. After centrifugation, the pellet was resuspended and dialyzed in 20 mM MES (pH 5.8). The GII196 protein was then purified using S Sepharose and butyl hydrophobic interaction chromotagraphy. The pure protein was dialyzed against 20 mM MES (pH 5.8). To produce the selenomethionine-substituted (SeMet-GII196) protein, pTRCGII196 was transferred in DL41 and grown in LeMaster medium complemented by selenomethionine (Lemaster and Richards, 1985). Production and purification of SeMet-GII196 were as for wild-type.

Crystallization and Structure Determination

GII196 (the receptor binding domain of PapGII) was concentrated to 4–5 mg/ml. The binary complex was formed by adding a 2 molar excess of an analog of GbO4 containing a βO-linked trimethylsilylethyl (TMSEt) group instead of ceramide. Hence, the chemical composition of the GbO4 coumpound used in this study is: GalNAcβ1-3Galα1-4Galβ1-4GlcβOTMSEt. Crystals of the binary complex were obtained by the hanging drop vapor diffusion method using a reservoir solution containing 100 mM sodium citrate (pH 5.2), and 0.5 M to 1.1 M ammonium acetate. Crystals of the unbound form were grown by vapor diffusion against a reservoir containing 4% (w/v) PEG20,000. Apo and complex crystals were cryoprotected by the sequential addition of glycerol to a final concentration of 30%. Both apo and complex crystals were in space group I222 with unit cell dimensions a = 55.1 Å, b = 79.7 Å, and c = 158.0 Å for the complex crystals and a = 55.1 Å, b = 78.2 Å, and c = 158.4 Å for the apo form crystals. Both apo and complex crystals contained one complex per asymmetric unit. SeMet-GII196 crystallized under similar conditions and the complex SeMet-containing crystals were isomorphous to those grown from the wild-type protein.

SeMet-GII196 contains three selenomethionines. MAD data were collected to a resolution of 1.8 Å at three wavelengths using a single complex crystal (beamline 19BM, Structural Biology Center, Advanced Photon Source). Data were indexed, integrated, and reduced using DENZO and SCALEPACK (Otwinowski and Minor, 1997). Heavy atom positions were determined and experimental phases were calculated using the program SOLVE to yield an interpretable electron density map (Terwilliger, 1994). This map was further improved by solvent flattening using the program DM (CCP4, 1994). A representative region of the resulting electron density is shown in Figure 1A. This map was used to place all residues in the protein using the program O and a database of protein structures (Jones and Thirup, 1986; Jones et al., 1991). The receptor was also very well defined in the map (Figure 1A) except for the O-linked TMSEt which had no interpretable electron density. Hence, only the tetrasaccharide moiety was built. A model for the receptor was first constructed using InsightII (MSI, San Diego) and then placed in electron density using the program TURBO (Roussel and Cambillau, 1991). The resulting atomic model was refined using the program CNS (Brünger et al., 1998). CNS topology and parameter files for the tetrasaccharide GalNAcβ1-3Galα1-4Galβ1-4Glc were obtained from the heterocompound information center at the Uppsala soft-

ware factory (http://xray.bmc.uu.se/usf). The free R factor was monitored as an indicator of model improvement (Brünger, 1992). Water molecules were added conservatively. After bulk solvent and overall anisotropic B-factor correction, the refinement converged to a final R factor of 21.9% with an R free factor of 24.0% (30–1.8 Å resolution data; $|F|/\sigma|F| > 1.0$) with good stereochemistry (Table 1). The model includes residues 1–196, 1 molecule of GalNAcβ1-3Galα1-4Galβ1-4Glc, and 185 water molecules. All phi and psi angles lie in the allowed region of the Ramachandran plot with 95% in the most favored regions.

