

Sequence Diversity among Related Genes for Recognition of Specific Targets in DNA Molecules

JILL A. GOUGH† AND NOREEN E. MURRAY‡

European Molecular Biology Laboratory
Postfach 10.2209
6900 Heidelberg
Federal Republic of Germany

(Received 4 January 1983)

Escherichia coli strains K12 and B, and a new strain designated D, each encode a characteristic restriction and modification enzyme. These enzymes (*EcoK*, *EcoB* and presumably *EcoD*) comprise three subunits of which one, that encoded by the so-called specificity gene (*hsdS*), is responsible for recognition of the DNA sequence specific to that system. The other two subunits, encoded by *hsdR* and *hsdM*, are interchangeable between systems, and the available molecular evidence suggests that the *hsdR* and *hsdM* genes are highly conserved. The DNA sequence of a segment of the *hsd* region that includes the *hsdS* gene has been determined for each of the three strains. The *hsdS* gene varies in length from 1335 to 1425 base-pairs and the only regions showing obvious homology, one of about 100 base-pairs and a second of about 250 base-pairs, are highly conserved. The remainder of each *hsdS* gene shares little, or no, homology with either of the other related specificity genes. Thus, the specificity subunits, though components of a family of closely related enzymes with very similar functions, have remarkably dissimilar primary structure.

1. Introduction

The restriction enzymes of *Escherichia coli* K12 (*EcoK*) and B (*EcoB*) are complex, multifunctional enzymes (for reviews, see Modrich, 1979; Yuan, 1981; Endlich & Linn, 1981). Genetic analyses have shown that each enzyme is encoded by three chromosomally located genes. The relatedness of *EcoK* and *EcoB* is inferred from complementation tests, which indicate the exchange of subunits between the K and B enzymes. Such studies also led to the hypothesis that the product of the specificity gene, *hsdS* (*hsd* for *host specificity DNA*) is responsible for recognition of the DNA sequence specific to that system, that of a second gene, *hsdM*, together with *hsdS*, is required for modification, while the product of the third

† The Ludwig Institute of Cancer Research, Post Office, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia.

‡ Department of Molecular Biology, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh EH9 3JR, Scotland.

gene, *hsdR*, together with the other two, is essential for restriction (Arber & Linn, 1969; Boyer & Roulland-Dussoix, 1969; Glover & Colson, 1969; Hubacek & Glover, 1970).

The *EcoK* and *EcoB* restriction enzymes contain three non-identical subunits, the approximate molecular weights of which are 135,000, 60,000 and 50,000 (Eskin & Linn, 1972; Meselson *et al.*, 1972). In addition to their activities as restriction endonucleases, both enzymes are modification methylases and DNA-dependent ATPases. Although these enzymes bind to a specific nucleotide sequence, the target for methylation (modification), the site of restriction (cleavage) is non-specific, and is often some thousands of base-pairs away from the primary recognition target.

A new *hsd* system, (*hsdD*), related to K and B, has been identified among natural isolates of *E. coli* (see Murray *et al.*, 1982). Evidence for the D system was first obtained from hybridization experiments in which a DNA fragment from within the *hsd* genes of *E. coli* K12 was used as a probe for homologous sequences in chromosomal DNAs. Biological tests have confirmed that the *hsdSD* gene confers a novel specificity (J. A. Gough & N. E. Murray, unpublished results). Both genetic (van Pel & Colson, 1974; Bullas & Colson, 1975) and molecular evidence (Murray *et al.*, 1982) have shown that the K and B systems of *E. coli* are also related to the SB and SP systems of *Salmonella*.

Molecular experiments based on DNA sequence homology and on immunological cross-reactivity suggest a high degree of conservation of the structure of the *hsdM* and *hsdR* polypeptides in the *EcoK* and *EcoB* systems (Murray *et al.*, 1982). Furthermore, the hybridization of a DNA fragment from within the *hsdS* gene of *E. coli* K12 with chromosomal DNAs from *E. coli* K12, B and D, or even *Salmonella typhimurium* and *Salmonella potsdam*, shows that some sequence within the specificity genes is conserved (Murray *et al.*, 1982).

These related restriction enzymes constitute a family of structurally similar proteins that bind to different, but specific DNA sequences. The present diverse family is presumed to have a common ancestor, hence comparative nucleotide

TABLE I
Bacterial strains used in this work

Strain	Relevant genotype	Reference
C600	<i>hsdK</i> ⁺	Appleyard (1954)
NM477	<i>hsd</i> ∇5 derivative of C600	This work (see Fig. 1)
C(P2)	P2 lysogen of <i>E. coli</i> C	Bertani & Weigle (1953)
HB101	<i>recA</i> ⁻ <i>hsdS</i> ⁻	Boyer & Roulland-Dussoix (1969)
71-18	(<i>lac-pro</i>)∇ F' <i>lacZ</i> M15 <i>lacI</i> ^a	Messing <i>et al.</i> (1977)
NM522	<i>hsd</i> ∇5 derivative of 71-18	
WA960	<i>hsdB</i> ⁺	Wood (1966)
629†		
E166†		

† These strains were isolated by K. Cartwright in the Western Infirmary, Edinburgh; their DNA was a generous gift from K. Kaiser. 629 is shown to have the same specificity as *E. coli* B, while E166 has a novel specificity system.

sequence analysis of representative members of the family that dictate the different specificities. In these analyses, the specificity gene is transferred to a plasmid (Boyer, 1980) and, since the specificity polypeptide is the *hsdM* polypeptide, and possibly other polypeptides, regions involved in protein synthesis, whereas those involved in DNA target recognition.

We have therefore determined the specificity genes isolated from *E. coli* K12 (Bertani, 1953), B (Arber & Dussoix, 1962) and a strain having the B specificity.

2. Materials and Methods

(a) Phages, plasmids and DNA

*λ**hsdK* phages and plasmids pBg3, fragments of the *hsd* genes of *E. coli* K12 (Boyer, 1980).

The *hsd* genes of strains E166 and C600 were cloned into the *EcoRI* fragments of the vector pMB1. The recombinants can be monitored by the presence of the *lacZ* fragments (EMBL4) is a replacement of the *lacZ* phenotype of the recombinant. P2 (strain C(P2) in Table 1).

Plasmid vectors were either pBR322 (Messing & Vieira, 1982).

The M13 vector used for DNA sequencing was mp8 and mp9 (Messing & Vieira, 1982).

All *λ**hsd* phages, other than *λ**hsdK*, were derived from an *E. coli* K derivative in which a derivative of the *hsdS* (∇5) was transferred from C600 (see Fig. 1). Plasmids were propagated either in *E. coli* K12 derivative of the *lacZ* strain 71-18 (see Fig. 1) or in recombinants recognized, in either 71-18 or C600 (*E. coli* K-∇5 or *E. coli* C) were used to clone *hsd* genes and homologous chromosomes. The results are documented in Table 1.

