Microbial Pathogenesis 1990; **9**: 331–343

Complete genetic organization and functional aspects of the *Escherichia coli* S fimbrial adhesin determinant: nucleotide sequence of the genes *sfa* B, C, D, E, F

Thomas Schmoll,[1]* Joachim Morschhäuser,[1] Manfred Ott,[1] Birgit Ludwig,[1] Irma van Die[2] and Jörg Hacker[1]†

[1]*Institut für Genetik und Mikrobiologie, University of Würzburg, Röntgenring 11, D-8700 Würzburg, Germany and* [2]*Department of Molecular and Cellular Biology, University of Utrecht, Palualaan 8, CH 3584 Utrecht, The Netherlands*

(Received June 5, 1990; accepted in revised form August 14, 1990)

Schmoll, T. (Institut für Genetik und Mikrobiologie, University of Würzburg, Röntgenring 11, D-8700, Würzburg, Germany), J. Morschhäuser, M. Ott, B. Ludwig, I. van Die and J. Hacker. Complete genetic organization, and functional aspects of the *Escherichia coli* S fimbrial adhesin determinant: nucleotide sequence of the genes *sfa* B, C, D, E, F. *Microbial Pathogenesis* 1990; **9**: 331–343.

The S fimbrial adhesin (*sfa*) determinant of *E. coli* comprises nine genes situated on a stretch of 7.9 kilobases (kb) DNA. Here the nucleotide sequence of the genes *sfa*B and *sfa*C situated proximal to the main structural gene *sfa*A is described. *Sfa–Lac*Z fusions show that the two genes are transcribed in opposite directions. The isolation of mutants in the proximal region of the *sfa* gene cluster, the construction of *sfa–pho*A gene fusions and subsequent *trans*-complementation studies indicated that the genes *sfa*B and *sfa*C play a role in regulation of the *sfa* determinant. In addition the nucleotide sequence of the genes *sfa*D, *sfa*E and *sfa*F situated between the genes *sfa*A and *sfa*G responsible for S subunit proteins, were determined. It is suggested that these genes are involved in transport and assembly of fimbrial subunits. Thus the entire genetic organization of the *sfa* determinant is presented and compared with the gene clusters coding for P fimbriae (*pap*), F1C fimbriae (*foc*) and type I fimbriae (*fim*). The evolutionary relationship of fimbrial adhesin determinants is discussed.

Key words: *Escherichia coli*; S fimbrial adhesin (Sfa); genetic organization; gene regulation; nucleotide sequence.

Introduction

Adhesion of *Escherichia coli* to eukaryotic cells is a prerequisite for colonization of host tissues, leading to extraintestinal infections (urinary tract infection, sepsis, meningitis).[1,2] The adhesion is often mediated by bacterial cell appendages, called fimbrial adhesins. Receptors on the host cells are recognized by these organelles, which are proteinaceous heteropolymers, consisting mainly of a structural protein (major subunit). Minor subunits incorporated into the fimbrial rod confer the adhesive properties.[3,4] The various fimbrial adhesin types can be distinguished by their receptor

* Present address: Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, U.K.

† Author to whom correspondence should be addressed.

0882-4010/90/110331+13 $03.00/0

specifities.[4,5] In extraintestinal *E. coli*, type 1, P, S, and F1C fimbrial adhesins are the most commonly found.[2,4,5] The corresponding gene clusters have been cloned from the chromosomes of various *E. coli* strains and analysed in the last years[3,4,6-9] (for review see Hacker[4]).

S fimbrial adhesins (Sfa) mediate adhesion to sialic acid-containing receptors.[10] The *sfa* determinant has been cloned from a urinary tract isolate and analysed in detail.[6,11-14] It was shown that the S fimbrial adhesin complex consists of four proteins; SfaA representing the major fimbrillin protein and SfaG, SfaS and SfaH, the minor subunits. The corresponding genes were analysed and sequenced,[11,13] and SfaS was determined as the sialic-binding S-specific adhesin.[10,14] It was further shown that the expression of S fimbriae highly depends on environmental conditions.[15] In this report we present the nucleotide sequences of another five *sfa*-specific genes and show that the genes *sfa*B and *sfa*C situated proximally to the major subunit gene *sfa*A play a role in the regulation of the *sfa* determinant. Furthermore the entire genetic organization of the *sfa* gene cluster is presented and aspects of the evolution of fimbrial adhesins are discussed.

Results

Nucleotide sequence of sfa*B and* sfa*C*

In previous reports[11,13,14] we have determined the DNA sequences of the genes *sfa*A, *sfa*G, *sfa*S and *sfa*H coding for the major and minor subunits of the S fimbrial adhesin complex. The genes are situated between map position 2.3 and 2.8 kb (*sfa*A), and 6.6 and 9.0 kb (*sfa*G, S, H) in Fig. 1(a). In order to get a complete picture of the *sfa* gene cluster we have sequenced the DNA regions between 0.6 and 2.1 kb as well as between 2.8 and 6.6 kb. The sequencing strategy is given in Fig. 1(b).

As shown in Fig. 2 the sequence proximal to *sfa*A covers a stretch of 1.4 kb. This sequence contains two open reading frames (ORFs). One ORF starts at an ATG codon at position 663–665 and ends at a TAA stop codon at position 990–992. The second ORF runs in the opposite direction, and the coding sequence is therefore on the

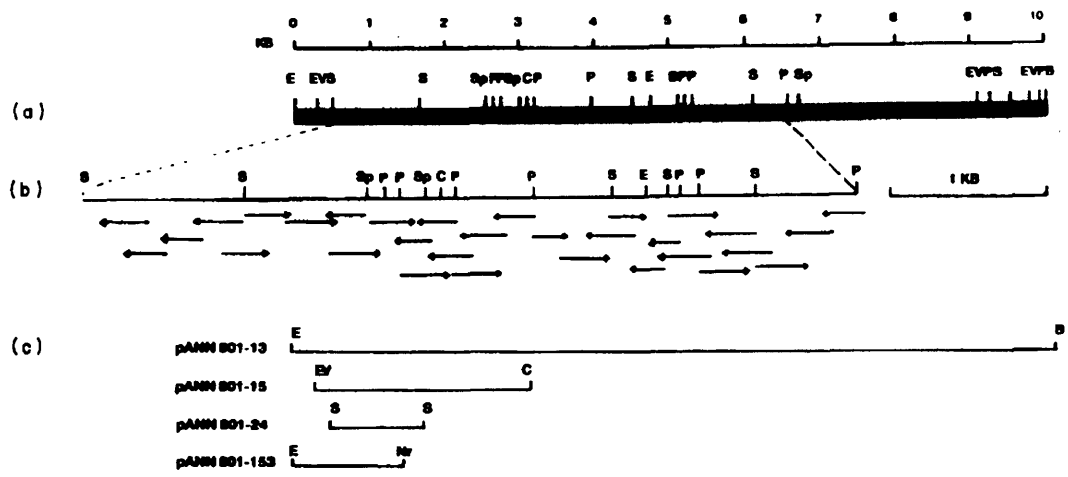

Fig. 1. (a) Restriction map of the *sfa* specific DNA region (see Hacker *et al.*[6]). (b) Strategy of sequencing of the proximal and internal part of the *sfa* region, comprising genes *sfa*C, B and *sfa*D, E, F, respectively (see Fig. 2 and Fig. 5). Arrows indicate the regions that have been sequenced. (c) Subclones of the *sfa*-region used. Restriction enzymes are abbreviated as follows: E, *Eco*RI; EV, *Eco*RV; S, *Sma*I; Sp, *Sph*I; P, *Pst*I; C, *Cla*I; B, *Bam*HI; Nr, *Nru*I.