For the unbound form, crystals diffracted to a resolution of 2.1 Å in the laboratory setting (Rigaku Raxis IV image plate mounted on a Rigaku RU200 rotating anode X-ray generator). The atomic model corresponding to the protein part of the refined complex structure was used directly in refinement against the unbound form data (Table 1) using simulated annealing (Brünger et al., 1987). Water molecules were then added in electron density. After bulk solvent and overall anisotropic B-factor correction, R and free R factor values are 23.2% and 25.4% (30–2.1 Å resolution data; $|F|/\sigma|F| > 1.0$) with good stereochemistry (Table 1).

Mutational Studies and Hemaglutination Assay

Mutations in residues involved in receptor binding were made in the full-length papGII gene from pDC1 (Clegg, 1982) by standard PCR methods and verified by sequence analysis. The ability of the mutated PapGII adhesins to be localized to the tip of P pili was verified by labeling for 1 hr C600 cells containing pKD320 (papCD-KEFG::kan) and the pTRCpapGII variants with [^{35}S]methionine and cysteine (100 μCi/ml) after growth in minimal M9/PO4 medium to an OD$_{600}$ of ~0.6 and inducing with 1 mM IPTG. Cells expressing Pap tips were pelleted and resuspended in phosphate-buffered saline, heated to 65°C for 30 min to release Pap tips, and pelleted. The Pap tip-containing supernatants were then immunoprecipitated using protein G Sepharose beads and anti-tip antibody. The beads were then pelleted, washed twice with RIPA Buffer (100 mM Na$_2$HPO$_4$ [pH 8.0], 150 mM NaCl), and twice with RIPA buffer + 1 M NaCl, and finally once again with RIPA buffer. SDS-PAGE and autoradiography of the immunoprecipitated samples confirmed that each of the mutated adhesins was able to be localized to Pap tips. Hemaglutinations, a measure of the ability of the PapGII being expressed at the distal end of P pili to bind to GbO4 present on red blood cells, were carried out using the strain HB101/pFJ29G::kan transformed with a pTRC99 vector encoding the wild-type or the indicated papGII mutants and human erythrocytes as described previously (Hultgren et al., 1990). In brief, the bacteria expressing P pili containing the PapG variants were resuspended in phosphate buffered saline (PBS) to an OD$_{600}$ = 1.0 and then pelleted and resuspended in 1/10 volume of PBS. These bacterial suspensions were serially diluted in 2-fold increments in v-bottom microtiter plates. PBS-washed human erythrocytes (OD$_{640}$ = 1.7) were added to each well and the plates incubated at 4°C for 2–16 hr. The ability of the various bacterial dilutions to prevent erythryocytes from settling over a given incubation period is a measure of the ability of the adhesin to bind to GbO4 present on the erythrocytes.

References

Brünger, A.T. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–474.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P.,

Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. *D54*, 905–921.

Brünger, A.T., Kuriyan, J., and Karplus, M. (1987). Crystallographic R-factor refinement by molecular dynamics. Science *235*, 458–460.

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

CCP4 (1994). Collaborative computational project number 4. The CCP4 suite programs for protein crystallography. Acta Crystallogr. *D50*, 760–763.

Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S.J., and Knight, S.D. (1999). X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. Science *285*, 1061–1066.

Clegg, S. (1982). Cloning of genes determining the production of mannose-resistant fimbriae in a uropathogenic strain of *Escherichia coli* belonging to serogroup O6. Infect. Immun. *38*, 739–744.

Finlay, B.B., and Falkow, S. (1997). Common themes in microbial pathogenicity revisited. Microbiol. Mol. Reviews *61*, 136–169.

Foxman, B. (1990). Recurring urinary tract infection: incidence and risks factors. Am. J. Public Health *80*, 331–333.

Guex, N., Diemand, A., and Peitsch, M.C. (1999). Protein modelling for all. TIBS *24*, 364–367.

Hedlund, M., Svensson, M., Nilsson, A., Duan, R.D., and Svanborg, C. (1996). Role of the ceramide-signaling pathway in cytokine responses to P-fimbriated *Escherichia coli*. J. Exp. Med. *183*, 1037–1044.