(b) Enzymes

Restriction endonucleases were purchased from Pharmacia Research Laboratories, or prepared by the method of E. Remaut; DNA polymerase I (Klenow fragment) from Boehringer Mannheim GmbH and dideoxynucleotides (ddNTPs) from Pharmacia Inc.; the M13 sequencing primer was synthesized in-house; ³²P]dNTPs (400 Ci/mmol) from the Radiochemical Centre.

(c) Purification of DNA

The DNA of phage *λ* was prepared by the method of Clewell & Helinski (1968) and purified from cleared bacteriophage lysates by CsCl/ethidium bromide (Clewell & Helinski, 1968).

al for restriction (Arber & Linn, 1969; Hubacek & Colson, 1969; Hubacek & Murray, 1982). The D system is composed of three non-identical subunits, with molecular weights of 35,000, 60,000 and 50,000 (Eskin & Murray, 1982). The D system is specific to their activities as restriction endonucleases and DNA-dependent methylases and DNA-dependent methyltransferases. The D system is specific to their activities as restriction endonucleases and DNA-dependent methyltransferases. The D system is specific to their activities as restriction endonucleases and DNA-dependent methyltransferases.

B, has been identified among *E. coli* strains. Evidence for the D system was first reported in which a DNA fragment from the D system was used as a probe for homologous sequences in other *E. coli* strains. It was confirmed that the *hsdSD* gene is present in strains of *E. coli* K12 (Murray, unpublished results), *E. coli* B (Arber & Colson, 1975) and molecular clones of the K and B systems of *E. coli* are available.

sequence homology and on the degree of conservation of the D system in the *EcoK* and *EcoB* systems. The D system is a DNA fragment from the D system of chromosomal DNAs from *E. coli* K12, *B. subtilis* and *Monella potsdam*, shows that some D system sequences are conserved (Murray *et al.*, 1982).

a family of structurally similar D system sequences. The present diverse D system sequences, hence comparative nucleotide

his work

Reference

- Appleyard (1954)
This work (see Fig. 1)
Bertani & Weigle (1953)
Boyer & Roulland-Dussoix (1969)
Messing *et al.* (1977)
Wood (1966)

General Infirmary, Edinburgh; their DNA was of the same specificity as *E. coli* B, while E166

sequence analysis of representative specificity genes could identify the changes that dictate the different specificities. Fortunately for subsequent physical analyses, the specificity gene is the smallest of the *hsd* genes (Sain & Murray, 1980) and, since the specificity polypeptide of each system must interact with the *hsdM* polypeptide, and possibly also the *hsdR* polypeptide, of heterologous systems, regions involved in protein-protein interactions must be conserved, whereas those involved in DNA target recognition should differ.

We have therefore determined and compared the DNA sequences of the *hsdS* genes isolated from *E. coli* K12 (Bertani & Weigle, 1953), D (Murray *et al.*, 1982), B (Arber & Dussoix, 1962) and a second, very recent, natural isolate of *E. coli* having the B specificity.

2. Materials and Methods

(a) Phages, plasmids and bacterial strains

*λ**hsdK* phages and plasmids pBg3, pBg6 and pRH3, derivatives of pBR322 carrying fragments of the *hsd* genes of *E. coli* K12, (see Fig. 1) were described by Sain & Murray (1980).

The *hsd* genes of strains E166 and 629 (see Table 1) were cloned using phage *λ* vectors. The vector used to clone *EcoRI* fragments (NM1039) is a *lacZ* replacement vector in which the recombinants can be monitored as *lacZ*⁻ derivatives, that used to clone *BamHI* fragments (EMBL4) is a replacement vector (Frischauf *et al.*, unpublished results) in which the *Spi*⁻ phenotype of the recombinants allows their selection on *E. coli* lysogenic for phage P2 (strain C(P2) in Table 1).

Plasmid vectors were either pBR322 (Bolivar *et al.*, 1977) or, more recently, pUC9 (Messing & Vieira, 1982).

The M13 vector used for DNA sequencing was initially mp7, but more generally the vectors were mp8 and mp9 (Messing & Vieira, 1982).

All *λ**hsd* phages, other than *λ**hsdK*, were grown in either an *E. coli* C strain or in NM477, an *E. coli* K derivative in which a deletion extending from within *hsdR* to beyond *hsdM* and *hsdS* (V5) was transferred from a *λ**hsd*-V5 phage to the bacterial chromosome (see Fig. 1). Plasmids were propagated either in the *recA* host HB101 or in NM522, an *hsd*-V5 derivative of the *lacZ* strain 71-18 (see Table 1). Similarly, M13 clones were recovered, and recombinants recognized, in either 71-18 or its *hsd*-V5 derivative. Hosts lacking *hsd* genes (*E. coli* K-V5 or *E. coli* C) were used to eliminate the possibility of recombination between cloned *hsd* genes and homologous chromosomal sequences. All bacterial strains used are documented in Table 1.

(b) Enzymes and chemicals

Restriction endonucleases were purchased from either New England Biolabs or Bethesda Research Laboratories, or prepared by H. Cambier; phage T4 DNA ligase was a gift from E. Remaut; DNA polymerase I (Klenow fragment) and nucleoside triphosphates were from Boehringer Mannheim GmbH and dideoxynucleoside triphosphates from P-L Biochemicals Inc.; the M13 sequencing primer was from Bethesda Research Laboratories and [³²P]dNTPs (400 Ci/mmol) from the Radiochemical Centre, Amersham.

(c) Purification and isolation of DNA

The DNA of phage *λ* was prepared as described by Wilson *et al.* (1977). Plasmid DNA was purified from cleared bacterial lysates by centrifugation to equilibrium in CsCl/ethidium bromide (Clewell & Helinski, 1969), although mini-scale preparations (Gough

