Mini-review

Analysis of the primary structure of P-fimbrillins of uropathogenic *Escherichia coli*

Irma van Die, Wiel Hoekstra and Hans Bergmans

Department of Molecular Cell Biology, University of Utrecht, 3584 CH Utrecht, The Netherlands

Introduction

The ability of uropathogenic *Escherichia coli* to adhere to uroepithelial cells is a prerequisite for successful colonization of the urinary tract by this organism, and the subsequent development of urinary tract infections.[1-3] The mechanism of adherence has been studied intensively during the last years. Many investigators support the view that P-fimbriae play an important role in this adhesion process. P-fimbriae recognize a disaccharide with the structure α-D-Galp(1–4) β-D-Galp in the uroepithelium.[4,5] This structure is also part of the P-blood-group antigen. Adherence of P-fimbriated cells to the P blood group antigen on the surface of human erytrocytes causes mannose resistant hemagglutination (MRHA).

Serologically P-fimbriae are a heterogeneous group, with serotypes F7-F13 discriminated.[6] Specific monoclonal antibodies can be raised against isolated P-fimbriae of different serotype, showing the presence of unique epitopes on these fimbriae.[7] P-fimbriae have a very complicated structure. They are built up from a thousand identical subunits (P-fimbrillins or major subunits) with an apparent molecular weight of 16–23 kDa. However, also minor proteins have been shown to occur in the fimbriae (< 1%), the exact way in which these proteins are associated with the fimbrial structure is not known, i.e. it is not clear whether these minor proteins are an integral part of the fimbrial structure, or are more or less tightly associated with the fimbriae.[8,9] It is supposed that one of these minor proteins represents the actual adhesin protein[8,9] and that the fimbriae function as carriers to expose the adhesin outside the LPS layer of the cell wall.[10]

The biogenesis of P-fimbriae requires the major and minor subunit proteins, and a number of accessory proteins supposedly involved in the translocation of the subunits to the cell surface and their assembly into fimbriae. It has been shown that the P-fimbrillin of one gene cluster can be assembled to fimbriae by the accessory proteins of other P-fimbrial gene clusters, and also minor proteins can be exchanged between the gene clusters.[11] Genetic analysis of the gene clusters involved in the biogenesis of P-fimbriae of various serotype[12-14] has shown that the genetic organization of the genes encoding the subunits and accessory proteins in the clusters is quite similar.

Given the serological difference of P-fimbriae, the different major subunits should contain variable regions that build up the specific epitopes. The fact that different subunits can use the same set of accessory proteins implies that the regions of the major subunits that interact with the accessory proteins are conserved. Also it is

feasible that specific regions exist that are involved in the interaction with the minor proteins. Therefore we expect that detailed knowledge of the structure of the P-fimbrillin and the minor proteins will help to get more insight in the mechanisms underlying the biogenesis of fimbriae. In this paper we summarize the data on the primary structure of four different P-fimbrillins and we present an analysis of these primary structures, focussing on the conserved and variable regions.

Nucleotide and amino acid sequences of the F7$_1$, F7$_2$, F11 and F13 fimbrillins

The nucleotide sequences of the P-fimbrillins of serotype F7$_1$, F7$_2$, F11 and F13 have been described before.[15–18] The deduced amino acid sequences of the mature fimbrillin proteins are shown in Fig. 1. The mature proteins in Fig. 1 are aligned for maximal homology. Non-homologous amino acids that are functionally similar are indicated by vertical lines.

This alignment already shows clearly conserved and less conserved regions. It is to be expected that conserved regions have a similar secondary structure in the protein, whereas the secondary structure of the non-conserved regions may be different. The secondary structure of the four subunit proteins was deduced from the primary stucture using the algorithm of Garnier *et al.*[19] implemented in the Beckman Microgenie computer programs. As these predictions are known to be unreliable, we estimated the prediction to be of value when it happened to be the same in three of the four proteins; so only in those stretches where the predicted secondary structure is similar in three of the four proteins is the predicted secondary structure indicated below the amino acid sequences in Fig. 1.