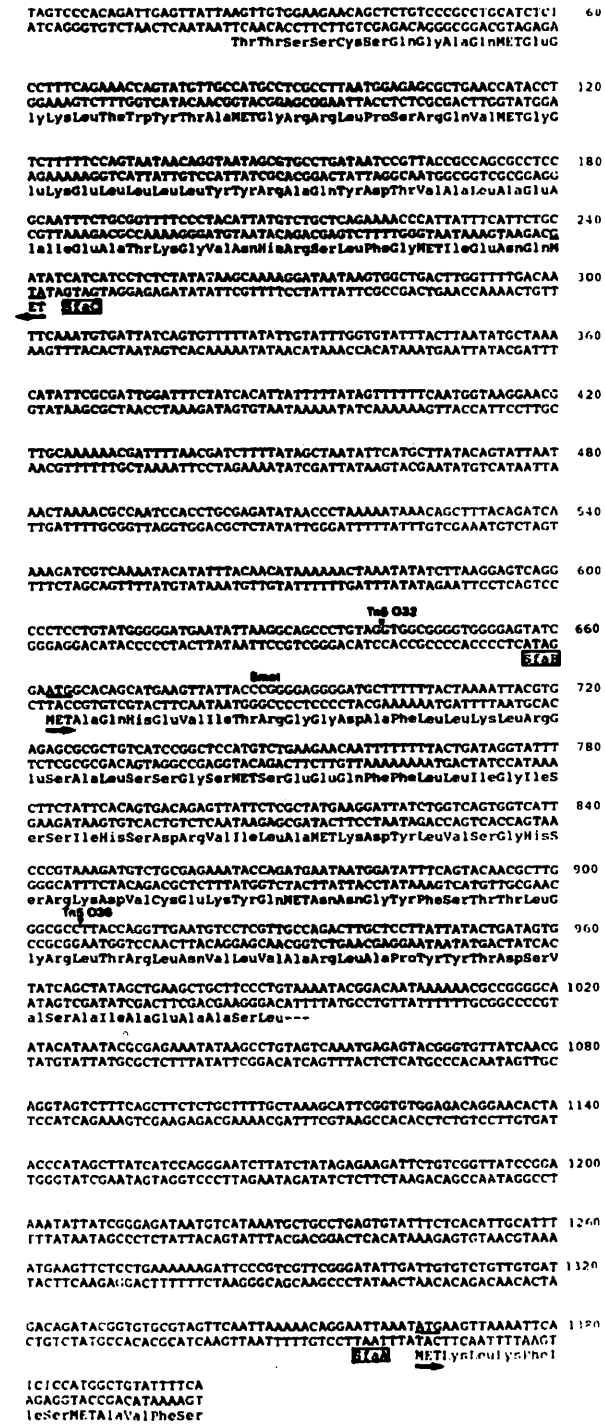

Fig. 2. Nucleotide sequence of the proximal part of the *sfa* determinant, comprising genes *sfa*B and *sfa*C. Amino acid residues are given below. The positions of integration of Tn*5* transposons are indicated by triangles.

antiparallel strand. Starting at an initiation codon at map position 238–240 the ORF stops at map position 23–21. The two ORFs which may correspond to two genes were preliminary termed *sfa*B and *sfa*C. The putative molecular masses of the two gene products SfaB and SfaC calculated from the nucleotide sequence were 12.0 kDa (SfaB) and 8.3 kDa (SfaC).

Localization of Tn5 mutants

Transposon mutants pANN 801-13/Tn*5*-014, pANN 801-13/Tn*5*-032 and pANN 801-13/Tn*5*-036 which carry Tn*5* elements inserted into the proximal *sfa* region located on the recombinant plasmid pANN 801-13 were isolated previously.[6] These mutants were negative for S fimbrial formation (Fim$^-$) as well as for S specific mannose-resistant hemagglutination (Mrh$^-$). It was shown by DNA sequence studies that the Tn*5*-mutant pANN 801-13/Tn*5*-041 carries the transposon inserted into the gene *sfa*A, 336 bp downstream of the first nucleotide of the gene (see Schmoll *et al.*[13]). In the mutants pANN 801-13/Tn*5*-036 and pANN 801-13/Tn*5*-032 Tn*5* transposons are inserted into the gene *sfa*B at position 906 and 23 bp upstream of the first codon of *sfa*B at position 640, respectively (see Fig. 2). The location and the phenotypic characterization of these mutants suggest that the genes situated in the proximal region of the *sfa* determinant play a role in regulation.

Isolation of Sfa–Lac*Z protein fusions*

In order to analyse the gene products SfaB and SfaC and to confirm the direction of transcription of the proximal part of the *sfa* determinant, Sfa–LacZ translational fusions were isolated. λp*lac*Mu3 phages were randomly inserted into *E. coli* K-12 clones carrying the recombinant plasmids pANN 801-15 and pANN 801-24 [Fig. 1(c)]. The map positions of the insertions and the orientations of the inserted *lac*Z genes were determined by suitable cleavages with restriction enzymes. The β-galactosidase activities of the strains harbouring the fused DNA were analysed. As indicated in Fig. 3, four Lac-positive SfaC–LacZ fusions were isolated. The orientation of insertion of the *lac*Z genes was from right to left while all insertions in the left to right orientation were LacZ-negative. Other LacZ-positive transductants were isolated following