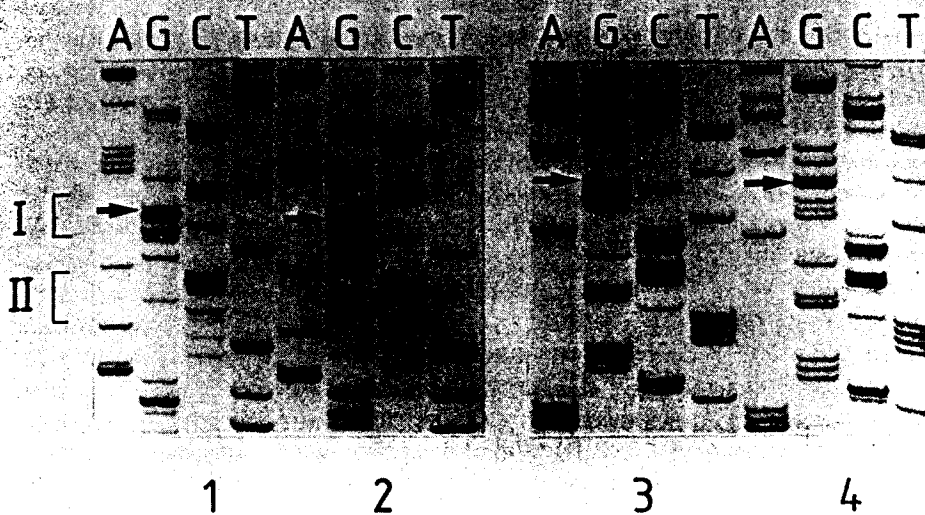


FIG. 1. Sequencing gels illustrating the resolution of strand-specific discrepancies. Experiments 1 and 2 use a clone of complementary sequence to that used in experiments 3 and 4. Two discrepancies, in close proximity, are detected. The ddNTP used in each reaction is indicated at the top of the gel. Experiments 2 and 4 differ from 1 and 3 in the use of dITP in place of dGTP. The brackets labelled I and II identify the corresponding regions of the complementary strands, and arrows draw attention to the relevant changes in the G track. The sequence for region I read from the gel of experiment 1 is 5'-G-C-G-G-C-C, whereas that deduced for this same region in experiment 3, where the opposite strand was used as template, is 5'-G-G-C-C-G-G-C-C (G-G-C-C-C-G-C-C when reading this gel). Similarly, region II: in experiment 1, the sequence read is 5'-A-C-C-G-C-C-C, whereas the opposite strand (region II, in experiment 3) in fact gave the sequence 5'-A-C-G-C-C-C (G-G-G-C-G-T when read from this gel). The "missing" G residues (indicated by arrows) were readily apparent when dITP was used in place of dGTP (experiments 2 and 4). The regions sequenced in this experiment are identified in Fig. 5.

& Murray, 1982) were made for analytical purposes. Fragments of DNA up to 1.4 kb† in length were eluted from polyacrylamide gels (Maxam & Gilbert, 1977) and larger fragments from agarose gels (Dretzen *et al.*, 1981).

DNA from single plaques of M13 was prepared by a modification of the method of Sanger *et al.* (1977). A single plaque was added to 1 to 2 ml of an early exponential culture of bacteria (either 71-18 or NM522), and the infected cells incubated at 37°C with aeration for 4 to 5 h. The lysate was clarified by spinning for 5 min in an Eppendorf centrifuge. The DNA was either extracted immediately or the phage were stored at 4°C for as long as several months. If stored, the supernatant was reclarified before proceeding. Phage were precipitated with polyethylene glycol P6000 (~4% (w/v) final concentration) and NaCl (~0.5 M), recovered by centrifugation and, after careful removal of the supernatant, resuspended in TE buffer (10 mM-Tris·HCl (pH 8.0), 1 mM-EDTA). The DNA was extracted with an equal volume of phenol saturated with buffer, recovered by centrifugation, precipitated with 1 M-sodium perchlorate, 30% (v/v) isopropanol, and washed with 70% (v/v) ethanol before suspension in 30 to 50 µl of TE buffer.

† Abbreviations used: kb, 10³ bases; bp, base-pairs.

SEQUENCE DIVERSITY

(d) Restriction

These reactions were as described by Murray

(e) DNA sequencing

The sequences of some purified DNA fragments were determined by the Maxam & Gilbert (1980) procedure of Maxam & Gilbert (1980). The appropriate [³²P]dNTP and the Klenow fragment (Klenow, 1975) were used. More generally, DNA fragments subcloned into a suitable vector (Sanger *et al.*, 1977) were sequenced by the chain-termination method of Sanger *et al.* (1977). The sequencing primer was a pair universal primer derived from pSP14 (Sanger & Coulter, 1975), 0.2 mm thick, thermostatically controlled (Ansorge, 1981).

"Compression" of bands, particularly in the G track, was observed in some experiments. Such compressions have been reported to occur in newly synthesized strand, which lead to ambiguities in the sequence (Kramer, 1979). Figure 1 illustrates the use of dITP in place of dGTP to resolve such ambiguities.

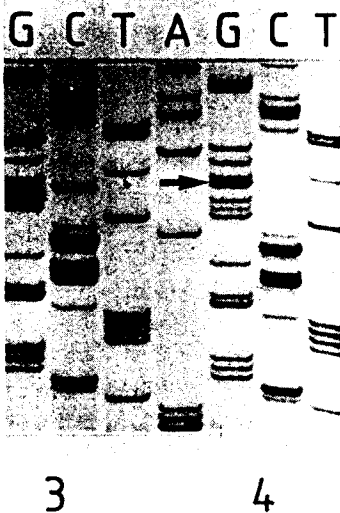
(f) Detection of overlapping

Matrices of spots of M13 phage supernatant were prepared on nitrocellulose filters (Schleicher and Schüll, 0.45 µm). The matrices were washed with 1.5 M-NaCl, 0.5 M-NaOH; 2 changes of 3 M-NaCl, 0.15 M-NaCl, 0.015 M-sodium citrate), blotted onto a membrane (Schleicher and Schüll, 1975). Radioactively labelled probes were made by labelling DNA as template and the primer fragment with [³²P]dATP. Prior to hybridization, the nitrocellulose filters were treated with 0.2% (w/v) bovine serum albumin, 0.2% (w/v) polyvinylpyrrolidone, 0.2% (w/v) bovine serum albumin, 0.2% (w/v) DNA/ml at 65°C for several hours. The filters were then incubated overnight at 65°C (4 × 10⁶ to 8 × 10⁶ c.p.m./ml). After hybridization, filters were washed extensively with 0.5 M-sulphate at 65°C. Autoradiography was carried out on intensifying screens.

3. Results

(a) Cloning and identification

The entire *hsd* region of *E. coli* K12 (Sain & Murray, 1980) was cloned (Fig. 2) and as the two non-overlapping fragments (Fig. 2: Sain & Murray, 1980). The positions of the restriction sites for some restriction enzymes were identified. The endonucleases *Hind*III and *Bgl*II were used. Each include part of the *hsdS* gene. The *hsd* genes from other strains of *E. coli* were also cloned. The DNAs of two recent natural isolates as well as *E. coli* B, have been shown to be related to those of *E. coli* K12 (Murray *et al.*, 1982), and the DNA from strain E166 was digested with *Bam*HI, the resulting fragments



1-specific discrepancies. Experiments 1 and 2 show discrepancies. Experiment 3 shows a discrepancy indicated at the top of the gel in place of dGTP. The brackets labelled 1 and 2 indicate complementary strands. Arrows draw attention to discrepancies in the sequence read from the gel of experiment 1 is 5' G C T A G C T 3' in experiment 3, where the opposite strand is 3' C G A C T G C 5' (when reading this gel). Similarly, for experiment 4, the opposite strand is 3' C G C C C C C 5' (G-G-G-C-G-T when read from the gel). Discrepancies were readily apparent when dTTP was used in this experiment are identified in

Fragments of DNA up to 1.4 kb† in length (Murray & Gilbert, 1977) and larger fragments

modification of the method of Sanger (1975) using an early exponential culture of *E. coli* K12 incubated at 37°C with aeration for 18 h in an Eppendorf centrifuge. The phages were stored at 4°C for as long as 6 months before proceeding. Phage were pelleted by centrifugation (1000g, 10 min, 4°C) and NaCl removed (w/v final concentration) and NaCl removed of the supernatant (1000g, 10 min, 4°C). The DNA was extracted with buffer, recovered by 20% isopropanol, and washed with 70% ethanol.