Hedlund, M., Wachtler, C., Johansson, E., Hang, L., Somerville, J.E., Darveau, R.P., and Svanborg, C. (1999). P fimbriae-dependent, lipopolysaccharide-independent activation of epithelial cytokine responses. Mol. Microbiol. *33*, 693–703.

Hooton, T.M., and Stamm, W.E. (1997). Diagnosis and treatment of uncomplicated urinary tract infection. Infect. Dis. Clin. N. Amer. *11*, 551–574.

Hultgren, S.J., Duncan, J.L., Schaeffer, A.J., and Amundsen, S.K. (1990). Mannose-sensitive haemagglutination in the absence of piliation in *Escherichia coli*. Mol. Microbiol. *4*, 1311–1318.

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.

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

Johnson, J.R., Russo, T.A., Brown, J.J., and Stapleton, A. (1998). PapG alleles of *Escherichia coli* strains causing first-episode or recurrent acute cystitis in adult women. J. Infect. Dis. *177*, 97–101.

Jones, T.A., and Thirup, S. (1986). Using known substructures in protein model building and crystallography. EMBO J. *5*, 819–822.

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.

Kihlberg, J., Hultgren, S.J., Normark, S., and Magnusson, G. (1989). Probing the combining site of the PapG adhesin of uropathogenic *Escherichia coli* bacteria by synthetic analogues of galabiose. J. Am. Chem. Soc. *111*, 6364–6368.

Klann, A.G., Hull, R.A., Palzkill, T., and Hull, S.I. (1994). Alanine-scanning mutagenesis reveals residues involved in binding of pap-3-encoded pili. J. Bacteriol. *176*, 2312–2317.

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.

Leffler, H., and Svanborg-Eden, C. (1980). Chemical identification of a glycosphingolipid receptor for *Escherichia coli* attaching to human urinary tract epithelial cells and agglutinating human erythrocytes. FEMS Microbiol. Lett. *8*, 127–134.

Lemaster, D.M., and Richards, F.M. (1985). H-^{15}N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. Biochemistry *24*, 7263–7268.

Lund, B., Lindberg, F., Marklund, B.I., and Normark, S. (1987). The PapG protein is the alpha-D-galactopyranosyl-(1–4)-beta-D-galactopyranose-binding adhesin of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. USA *84*, 5898–5902.

Martinez, J.J., Mulvey, M.A., Schilling, J.D., Pinkner, J.S., and Hultgren, S.J. (2000). Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. EMBO J. *19*, 2803–2812.

Nicholls, A., and Honig, B. (1991). A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzman equation. J. Comp. Chem. *12*, 435–445.

Otwinowski, O., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Meth. Enzymol. *276*, 307–326.

Pascher, I., Lundmark, M., Nyholm, P.G., and Sundell, S. (1992). Crystal structures of membrane lipids. Biochim. Biophys. Acta *1113*, 339–373.

Patton, J.P., Nash, D.B., and Abrutyn, E. (1991). Urinary tract infection: economic considerations. Med. Clin. N. Amer. *75*, 495–513.

Pernestig, A.K., Normark, S.J., Georgellis, D., and Melefors, O. (2000). The role of the AirS two-component system in uropathogenic *Escherichia coli*. Adv. Exp. Med. Biol. *485*, 137–142.

Roberts, J.A., Marklund, B.I., Ilver, D., Haslam, D., Kaack, M.B., Baskin, G., Louis, M., Mollby, R., Winberg, J., and Normark, S. (1994). The Gal(alpha 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.

Ronald, A.R., and Pattullo, A.L. (1991). The natural history of urinary infection in adults. Med. Clin. N. Amer. *75*, 299–312.

Roussel, A., and Cambillau, C. (1991). TURBO-FRODO. In Silicon Graphics Geometry Partners Directory (Silicon Graphics), pp. 86.