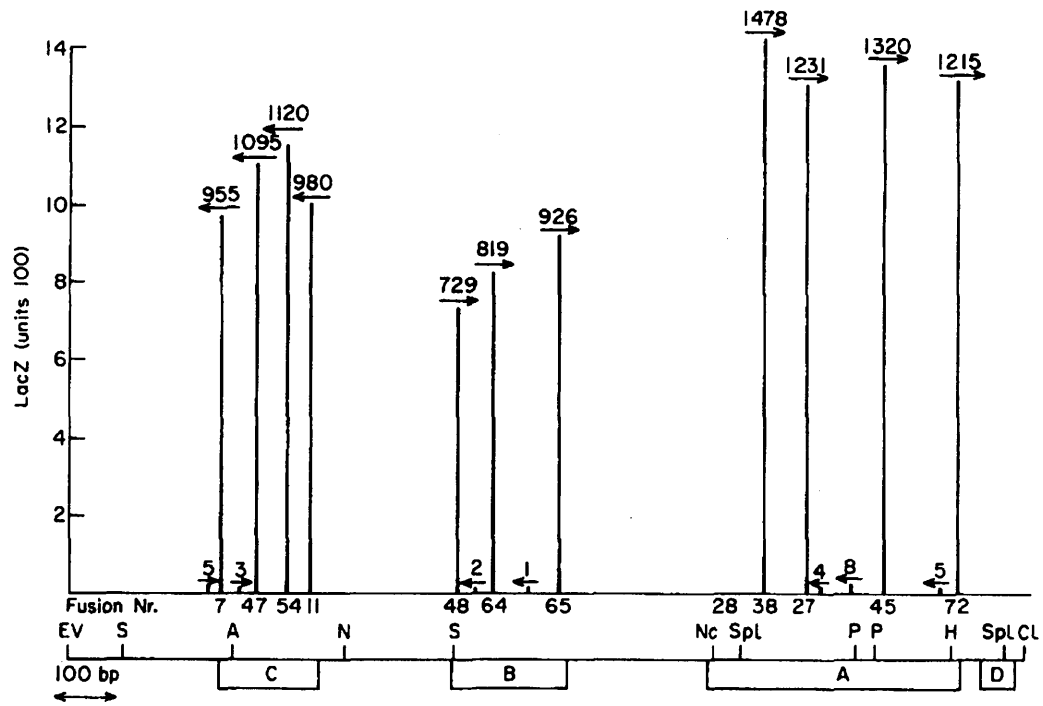

Fig. 3. Analysis of Sfa–LacZ translational fusions of genes *sfa*C, B, and A following insertion of λp*lac* Mu3 phages. The map position of the *sfa* specific genes is given in the lower part. Orientations of the fusions are indicated by an arrow. The amount of β-galactosidase activity is shown graphically, with LacZ values given on top and the fusion numbers indicated below. MC 4100 was used as the host strain. Restriction enzymes are abbreviated as follows: EV, *Evo*RV; S, *Sma*I; A, *Ava*I; N, *Nru*I; Nc, *Nco*I; Spl, *Sph*I; P, *Pst*I; H, *Hin*cII; Cl, *Cla*I.

insertion of λp*lac*Mu3 phages into the coding regions of *sfa*B and *sfa*A. In all cases the direction of the *lac*Z reading frame was from left to right, i.e. the opposite direction to that detected for the SfaC–LacZ fusions. Right to left insertions were LacZ-negative. These results clearly show that the *sfa*C-specific and *sfa*B-specific ORFs are translated *in vivo*. The data further demonstrate that the transcription of *sfa*C runs in the opposite direction to *sfa*B and *sfa*A, which is in good agreement with the sequence data (see Fig. 2).

Cloning of sfa *specific sequences into the promoter probe vector pCB267*

To get information on the presence of promoter regions situated in front of the genes *sfa*C, *sfa*B and *sfa*A, suitable DNA fragments were cloned into the promoter probe vector pCB267.[16] The strength of putative promoter regions was measured on the basis of the production of alkaline phosphatase (PhoA units).[17] As indicated in Fig. 4 the construct pTTS 267-69 (*sfa*A–*pho*A, B^-, C^-) produced 80 units PhoA (Fig. 4, row 5) indicating the existence of a weak but clearly detectable promoter in front of the *sfa*A gene. The activity of the promoters in front of *sfa*B and *sfa*C was determined by the use of the recombinant plasmids pTTS267-1/5 (*sfa*B–*pho*A, A^-, C^-) and pTTS267-1/8 (*sfa*C–*pho*A, A^-, B^-) which were able to produce PhoA values of 488 and 429 units (Fig. 4, rows 6, 8). In all cases the strains used as controls were negative for PhoA production. The putative promoter regions in front of *sfa*C, *sfa*B and *sfa*A were preliminarily termed P_A, P_B and P_C. A computer search for promoter consensus sequences revealed good −10 and −35 boxes in front of the genes *sfa*C and *sfa*B at map positions 272 to 300 bp and 561 to 589 bp, respectively (Fig. 2).

Role of SfaB and SfaC in expression of the sfa *determinant*

To define the role of SfaB and SfaC in the expression of the *sfa* gene cluster we used *sfa–pho*A constructs in *trans*-complementation tests with $SfaA^+$, B^+, C^+ (pANN801-15) and $SfaC^+$ (pANN801-153) plasmids. The different DNA constructs were able to co-replicate in the same cell because the *sfa–pho*A hybrid plasmids carry pBR322 sequences whereas pANN801-15 and pANN801-153 used for complementation consist of the vector molecule pACYC184 together with *sfa*-specific insert DNA.

As shown in Fig. 4 (row 5) the gene products SfaB and SfaC have no effect on the promoter P_A. In contrast SfaB and SfaC positively influence the activity of the promoters P_B and P_C. Using a $SfaC^-$, $SfaB^+$, $SfaA^-$ PhoA specific construct (pTTS267-282, Fig. 4, row 4) the SfaC coding DNA increases the PhoA production after *trans*-complementation by a factor of two. SfaC, however, has no effect on its own promoter P_C (data not shown). As also indicated in Fig. 4, SfaB has a more pronounced effect on P_B (row 3) and P_C (row 8). Since SfaB and SfaC do not inflence P_A (Fig. 4, row 5, see above) the increase of expression of *sfa*A by the action of SfaB and SfaC must result from a transcript which starts at promoter P_B (Morschhäuser and Hacker, unpublished data). The major subunit SfaA has no influence on the regulation of the *sfa* determinant (Fig. 4, Schmoll and Hacker, unpublished data). Thus both gene products, SfaB and SfaC, are able to stimulate the expression of the *sfa* determinant in *trans*.

