Pilins of fimbrial adhesins of different member species of Enterobacteriaceae are structurally similar to the C-terminal half of adhesin proteins

Jean-Pierre Girardeau*, Yolande Bertin

Laboratoire de Microbiologie, INRA, Centre de Recherches de Clermont-Ferrand-Theix, 63122 Saint-Genès-Champanelle, France

Received 7 October 1994; revised version received 24 November 1994

Abstract The structural relatedness of pilins and the C-terminal half of adhesin proteins in different member species of *Enterobacteriaceae* was deduced from their two-dimensional sequence analysis using the hydrophobic cluster analysis (HCA) and secondary structure predictions from the profile network HeiDelberg program (PHD). Despite a large evolutionary distance between the two protein families, we show that pilins and the C-terminal domain of adhesins have a similar folding that can serve as modules for pilus assembly.

Key words: Structural similarity; Fimbrial adhesin; Hydrophobic cluster analysis; Protein structure prediction

1. Introduction

The adherence of bacteria to epithelial host cells is an important event in the pathogenesis of intestinal, respiratory and urinary tract infections. Although afimbrial adhesins have been identified, the adhesion of bacterial cell to the host cell surface is usually mediated by fimbriae or pili [1,2]. Pili are heteropolymers composed of one major subunit (pilin) and two or three minor subunits involved in initiation and elongation of the fiber and in the association with the adhesin protein, which mediates the specific binding of bacteria to the host cell surface [2,3]. Adhesin proteins are located either exclusively at the distal tip of the pili (Pap pili), or along the length of the pilus filaments (type I pili) [4,5]. Pilins, generally composed of 160 to 180 amino acid residues, contain two cysteine residues in the amino terminal half of the protein which may form intra-molecular disulphide bridges [6]. From Pap pilin sequence analyses, four regions are conveniently identified, the N-terminal region (R1 to R21), the Cys–Cys loop (R22 to R61), the variable region (R63 to R153) and the COOH-terminal region (R154 to terminus) [6]. Protein sequence analyses indicate variable homology (20–85%) between various pilins. The NH_2 and COOH termini of the protein which are the most conserved are involved in fimbrial polymerization through subunit-subunit or subunit-chaperone interactions [3,7].

Proteins corresponding in size and function to the *E. coli* type I adhesin (FimH) have been identified in various fimbriae expressed by different member species of the family *Enterobacteriaceae* [8–12]. Adhesin proteins, generally composed of 260 to 340 amino acid residues, contain two Cys–Cys loops in similar positions and share homology (20–40%) in amino acid sequence, indicating an apparent conservation in overall molecular structure. Although the adhesins are about twice the size of other fimbrial proteins, significant sequence agreement between pilins and the C-terminal half of adhesins has been reported [9,11], but whether or not adhesins and pilins support an overall homology in structure is not known.

Traditional sequence comparison methods failed to establish a satisfying alignment between distantly related proteins (< 20% identity). For this reason they are not successfully used in secondary structure predictions using multiple sequence alignments. In this paper, we describe a two-dimensional sequence analysis of adhesins and pilins using hydrophobic cluster analysis (HCA) [13,14] and secondary structure predictions produced by the Profil Network HeiDelberg program (PHD) [15]. We show that pilins and the C-terminal half of adhesins of different member species of *Enterobacteriaceae* are structurally similar.

2. Materials and methods

The 10 pilin and the 8 adhesin protein sequences were from Genbank and NBRF sequence data banks (Table 1). HCA plots with automatic hydrophobic cluster contouring were drawn using the plot program from Doriane S.A. (France) run on a Macintosh computer. This method can detect similarities in the secondary folding of protein domains even for weak sequence identities (<10%). Briefly, HCA is a protein sequence comparison method based on helical representation of the sequences where size, shape and orientation of hydrophobic clusters are primarily compared.