Based on these data four regions can be discriminated in the P-fimbrillin (Table 1, Fig. 1): the N-terminal part, the region between the two cysteine residues in the protein (the cys–cys loop), a variable region, and the C-terminal part. The N- and C-terminal parts of the proteins are very homologous. In the cys–cys loop the homology on amino acid level is clearly lower, but the predicted secondary structure in the major part of this region is similar. Clearly less homology is found in the variable region; only in this region gaps are needed for maximal alignment of the sequences. In this part of the protein five hypervariable regions are found: stretches of amino acids that are different among the proteins, both in amino acid composition and predicted secondary structure; besides, in these hypervariable regions some of the proteins have additional amino acids.

For all P-fimbrillins the hydrophilicity values were determined according to Hopp & Woods.[20] It was observed that the highly conserved N- and C-terminal parts of the proteins consist mainly of hydrophobic aminoacids, the C-terminus being highly hydrophobic (mean hydrophilicity values ranging from −0.8 to −1.1 in the region starting at amino acid residue 155). In the cys–cys loop two hydrophilic maxima were found but no pronounced hydrophobic regions. In the variable region hydrophilic as well as hydrophobic stretches were found. It is remarkable that the hydrophobic regions in the P-fimbrillins correspond very well to the conserved amino acid stretches as deduced in Fig. 1 (Table 2a). The hydrophilic maxima in the variable region are found mainly in or very close to the hypervariable regions (Table 2b). The cys–cys loop contains two hydrophilic maxima in a conserved region.

Discussions and conclusions

Based on the data described above a schematic model for P-fimbrillins can be drawn (Fig. 2), showing the conserved, variable and hypervariable regions. The hypervariable

1   **28**

```
F7₂   ALA PRO THR |ILE PRO GLN GLY GLN GLY| LYS VAL |THR PHE| ASN GLY THR VAL VAL ASP ALA PRO CYS |GLY| ILE |ASP ALA GLN |SER
F7₁   ALA ALA SER |ILE PRO GLN GLY GLN GLY| GLU VAL |SER PHE| LYS GLY THR VAL VAL ASP ALA PRO CYS |GLY| ILE |GLU THR GLN |SER
PapA  ALA PRO THR |ILE PRO GLN GLY GLN GLY| LYS VAL |THR PHE| ASN GLY THR VAL VAL ASP ALA PRO CYS |SER| ILE |SER GLN LYS |SER
F11   ALA PRO THR |ILE PRO GLN GLY GLN GLY| LYS VAL |THR PHE| ASN GLY THR VAL VAL ASP ALA PRO CYS |SER| ILE |SER GLN LYS |SER

      T  B  B  B  r  r  r  T  T  T  r  r  B  B  B  B  B  B  B  B  B  B        T  T  T  T  T
```

29   **56**

```
F7₂   ALA ASP |GLN| SER |ILE ASP PHE GLY GLN| VAL SER LYS LEU |PHE LEU| GLU ASN ASP |GLY| GLU SER GLN |PRO| LYS SER PHE ASP |ILE
F7₁   ALA LYS |GLN| GLU |ILE ASP PHE GLY GLN| ILE SER LYS SER |PHE LEU| GLN GLU GLY |GLY| GLU THR GLN |PRO| LYS ASP LEU ASN |ILE
PapA  ALA ASP |GLN| SER |ILE ASP PHE GLY GLN| LEU SER LYS SER |PHE LEU| GLU ALA GLY |GLY| VAL SER LYS |PRO| MET ASP LEU ASP |ILE
F11   ALA ASP |GLN| SER |ILE ASP PHE GLY GLN| LEU SER LYS SER |PHE LEU| GLU ALA GLY |GLY| THR SER LYS |PRO| MET ASP LEU ASP |ILE

      T  T  r  r        B  T  T  T  r  r  A  A  A  A  A  A  r  r        r  r
```