Nucleotide sequence of sfa*D*, sfa*E and* sfa*F*

To complete the picture of the genetic organization of the *sfa* determinant the DNA region between the 3'-end of *sfa*A at position 2.8 kb and the 5'-end of *sfa*G at position 6.6 kb (Fig. 1) was sequenced. As indicated in Fig. 5, three open reading frames were found. The first ORF (left to right direction) starts at position 118–120 with an ATG codon and stops at position 655–657 with TAA. Downstream, in near vicinity, a

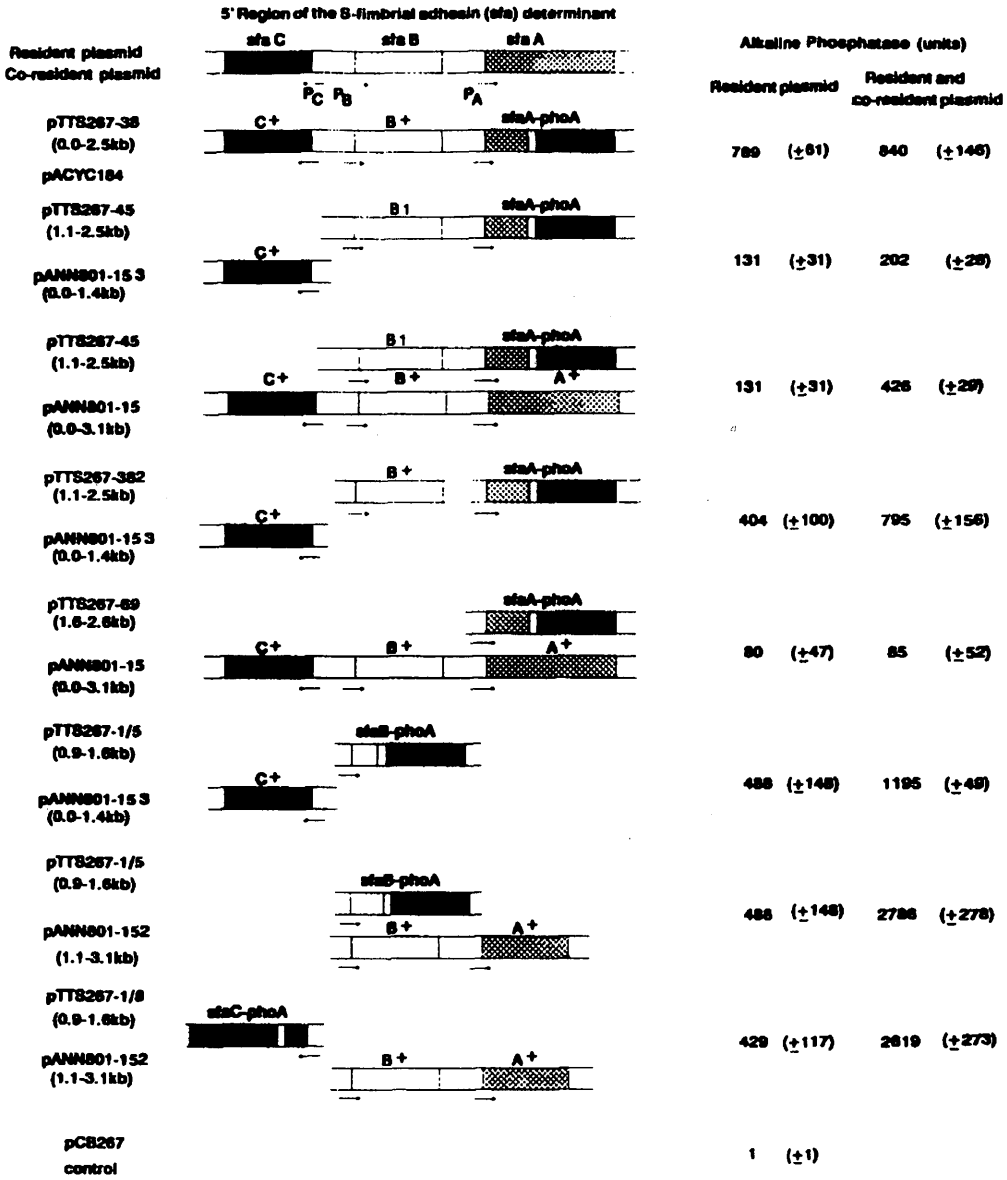

Fig. 4. Transcriptional fusions of *sfa* specific sequences and the *pho*A gene. The coordinates of the *sfa* specific sequences are indicated in Fig. 1. For *trans*-complementations the plasmids pANN801-15, pANN801-153 and pACYC184 (as a control) were used. CC118 was used as the host strain and CC118 (pCB267) as a *pho*A-negative control. PhoA units are given as mean ± standard deviation.

second ORF begins at position 698–701 with ATG, and ends at position 1391–1393 with a TAA stop codon. The third ORF starts at position 1464–1466 with ATG and stops at position 4076–4078, ending with TAA.

These three ORFs are estimated to represent the *sfa*-specific proteins SfaD, E, and F. By calculation of molecular masses of these ORFs from sequence data, proteins of 17.4, 23.5 and 90.4 kDa should be encoded by this stretch of DNA. Minicell analysis of specific subclones comprising this particular *sfa*-region exhibited proteins, corresponding in size to the molecular weights deduced from nucleotide sequence (data not shown). Although there is no information on processing of precursors into mature proteins for these ORFs, we speculate that all three proteins contain leader sequences (see Fig. 5).