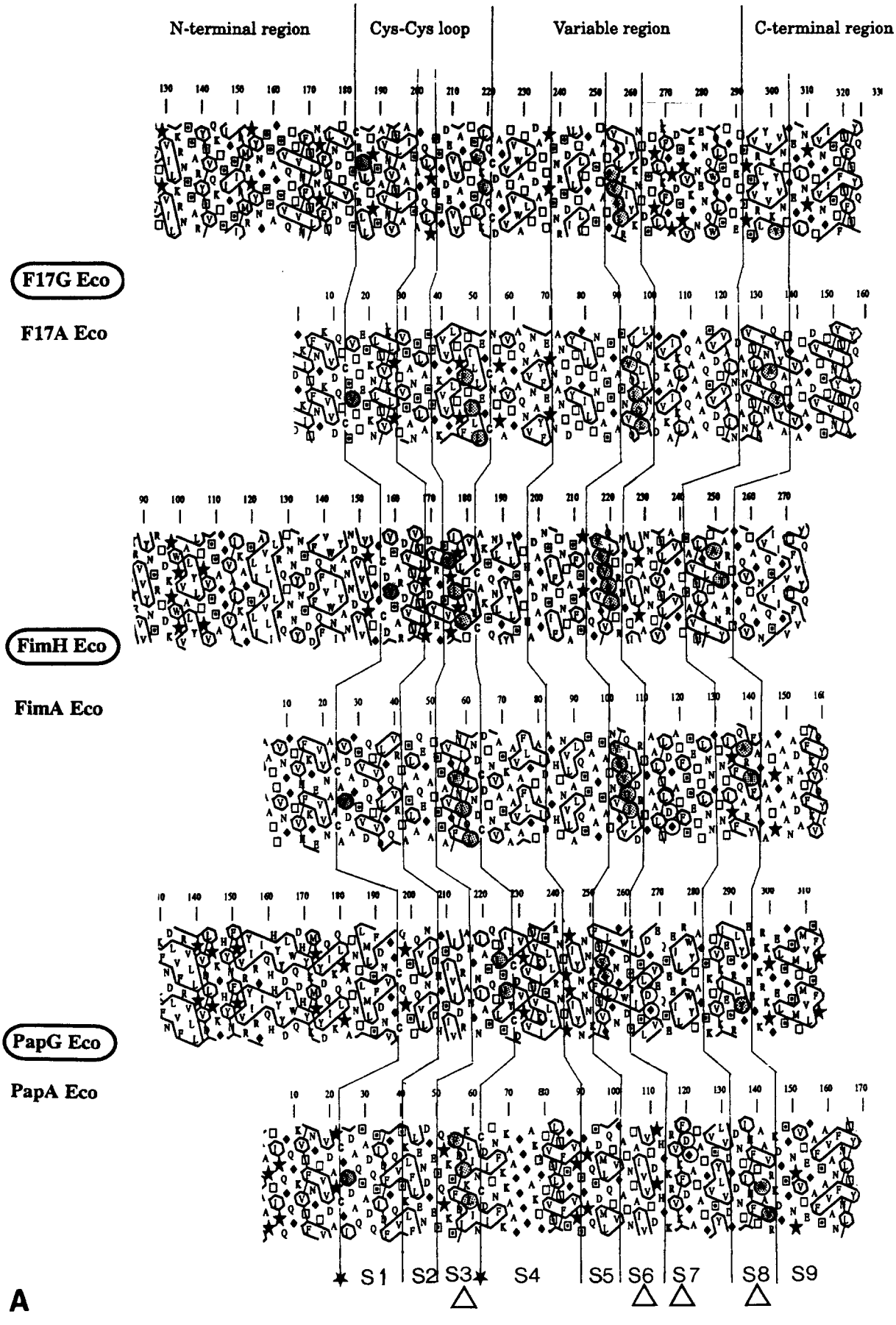
For HCA alignments, a numerical value (HCA score) can be calculated from clusters to assess the alignments [13,18]. HCA strategies and applications have been reviewed elsewhere [13,14]. The PileUp algorithm [16] provided by the GCG package from the University of Wisconsin [17] was used for calculating the percentage of residue identity. The secondary structure predictions for multiple sequence alignments (MaxHom) were produced by the PHD profile network method provided by the software facilities 'PredicProtein' of EMBL HeiDelberg [15].

3. Results

The HCA plots of pilins and the C-terminal half of adhesins originating from the same parental fimbriae (Fig. 1A), or from different parental fimbriae (Fig. 1B), were compared to detect similarities in shape, orientation and distribution of hydrophobic clusters over the sequences. When the two conserved cysteines of pilin sequences were positioned close to the second cysteine pair of adhesin sequences, topological similarities between hydrophobic clusters, loops connecting these structures and groups of identical amino acids were recognized. The occurrence of conserved segments (S1, S3 and S6 to S9) alternating with more variable ones (S2, S4, S5), clearly demonstrated

*Corresponding author. Fax: (33) 73 62 45 81.