(d) Restriction and ligation of DNA

These reactions were as described by Murray *et al.* (1982).

(e) DNA sequence determination

The sequences of some purified DNA fragments, and some plasmids linearized by digestion with an appropriate restriction enzyme, were determined by the chemical procedures of Maxam & Gilbert (1980). The fragments were 3'-terminally labelled using the appropriate [α - 32 P]dNTP and the Klenow fragment of DNA polymerase I.

More generally, DNA fragments subcloned in M13 vectors (Sanger *et al.*, 1980; Messing *et al.*, 1981) were sequenced by the chain-termination method using dideoxynucleoside triphosphates (Sanger *et al.*, 1977). The sequencing reactions were primed with the 26 base-pair universal primer derived from pSP14 (Anderson *et al.*, 1980), and were resolved on 0.2 mm thick, thermostatically controlled, 8% (w/v) polyacrylamide gels (Garoff & Anson, 1981).

"Compression" of bands, particularly in G+C-rich regions, can lead to strand-specific ambiguities. Such compressions have been attributed to secondary structure within the newly synthesized strand, which leads to its anomalous migration through a polyacrylamide gel. dTTP in place of dGTP can resolve such discrepancies: I·C base-pairing is weaker than G·C, so that secondary structure formation is less likely (Mills & Kramer, 1979). Figure 1 illustrates the use of this method.

(f) Detection of overlapping cloned sequences by hybridization

Matrices of spots of M13 phage supernatants (3 to 5 μ l) were applied to nitrocellulose filters (Schleicher and Schüll, 0.45 μ m). The filters were treated successively in 0.5 M-NaOH; 1.5 M-NaCl, 0.5 M-NaOH; 2 changes of 3 M-NaCl, 1 M-Tris·HCl (pH 7.5); 2 \times SSC (SSC is 0.15 M-NaCl, 0.015 M-sodium citrate), blotted dry and baked for ~2 h at 80°C under vacuum.

Radioactively labelled probes were made from appropriate M13 clones using the phage DNA as template and the primer fragment derived from pHM235 (Hu & Messing, 1982).

Prior to hybridization, the nitrocellulose filters were soaked in 2 \times SSC, 0.1% (w/v) sodium dodecyl sulphate, 0.2% (w/v) Ficoll (Sigma Chemical Company), 0.2% (w/v) polyvinylpyrrolidone, 0.2% (w/v) bovine serum albumin, 50 μ g of denatured salmon sperm DNA/ml at 65°C for several hours. The filters were hybridized in the same buffer by incubation overnight at 65°C (4×10^6 to 8×10^6 cts/min total; 10^6 to 2×10^6 cts/min per ml). After hybridization, filters were washed extensively in 2 \times SSC, 0.1% (w/v) sodium dodecyl sulphate at 65°C. Autoradiography was generally for 2 to 3 h at -70°C between 2 intensifying screens.

3. Results and Discussion

(a) Cloning and identification of the *hsd* genes of *E. coli* strains

The entire *hsd* region of *E. coli* K12 has been cloned in a λ phage (λ *hsdRMS* in Fig. 2) and as the two non-overlapping segments in plasmids pBg3 and pBg6 (see Fig. 2; Sain & Murray, 1980). The positions of the *hsd* genes relative to the targets for some restriction enzymes were identified. The *hsdS* gene includes targets for the endonucleases *Hind*III and *Bgl*II, and plasmids pBg3 and pBg6 (Fig. 2(a)) each include part of the *hsdS* gene. These plasmids were used as probes to identify the *hsd* genes from other strains of *E. coli*.

The DNAs of two recent natural isolates of *E. coli*, E166 and 629 (see Table 1), as well as *E. coli* B, have been shown to share homology with the *hsd* genes of *E. coli* K12 (Murray *et al.*, 1982), and the *hsd* genes from these strains were cloned. DNA from strain E166 was digested with *Eco*RI, that from 629 with *Eco*RI, or with *Bam*HI, the resulting fragments of DNA were cloned using phage λ vectors and

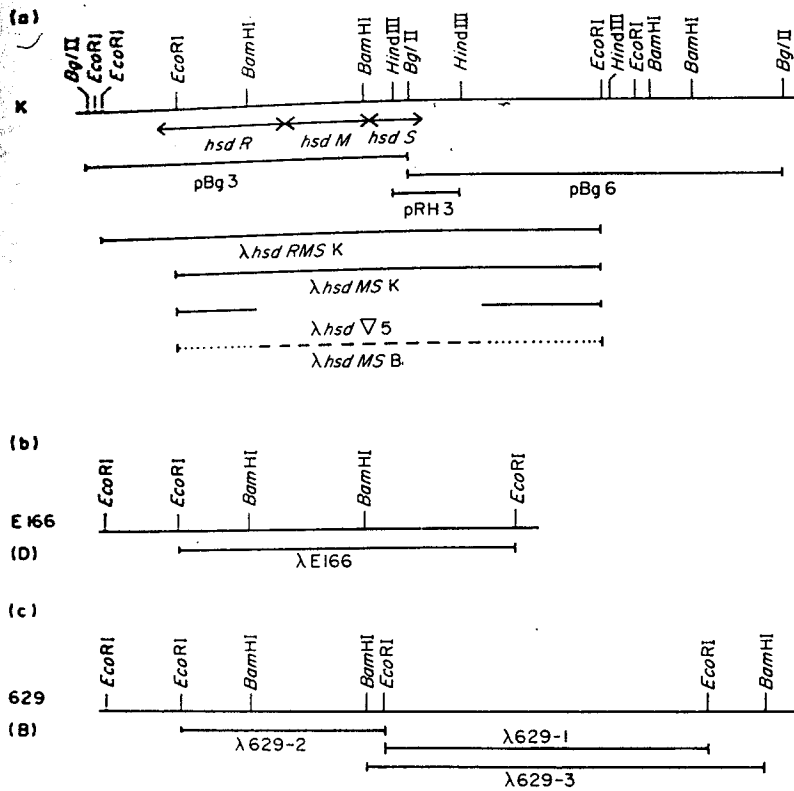


FIG. 2. The *hsd* regions of *E. coli* K12, E166 (D) and 629 (B). (a) The positions of the *hsd* K genes *M* and *S* are indicated relative to the sites for some restriction enzymes. Below this map, fragments of the *hsd* region within plasmid and phage derivatives are identified. *λhsd* *M S B* was made from *λhsd* *∇5*, by rescue of the B region from WA960 (see Table 1). The genetic exchanges could have occurred anywhere within the dotted region, and the broken line identifies DNA that must be derived from the *hsd*B region of WA960. (b) The restriction targets within the *hsd*D region and the fragment of chromosomal DNA in the *λhsd* E166 phage. (c) The restriction targets within the *hsd* region of strain 629 and the fragments of chromosomal DNA within the derivative *λhsd* 629 phages.