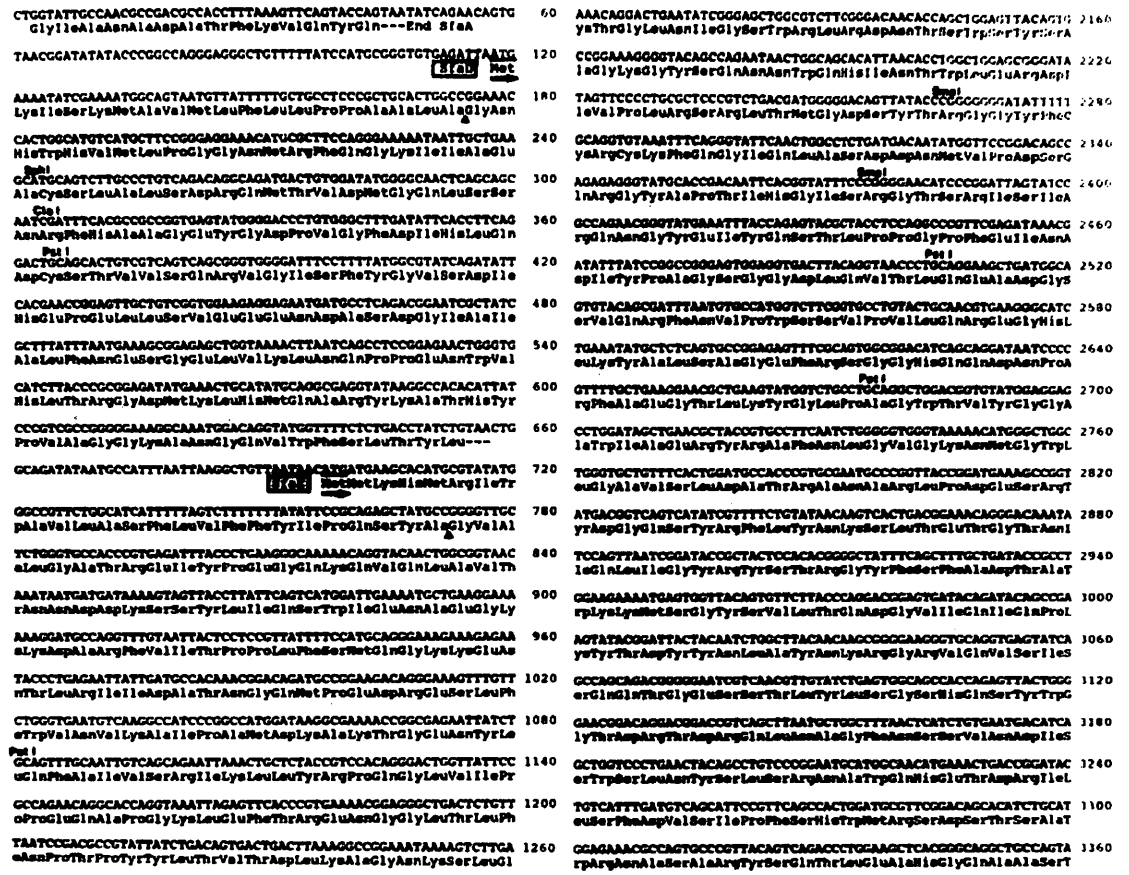

Fig. 5. Nucleotide sequence of genes *sfa*D, *sfa*E, and *sfa*F. The amino acid residues are given below. The putative cleavage sites of the leader sequences are marked by a triangle. Important restriction sites are indicated.

Discussion and conclusions

In this report, we have presented data on the genetic organization of the entire *E. coli* S fimbrial adhesin determinant (*sfa*) and demonstrated that nine *sfa*-specific genes are situated on a 7.9 kb DNA region. In Fig. 6 in *sfa* gene cluster is compared with the

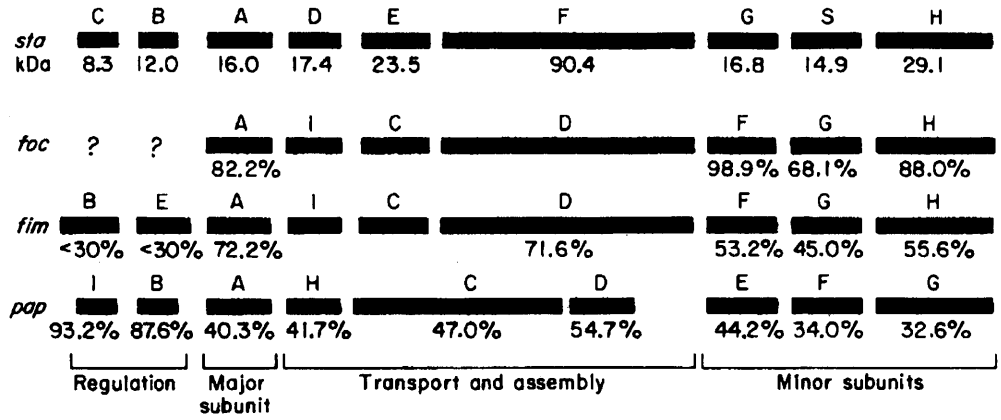

Fig. 6. Comparison of gene clusters encoding S fimbrial adhesin (*sfa*), F1C fimbriae (*foc*), type 1 fimbriae (*fim*), and P fimbriae, serotype F13 (*pap*). Cistrons, designated by capital letters, are shown. The indicated percentages refer to the degree of homology of the proteins to the corresponding Sfa gene products on basis of identical amino acids or amino acids of similar properties (see Schmoll *et al.*[11]). The PapD protein is compared to the SfaE product (see text). Molecular masses of mature Sfa specific proteins are given (kDa). Functional regions are indicated.

determinants coding for F1C fimbriae, P fimbriae and type I fimbriae termed *foc, pap* and *fim* respectively[3,7–9] (for review see Hacker[4]). It is obvious from the literature that the genetic composition of different adhesin determinants is similar. Different regions of the determinants that code for the major fimbrillin (pilin) subunits and for minor fimbrillin proteins, including the corresponding adhesin proteins, can be distinguished. In addition, sequences necessary for transport and assembly functions as well as regions involved in the regulation of transcription can be identified.

Sequence comparison between the *sfa* and *pap* gene clusters shows that there is strong homology between the regulatory genes (*papI*, B and *sfa*C, B) while, interestingly, the remaining portions do not exhibit any significant sequence similarities (Fig. 6). P fimbrial gene clusters coding for serotypes other than the *pap* determinant also exhibit strong homology to this particular *sfa*-region.[18,19] Besides sequence homology between P-fimbrial regulatory elements and the proximal *sfa*-region, a similar arrangement of the direction of transcription of the respective genes can be observed.[19–21] The relationship of the *pap* and *sfa* regulatory region is further corroborated by the fact that the genes *sfa*B, *sfa*C and *pap*B, *papI* are able to *trans*-complement each other.[22] A divergent promoter arrangement similar to the *pap* and *sfa* determinants has also been described for regulatory elements of the K99 fimbriae determinant (*fan*), although no sequence homology to *pap* or *sfa* genes exists.[23] For the F1C determinant (*foc*), it has been shown by genetic analysis and by DNA–DNA hybridization studies[24–26] that over the whole gene cluster a strong homology to the *sfa* determinant can be observed, including the proximal part of the gene clusters, suggesting a homologous regulatory region.

In addition we have described three promoter regions which are situated in the proximal part of the *sfa* determinant. As already mentioned above, two of these, P_B and P_C, show very good promoter consensus sequences[27] (see Fig. 2). One promoter, P_A, is situated in front of the *sfa*A structural gene. Such a promoter is missing in front of the structural P fimbrial gene *pap*A[19–21,28] but is present in front of the type I specific structural gene *fim*A.[29,30] It is interesting to note that the two regulatory genes *fim*B and *fim*E of the type I fimbrial determinant do not exhibit any homology to the regulatory genes of S fimbriae *sfa*B and *sfa*C,[30] but the fimbrial structural genes *sfa*A

and *fim*A are very similar. We speculate that the presence of promoters in front of *sfa*A and *fim*A may reflect this relatedness between type I and S subunit genes.