SSDI 0014-5793(94)01340-3


N-terminal region
Cys-Cys loop
Variable region
C-terminal region
F17G Eco
F17A Eco
FimH Eco
FimA Eco
PapG Eco
PapA Eco
S1
S2
S3
S4
S5
S6
S7
S8
S9
A

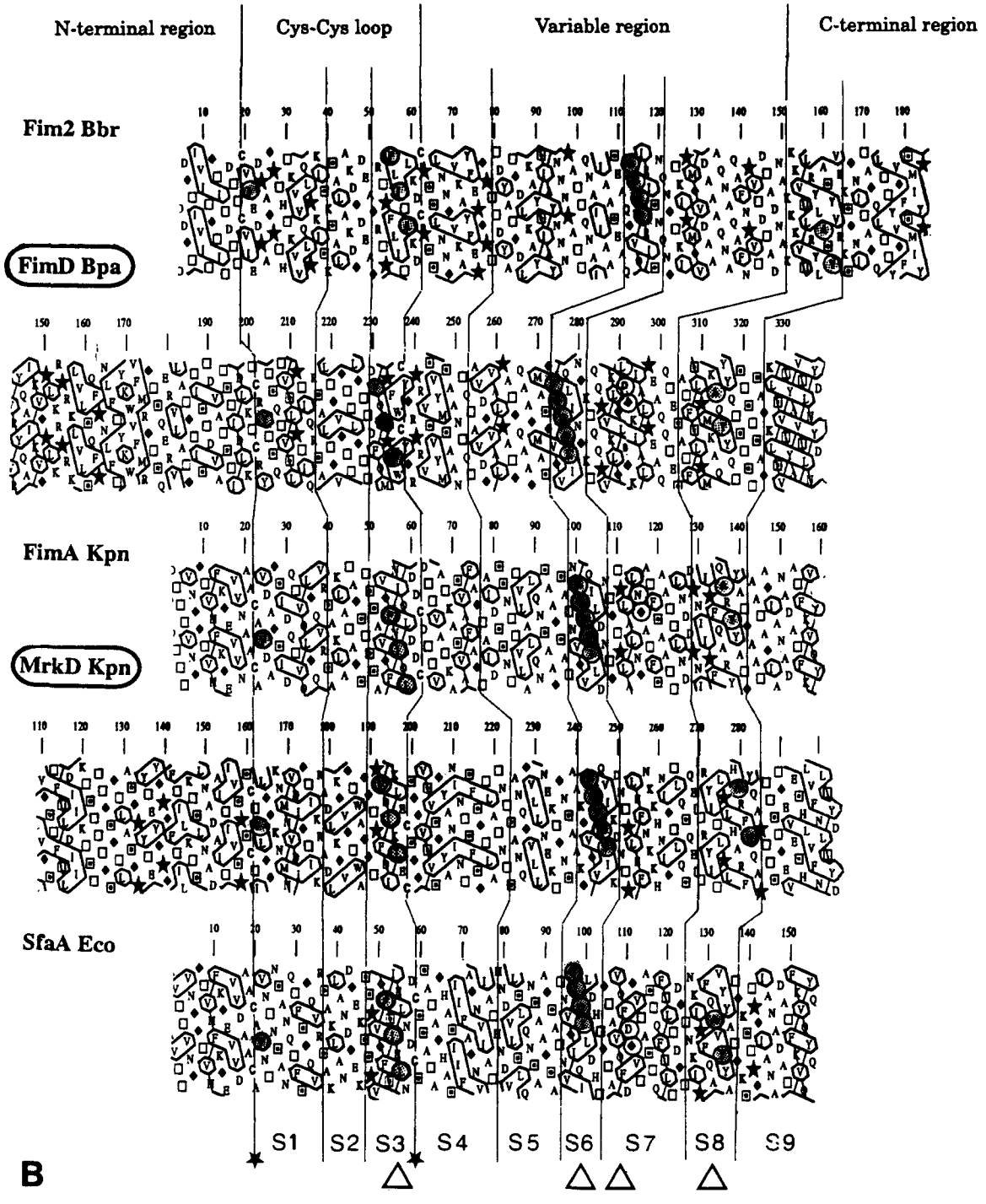


Fig. 1. Alignment of HCA plots of pilins with HCA plots of the C-terminal half of adhesin proteins. The different clusters are delineated by vertical lines which, when joined, show the correspondence between the major segments (S1 to S9). Particular conserved features including the four consensus (Fig. 2) are indicated (①) in segment S3, S6, S7 and S8. Groups of identical amino acids are shaded inside open circles. The four regions identified on pilin sequences are according to Van Die et al. [6]. The one letter amino acid code is used except for proline (②), glycine (③), threonine (④) and serine (⑤). Alignment of HCA plots of pilin and adhesin, proteins originating from the same parental fimbriae (A) are similar to alignment of HCA plots obtained with pilins and adhesins originating from different parental fimbriae (B). Adhesins are boxed.