57   **84**

```
F7₂   LYS |LEU| ILE |ASN CYS ASP ILE THR| ASN PHE |LYS| LYS ALA ALA GLY |GLY| GLY GLY |ALA LYS THR |GLY| THR |VAL| SER |LEU| THR |PHE
F7₁   LYS |LEU| VAL |ASN CYS ASP ILE THR| ASN LEU |LYS| GLN LEU GLN GLY |GLY|  -  ALA |ALA LYS LYS |GLY| THR |VAL| SER |LEU| THR |PHE
PapA  GLU |LEU| VAL |ASN CYS ASP ILE THR| ALA PHE |LYS|  -   -   -  GLY |GLY| ASN GLY |ALA LYS LYS |GLY| THR |VAL| LYS |LEU| ALA |PHE
F11   GLU |LEU| VAL |ASN CYS ASP ILE THR| ALA PHE |LYS|  -   -   -  GLN |GLY| GLN ALA |ALA LYS ASN |GLY| LYS |VAL| GLN |LEU| SER |PHE

           A  T        B  B  B  B  B                       r  r  r        B  B  B  B
```

85   **112**

```
F7₂   SER |GLY| VAL PRO SER GLY PRO GLN SER ASP MET |LEU| GLN |THR| VAL |GLY| ALA |THR| ASN |THR ALA ILE VAL VAL| THR ASP PRO HIS
F7₁   SER |GLY| VAL PRO ALA GLU ASN ALA ASP ASP MET |LEU| GLN |THR| VAL |GLY| ASP |THR| ASN |THR ALA ILE VAL VAL| THR ASP SER SER
PapA  THR |GLY| PRO ILE VAL ASN GLY HIS SER ASP GLU |LEU| ASP |THR| ASN |GLY| GLY |THR| GLY |THR ALA ILE VAL VAL| GLN  -  GLY ALA
F11   THR |GLY| PRO GLN VAL THR GLY GLN ALA GLU GLU |LEU| ALA |THR| ASN |GLY| GLY |THR| GLY |THR ALA ILE VAL VAL| GLN  -  ALA ALA

      B        r  r        T              B  B              r  r  B  B  B  B  B  B        r  r
```

113   **140**

```
F7₂   GLY LYS |ARG| VAL |LYS| PHE ASP GLY |ALA THR ALA THR GLY VAL SER TYR |LEU| VAL ASP |GLY| ASP |ASN| THR ILE |HIS| PHE |THR ALA
F7₁   GLY LYS |ARG| VAL |LYS| PHE ASP GLY |ALA THR GLU THR GLY ALA SER ASN |LEU| ILE ASN |GLY| ASP |ASN| THR ILE |HIS| PHE |THR ALA
PapA  GLY LYS |ASN| VAL |VAL| PHE ASP GLY | -  SER GLU GLY ASP ALA ASN THR |LEU| LYS ASP |GLY| GLU |ASN| VAL LEU |HIS| TYR |THR ALA
F11   GLY LYS |ASN| VAL |SER| PHE ASP GLY | -  THR ALA GLY ASP ALA TYR PRO |LEU| LYS ASP |GLY| ASP |ASN| VAL LEU |HIS| TYR |THR ALA

           T        B                                          B        T  T  T        B  B  B  B  B
```

