Mol Gen Genet (1985) 200:385–392

Analysis of the flanking regions from different haemolysin determinants of *Escherichia coli*

S. Knapp[1], I. Then[1], W. Wels[1], G. Michel[1], H. Tschäpe[2], J. Hacker[1], and W. Goebel[1]

[1] Institut für Genetik und Mikrobiologie, Universität Würzburg, Röntgenring 11, D-8700 Würzburg, Federal Republic of Germany
[2] Institut für Experimentelle Epidemiologie, Burgstr. 37, DDR-3700 Wernigerode, German Democratic Republic

Summary. The haemolysin (*hly*) determinant of the plasmid pHly152 contains an IS*2* element at 469 bp upstream of the *hlyC* gene. The sequence at the other (right-hand) end (RS) also shows multiple hybridization with the plasmid pHly152 and the chromosome of some *Escherichia coli* strains but the nucleotide sequence of this region does not reveal the typical properties of an IS element. Similar arrangements in the regions flanking the *hly* determinant are also found on various Hly plasmids from uropathogenic *E. coli* strains. Chromosomal *hly* determinants lack both flanking sequences (IS*2* and RS) in the immediate vicinity of the *hly* genes. The sequences immediately upstream of the *hlyC* gene have been determined from several chromosomal *hly* determinants and compared with the corresponding sequence of the *hly* determinant of the plasmid pHly152. We show that these sequences, which contain one promoter (left promoter, $phly_L$) in all *hly* determinants tested, vary considerably although common sequence elements can still be identified. In contrast, only relatively few nucleotide exchanges have been detected in the adjacent structural *hlyC* genes. The A+T content of the 200 bp sequence upstream of *hlyC* is very high (72 mol% A+T) but even the structural *hly* genes show a considerably higher A+T content (about 60 mol%) than the *E. coli* chromosome on average (50 mol% A+T) suggesting that the *hly* determinant may not have originated in *E. coli*.

Introduction

The four genes comprising the *hly* determinant of *Escherichia coli* are either located on large transmissible plasmids or inserted in the chromosome (Goebel et al. 1974; Minshew et al. 1978; Berger et al. 1982). It has therefore been suggested that the *hly* genes are part of a transposable element (Hacker et al. 1983a; Zabala et al. 1982). Previous data (Knapp et al. 1984) suggested that the flanking sequences of the *hly* determinant of plasmid pHly152 have properties reminiscent of insertion (IS) elements. They can occur in multiple copies on Hly plasmids and on the chromosome of uropathogenic and faecal *E. coli* strains (Knapp et al. 1984). Yet it has been shown that the sequences immediately flanking the *hly* genes of pHly152 are not identical (Knapp et al. 1984) and hence the structure of the *hly* determinant is not that of a transposon. Furthermore it has been demonstrated that the boundaries of several chromosomal *hly* determinants are not the same as those of the pHly152-encoded *hly* genes (Hacker et al. 1983a; Knapp et al. 1984). Recent data (Zabala et al. 1984) indicate that sequences related to the insertion element IS*91* are located in the vicinity of the pHly152-encoded *hly* determinant but the precise location and structure of these sequences has not yet been worked out.

A knowledge of the flanking regions of the *hly* genes is also important for an understanding of the regulation of haemolysin synthesis since several lines of evidence suggest that one of the promoters which control the expression of the *hly* genes is located at the left-hand end, $phly_L$. By inserting Tn*5* into the *hly* genes it has been shown that $phly_L$ probably regulates the expression of *hlyC*, *hlyA* and $hlyB_a$ (Wagner et al. 1983; Juarez et al. 1984). One other promoter, $phly_R$, still rather poorly defined, seems to be responsible for the expression of the $hlyB_b$ gene. Differential expression of the haemolysin determinants may be partly responsible for the differences in virulence observed among some cloned plasmid- and chromosome-encoded haemolysins (Welch et al. 1981; Hacker et al. 1983b).

We describe the further analysis of the flanking regions of the *hly* determinant of plasmid pHly152 which is derived from an *E. coli* strain of animal origin and compare the flanking regions with the corresponding sequences of *hly* determinants located on plasmids or on the chromosomes of uropathogenic *E. coli* strains.

Materials and methods

Bacteria, plasmids and recombinant DNAs. The *E. coli* wild-type strain 536 (O6:K15:H31) and the mutants 536-225 and 536-14 have been described previously (Hacker et al. 1983a; Knapp et al. 1984). For transformation and conjugation the *E. coli* K12 strains HB101 (H. Boyer) and 20144 (G. Schmidt) were used. The haemolysin plasmid pHly152 has been described in detail previously (Smith and Halls 1967; Noegel et al. 1981). The three haemolysin plasmids pHly33, pHly3451 and pHly7380 were isolated from uropathogenic *E. coli* strains in Wenigerode (GDR) and analysed by standard methods (Tietze and Tschäpe 1983); they are listed in Table 1. The recombinant DNAs used are listed in Table 2.

Offprint requests to: S. Knapp

Dedicated to Prof. F. Lingens on the occasion of his 60th anniversary

Table 1. Haemolysin plasmids

Plasmid	Inc group	Size (Md)	Source	Reference
pHly152	I_2	41	Faecal, animal	Smith and Halls 1967
pHly33	F_{VI}	140	Uropathogenic human	This study
pHly3451	F_{VI}	145	Uropathogenic human	This study
pHly7380	I_1	n.t.[a]	Uropathogenic human	This study