the structural similarity of these proteins. A first assessment of the significance of this structural relatedness was carried out by calculating the pairwise HCA scores (Table 2). Despite weak sequence identity (~20%) and differences in size, the high HCA score (65–73%) indicated that, overall, pilins and the C-terminal domain of adhesins are structurally similar. HCA score values of 65% are found among proteins having the same overall fold but with significant structural divergence. No structurally-related sequences have been found to display an overall HCA score less than 60% [19].

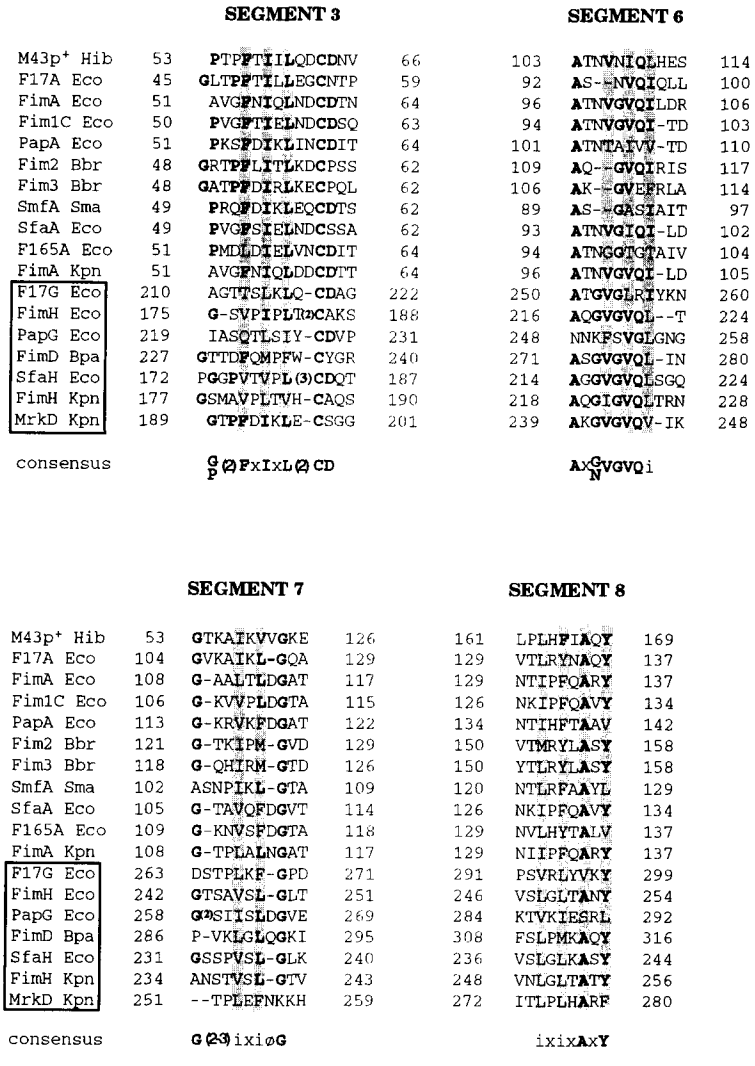


Fig. 2. Sequence of the four consensus included in the hydrophobic clusters used to anchor the alignment based on HCA method. Conserved residues are shown in bold and hydrophobes are shaded. Consensus are indicated below the sequences (i = hydrophobic, ø = polar or charged, x = any residues). Adhesins are boxed.

A second alignment was performed by comparing all the HCA plots which allowed the identification of groups of identical (or conserved) amino acids in segments S3, S6, S7 and S8 (Fig. 2). These conserved motifs which are major anchors for the alignment of sequences, are included in conserved hydrophobic clusters displaying the typical 'zig-zag' shape strongly associated with amphipatic β-strands [13,14]. Positioned close to the second cysteine of pilins (segment S3) the motif **FxIxLxxC** is very frequent, constituting a clearly recognized marker for the Cys–Cys loop within the pilins family, and to a lesser extent for adhesins. However, the amphipathicity of the motif was always conserved whatever the sequence origin (Fig. 2).