Sauer, F.G., Fütterer, K., Pinkner, J.S., Dobson, K.W., Hultgren, S.J., and Waksman, G. (1999). Structural basis of chaperone function and pilus biogenesis. Science *285*, 1058–1061.

Sauer, F.G., Knight, S.D., Waksman, G., and Hultgren, S.J. (2000). PapD-like chaperones and pilus biogenesis. Semin. Cell. Dev. Biol. *11*, 27–34.

Schilling, J.D., Mulvey, M.A., Vincent, C.D., Lorenz, R.G., and Hultgren, S.J. (2001). Bacterial invasion augments epithelial cytokine responses to *Escherichia coli* through a lipopolysaccharide-dependent mechanism. J. Immunol. *166*, 1148–1155.

Soto, G.E., and Hultgren, S.J. (1999). Bacterial adhesins: common themes and variations in architecture and assembly. J. Bacteriol. *181*, 1059–1071.

Stapleton, A.E., Stroud, M.R., Hakomori, S.I., and Stamm, W.E. (1998). The globoseries glycosphingolipid sialosyl galactosyl globoside is found in urinary tract tissues and is a preferred binding receptor in vitro for uropathogenic *Escherichia coli* expressing pap-encoded adhesins. Infect. Immun. *66*, 3856–3861.

Striker, R., Nilsson, U., Stonecipher, A., Magnusson, G., and Hultgren, S.J. (1995). Structural requirements for the glycolipid receptor of human uropathogenic *Escherichia coli*. Mol. Microbiol. *16*, 1021–1029.

Stromberg, N., Marklund, B.I., Lund, B., Ilver, D., Hamers, A., Gaastra, W., Karlsson, K.A., and Normark, S. (1990). Host-specificity of uropathogenic *Escherichia coli* depends on differences in binding specificity to Gal alpha 1–4Gal-containing isoreceptors. EMBO J. *9*, 2001–2010.

Stromberg, N., Nyholm, P.G., Pascher, I., and Normark, S. (1991). Saccharide orientation at the cell surface affects glycolipid receptor function. Proc. Natl. Acad. Sci. USA *88*, 9340–9344.

Stroud, M.R., Stapleton, A.E., and Levery, S.B. (1998). The P histo-blood group-related glycosphingolipid sialosyl galactosyl globoside as a preferred binding receptor for uropathogenic Escherichia coli: isolation and structural characterization from human kidney. Biochemistry *37*, 17420–17428.

Svanborg, C., and Godaly, G. (1997). Bacterial virulence in urinary tract infection. Infect. Dis. Clin. N. Amer. *11*, 513–525.

Svanborg, C., Hedlund, M., Connell, H., Agace, W., Duan, R.D., Nilsson, A., and Wullt, B. (1996). Bacterial adherence and mucosal cytokine responses. Receptors and transmembrane signaling. Ann. NY Acad. Sci. *797*, 177–190.

Terwilliger, T.C. (1994). MAD phasing: treatment of dispersive differences as isomorphous replacement information. Acta Crystallogr. *D50*, 17–23.

Thanassi, D.G., and Hultgren, S.J. (2000). Assembly of complex organelles: pilus biogenesis in gram-negative bacteria as a model system. Methods *20*, 111–126.

Warren, J. (1996). Clinical presentations and epidemiology of urinary tract infections. In Urinary Tract Infections: Molecular Pathogenesis and Clinical Management, H.L.T. Mobley and J. W. Warren, eds. (Washington, DC: ASM Press), pp. 3–27.

Zhang, J.P., and Normark, S. (1996). Induction of gene expression in *Escherichia coli* after pilus-mediated adherence. Science *273*, 1234–1236.

Accession Numbers

The apo and GbO4-bound forms of the PapG receptor binding domain reported in this paper have been deposited in the Protein Data Bank with entry codes 1J8R and 1J8S.