[a] n.t., not tested

Table 2. Recombinant DNAs

Recombinant plasmid	Characteristics	Reference
pANN104	Cm^r, *hlyC*, part of *hlyA* from pHly152	Noegel et al. (1981)
pANN202	Cm^r, *hlyC*, part of *hlyA* from pHly152	Noegel et al. (1981)
pANN215	Cm^r, part of *hlyA*	Noegel et al. (1981)
pANN505	$hlyB_b$ downstream region of pHly152 and part of $hlyB_b$	Hacker et al. (1983a)
pANN531	Ap^r, *hlyC, hlyA,* $hlyB_a$, $hlyB_b$, from *Escherichia coli* 764 (018), cosmid	Berger et al. (1982)
pANN541	Ap^r, *hlyC, hlyA,* $hlyB_a$, $hlyB_b$, from *E. coli* 341 (075), cosmid	Berger et al. (1982)
pANN522	Ap^r, *hlyI* (*C, A,* B_a, B_b) from *E. coli* 536[a] cosmid	Knapp et al. (1984)
pANN473	Ap^r, *hlyII* (*C, A,* B_a, B_b) from *E. coli* 536[a] cosmid	Knapp et al. (1984)
pANN912	Ap^r, $hlyB_b$, downstream region of *hlyI* from *E. coli* 536, cosmid	This study
pANN762	Ap^r, *hlyC* downstream region of *hlyI* from *E. coli* 536, cosmid	This study

[a] Strain 536 of *E. coli* harbours two different haemolysin determinants (*hlyI* and *hlyII*) on the chromosome (for details see Knapp et al. 1984)

Media, chemicals and enzymes. Cultures were grown in enriched nutrient broth or in alkaline broth extract. Radiochemicals were purchased from New England Nuclear Corporation, USA; antibiotics used were a gift from Bayer, Leverkusen, FRG. All other chemicals were obtained from Merck, Darmstadt, FRG. Restriction enzymes and T4 ligase were purchased from Biolabs, Richmond, USA. DNA polymerase I was obtained from Boehringer, Mannheim, FRG.

Isolation of chromosomal and plasmid DNA. Chromosomal DNA was isolated as described earlier (Hacker et al. 1983a). Plasmid DNA from clones carrying recombinant DNA was screened by the cleared lysate procedure (Birnboim and Doly 1979) and preparative DNA isolation was achieved as described (Grinstedt et al. 1978).

Nick translation. Plasmids were labelled by nick translation with a mixture of all four α-^{32}P-labelled dNTPs as described previously (Rigby et al. 1977) and purified by ethanol precipitation.

Cleavage with restriction enzymes and electrophoresis of chromosomal DNAs. The chromosomal DNA was treated with appropriate restriction enzymes and the resulting fragments were separated by agarose gel electrophoresis using 0.7–1.0% gels as described previously (Hacker et al. 1983a).

Hybridization and autoradiography. The transfer of DNA fragments from agarose gels to nitrocellulose filters and the washing and autoradiography of the filters after hybridization were performed as described by Southern (1975). The filters were hybridized in 50% formamide at 43° C overnight. Stringent conditions were used for the washing procedure.

Conjugation. Conjugation of Hly plasmids was performed on plates as described previously (Goebel et al. 1974).

Transformation. The *E. coli* strains were transformed by the $CaCl_2$ method (Lederberg and Cohen 1974).

Haemolysin production. Erythrocyte lysis was detected on meat agar plates containing washed human erythrocytes and confirmed in a liquid assay (Wagner et al. 1983).

Cosmid cloning procedure. Recombinant DNA was packaged in vitro as described previously (Berger et al. 1982; Collins and Brüning 1978).

Oligonucleotide synthesis. Oligonucleotides were synthesized in an Applied Biosystems 380A DNA synthesizer by the phosphoramidite method of Beaucage and Caruthers (1981). Oligonucleotides were purified on a polyacrylamide gel.

DNA sequencing was performed by the Sanger technique using the M13 mp8/mp9 systems as described (Hu and Messing 1982). For sequencing the regions 5′ upstream of *hlyC*, of the *hly* determinants, a synthetic 15mer oligonucleotide was used starting with the third nucleotide of the coding region of *hlyC*.

Results

Arrangement of the nucleotide sequence of the hly determinant of plasmid pHly152 upstream of hlyC

We have previously reported that a sequence of about 1,400 base pairs (bp) flanks the left-hand end (*hlyC* proximal end) of the *hly* determinant of the plasmid pHly152 and occurs in multiple copies in the chromosome of various *E. coli* strains (Knapp et al. 1984). A comparison of the restriction sites identified within this sequence (Fig. 1) with known IS elements of *E. coli* suggests its identity with IS*2*. In order to prove this directly we have sequenced about 250 bp starting from the *Hin*dIII site. As shown in Fig. 1 this sequence is, with the exception of a few base pairs, identical to the sequence reported previously for IS*2* (Ghosal et al. 1979). The restriction sites within this sequence in pHly152 suggest that a full-length IS*2* element is located 469 bp upstream of the *hlyC* gene (Wels et al., to be published). The region between this IS*2* and *hlyC* of pHly152 carries the left pro-