λhsd phages were detected by plaque hybridization (Benton & Davis, 1977) with pBq3 and pBq6 (Fig. 2(a)) as probes. The *λhsd* phages derived from E166 contain a single 9 kb *Eco*RI fragment that shares homology with both pBq3 and pBq6 (Fig. 2(b)). Genetic tests (J. A. Gough & N. E. Murray, unpublished results) have shown that the *hsd*S gene within this fragment confers a novel specificity, which was designated D. Strain 629 seems likely to confer the B specificity, since the modification imposed by this strain provides protection against restriction by the classical B system (J. A. Gough & N. E. Murray, unpublished results). Two types of *λhsd* clones were made from *Eco*RI digests of strain 629 DNA (see Fig. 2(c)): first, *λhsd* 629-1 (containing an 8.5 kb fragment), shares homology with both pBq3 and pBq6, while the other, *λhsd* 629-2 (containing a 5.5 kb fragment), shares extensive homology with pBq3. An overlapping *Bam*HI fragment was also cloned

hsd 629-3 in Fig. 2(c)), and sequence analysis of two *Eco*RI fragments were contiguous in the map. The *E. coli* B specificity genes (*hsd*B) were identified in a sensitive strain whose *hsd* genes derive from strain WA960 (Wood, 1966). These *hsd* genes were cloned into a phage (see Fig. 2(a)). First, a derivative of *λhsd* was made in which deletion removed the *hsd*S gene together with the *hsd*M gene (Fig. 2(a)). A *λhsd*B phage was then made by packaging the *hsd*S gene from the chromosome of WA960 into the resulting *λhsd*B phage (NM1130) is, the *hsd* regions of *E. coli* K12 (designated B) include a 2 kb *Eco*RI fragment, which

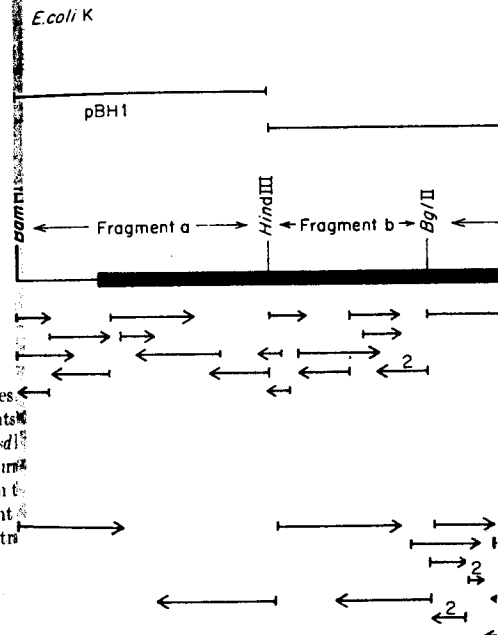
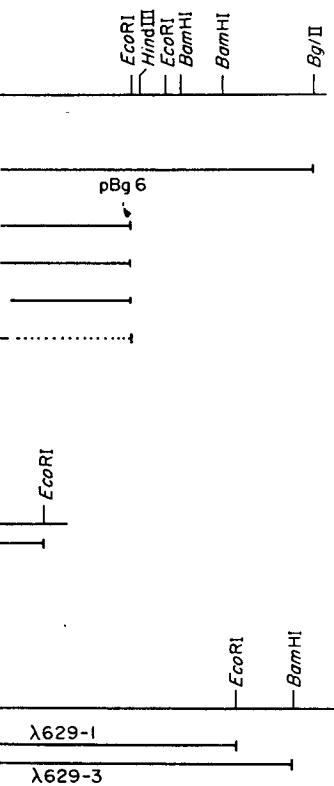


FIG. 3. Sequencing strategy for the *hsd* region of *E. coli* K12. The map identifies the *hsd*S gene. The plasmids used as probes are pBq3 and pBq6. Below the map, each arrow indicates a segment of DNA to be sequenced by the chemical method (Maxam & Gilbert, 1980). The order of sequencing is identified by the direction of the arrow. The first sequencing was done with *Hind*III and *Fnu*4HI, and fragment b by *Hpa*II. The second sequencing was done with *Hpa*II, *Hha*I, and *Hae*III. The substrates for the Maxam & Gilbert method. Fragments a and b were sequenced with the Maxam & Gilbert method. Fragment c, the 1.4 kb *Bgl*III-*Hpa*I fragment, was sequenced with each of the following endonucleases: *Hpa*II, *Hha*I, and *Hae*III. The contiguity of fragments a and b, and the overlapping fragments.



(B). (a) The positions of the *hsd* K genes and restriction enzymes. Below this map, fragments identified. *λ**hsd* M S B was made from *λ**hsd* B. The genetic exchanges could have occurred. This map identifies DNA that must be derived from the *hsd* region and the fragment targets within the *hsd* region of strain derivative *λ**hsd* 629 phages.

ization (Benton & Davis, 1977) with *λ* phages derived from E166 containing homology with both pBg3 and pBg6. Murray, unpublished results) have not confer a novel specificity, which confer the B specificity, since the protection against restriction by the phage (Murray, unpublished results). Two types of strain 629 DNA (see Fig. 2(c)); the phage (Murray, unpublished results), shares homology with both pBg3 (containing a 5.5 kb fragment), sharing a *Bam*HI fragment was also cloned

*λ**hsd* 629-3 in Fig. 2(c)), and sequence analysis of this fragment confirmed that the two *Eco*RI fragments were contiguous in the *E. coli* chromosome.

The *E. coli* B specificity genes (*hsd*B) were also cloned from strain WA960, a *λ*-sensitive strain whose *hsd* genes derive from the original, *λ*-resistant *E. coli* B (Wood, 1966). These *hsd* genes were cloned by manipulation *in vivo* of a *λ**hsd*K phage (see Fig. 2(a)). First, a derivative of *λ**hsd*K was isolated in which a large deletion removed the *hsd*S gene together with some flanking sequences (*λ**hsd*K-∇5 in Fig. 2(a)). A *λ**hsd*B phage was then made by homology-dependent rescue of the *hsd*S gene from the chromosome of WA960. The provenance of the *hsd*S gene in the resulting *λ**hsd*B phage (NM1130) is, therefore, the classical *E. coli* B strain.