The motif Ax(G/N)VGVQi (i = hydrophobic residue, x = any residue) mainly observed in segment S6 of pilin and adhesin proteins constitutes a second marker and provided good evidence to support the structural relatedness of these domains.

Segments S2, S4 and S5 which were much more divergent (Fig. 1) corresponded in position to the hypervariable regions previously described as containing the dominant immunogenic epitope for the Pap pilins [6,20]. In addition to the sequence identity and the similarity of all HCA plots, the HCA scores of all members of the pilin family range from 75 to 95% (Table 2), values which are found among proteins sharing a similar 3D fold [13,19].

Introducing gaps and arranging sequences according to coinciding hydrophobic clusters showed that secondary structure predictions of pilins and the C-terminal half of adhesins according to the Rost and Sander method [15] are markedly alike (Fig. 3). Pilin and adhesin sequences could be aligned in a similar secondary structure prediction pattern. Conserved motifs identified in segments S3, S6, S7 and S8 of the HCA plots were highly predicted with β strand structures. Regions predicted as coil and turn connecting the predicted β strands were also remarkably alike in position and agreed well with the hypervariable regions (HV1 to HV5) previously described for Pap pilins [21].

PapA and PapG exibited the least homology among the pilin and adhesin proteins compared here (Table 2). However, the

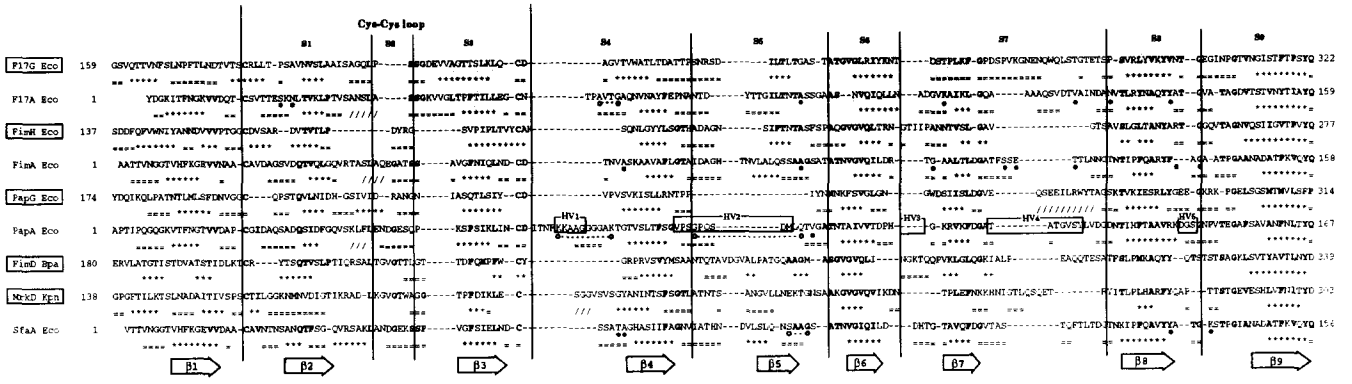


Fig. 3. Alignment of pilins and C-terminal half of adhesins, based on HCA alignment reported on Figure 1. Gaps (−) have been introduced to arrange sequences according to coinciding hydrophobic clusters. Vertical lines delineate segments (S1–S9) as on HCA plots. Only secondary structure predictions with high prediction potential are reported (Score >6 on scale ranging from 0 to 9; == loop, *β strand, ///α-helix). Identical residues between the different proteins are shown in bold and the four consensus observed in sequences (Fig. 2) included in hydrophobic clusters used as anchors for alignments are shaded. The previously identified [20] hypervariable regions (HV1 to HV5) of Pap pilins are boxed. Positions of insertions (●) or deletions (○---○) which occurred in the multiple sequence alignment produced from the 10 pilin sequences by the MaxHom program used in the PHD method are indicated on F17A, FimA, PapA and SfaA sequences. Open arrows below the sequences indicate the position of β strands (β1 to β9) predicted on the basis of the PHD secondary structure prediction method [15]. Adhesins are boxed.

majority of divergences occurred principally in segments S4 and S5 spanning the hypervariable region (R65–R79) previously described [20] as that exhibiting the greatest degree of heterogeneity among the Pappilin sequences.