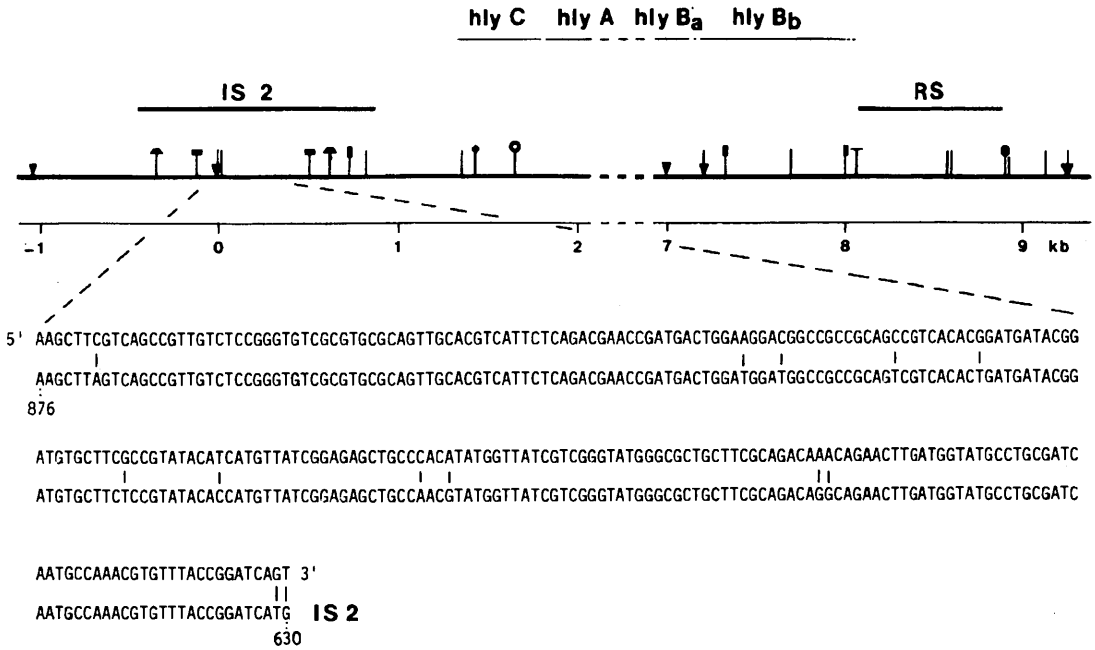

Fig. 1. Physical map of the *hlyC* and $hlyB_b$ proximal regions and nucleotide sequence downstream of the *Hin*dIII restriction site at map coordinate 0 of the haemolysin (*hly*) determinant of pHly152 and comparison of this sequence to the IS*2* sequence from bp 876 to 630 (Ghosal et al. 1979). The coordinates of the *hly* determinant have been defined previously (Goebel and Hedgpeth 1982). The size of the restriction fragments is given in kilobase pairs (kb). The *bars* above the map represent the lengths and the locations of the flanking sequences IS*2* and RS (right-hand sequence). Differences between the determined left-hand sequence and IS*2* are indicated as small *vertical dashes*. Restriction sites for *Hin*fI and *Bgl*I are only given for that part of the *hly* map corresponding to IS*2*. Symbols: ↑, *Ava*I; ♀, *Bam*HI; ↑, *Bgl*I; ♀, *Bgl*II; ▾, *Eco*RI; |, *Hae*III; ↑, *Hin*cII; ↓, *Hin*dIII; ↑, *Hin*fI; ⊤, *Pst*I

moter, $phly_L$, which regulates the expression of the *hly* genes, *hlyC*, *hlyA* and possibly $hlyB_a$ (Wagner et al. 1983; Juarez et al. 1984).

Nucleotide sequence of the flanking region downstream of $hlyB_b$ of plasmid pHly152

The right-hand flanking sequence of the *hly* determinant of pHly152 has also been shown to cause multiple hybridizations with the chromosomes of various *E. coli* strains (Hacker et al. 1983a; Knapp et al. 1984). The length of this sequence was determined to be 850 bp, located between the *Pst*I and the *Bgl*II sites at coordinates 8.05 kb and 8.95 kb of the *hly* determinant of pHly152 (Hacker et al. 1983a; Knapp et al. 1984). A substantial portion of this nucleotide sequence has been determined (Michel 1984) and compared with sequences reported for other IS elements of *E. coli* but identity to known IS sequences could not be demonstrated. Furthermore we could not detect any of the characteristic features of an IS element such as duplication of the target site and inverted repeats at the ends of this sequence.

Similar flanking sequences are found near hly determinants from other Hly plasmids but not near those of chromosomal origin

Previous hybridization data (de la Cruz et al. 1980) obtained with different Hly plasmids indicated that, in some, the region between the *Bam*HI site in *hlyC* and the *Hin*dIII site in IS*2* is identical to that of pHly152. This suggests that the nucleotide sequence upstream of *hlyC* in these plasmids and pHly152 is identical. These Hly plasmids were derived from *E. coli* strains isolated from animal sources. We hybridized three plasmids, pHly33, pHly3451 and pHly7380, derived from *E. coli* strains isolated from pyelonephritis patients, with specific probes for the flanking regions of the *hly* determinant of pHly152. The three Hly plasmids were transferred into *E. coli* K12 strain 20144 and hybridization was performed with total DNA from the resultant plasmid-bearing strains using as a control the plasmid-free parental strain. Total DNA was cleaved with either *Sal*I or *Eco*RI and hybridized with two probes containing the right-hand or left-hand flanking sequences of the pHly152-encoded *hly* genes (probes a and c, Fig. 2) and with a probe representing an internal fragment of the same *hly* determinant (probe b, Fig. 2). There is no *Sal*I site within the known *hly* determinants (Hacker and Hughes 1985) and all three probes should therefore hybridize with the same *Sal*I fragment if the flanking regions of the *hly* genes analysed are similar to those of the pHly152-encoded *hly* gene sequence. As shown in Fig. 2A both probes containing the flanking sequence hybridize with the same *Sal*I fragment as the probe containing the structural *hly* gene. This indicates that flanking sequences identical, or similar to those of the pHly152-encoded *hly* determinant are also located in the vicinity of the *hly* determinants of the three plasmids isolated from uropathogenic strains. In addition, several other *Sal*I fragments hybridize with the probes containing the flanking sequences (particularly with the left-hand flanking sequence containing part of IS*2*) suggesting that these sequences are present in additional copies on the plasmid and/or the chromosome of these strains.