The *hsd* regions of *E. coli* K12 (designated K), E166 (D), 629 and B (see Fig. 2) all include a 2 kb *Eco*RI fragment, which for K has been shown to contain the

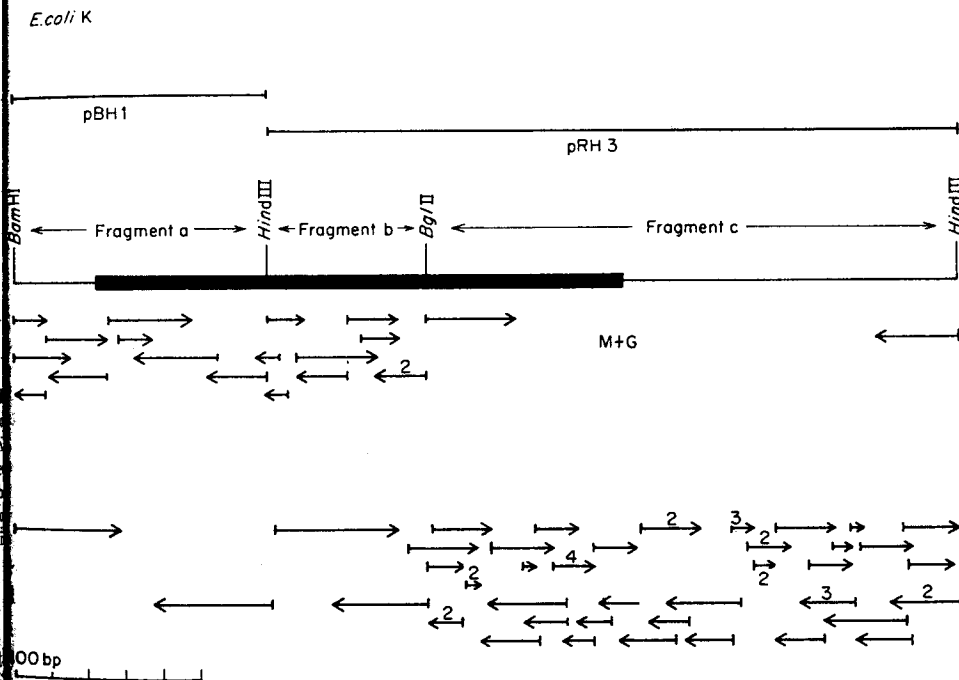


FIG. 3. Sequencing strategy for the *hsd* region of *E. coli* K12. The heavy line within the restriction map identifies the *hsd*S gene. The plasmids used as a source of DNA fragments are indicated above the map. Below the map, each arrow indicates a segment of DNA whose sequence was determined either by the chemical method (Maxam & Gilbert, 1980), identified as M+G, or by the chain termination method of Sanger *et al.* (1977) with fragments cloned in M13 vectors as templates. The strand sequenced is identified by the direction of the arrow and the number of times a fragment was sequenced, when more than one, is indicated above the arrow. Fragment a was digested with *Acc*I, *Hin*FI and *Fnu*4HI, and fragment b by *Hpa*II, *Hha*I and *Fnu*4HI to provide purified fragments as substrates for the Maxam & Gilbert method. Fragments a and b were also cloned using M13 vectors mp8 and mp9, and sequence was determined from the *Bam*HI, *Hind*III and *Bgl*II targets by the chain termination method. Fragment c, the 1.4 kb *Bgl*II-*Hind*III fragment, was subcloned in mp7 following digestion with each of the following endonucleases; *Bst*NI, *Dde*I, *Hin*FI, *Hpa*II, *Alu*I, *Sau*3AI, *Taq*I and *Hae*III. The contiguity of fragments a and b, and b and c was confirmed by the sequence analysis of overlapping fragments.

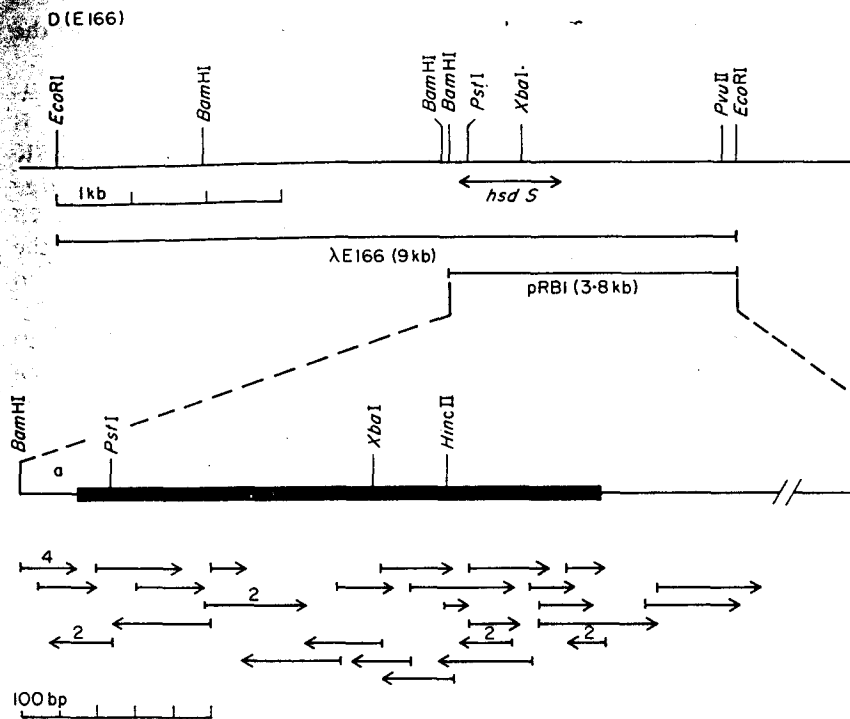


FIG. 4. The sequencing strategy for the *hsd* region of E166 (D). λ *hsd* E166 and its derivative plasmid pRB1 are indicated below the restriction map of the *hsdD* region. Within the enlarged map, the *hsdS* gene is identified by the heavy line. Sequence determination was initiated from fragment a and then proceeded by a series of overlapping fragments. (Clones containing overlapping sequences were detected by hybridization (Hu & Messing, 1982). Subclones using mp8 as vector were generated from the *Bam*HI-*Pvu*II fragment of pRB1, after digestion with each of the following enzymes: *Rsa*I, *Taq*I, *Hae*III, *Sau*3AI, *Dde*I, *Bst*XI and *Alu*I. Additional sequence from the *Pst*I, *Xba*I and *Hinc*II sites was obtained after cloning the *Pst*I-*Xba*I and *Xba*I-*Hinc*II fragments.

beginning of the *hsdR* gene, and a 3.1 kb *Bam*HI fragment, which for *hsdK* spans the 3' end of *hsdR* and most of *hsdM* (Sain & Murray, 1980). These conserved fragments were used to align the *hsdR* and *M* genes of the four strains. If the gene order is the same for each specificity system, the *hsdS* gene of strains E166 and D, like that of K, is expected to lie to the right of *hsdR* and *hsdM*; that is, to the right of the 3.1 kb *Bam*HI fragment within a region showing divergence in the distribution of *Hind*III and *Eco*RI sites (Fig. 2). Fragments of DNA derived from various λ *hsd* phages, and presumed to include part, or all, of an *hsdS* gene, were subcloned in plasmids (see Figs 3, 4 and 5).