4. Discussion

Alignment of pilin proteins (major subunits) with the C-terminal half of adhesin proteins reported in Fig. 1 and 3 led to the following remarks: (i) the general distribution of hydrophobic clusters and loops is as similar between pilin and adhesin proteins as is the distribution of hydrophobic clusters and loops within members of each protein family; (ii) the positions of conserved cysteine residues in pilins (N-terminal Cys–Cys pair) and the C-terminal half of adhesins (C-terminal Cys–Cys pair) are consistent with a disulphide bridge involved in maintenance of protein structure; (iii) motifs **FxIxLxxC** (Segment S3) and **Ax(G/N)VGVQi** (Segment S6) constitute clearly recognized markers of the pilin family and to a lesser extent of the C-terminal domain of adhesins; (iv) these groups of conserved residues may be critical residues in the function or maintenance of the structural integrity of the protein; (v) most of these conserved motifs (or residues) were found concentrated in β strands predicted with the PHD profile network method, while the variable residues were usually situated in regions predicted as loops and turns connecting β strands.

Despite the interspecies variations and the considerable amino acid sequence and antigenic variations in pilin proteins (major subunit), the similarities of their HCA plots, the high HCA score, the conserved cysteine pair at equivalent positions and the presence of groups of identical amino acids located in the predicted β strands are highly consistent with similarity in the overall three dimensional folding of pilins of different member species of *Enterobacteriaceae*. The conserved motifs and significant sequence agreement suggest that the genes encoding for pilins and adhesins have evolved from a common ancestral

Table 1
Origins and characteristics of pilin and adhesin sequences used in this study

Abbreviation	Organism	Subunit	Length (residues)	Accession number	
				NBRF	GenBank
F17A Eco	*E. coli*	Pilin	159	A27625	
F17G Eco	*E. coli*	Adhesin	322	A42359	
FimA Eco	*E. coli* (type I)	Pilin	158	A23180	
FimH Eco	*E. coli* (type I)	Adhesin	277	S09563	
Fim1C Eco	*E. coli* (type 1C)	Pilin	156	A22795	
PapA Eco	*E. coli* (F7-2)	Pilin	167	A03496	
PapG Eco	*E. coli* (F13)	Adhesin	314	G27743	
Fim2 Bbr	*B. bronchiseptica* (type 2)	Pilin	202	S36449	
Fim3 Bbr	*B. bronchiseptica* (type 3)	Pilin	182	S36450	
FimD Bpa	*B. parapertussis*	Adhesin	339		X75812
SmfA Sma	*S. marcescens*	Pilin	153	A31096	
SfaA Eco	*E. coli*	Pilin	156	S00352	
SfaH Eco	*E. coli*	Adhesin	268		X16664
F165G Eco	*E. coli*	Adhesin	313		L07092
FimA Kpn	*K. pneumoniae* (type I)	Pilin	159	JT0284	
FimH Kpn	*K. pneumoniae* (type 1)	Adhesin	279	A32801	
MrkD Kpn	*K. pneumoniae* (type 3)	Adhesin	303	B32801	
M43 Hib	*H. influenzae*	Pilin	191		M64334

Table 2
Comparison of scores between pilins and the C-terminal half of adhesins

Sequences	*FimA Kpn*	*F17A Eco*	*PapA Eco*	*SfaA Eco*	**F17G Eco** (163–322)*	**FimH Eco** (119–277)	**PapG Eco** (147–314)	**FimD Bpa** (169–339)	**MrkD Kpn** (133–303)
FimA Eco	82	28	33	60	**22**	**20**	**12**	**19**	**21**
	(95)	(77)	(70)	(88)	(69)	(65)	(60)	(64)	(70)
FimA Kpn		31	29	57	**22**	**20**	**12**	**23**	**24**
		(78)	(69)	(85)	(68)	(65)	(61)	(65)	(73)
F17A Eco			26	28	**22**	**24**	**12**	**24**	**26**
			(64)	(76)	(70)	(70)	(54)	(70)	(66)
PapA Eco				28	**16**	**15**	**15**	**12**	**13**
				(65)	(67)	(63)	(63)	(64)	(63)
SfaA Eco					**24**	**17**	**16**	**20**	**20**
					(70)	(65)	(56)	(64)	(60)
F17G Eco (163–322)						20 (70)	16 (62)	19 (71)	21 (70)
FimH Eco (119–277)							15 (62)	20 (69)	21 (69)
PapG Eco (147–314)								13 (62)	15 (56)
FimD Bpa (169–339)									26 (68)