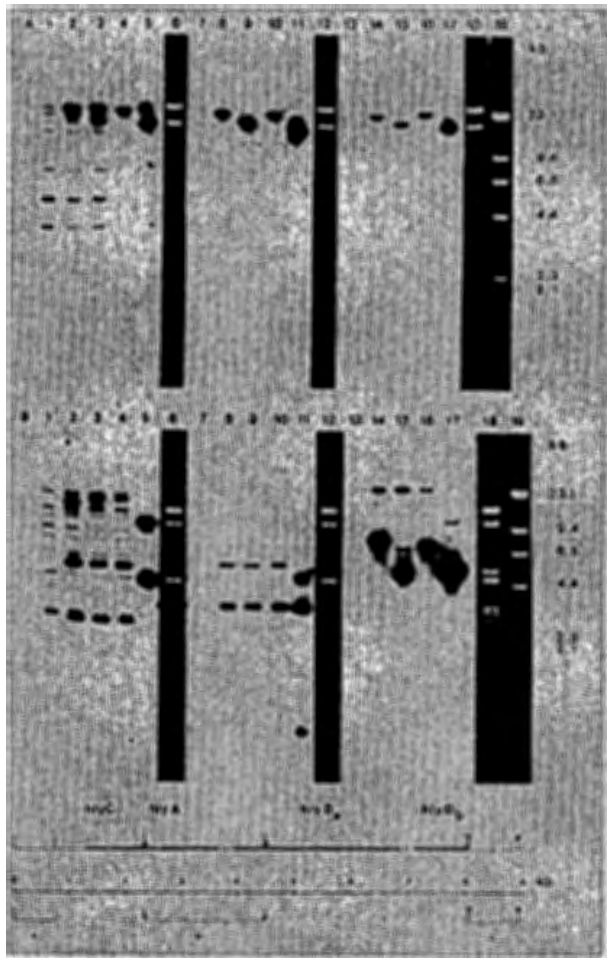

Fig. 2A, B. Hybridization pattern of **A** *Sal*I- and **B** *Eco*RI-cleaved total DNA of *Escherichia coli* 20144 with no plasmid, lanes 1, 7 and 13, or harbouring: pHly33, lanes 2, 8 and 14; pHly3451, lanes 3, 9 and 15; pHly7380, lanes 4, 10 and 16. Hybridization patterns of **A** *Sal*I- and **B** *Eco*RI-cleaved plasmid DNA of pHly152, lanes 5, 11 and 17, and restriction fragments, lanes 6, 12 and 18. As size markers, *Hin*dIII fragments of λ DNA were used (lane 19). The fragments (a, b, c) used for the hybridization as nick-translated, α^{32}P-labelled DNA probes are indicated on the map below. Probe a (lanes 1–5) was isolated from pANN104; probe b (lanes 7–11) was isolated from pANN215, probe c (lanes 13–17) was isolated from pANN505 (see Table 2). The *thick line* in the map represents the structural *hly* genes of pHly152 whereas the *thin line* indicates flanking regions. The map coordinates are those defined by Goebel and Hedgpeth 1982. For symbols see Fig. 1

To narrow down further the flanking regions relative to the structural *hly* genes, we cleaved total DNA of the strains habouring the three plasmids with *Eco*RI and hybridized these DNAs with the three probes used in the previous experiment. Plasmid pHly152 DNA was treated identically as a control.

The IS*2*-specific probe (probe a, Fig. 2) hybridized in all cases with three plasmid-specific *Eco*RI-fragments of more than 17 kb, 5.3 kb and 2.9 kb, whereas pHly152 hybridized to only two *Eco*RI fragments of 10.2 kb and 4.9 kb (Fig. 2B). The *hly* gene-specific probe, *Hind*III-C, (probe b, Fig. 2) hybridized with the 5.3 kb *Eco*RI fragment. It also hybridized as expected with a small 0.53 kb fragment, which is part of the structural *hlyA* gene, and the adjacent *Eco*RI fragment of 3.2 kb, identical to *Eco*RI-G of pHly152 (Noegel et al. 1981) which is again conserved in all four Hly plasmids. There is no background hybridization between the *hly* gene-specific probe and the chromosomal DNA of strain 20144 alone (Fig. 2). Probe c (Fig. 2), specific for the right-hand flanking 850 bp sequence, hybridized in all three plasmids with those *Eco*RI fragments which follow the 3.2 kb *Eco*RI fragment, suggesting that the *hly* determinants of these Hly plasmids are also flanked at the right-hand end of the *hly* genes by this 850 bp sequence. A summary of the hybridization data is given in Fig. 3. These data suggest that all three Hly plasmids from uropathogenic *E. coli* strains have the same arrangement of flanking sequences as pHly152 which derives from an animal *E. coli* strain.

In contrast to the plasmid *hly* determinants, none of several chromosomal *hly* determinants from haemolytic uropathogenic *E. coli* strains carry an IS*2* element in the vicinity of the *hlyC* gene. No hybridization of large *Sal*I fragments, which contain all four structural *hly* genes and flanking sequences of more than 1,000 bp at both ends, has been found (Knapp et al. 1984).

The haemolytic uropathogenic *E. coli* strain 536, which harbours two chromosomally encoded *hly* determinants (Knapp et al. 1984) also seems to lack IS*2* in the immediate vicinity of either of the two *hlyC* genes. As shown in Fig. 4A the IS*2*-specific probe (probe a) hybridizes, albeit only weakly, with the *Eco*RI fragment which contains the *hlyC* structural gene and its flanking region. This weak hybridization signal cannot be explained as yet but it is certainly not caused by a complete IS*2* sequence as indicated by a comparison of hybridization intensities between this and the other IS*2* homologous fragment at about 4.0 kb. The DNA of mutant 536-225, retaining *hlyI* only (Knapp et al. 1984), does not hybridize at all with the IS*2*-specific probe a. These data, together with the observation that the same *Eco*RI DNA fragments of mutant 536-14, retaining *hlyII*, hybridize with the IS*2*-specific probe a as in the wild-type strain 536 (Fig. 4A), indicate that the complete IS*2* copy on the 4.0 kb *Eco*RI fragment, and the weakly hybridizing, IS*2*-related sequence, are located on the 70 kb region containing *hlyI* (Knapp et al. 1984; Knapp, unpublished data).