(b) Sequence determination

(i) *hsdK*

The *hsdS* gene includes targets for the endonucleases *Hind*III and *Bgl*II, and is flanked by *Bam*HI and *Hind*III targets (see Fig. 2(a)). The entire sequence

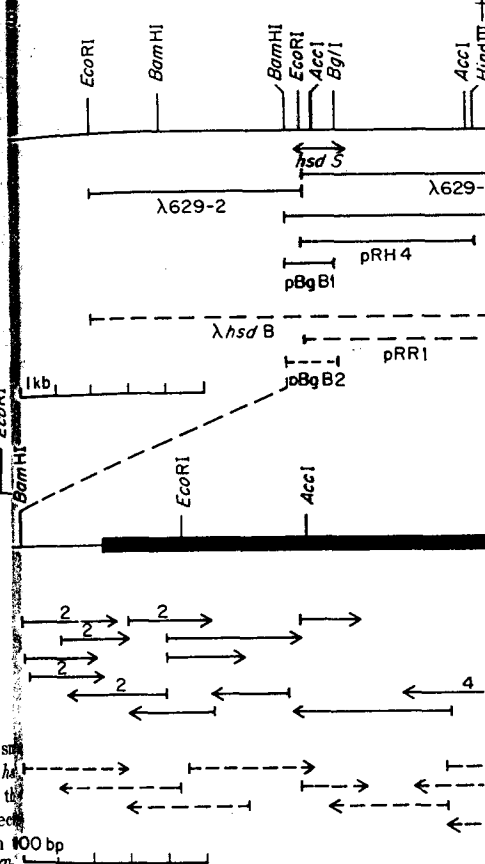
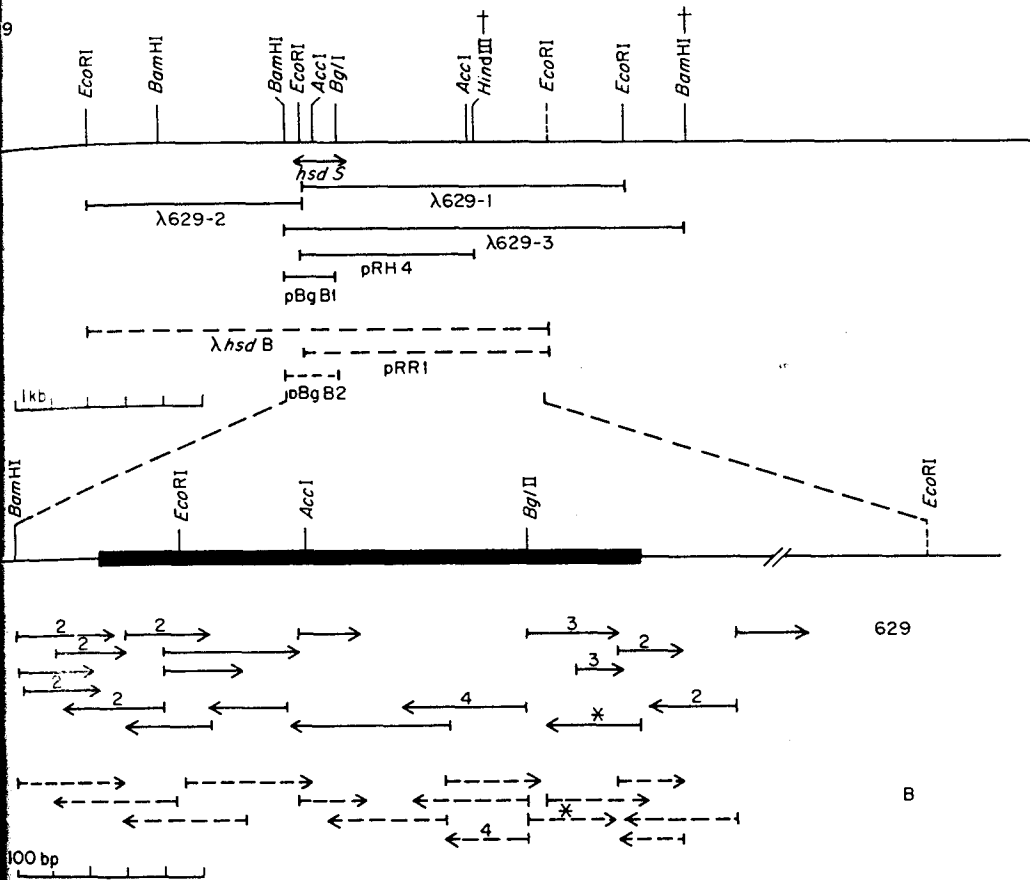
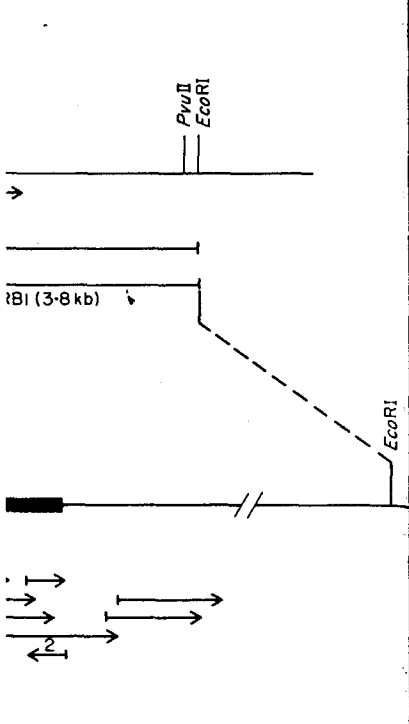


FIG. 5. The sequencing strategy for the *hsd* regions of strains 629 and *E. coli* B. The λ *hsd*629 and λ *hsd*629-1 phages and their derivative plasmids are indicated by broken lines. The *E. coli* B *hsdS* gene is indicated by a solid line. The λ *hsd*K phage (see Fig. 1) targets found in 629 not in K are indicated by †. Libraries of subclones in mp8 were made from the *hsdS* gene and from a purified fragment of pBgB1 including a variety of different endonucleases was used. Sequences were identified by hybridization with a probe. In addition, sequence was deduced from the *hsdS* gene and mp9. A similar strategy was used for B. t. subcloned in plasmids (see Fig. 1).

(2528 bp) of the three fragments defined in Fig. 3, particularly the legend, for detailed sequence of the *hsdS* gene was first determined by the degradation method of Maxam & Gilbert using purified DNA fragments as substrates.



166 (D). *λ*hsd E166 and its derivative plasmid region. Within the enlarged map, the *hsd* mutation was initiated from fragment a and the containing overlapping sequences were detected using mp8 as vector were generated from each of the following enzymes; *Rsa*I, *Taq*I, *Pst*I, *Xba*I and *Hinc*II sites were used to generate fragments.