141   **168**

```
F7₂   ALA VAL ARG |LYS| ASP GLY  -  SER |GLY| ASN PRO |VAL| THR |GLU GLY ALA PHE SER ALA VAL ALA| ASN |PHE ASN LEU |THR| TYR GLN
F7₁   PHE VAL LYS |LYS| ASP ASN  -  SER |GLY| LYS ASN |VAL| ALA |GLU GLY ALA PHE SER ALA VAL ALA| ASN |PHE ASN LEU |THR| TYR GLN
PapA  VAL VAL LYS |LYS| SER SER ALA VAL |GLY| ALA ALA |VAL| THR |GLU GLY ALA PHE SER ALA VAL ALA| ASN |PHE ASN LEU |THR| TYR GLN
F11   LEU VAL LYS |LYS|  -   -  ALA ASN |GLY| GLY THR |VAL| SER |GLU GLY ALA PHE SER ALA VAL ALA| THR |PHE ASN LEU |SER| TYR GLN

      B  B  A  A              r              A  A  A  A  A  r  r  B  B  r  B  B
```
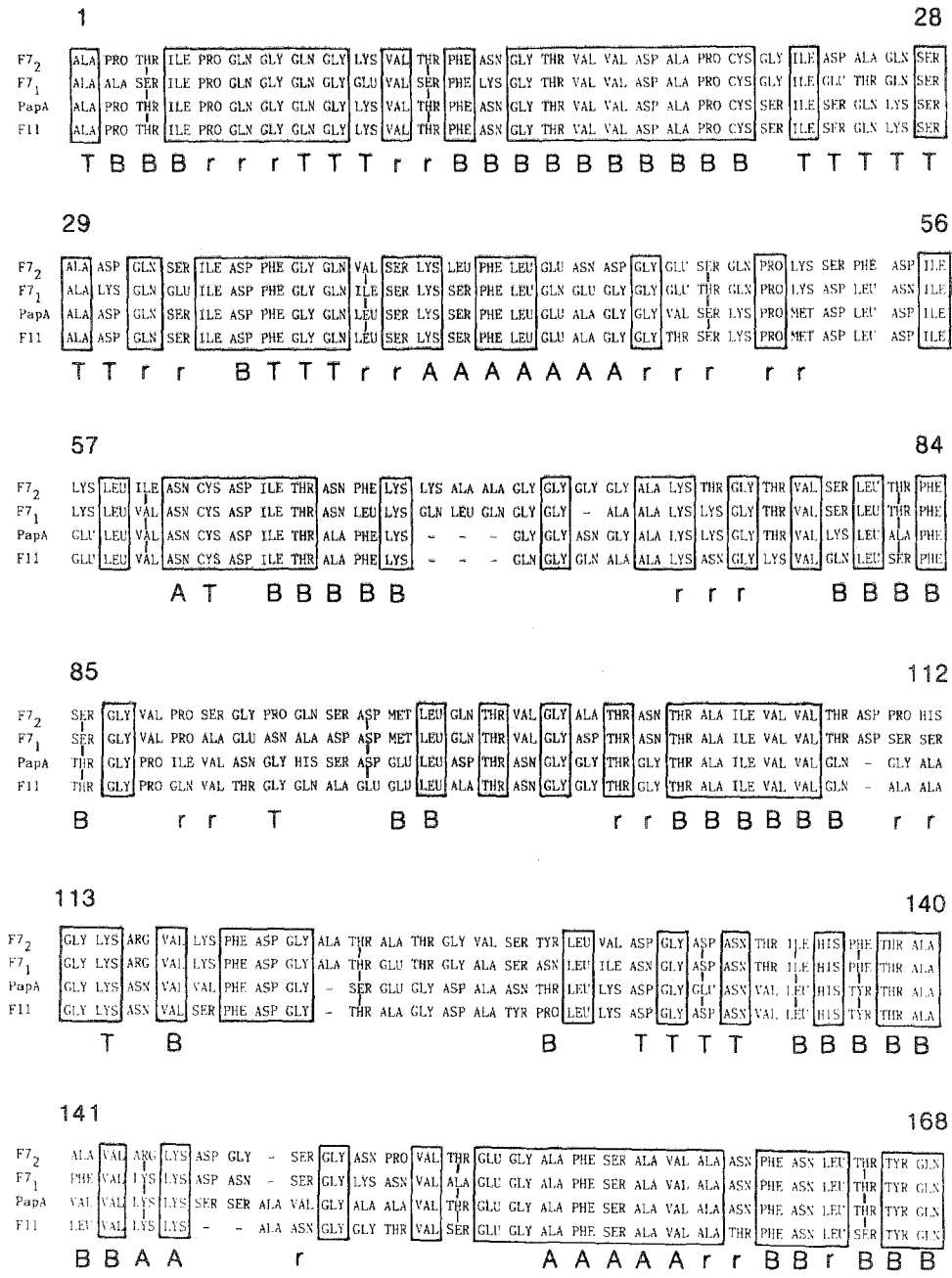
Fig. 1. Amino acid sequences of the mature major fimbrial subunit proteins of the F7₂, F7₁, Pap and F11 fimbriae, aligned for maximal homology. Homologous amino acids are boxed; functionally similar amino acids are indicated by vertical lines. The letters A, B, T and r correspond to the predicted secondary structure of the fimbrillins A = α-helix, B ≈ β sheet, T = turn and r = random coil.