It has been shown that the right-hand 850 bp flanking sequence (RS) of pHly152 hybridizes with several restriction fragments of *E. coli* 536. The two cosmid DNAs pANN522 and pANN473, carrying the (*hlyI*) or (*hlyII*) determinants respectively of *E. coli* 536, do not hybridize with the RS-specific probe (Fig. 4B, probe c) indicating that the RS sequence is not part of the right-hand flanking regions of these two chromosomal *hly* determinants. The cosmid clones obtained from a gene bank of *E. coli* 536, which overlap with pANN522 either at the right-hand end (pANN912) or at the left-hand end (pANN762), hybridize with probe c (Fig. 4B) in restriction fragments which map approximately 25 kb upstream of the 5′ end of *hlyC* and 14 kb downstream of the putative 5′ end of the $hlyB_b$ gene in the *hly* determinant (data not shown). In addition, there is evidence that these chromosomal sequences are more complex than the 850 bp $hlyB_b$-flanking sequence (RS) of the Hly plasmids and preliminary evidence suggests that RS is only part of these chromosomal sequences (S. Knapp, unpublished data).

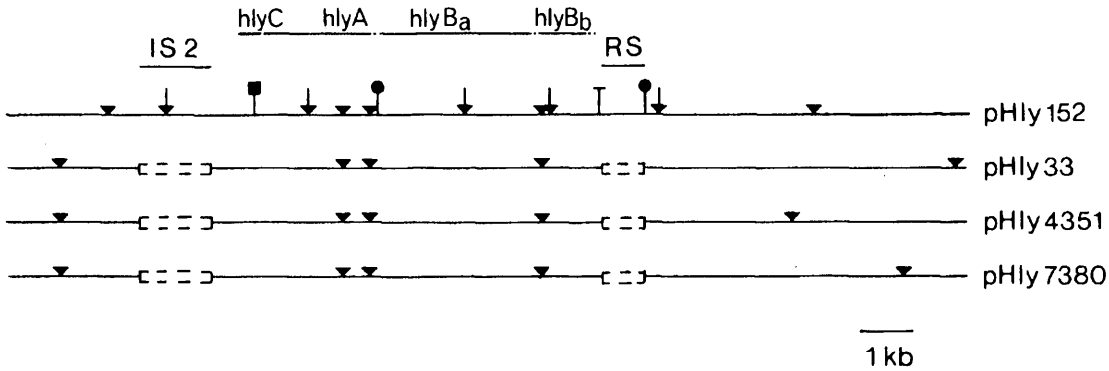

Fig. 3. Summary of the hybridization data obtained with plasmids pHly152, pHly33, pHly3451 and pHly7380. The small *dashed boxes* represent copies of IS*2* and RS respectively. The precise map coordinates of both sequences within the indicated *Eco*RI fragments of these plasmids were not determined. For symbols see Fig. 1

The nucleotide sequence upstream of hlyC (promoter phly$_L$-containing sequence) varies considerably between the plasmid and chromosomal hly determinants

Previous experiments have indicated that the sequence between *hlyC* and IS*2* of pHly152 carries a promoter (p*hly*$_L$) regulating the gene expression of hlyC, hlyA and *hlyB*$_a$ (Wagner et al. 1983; Juarez et al. 1984). Since the arrangement of the four structural genes has been shown to be identical in both plasmid and chromosomal *hly* determinants, we assume that the region upstream of *hlyC* also contains a transcriptional control region for the corresponding chromosomally encoded *hly* genes. The abovementioned hybridization data suggest that the flanking regions of these two types of *hly* determinants may not be conserved, in contrast to the structural genes (Müller et al. 1983; Welch et al. 1983). Using synthetic oligonucleotides derived from the known sequence of the *hlyC* gene of pHly152 (Juarez et al. 1984) we sequenced the *hlyC*-proximal upstream region of several chromosomal *hly* determinants. These parts of the chromosomal *hly* determinants were cloned into M13mp9 using suitable fragments which are located upstream of the highly conserved *Bam*HI site of *hlyC* (Müller et al. 1983; Welch et al. 1983; Stark and Shuster 1982; Mackman and Holland 1984). Figure 5 shows that the corresponding sequences upstream of the *hlyC* structural gene vary considerably between the hly determinant of pHly152 and the two chromosomal *hly* determinants. Our previously published nucleotide sequence of pHly152 has been corrected in some positions, (cf. Juarez et al. 1984). These chromosomal sequences, isolated from two completely different strains (06 and 075 serogroups) and representing expression sites of *hly* determinants which exhibit different levels of haemolysin synthesis and virulence (Hacker et al. 1983b) are rather similar. The differences between them are small insertions and single base pair exchanges (Fig. 5). In contrast, both chromosomal sequences (and also other *hlyC* upstream sequences of chromosomal origin – data not shown) differ widely from the corresponding plasmid sequence. There are still some common elements between these plasmid and chromosomal sequences which suggest a common origin (particularly in the first 20 bp immediately upstream of *hlyC*) but the divergence of these sequences is much more pronounced than that of the *hlyC* genes. The differences which we could detect within the structural part of plasmid- and chromosome-encoded *hlyC* sequences are relatively minute (Fig. 6). Most alterations represent single base pair exchanges which are either silent or lead to amino acid exchanges which are not likely to alter significantly the structure of the HlyC proteins (with the exception of an arginine to threonine exchange close to the C-terminal end in both chromosomal HlyC proteins).

One remarkable feature of these *hlyC*-upstream sequences is their high A+T content which amounts to 72 mol% A+T in the 200 bp upstream of *hlyC*. The A+T content even of the structural *hly* genes is unusually high for *E. coli*: *hlyC* has an A+T content of 62 mol%, *hlyA* of 59.5 mol% and the *hlyB*$_a$ and *hlyB*$_b$ genes not yet completely sequenced are also close to 60 mol% A+T. An average A+T content of 50 mol% is usually exhibited by *Escherichia coli* in its chromosome and most of its plasmids.