The relatedness of *sfa*- and *fim*-specific subunit genes is restricted to the cistrons encoding major fimbrillins (see Fig. 6). The genes coding for the S specific minor subunit proteins show a high degree of homology to the corresponding region of the *foc* determinant[11] (Van Die *et al.*, unpublished data) but they differ from those of the *pap* and *fim* gene clusters (for review see Hacker[4]). These data confirm our previous observations[24–26] that S and F1C fimbrial gene clusters represent a family of adhesin determinants, as also described for the P/Prs group[35,36] and the MS-agglutinating type I family.[5]

In addition to the regions coding for major and minor subunits and for regulatory proteins, at least three proteins are encoded by the central parts of P, S and type I determinants.[9,31–34] The Sfa specific stretch of DNA between the genes *sfa*A and *sfa*G also codes for three gene products (see Fig. 5): SfaD (17.4 kDa), which forms a subunit like structure, SfaE (23.5 kDa) and SfaF (90.4 kDa). Proteins similar in size to SfaF, encoded by the type I and P fimbrial determinants, represent outer membrane proteins which apparently anchor the corresponding fimbrial adhesin complexes to the cell wall.[9,31,32] SfaF, which exhibits 71% homology to the anchor protein of type I fimbriae, FimD, may perform a similar function. The high degree of homology between SfaF and FimD may be the reason for the hybridization observed following DNA–DNA hybridizations between S and type I fimbrial gene clusters.[24]

As shown recently, the proteins PapD and PapH of P fimbriae are also involved in biogenesis of the fimbrial adhesin structure. PapH, which forms a subunit like structure, is situated at the base of the fimbriae and terminates fimbrial biogenesis.[33] In contrast, PapD functions as a transporter responsible for the transfer of subunit proteins through the periplasmic space.[34] From the sequences given in Fig. 5, we suggest that the subunit-like protein SfaD may share common features with PapH while SfaE could be equivalent to PapD. The corresponding genes, however, are situated at different positions in the determinants. The *pap*D gene is situated downstream of the large gene *pap*C, while *sfa*E (and also *fim*C) is encoded by a DNA region upstream of the large gene responsible for the anchor protein (Fig. 6).

The highly conserved arrangement of cistrons of the various fimbrial adhesin determinants is an interesting feature, since there is only limited overall sequence homology, with the exception of the *sfa* and *foc* gene clusters. When the determinants are divided into functional regions (see Fig. 6), it is apparent that sequence homology is restricted to certain parts, which could be defined as 'cassettes'. The *sfa* and *foc* determinants seem to be closely related. However, only the regulatory cassettes of *sfa* and *pap*, and the major subunits and large anchor proteins of *fim* and *sfa*, seem to be closely related. These observations lead to speculations about the evolution of fimbrial adhesin gene clusters. Possibly diversity has been generated by combining cassettes of a few ancestral determinants. By this mechanism new fimbrial determinants are created which are selected according to their usefulness for the strains. This hypothesis would explain the wide spectrum of different adhesin determinants, described so far.

Materials and methods

Bacterial strains, plasmids and phages. The *sfa* determinant was originally cloned from *E. coli* strain 536, a urinary tract isolate (O6 : K15 : H31), described elsewhere.[6,37] Plasmid pANN 801-13 carries the entire *sfa* gene cluster (Fig. 1) as described.[6,11–15] Plasmids pANN 801-15, and pANN 801-24 have been described recently.[6,12] Plasmid pANN 801-153 consists of the vector pACYC184 and a 1.4 kb *Eco*I–*Nru*I fragment coding for SfaC (see Fig. 1). The bacterial strains,

Table 1 Bacterial strains, cloning vectors and phages used

Designation	Description	Reference
Escherichia coli K-12 strains		
HB101	F⁻, *hsd*S20 (r_B-, m_B-), *rec*A13, *ara*-14, *pro*A2, *lac*Y1, *gal*K2, *rps*L20 (Sm^r), *xyl*-5, *mtl*-1, *sup*E 44	38
JM103	Δ(*lac, pro*), *thi, rps*L20, *sup*E, *end*A, *sbc*B, *hsd*R, F′ (*tra*D36, *pro*A, B, *lac*Iq)	39
MC4100	F⁻, *ara*D139, Δ(*arg*F⁻, *lac*) U169 *rps*L150, *rel*A1, *deo*C1, *pts*F25, *rbs*R, *flb*B5301	40
MBM 7014	F⁻, *ara*C_{am}, *ara*D, Δ(*arg*F⁻ *lac*) U169 trp_{am}, *mal*B_{am}, *rps*L, *rel*A, *thi*	40
CC118	*ara*D 139, Δ(*ara, leu*) 7697, Δ*lac*X74 *pho*A 20, *gal*E, *gal*K, *thi, rps*E, *rpo*B, *arg*E_{am}, *rec*A1	16
Cloning vectors		
pCB267	*bla*, '*lac*Z, '*pho*A	16
pBR322	*bla, tet*	41
pACYC184	*cat, tet*	42
pUC8	*bla, lac*Z	43
M13-mp18/mp19	DNA sequencing	39
Phages		
λ*plac*Mu3	*att*⁻, *lac*Z*lac*Y, MuA⁺, *imm*21, Mucs	44,45
λpMu507-3	*imm*21, S_{am}7, MuA⁺E⁺, *sup*F	44,45

cloning vectors and phages used are listed in Table 1. The phages are described by Bremer *et al.*[44,45] All *E. coli* strains were K-12 derivatives.

Media and reagents. Bacterial strains were grown on Luria Bertani (LB) agar plates or in liquid LB medium, as described.[6] For maintenance of plasmids 50 μg ampicillin per ml, or 20 μg chloramphenicol per ml were added. For detection of the Lac phenotype, LB medium was supplemented with IPTG (isopropylthiogalactoside, 0.05 mM) and X-Gal (5 bromo-4-chloro-3-indoloyl-galactoside, 0.01%) or MacConkey agar was used. The production of alkaline phosphatase was detected on XP-plates containing 20 μg of 5-bromo-4-chloro-3-indolyl phosphate per ml. Restriction enzymes, T4 Ligase and Klenow enzyme were purchased from Boehringer (Germany). Antibiotics were a gift from Bayer (Germany). Reagents for growth media were obtained from Oxoid (Germany). All other chemicals were obtained from Sigma (Germany).

Recombinant DNA techniques. Plasmid DNA was isolated by the method of Birnboim and Doly.[46] Large-scale preparations were further purified over CsCl gradients, as described.[37] DNA was cleaved with restriction enzymes according to the manufacturer's instructions, and separated on 1% agarose gels, as described by Maniatis *et al.*[47] Isolation of DNA fragments from agarose gels was performed by the freeze-squeeze method according to Thuring *et al.*[48] Ligation of DNA fragments was carried out, as described.[47] Competent cells for DNA transformation were prepared by $CaCl_2$ method.[49]

Construction of recombinant plasmids. The recombinant plasmids indicated in Fig. 1 were constructed by ligating suitable *sfa*-specific fragments directly into pACYC184 (pANN801-15 and pANN801-153) or into the *Sma*I site of the pUC8 polylinker sequence (pANN801-24). Cloning vector pCB267[16] and strain CC118 were used for the construction of *sfa*–phoA fusions. As indicated in Fig. 4, suitable *sfa* specific DNA fragments ligated into the cloning vector pCB267 resulted in the *sfa–pho*A recombinant DNAs. The recombinant plasmid pTTS267-45 carries the linker sequence from pCB267 inserted into the *Sma*I-site at map position 1.65 (Fig. 1). Although the *sfa*B gene was out of frame, the putative promoter P_B, situated in front of *sfa*B, was still active in this recombinant DNA.