Table 1 Homology in the different regions of the protein sequences among the $F7_1$, $F7_2$, F11 and F13 fimbrillins

Regions in the P-fimbrillin	Amino acids (Fig. 2)	% homology based on amino acid similarity	% homology based on predicted secondary structure
N-terminus	1–21	76	100
cys–cys loop	22–61	50	75
variable region	62–153	40	50
C-terminus	154–168	87	87

Table 2 Hydrophobic and hydrophilic regions in P-fimbrillins correlated with the observed sequence conservation in these proteins.

(a) Part of fimbrillin	Hydrophobic regions[a] in the fimbrillins of serotype				Main conserved regions (Fig.2)
	$F7_1$	$F7_2$	F11	F13	
cys–cys loop	—	—	—	—	24–50
variable region	—	—	—	—	60–68
	78–85	77–85	80–85	79–89	75–86
	102–106	101–106	101–108	101–108	102–110
	—	121–127	—	—	—
	134–139	134–139	133–139	133–139	131–145

(b) Part of fimbrillin	Hydrophilic regions[a] in fimbrillins of serotype				Hypervariable regions (Fig. 2)
	$F7_1$	$F7_2$	F11	F13	
cys–cys loop	25–31	25–30	25–30	25–31	—
	43–53	43–53	48–54	48–54	—
variable region	71–77	—	73–77	68–77	68–75
	88–94	—	90–95	90–96	87–96
	109–116	109–116	—	—	109–113
	—	—	128–132	118–132	121–129
	140–146	139–146	—	140–145	145–149

[a]hydrophobic and hydrophilic regions: stretches of at least 4 amino acids with a hydrophilicity value of < -0.5 and $> +0.5$ respectively.

regions that coincide with the hydrophilic maxima are likely to be responsible for the serological diversity of the P-fimbriae. Among P-fimbriae a number of different serotypes have been distinguished, although cross reactivity was found among a number of them. The cross reactive epitopes could be located in the (conserved) hydrophilic maxima that are shown to occur in the cys–cys loop. However, the crossreactivity could also be due to the presence of minor proteins in the P-fimbriae. It has been shown by De Ree *et al.* that all their monoclonals that recognized epitopes on different P-fimbrial types, were directed against one of the minor proteins.[21] All other isolated monoclonals were specific only for one P-fimbrial type. These results argue that crossreactivity is due to the presence of minor proteins, however, they don't exclude the possibility that common epitopes could also occur.

De Ree *et al.*[22] studied the serological response of patients with pyelonephritis to

Fig. 2. Homologous regions in the P-fimbrial major subunit genes. The degree of shading indicates regions with high homology (black) through regions without homology (white).

P-fimbriae. Sera of these patients contained antibodies directed against the P-fimbrillins, but not against the actual adhesins. This antigenic response of the host might act as a selective pressure, causing serological diversity of the P-fimbrillins. Given the low immunogenicity of the minor proteins it may be predicted that the composition of these proteins is more constant, due to lack of selective pressure. This might enable the pathogen to conserve an adhesin with a constant specificity. The combination of an immunogenic variabie fimbrillin and a constant specific adhesin supplies the pathogen with a mechanism to evade the host immune system, a mechanism already proposed by Normark et al.[23] This mechanism might apply to the many uropathogenic E. coli strains that can express more than one P-fimbrial type by phase variation,[24,25] but also to the case of reinfection with another uropathogen with a different P-fimbrial type. P-fimbrial gene clusters as we find them now may have originated from a primordial gene cluster. The hypervariable regions evolved then due to the selective pressure described above. The conserved regions could be conserved either because there was no selective advantage for mutations in this region, or because these regions were essential for the biogenesis of fimbriae, i.e. for translocation of the subunit protein to the cell surface, for interaction with the accessory proteins or for interaction of the fimbrial subunits with each other. Both mechanisms for conservation are likely to have operated, and their results are difficult to separate. The complete homology of the F11 and Pap coding sequences up to base 150, including regions where there is only partial homology with the $F7_1$ and $F7_2$ coding sequences, may be seen as a case of conservation due to lack of selective pressure. On the other hand, the regions that have been conserved among all four proteins show codon degeneracy to a large extent. Of the 86 amino acids completely conserved (boxed in Fig. 1), 61 (71%) show codon degeneracy at the DNA level. This would argue that selective pressure has been operative in conserving these amino acids.