Discussion

The well characterized haemolysin determinant of plasmid pHly152 is flanked by two relatively short sequences of different size (1,400 bp and 850 bp) which are frequently found in multiple copies in the chromosome of *E. coli* strains (Hacker et al. 1983a; Knapp et al. 1984) suggesting that they may represent IS elements. We have shown here that the sequence at the left-hand end (*hlyC*-proximal) is indeed identical to IS*2*, based on similar restriction sites and the identity of a partial nucleotide sequence of the left-hand *hly*-flanking sequence with the reported IS*2* sequence (Ghosal et al. 1979). The 850 bp sequence at the right-hand end (*hlyB*$_b$-proximal; RS) has also been partially sequenced. As already expected from the complete lack of hybridization between these two sequences flanking the *hly* determinant of pHly152 (Knapp et al. 1984), there is no similarity in their nucleotide sequences. Nucleotide analysis of the right-hand sequence (RS) has not revealed any characteristic features of an IS element such as duplicated target sites or inverted repeats at the ends, although both ends should be contained within the determined sequence. This suggests that RS may be only part of an IS element. This assumption is supported by the observation that a recently cloned chromosomal fragment of *E. coli* 536 hybridizes strongly with RS from the pHly152-encoded *hly* determinant. The sequence responsible for the hybridization is considerably larger in size (about 2 kb) than RS (850 bp; Knapp et al., unpublished results). Zabala et al. (1984) have

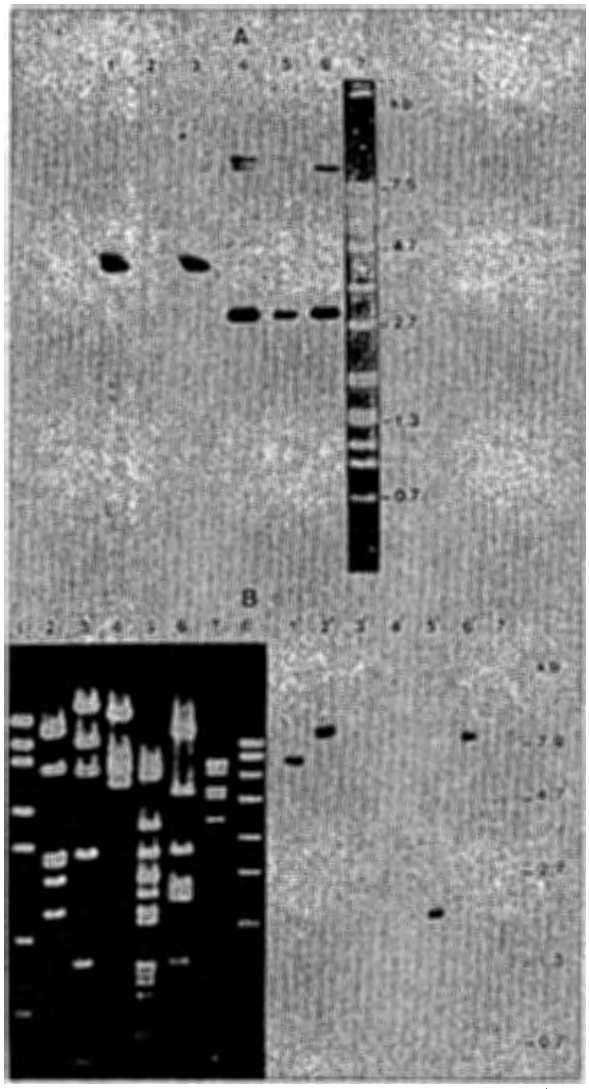

Fig. 4. **A** Hybridization patterns of *Eco*RI-cleaved chromosomal DNA of *Escherichia coli* 536 (lanes 1 and 4), mutant 536–225 (lanes 2 and 5) and mutant 536–14 (lanes 3 and 6) obtained with nick-translated, α^{32}P-labelled DNA probes a and b (see Fig. 2). Probe a, lanes 1–3; probe b, lanes 4–6 **B** Restriction and hybridization patterns of cosmid clones pANN762 (lane 1 and 1′, *Bgl*II; lane 2 and 2′, *Hin*dIII); pANN522 (lane 3 and 3′, *Eco*RI; lane 4 and 4′, *Bam*HI); pANN912 (lane 5 and 5′, *Pst*I; lane 6 and 6′, *Eco*RI); pANN473 (lane 7 and 7′, *Bam*HI). Hybridization was with nick-translated, α^{32}P-labelled probe c DNA (see Fig. 2). As size markers, *Eco*RI fragments of phage SPPI were used (lane 7, **A**; lane 8, **B**)

```
-200  ACAGGAATGA GTAATTATTT ATGCTTGATG TTTTTTGA.C TCTTGCTTTT  p-152
      ATAAGCATAT GTAATGATTT TCGTTTGCTT TTTTTAAACC TGCTACCGCA  c-06
      ATAAGCATAT GTAAGGATTT TGGTTTGCTT TTTTT.AACC TGCCACCGCA  c-075
                        C

-150  TATAGTTATT ATTTTTAAGT TAGTCAGCGC AAT.AAA... AACTTGCTTT
      ATGAATGCTT TTTTTAATGT GAATGTGCGT TATCAAACTC AAATGGCAAG
      ATGAATGCTT TTTTT.ATGT TAATGTGCGT TATGAAACT. AAAT.GCAAG

-100  TA...ATATT AATGGAGTTA TGACATTAAA CGGAAGAAAC ATAAAGGCAT
      AAACGATATT TAAAGGCTTA AGGCTTAATA TTCGTTCTTC TCACAGAC.T
      AAAC.ATATT TA.......A AGCATTAATA .TCGTTC.TC TCACAGAC.T

-50   ATTTTTGCCA CAATATTTAA TCATATAATT TAAGTTGTAG TGAGTTTATT
      CCGTTT..CA C.TTATTCAA GAATATAATT TAATTTATAG TGAGCTTATT
      CCGCTT..TA C.TTATTCAA GAATATAATT TAATTTATAG TGAGCTTATT

1     ATGAATATAA ACAAACCATT AGAGATTCTT GGGCATGTAT CCTGGCTATG
      ATGAATATAA ACAATCCATT AGAGATTCTT GGGCATGTAT CCTGGCTATG
      ATGAATATAA ACAATCCATT AGAGATTCTT GGGCATGTAT CCTGGCTATG
```