DNA sequencing and oligonucleotide synthesis. DNA sequencing was performed by the dideoxy chain termination method of Sanger *et al.*,[50] using α^{32}P-ATP (Amersham-Buchler,

Germany). Either single stranded recombinant M13 phage DNA or double stranded DNA of recombinant pUC plasmids were used.[43] Cloning of DNA fragments in M13 phage vector and isolation of single stranded DNA were carried out as described.[39,47] The primers used were synthetic oligodeoxyribonucleotides, prepared by an automated phosphoramidite coupling method.[51] Sequencing was carried out with a T7 sequencing kit (Pharmacia, Germany), including universal primer.[43] As a recipient strain for recombinant DNA, JM 103 was used.

Determination of the insertion points of Tn5 mutants. In order to determine the exact position of Tn*5* insertions into the pANN801-13, derivatives pANN801-13/Tn*5*-014, pANN801-13/Tn*5*-032, pANN801-13/Tn*5*-036 were sequenced. The DNA was first cleaved with the restriction enzyme *Hpa*I and suitable DNA fragments carrying 195 base pairs Tn*5*-specific sequence together with the *sfa*-specific region were ligated into the vectors pUC18 and pUC19. The DNA sequences of the insertion points were determined by the plasmid sequence technique with the help of the DNA sequencing kit from Pharmacia.

Computer analysis. General compilation and analysis of DNA sequences was performed with the UWGCG programs obtained from Devereux.[52] For comparison of proteins the software package PC Gene was used (Intelli Genetics, Switzerland).[53]

Preparation of phage stock solutions. As indicated by Bremer *et al.*[44,45] nearly 10^8 phage particles were mixed with 10^8 bacteria (λp*lac*Mu3 was mixed with strain MC 4100, λpMu507-3 with strain MBM 7014), spotted on soft agar plates and incubated for 12 h at 37°C. The plates were then incubated with 5 ml SM-buffer for 4 h at 4°C. The supernatant which contains the phage particles, was removed and stored at 4°C.

Transposition assays of λplac*Mu phages.* The λp*lac*Mu phage transposition was performed according to Bremer *et al.*[44,45] Briefly, the plasmid-bearing strain MC 4100 was infected with a mixed stock of λp*lac*Mu3 and λpMu507-3 and plated on MacConkey agar plates. After overnight incubation at 37°C the cells were scraped from the plates, washed, and then grown to exponential phase. A phage lysate was prepared by induction with UV-irradiation and used to transduce a K-12 Lac^- strain with selection of Lac^+ and the plasmid-borne resistance. The position of insertion was determined by suitable digestions of the recombinant DNA consisting of phage and plasmid specific sequences.

Enzyme assays. The β-galactosidase units were determined by the method of Miller.[54] A quantitative assay for alkaline phosphatase was carried out according to Michaelis *et al.*[17]

The authors wish to thank M. Wuenscher for critical reading of the manuscript and H. Kurz for editorial assistance. The work was supported by the Deutsch Forschungsgemeinschaft (DFG Ha 1434/1-7).

References

1. Svanborg-Eden C, Hagberg L, Hanson LA *et al.* Bacterial adherence—a pathogenic mechanism in urinary tract infections caused by *Escherichia coli.* Prog Allergy 1983; 33: 175–88.
2. Korhonen TK, Valtonen MV, Parkkinen J *et al.* Serotypes, hemolysin production, and receptor recognition of *Escherichia coli* strains associated with neonatal sepsis and meningitis. Infect Immun 1985; 48: 486–91.
3. Lindberg F, Lund B, Normark S. Gene products specifying adhesion of uropathogenic *Escherichia coli* are minor components of pili. Proc Natl Acad Sci USA 1986; 83: 1981–5.
4. Hacker J. Genetic determinants coding for fimbriae and adhesins of extraintestinal *Escherichia coli.* Curr Top Microbiol Immunol 1990; 151: 1–27.
5. Ørskov F, Ørskov I. Serology of *Escherichia coli* fimbriae. Prog Allergy 1983; 33: 80–105.
6. Hacker J, Schmidt G, Hughes C, Knapp S, Marget M, Goebel W. Cloning and characterization of genes involved in the production of mannose-resistant, neuraminidase-susceptible (X) fimbriae from a uropathogenic O6:K15:H31 *Escherichia coli* strain. Infect Immun 1985; 47: 434–40.
7. Van Die I, van Geffen B, Hoekstra W, Bergmans H. Type 1C fimbriae of a uropathogenic *Escherichia coli* strain. Cloning and characterization of genes involved in the expression of the 1C antigen and nucleotide sequence of the subunit gene. Gene 1984; 34: 187–96.
8. Eisenstein BI, Sweet DS, Vaughn V, Friedman DI. Integration host factor is required for the DNA