Little is known about the way the minor proteins are associated with the P-fimbriae. It is known that the same accessory proteins that are needed for the assembly of the P-fimbrillins into fimbriae, are involved in expression of the minor proteins at the cell surface.[11] Sequence data derived from the minor proteins of the F13 fimbrial gene cluster (PapE and PapF), show that the minor proteins have most of the general features characteristic of E. coli fimbrillins.[9] However, the observed sequence homology between the minor proteins and the P-fimbrillins is rather low. Most resemblances with known fimbrial subunits are found in the N-terminal part; here some homology is found and also a stretch of alternating hydrophilic and hydrophobic aminoacid residues is observed, similar to the one that was described by Klemm[26] as a homologous part of a number of fimbrillins.

The limited homology among the minor proteins and the P-fimbrillins is difficult to explain when one expects the minor proteins to be integrated in the fimbriae as 'normal' subunits, in between the P-fimbrillins. On the other hand, the resemblances to other fimbrial subunits appear to be significant.[9] This could mean, that the requirements for a protein to be assembled to fimbriae are not very stringent, which seems unlikely. Alternatively it might indicate that the minor proteins are not integrated in the head-to-tail fashion envisaged for the major subunits, but are in some other, possibly more loosely, way associated with them.

References

1. Van den Bosch JF, Verboom-Sohmer U, Postma P, De Graaff J, MacLaren DM. Mannose sensitive and mannose resistant adherence to human uroepithelial cells and urinary virulence of *Escherichia coli*. Infect Immun 1980; 29: 226–33.
2. Svanborg-Eden C, Hagberg L, Hanson LA, Korhonen TK, Leffler H, Olling S. Adhesion of *Escherichia coli* in urinary tract infection. In: Elliot K, ed. Adhesion and microorganism pathogenicity. Ciba Foundation Symposium. London: Pitman Medical, 1981; 161–87.
3. Hagberg L, Jodal U, Korhonen TK, Lidin-Jason G, Lindberg U, Svanborg-Eden C. Adhesion, Hemagglutination and virulence of *Escherichia coli* causing urinary tract infections. Infect Immun 1981; 31: 564–70.
4. Kallenius G, Mollby R, Svenson SB, Winberg J, Hultberg H. Identification of a carbohydrate receptor recognized by uropathogenic *Escherichia coli*. Infection 1980; 8: 228–93.
5. Leffler H, Svanborg-Eden C. Chemical identification of a glycosphingolipid receptor for *Escherichia coli* attaching to human urinary tract epithelial cells and agglutinating erythrocytes. FEMS Microbiol Lett 1980; 8: 127–34.
6. Orskov I, Orskov F. Serology of *Escherichia coli* fimbriae. Prog Allergy 1983; 33: 80–105.
7. De Ree JM, Schwillens P, Van den Bosch JF. Monoclonal antibodies that recognize the P-fimbriae $F7_1$, $F7_2$, F9 and F11 from uropathogenic *Escherichia coli*. Infect Immun 1985; 50: 900–4.
8. Lindberg FP, Lund B, Normark S. Genes of pyelonephritogenic *Escherichia coli* required for digalactosidase-specific agglutination of human cells. EMBO J 1984; 3: 1167–73.
9. Lindberg F, Lund B, Normark S. Gene products specifying adhesion of uropathogenic *Escherichia coli* are minor components of pili. Proc Natl Acad Sci 1986; 83: 1891–5.