Fig. 5. Comparison of the nucleotide sequences 5′ upstream of the *hlyC* genes of the *hly* determinant of plasmid pHly152 (p-152), the chromosomal *hly* determinant of *Escherichia coli* 536 (c-06; *hly*II, Knapp et al. 1984) and *E. coli* 341 (c-075). DNA fragments from plasmids pANN202, pANN541 and pANN473 carrying the appropriate *hlyC* 5′ upstream regions of pHly152, *E. coli* 536 (*hlyII*) and *E. coli* 341 were isolated for the construction of M13mp9 recombinant clones. As primer a synthetic 15mer oligonucleotide from nucleotides +3 to +17 (*underlined* in the sequence) was used. Positions indicating differences in the nucleotide sequences of these regions between the various *hly* determinants are marked by a *point*. The start codon *hlyC* (Juarez et al. 1984), is *underlined*

recently reported that the right-hand end of the *hly* determinant of pHly152 forms a stable heteroduplex with part of IS*91*. This sequence is not identical to RS but appears to be located further downstream on pHly152 (Zabala et al. 1984; Knapp et al., unpublished results).

Sequences similar to those found in the flanking regions of the *hly* determinant of plasmid pHly152 seem to be present in other plasmid-encoded *hly* determinants. The restriction patterns and hybridization data obtained from *hly* determinants of several Hly plasmids from *E. coli* strains from animal sources have shown high similarity, not only in the four structural genes but also in the flanking regions (De la Cruz et al. 1980). We have extended this work by performing a comparative study that includes three rather rarely occurring Hly plasmids from uropathogenic *E. coli* strains and have found by hybridization with specific probes for IS*2* and RS that both sequences are present at similar distances from *hlyC* and $hlyB_b$, respectively, at the two ends of these *hly* determinants. In contrast, none of the chromosomal hly determinants that we have analysed carries an IS*2* element close the the left-hand end, and sequences related to RS of pHly152 were, if present, farther away from the $hlyB_b$ structural gene than in the plasmids analysed. These apparent differences in the flanking regions of the two types of *hly* determinants indicate a considerable evolutionary divergence of these determinants and argue against a frequent transposition event of the *hly* determinant as described for many antibiotic resistance genes. This is further supported by the high variability in the *hlyC*-proximal region shown to carry one left promoter regulating expression of *hlyC*, *hlyA* and possibly $hlyB_a$ (Wagner et al. 1983; Juarez et al. 1984). These sequences have been shown to differ considerably between plasmid pHly152 (and other Hly plasmids) and two chromosomal *hly* determinants of different uropathogenic strains. Both types of *hlyC* upstream sequences still have some elements in common which may suggest a common ancestral origin but the extent of

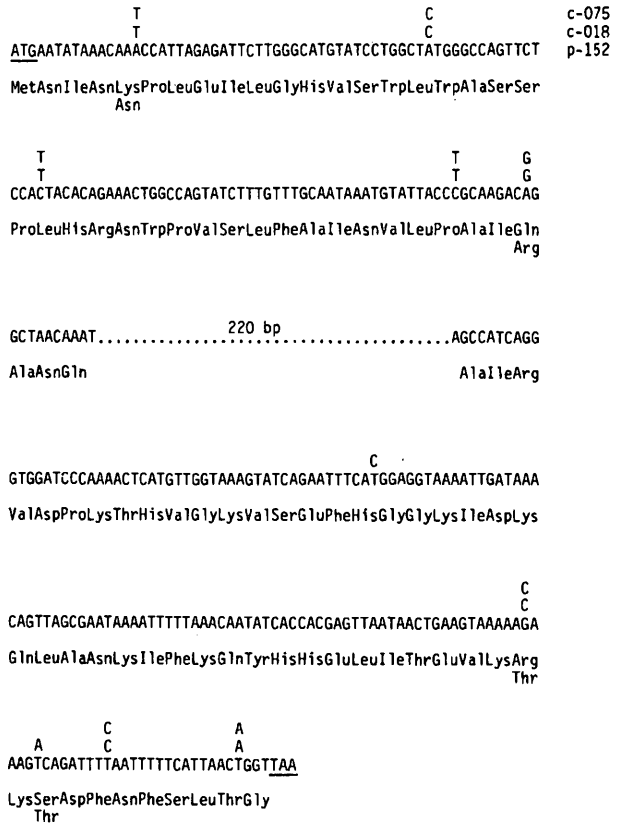

Fig. 6. Comparison of the nucleotide sequences of the *hlyC* genes from the pHly152-encoded *hly* determinant (p-152) and two chromosomal *hly* determinants (c-018 and c-075). Only the altered nucleotides and amino acid exchanges are indicated. DNA fragments from plasmids pANN202, pANN531 and pANN541 were used for the construction of M13 recombinant clones carrying the appropriate *hlyC* genes of pHly152, *Escherichia coli* 764 and *E. coli* 341

variation in the nucleotide sequence suggests the divergence of these sequences a long time ago.

It is interesting to note that not only the varying sequences in the left-hand promoter region but also the more conserved structural genes of the *hly* determinants from *Escherichia coli* show a remarkably high A+T content which exceeds 70 mol% in the first 200 bp upstream of *hlyC* and reaches an average of about 60 mol% in the two completely sequenced structural genes, *hlyC* and *hlyA* (Hess, Vogel and Goebel, to be published). The other two *hly* genes, B_a and B_b, not yet completely sequenced, seem to exhibit a similarly high A+T content. This nucleotide composition is much higher than the average A+T content of the *E. coli* chromosome and its plasmids. This suggests that the haemolysin determinant found today in *E. coli* may have originated in a different biological system. This assumption is further supported by the unusual codon usage for the HlyC and HlyA proteins (Hess, Vogel and Goebel, to be published).