For each entry, the HCA score is given in parentheses, and the sequence identity level (%) is given above from PileUp algorithm [16]. The HCA score was manually calculated as described by Lemesle-Varloot et al. [13]. Random sequences would score around 37 ± 6% [18]. The compared C-terminal regions of adhesins (in bold) are shown in parentheses (*).

determinant. However, given of the lack of DNA sequence agreement between genes encoding for these fimbrial polypeptides (not shown) the possibility of convergent evolution among these proteins cannot be ruled out. Recently, in the Pap pilus model, Slonim et al. [22] developed a hypothesis of pilus subunit proteins folding into domains that can serve as modules for pilus assembly. Our results argue that the C-terminal half of adhesins, which fold as pilin subunits, can serve as an assembly module required for adhesin presentation at the fibrillum tip.

References

[1] Hacker, J. (1992) Can. J. Microbiol. 38, 720–727.
[2] Hultgren, S.J., Abraham, S., Caparon, M., Falk, P., St. Geme III, J.W. and Normark, S. (1993) Cell 73, 887–901.
[3] Hultgren, S.J. and Normark, S. (1991) Annu. Rev. Microbiol. 45, 383–415.
[4] Hansen, M.S. and Brinton, C.C. (1988) Nature 332, 265–268.
[5] Jacob-Dubuisson, F., Heuser, J., Dobson, K., Normark, S. and Hultgren, S.J. (1993) EMBO J. 12, 837–847.
[6] Van Die, I., Hoekstra, W. and Bergmans, H. (1987) Microbial Pathol. 3, 149–154.
[7] Kuehn, M.J., Ogg, D.J., Kihlberg, J., Slonim, L.N., Flemmer, K., Bergfors, T. and Hultgren, S.J. (1993) Science 262, 1234–1241.
[8] Lund, B., Lindberg, F. and Normark, S. (1988) J. Bacteriol. 170, 1887–1894.
[9] Gerlach, G.F., Clegg, S. and Allen, B.L. (1989) J. Bacteriol. 171, 1262–1270.
[10] Schmoll, T., Hoschützky, H., Morschläuser, J., Lottspeich, F., Jann, K. and Hacker, J. (1989) Mol. Microbiol. 3, 1735–1744.
[11] Lintermans, P., Bertels, A., Schlicker, C., Deboeck, F., Charlier, G., Pohl, P., Norgren, M., Normark, S., Van Montagu, M. and De Greve, H. (1991) J. Bacteriol. 173, 3366–3373.
[12] Willems, R.J.L., Geuijen, C., Van Der Heide, H.G.J., Matheson, M., Robinson, A., Versluis, L.F., Ebberink, R., Theelen, J. and Mooï, F.R. (1993) Mol. Microbiol. 9, 623–634.
[13] Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Mornon, J.P. (1990) Biochimie 72, 555–574.
[14] Woodcock, S., Mornon, J.P. and Henrissat, B. (1992) Protein Eng. 5, 629–635.
[15] Rost, B. and Sander, C. (1993) Proc. Natl. Acad. Sci. USA 90, 7558–7562.
[16] Higgins, D.G. and Sharp, P.M. (1989) CABIOS 5, 151–153.
[17] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nuceic Acids Res. 12, 387–395.
[18] Schneider-Poetsch, H.A.W., Sensen, C. and Hanelt, S. (1991) NaturForsch. 46c, 750–758.
[19] Törrönen, A., Kubicek, C.P. and Henrissat, B. (1993) FEBS Lett. 321, 135–139.
[20] Denich, K., Blyn, L.B., Craiu, A., Braaten, B.A., Hardy, J., Low, D.A. and O'Hanley, P.D. (1991) Infect. Immunol. 59, 3849–3858.
[21] Van Die, I., Riegman, N., Gaykema, O., van Megen, I., Hoekstra, W., Bergmans, H., de Ree, H. and van den Bosch, H. (1988) FEMS Microbiol. Lett. 49, 95–100.
[22] Slonim, L.N., Pinker, J.S., Bränden, C.I. and Hultgren, S.J. (1992) EMBO J. 11, 4747–4756.