10. Van Die I, Zuidweg E, Hoekstra W, Bergmans H. The role of fimbriae of uropathogenic *Escherichia coli* as carrier of the adhesin involved in mannose resistant hemagglutination. Microbial Path 1986; 1: 51–6.
11. Van Die I, Van Megen I, Zuidweg E, Hoekstra W, De Ree H, Van den Bosch JF, Bergmans H. Functional relationship between the gene clusters encoding $F7_1$, $F7_2$, F9 and F11 fimbriae of human uropathogenic *Escherichia coli*. J Bacteriol 1986; 167: 407–10.
12. Norgren M, Normark S, Lark D, O'Hanley P, Schoolnik G, Falkow S, Svanborg-Eden C, Båga M, Uhlin BE. Mutations in *Escherichia coli* cistrons do not abolish Pap pili fiber formation. EMBO J 1984; 3: 1159–65.
13. Van Die I, Van Megen I, Hoekstra W, Bergmans H. Molecular organization of the genes involved in the production of $F7_2$ fimbriae, causing mannose resistant hemagglutination, of a uropathogenic *Escherichia coli* 06:K2:H1:F7 strain. Mol Gen Genet 1984; 194: 528–33.
14. Van Die I, Spierings G, Van Megen I, Zuidweg E, Hoekstra W, Bergmans H. Cloning and genetic organization of the gene cluster encoding $F7_1$ fimbriae of a uropathogenic *Escherichia coli* and comparison with the $F7_2$ gene cluster. FEMS Microbiol Lett 1985; 28: 329–34.
15. Rhen M, Van Die I, Rhen V, Bergmans H. Comparison of the nucleotide sequences of the genes encoding the KS71A and $F7_1$ fimbrial antigens of uropathogenic *Escherichia coli*. Eur J Biochem 1985; 151: 573–7.
16. Van Die I, Bergmans H. Nucleotide sequence of the gene encoding the $F7_2$ fimbrial subunit of a uropathogenic *Escherichia coli* strain. Gene 1984; 32: 83–90.
17. Van Die I, Dijksterhuis M, De Cock H, Hoekstra W, Bergmans H. Structural variation of P-fimbriae from uropathogenic *Escherichia coli*. In: Lark D, ed. Protein carbohydrate interactions in biological systems. London: Academic Press, 1986; 39–46.
18. Båga M, Normark S, Hardy J, O'Hanley P, Lark D, Olsson O, Schoolnik G, Falkow S. Nucleotide sequence of the PapA gene encoding the PapA pilus subunit of human uropathogenic *Escherichia coli*. J Bacteriol 1984; 157: 330–3.
19. Garnier J, Osguthorpe DJ, Robson B. Analysis and the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 1978; 120: 97–120.
20. Hopp TP, Woods KR. Prediction of protein antigenic determinants from aminoacid sequences. Proc Natl Acad Sci USA 1981; 78: 3824–8.
21. De Ree JM, Schwillens P, Van den Bosch JF. Monoclonal antibodies raised against Pap fimbriae recognize minor component(s) involved in receptor binding. Microbial Path 1987; 2: 113–21.
22. De Ree JM. PhD Thesis (1986) Maastricht, The Netherlands.
23. Normark S, Lindberg F, Lund B, Båga M, Ekbäck G, Göransson M, Mörner S, Norgren M, Marklund BI, Uhlin BE. Minor pilus components acting as adhesins. In: Lark D, ed. Protein carbohydrate interactions in biological systems. London: Academic Press, 1986; 3–12.
24. Rhen M, Mäkelä PH, Korhonen TK. P-fimbriae of *Escherichia coli* are subject to phase variation. FEMS Microbiol. Lett. 1983; 19: 267–271.
25. Nowicki B, Rhen M, Vaisänen-Rhen V, Pere A and Korhonen TK. Immunofluorescence study of fimbrial phase variation in *Escherichia coli* KS71. J Bacteriol 1984; 160: 691–5.
26. Klemm P. Fimbrial adhesins of *Escherichia coli*. Rev Infect Dis 1985; 7: 321–40.