Acknowledgements. The authors wish to thank M. Vogel for help and advice, E. Appel for typing the manuscript, T. Chakraborty for helpful suggestions and discussions and S. Kathariou for critical reading of the manuscript. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (SFB 105-A12).

References

Beaucage SL, Caruthers MH (1981) Deoxynucleoside phosphoramidites: a new class of key intermediates for deoxypolynucleotide synthesis. Tetrahedron Lett 22:1859–1862

Berger H, Hacker J, Juarez A, Hughes C, Goebel W (1982) Cloning of the chromosomal determinants encoding hemolysin production and mannose resistant hemagglutination in *Escherichia coli*. J Bacteriol 152:1241–1247

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acid Res 7:1513–1523

Collins J, Brüning HJ (1978) Plasmids useable as gene cloning vectors in an in vitro packaging by coliphage lambda: "cosmids". Gene 4:85–107

De la Cruz F, Müller D, Ortiz JM, Goebel W (1980) A hemolysin determinant common to *Escherichia coli* Hly plasmids of different incompatibility groups. J Bacteriol 143:825–833

Ghosal D, Sommer H, Saedler H (1979) Nucleotide sequence of the transposable DNA element IS2. Nucleic Acids Res 6:1111–1112

Goebel W, Hedgpeth J (1982) Cloning and functional characterization of the entire plasmid-encoded hemolysin determinant of *Escherichia coli*. J Bacteriol 151:1290–1298

Goebel W, Royer-Pokora B, Lindenmair W, Bujard H (1974) Plasmids controlling synthesis of hemolysis in *Escherichia coli*: molecular properties. J Bacteriol 118:964–973

Grinstedt J, Bennett PM, Higginson S, Richmond MH (1978) Regional preference of insertion of TN*501*and Tn*802* into RP1 and its derivatives. Mol Gen Genet 166:313–320

Hacker J, Knapp S, Goebel W (1983a) Spontaneous deletions and flanking regions of the chromosomally inherited hemolysin determinant of an *Escherichia coli* 06 strain. J Bacteriol 154:1145–1152

Hacker J, Hughes C, Hof H, Goebel W (1983b) Cloned hemolysin genes from *Escherichia coli* that cause urinary tract infection determine different levels of toxicity in mice. Infect Immunol 42:57–63

Hacker J, Hughes C (1985) Genetics of *Escherichia coli* hemolysin. Curr Top Microbiol Immunol, in press

Hu N, Messing J (1982) The making of strand-specific M13 probes. Gene 17:271–277

Juarez A, Hughes C, Vogel M, Goebel W (1984) Expression and regulation of the plasmid-encoded hemolysin determinant of *Escherichia coli*. Mol Gen Genet 197:196–203

Knapp S, Hacker J, Then I, Müller D, Goebel W (1984) Multiple copies of hemolysin genes and associated sequences in the chromosome of uropathogenic *Escherichia coli* strains. J Bacteriol 159:1027–1033

Lederberg EM, Cohen SN (1974) Transformation of *Salmonella typhimurium* by plasmid deoxyribonucleic acid. J Bacteriol 119:1072–1074

Mackman N, Holland B (1984) Functional characterization of a cloned haemolysin determinant from *E. coli* of human origin, encoding information for the secretion of a 107 K polypeptide. Mol Gen Genet 196:129–134

Michel G (1984) Feinkartierung und Sequenzierung eines Endes der Plasmidkodierten Hämolysindeterminante von *Escherichia coli*. Diplom Thesis, University Würzburg

Minshew BH, Jorgensen J, Counts GW, Falkow S (1978) Association of hemolysin production, hemagglutination of human erythrocytes, and virulence for chicken embryos of extraintestinal *Escherichia coli* isolates. Infect Immunol 20:50–54

Müller D, Hughes C, Goebel W (1983) Relationship between plasmid and chromosomal hemolysin determinants of *Escherichia coli*. J Bacteriol 153:846–851

Noegel A, Rdest U, Goebel W (1981) Determination of the functions of hemolytic plasmid pHly152 of *Escherichia coli*. J Bacteriol 145:233–247

Rigby PW, Dieckman JM, Rhodes C, Berg P (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J Mol Biol 113:237–251

Smith HW, Halls S (1967) The transmissible nature of the genetic factor in *Escherichia coli* that controls haemolysin production. J Gen Microbiol 47:153–161

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Stark JM, Shuster CW (1982) Analysis of hemolytic determinants of plasmid pHly152 by Tn*5* mutagenesis. J Bacteriol 152:963–967

Tietze E, Tschäpe H (1983) Plasmid pattern analysis of natural bacterial isolates and its epidemiological implication. J Hyg Camb 90:475–488

Wagner W, Vogel M, Goebel W (1983) Transport of hemolysin across the outer membrane of *Escherichia coli* requires two functions. J Bacteriol 154:200–210

Welch RA, Falkow S (1984) Characterization of *Escherichia coli* hemolysins conferring quantitative differences in virulence. Infect Immunol 43:156–160

Welch RA, Dellinger EP, Minshew B, Falkow S (1981) Hemolysin contributes to virulence of extraintestinal *Escherichia coli* infections. Nature 294:665–667

Welch RA, Hull R, Falkow S (1983) Molecular cloning and physical characterization of a chromosomal hemolysin from *Escherichia coli*. Infect Immunol 42:178–186

Zabala JC, De la Cruz F, Ortiz JM (1982) Several copies of the same insertion sequence are present on alpha-hemolytic plasmids belonging to four different inompatibility groups. J Bacteriol 151:472–476

Zabala JC, Garcia-Lobo JM, Diaz-Aroca E, De la Cruz F, Ortiz M (1984) *Escherichia coli* alpha-hemolysin synthesis and export genes are flanked by a direct repetition of IS*91* – like elements. Mol Gen Genet 197:90–97

Communicated by E. Bautz

Received March 21 / June 5, 1985