inversion that controls phase variation in *Escherichia coli*. Proc Natl Acad Sci USA 1987; 84: 6506–10.
9. Klemm P, Christiansen G. The *fimD* gene required for cell surface localization of *Escherichia coli* type 1 fimbriae. Mol Gen Genet 1990; 220: 334–8.
10. Moch T, Hoschützky H, Hacker J, Kröncke K, Jann K. Isolation and characterization of the alpha-sialyl-beta-2,3-galactosyl-specific adhesin from fimbriated *Escherichia coli*. Proc Natl Acad Sci USA 1987; 84: 3462–6.
11. Schmoll T, Hoschützky H, Morschhäuser J, Lottspeich F, Jann K, Hacker J. Analysis of genes coding for the sialic-acid-binding adhesin and two other minor fimbrial subunits of the S-fimbrial adhesin determinant of *Escherichia coli*. Mol Microbiol 1989; 3: 1735–44.
12. Hacker J, Schmoll T, Ott M *et al.* Genetic structure and expression of virulence determinants from strains of *Escherichia coli*. In: Kass E, Svanborg-Eden C, eds. Studies in infectious disease research. Host parasite interactions in urinary tract infection. Chicago: University of Chicago Press, 1989; 144–56.
13. Schmoll T, Hacker J, Goebel W. Nucleotide sequence of the *sfaA* gene coding for the S-fimbrial protein subunit of *Escherichia coli*. FEMS Microbiol Lett 1987; 41: 229–35.
14. Morschhäuser J, Hoschützky H, Jann K, Hacker J. Functional analysis of the sialic-acid binding adhesin SfaS of pathogenic *Escherichia coli* by site-specific mutagenesis. Infect Immun 1990; 58: 2133–8.
15. Schmoll T, Ott M, Oudega B, Hacker J. Use of a wild-type gene fusion to determine the influence of environmental conditions on the expression of the S fimbrial adhesins in an *Escherichia coli* pathogen. J Bacteriol 1990; 172: 5103–11.
16. Schneider K, Beck CF. Promoter-probe vectors for the analysis of divergently arranged promoters. Gene 1986; 42: 37–48.
17. Michaelis S, Inouye H, Oliver D, Bechwith J. Mutations that alter the signal sequence of alkaline phosphatase in *Escherichia coli*. J Bacteriol 1983; 154: 366–74.
18. Rhen M, Väisänen-Rhen V. Nucleotide sequence analysis of a P fimbrial regulatory element of the uropathogenic *Escherichia coli* strain KS71 (O4:K12). Microbial Pathogenesis 1987; 3: 387–91.
19. Blyn LB, Braaten BA, White-Ziegler CA, Rolfson DH, Low DA. Phase-variation of pyelonephritis-associated pili in *Escherichia coli*, evidence for transcriptional regulation. EMBO J 1989; 8: 613–20.
20. Båga M, Göransson S, Normark S, Uhlin BE. Transcriptional activation of a *pap* virulence operon from uropathogenic *Escherichia coli*. EMBO J 1985; 4: 3887–93.
21. Båga M, Göransson S, Normark S, Uhlin BE. Processed mRNA with differential stability in the regulation of *Escherichia coli* pilin expression. Cell 1988; 52: 197–206.
22. Göransson M, Forsman K, Uhlin BE. Functional and structural homology among regulatory cistrons of pili-adhesin determinants in *Escherichia coli*. Mol Gen Genet 1988; 212: 412–17.
23. Roosendahl E, Boots M, de Graaf FK. Two novel genes *fan*A and *fan*B, involved in the biogenesis of K99 fimbriae. Nucleic Acids Res 1987; 15: 5973–84.
24. Ott M, Schmoll T, Goebel W, van Die I, Hacker J. Comparison of the genetic determinant coding for the S-fimbrial adhesin (*sfa*) of *Escherichia coli* to other chromosomally encoded fimbrial determinants. Infect Immun 1987; 55: 1940–3.
25. Ott M, Hoschützky H, Jann K, van Die I, Hacker J. Gene clusters for S-fimbrial adhesin (*sfa*) and F1C fimbriae (*foc*) of *Escherichia coli*: comparative aspects of structure and function. J Bacteriol 1988; 170: 3983–90.
26. Riegman N, Kusters R, van Veggel H *et al.* F1C fimbriae of uropathogenic *Escherichia coli*: genetic and functional organization of the *foc* gene cluster and identification of minor subunits. J Bacteriol 1990; 172: 114–20.
27. McClure W. Mechanism and control of transcription in procaryotes. Annu Rev Biochem 1985; 54: 171–204.
28. Båga M, Normark S, Hardy I *et al.* Nucleotide sequence of the *pap*A gene encoding the Pap pilus subunit of human uropathogenic *Escherichia coli*. J. Bacteriol 1984; 157: 330–3.
29. Klemm P. The *fim*A gene encoding the type I fimbrial subunit of *Escherichia coli*: nucleotide sequence and primary structure of the protein. Eur J Biochem 1984; 143: 395–9.
30. Klemm P. Two regulatory *fim* genes, *fim*B and *fim*E, control the phase variation of type I fimbriae in *Escherichia coli*. EMBO J 1986; 5: 1389–93.
31. Norgren M, Baga M, Tennent JM, Normark S. Nucleotide sequence, regulation and functional analysis of the *pap*C gene required for cell surface localization of Pap pili of uropathogenic *Escherichia coli*. Mol Microbiol 1987; 1: 169–78.
32. Riegman N, Acton D, Sakkers R, Van Die I, Hoekstra W, Bergmans H. Functional analysis of the *fso*C gene product of the F7$_1$ (*fso*) fimbrial gene cluster. Mol Microbiol 1990; 4: 101–6.
33. Baga M, Norgren M, Normark S. Biogenesis of *E. coli* Pap pili: PapH, a minor subunit involved in cell anchoring and length modulation. Cell 1987; 49: 241–51.
34. Lindberg F, Tennent JM, Hultgren SJ, Lund B, Normark S. PapD a periplasmic transport protein in P-pilus biogenesis. J Bacteriol 1989; 171: 6052–8.
35. Van Die I, Hoekstra W, Bergmans H. Analysis of the primary structure of P-fimbrillins of uropathogenic *Escherichia coli*. Microbial Pathogenesis 1987; 3: 149–54.
36. Lund B, Marklund BI, Strömberg N, Lindberg F, Karlsson KA, Normark S. Uropathogenic *Escherichia coli* can express serologically identical pili of different receptor binding specificities. Mol Microbiol 1988; 2: 255–63.

37. Berger H, Hacker J, Juarez A, Hughes C, Goebel W. Cloning of the chromosomal determinants encoding hemolysin production and mannose-resistant haemagglutination in *Escherichia coli*. J Bacteriol 1982; 157: 330–3.
38. Boyer HW, Roulland-Dussoix D. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J Mol Biol 1969; 41: 459–72.
39. Hu N, Messing J. The making of strand-specific M13 probes. Gene 1982; 17: 271–7.
40. Casadaban MJ, Chou J, Cohen SN. *In vitro* gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initial signals. J Bacteriol 1980; 143: 971–80.
41. Bolivar F, Rodriguez RL, Green PJ, Betlach MC, Heyneker H, Boyer HW. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 1977; 2: 95–113.
42. Chang ACY, Cohen SN. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic plasmid. J Bacteriol 1978; 134: 1141–56.
43. Vieira J, Messing J. The pUC plasmids, an M13 mp7-derived system for insertion and mutagenesis and sequencing with synthetic universal primers. Gene 1982; 19: 259–68.
44. Bremer E, Silhavy TJ, Weisemann JM, Wienstock GM. λ*plac*Mu: a transposable derivative of bacteriophage lambda for creating *lac*Z protein fusions in a single step. J Bacteriol 1984; 158: 1084–93.
45. Bremer E, Silhavy TS, Weinstock GM. Transposable λ*plac* Mu bacteriophages containing *lac*Z operon fusions and Kanamycin resistance insertions in *Escherichia coli*. J Bacteriol 1985; 162: 1092–9.
46. Birnboim HC, Doly J. A rapid extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 1979; 7: 1513–23.
47. Maniatis T, Fritsch EF, Sambrook S. Molecular cloning. A laboratory manual. Cold Spring Harbor, Cold Spring Harbor Laboratory Press, 1982.
48. Thuring RWF, Sander JP, Borst B. A freeze-squeeze method for recovering long DNA from agarose gels. Anal Biochem 1975; 66: 213–20.
49. Lederberg E, Cohen SN. Transformation of *Salmonella typhimurium* by plasmid deoxyribonucleic acid. J Bacteriol 1974; 119: 1072–4.
50. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977; 74: 5463–7.
51. Beaucage SL, Caruthers MH. Deoxynucleoside phosphoramidites: a new class of key intermediate for deoxypolynucleotide synthesis. Tetrahedron Lett 1981; 22: 1859–62.
52. Devereux J. UWGCG computer programs. University of Wisconsin, 1984.
53. Intelli Genetics. Genofit software package. Geneva, 1986.
54. Miller J. Experiments in molecular genetics: a laboratory manual. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1972.