BamHI fragment, which for *hsdK* spacer (Murray & Murray, 1980). These conserved genes of the four strains. If the gene, the *hsdS* gene of strains E166, 629 is to the right of *hsdR* and *hsdM*; that is, in a region showing divergence in the sequence (see Fig. 2). Fragments of DNA derived from the part, or all, of an *hsdS* gene, were

mination

endonucleases *Hind*III and *Bgl*II, and the entire sequence was determined (see Fig. 2(a)). The entire sequence

Fig. 5. The sequencing strategy for the *hsd* regions of 629 and B. One restriction map is given for the *hsd* regions of strains 629 and *E. coli* B. The *λ*hsd629 phages and their derivative plasmids are indicated below the restriction map. The *λ*hsdB phage (see Fig. 1 for the origin of this phage) and its derivative plasmids are indicated by broken lines. The *Eco*RI target, similarly marked by a broken line, is from the *λ*hsdK phage. † targets found in 629 not in B. Libraries of subclones in mp8 were made from the purified 4.4 kb *Eco*RI-*Hind*III fragment of pRH4 and from a purified fragment of pBgB1 including the 1.3 kb *Bam*HI-*Bgl*II sequence. In each case a variety of different endonucleases was used. Sequencing was initiated from a fragment located downstream of the 3' end of the *hsdS* gene and proceeded, as for *hsdD*, by a series of overlapping sequences. The starting clones were identified by hybridization with a subclone from the *hsdK* region as a probe. In addition, sequence was deduced from the *Bam*HI-*Bgl*II fragment after cloning in mp8 and mp9. A similar strategy was used for B, but depended on a 4 kb *Acc*I fragment derived from pRR1 and on the *Bam*HI-*Bgl*II fragment of pBgB2. The starting clones were identified by hybridization with a subclone from the 629 region. The asterisks indicate fragments sequenced using dITP in place of dGTP (see Fig. 1).

(2528 bp) of the three fragments defined by these targets was determined (see Fig. 3, particularly the legend, for details of the strategy). Most of the nucleotide sequence of the *hsdS* gene was first determined by the partial chemical degradation method of Maxam & Gilbert (1980), with either linearized plasmid or purified DNA fragments as substrates. This sequence was confirmed and extended

by the chain termination method with fragments of DNA subcloned in M13 phage vectors as templates (Sanger *et al.*, 1980; Messing *et al.*, 1981). M13 clones derived from the right-hand (1.4 kb) *Bgl*III-*Hind*III fragment (fragment c in Fig. 3) were sequenced at random (the so-called "shot-gun" method). Consequently, as shown in Figure 3, the nucleotide sequence of most of this region was determined several times.

(ii) *hsdD*

The specificity gene of E166, shown by genetic tests to be within the 9 kb *Eco*RI fragment of the λ *hsdD* phage, should, by analogy with K, be within a derivative 3.8 kb *Bam*HI-*Eco*RI fragment (see Fig. 2(b)). This fragment, amplified in plasmid pRBI (see Fig. 4), was digested with a variety of restriction endonucleases and the resultant fragments subcloned in M13 vectors for sequencing (see the legend to Fig. 4 for further details). Initially, the 5'-terminal *Bam*HI-*Pst*I fragment (fragment a in Fig. 4) was subcloned in both orientations and its entire sequence determined. This sequence shares extensive homology with a segment of the *hsdK* region, thereby aligning the DNA sequence of D with that of K. Labelled probes, made from the two *Bam*HI-*Pst*I subclones, were then used to detect by hybridization other subclones containing overlapping sequences (Hill & Messing, 1982). A contiguous sequence covering the entire *hsdS* gene was thus assembled from a series of overlapping fragments.

(iii) *hsdB*

The sequence of the *hsdS* genes was determined for DNA fragments derived from both strains 629 and WA960. The strategy (see Fig. 5) again depended on libraries of subclones in M13 phage vectors. Sequence determination was initiated from a defined point sharing homology with *hsdK*, and extended by the sequential detection and analysis of overlapping clones.

(c) DNA sequences, genes and their products

The nucleotide sequences for the K, D and B *hsdS* regions are shown in Figures 6, 7 and 8, respectively: that for *hsdB* is the same whether the ancestry of the DNA was *E. coli* 629 or the classical B strain.

A single, long, open reading frame within the DNA sequence of the K region identifies the *hsdS* gene. As required, this reading frame traverses both the *Hind*III (Fig. 6, position 680) and the *Bgl*III (Fig. 6, position 1086) targets, and the molecular weight of the polypeptide encoded by this gene is calculated to be 51,336, in agreement with the value of 50,000 predicted from the electrophoretic mobility of the *hsdS* gene product (Sain & Murray, 1980). The *Bam*HI target at the beginning of the K sequence is known to be within the *hsdM* gene (Sain & Murray, 1980). The only open reading frame extending rightwards from this site which also extends leftwards (data not shown), terminates within the initiation codon of the *hsdS* gene. Such an overlap of termination (UGA) and initiation codons (U-G-A-U-G) was first found for the B and A genes of the *trp* operon of *E. coli* (Platt & Yanofsky, 1975) and has been shown to allow translational coupling of gene expression (Oppenheim & Yanofsky, 1980). DNA sequences (e.g.

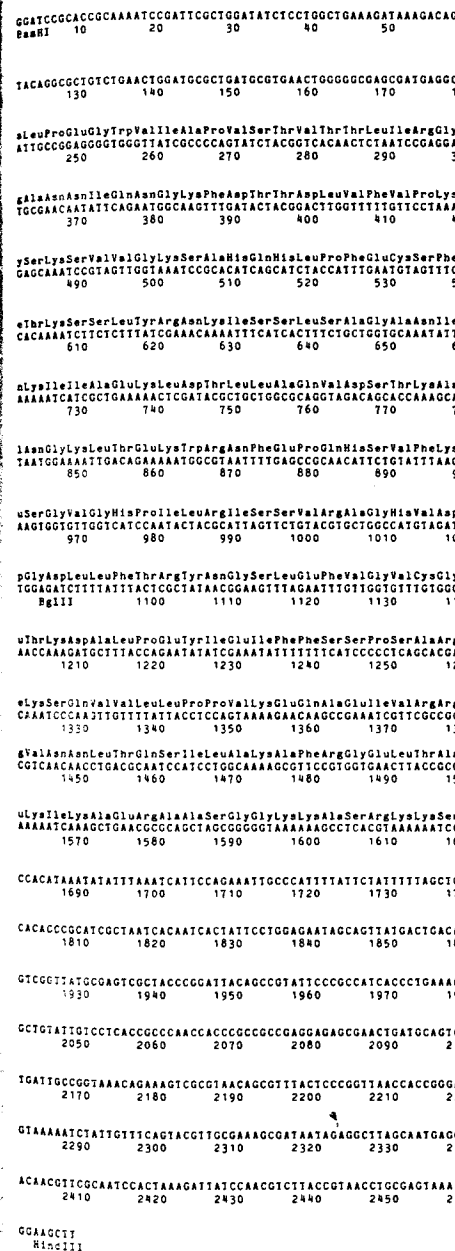


Fig. 6. DNA sequence of the *hsdK* region. The end. The amino acid sequence of the *hsdS* gene is shown above the nucleotide sequence. Hyphens have been omitted for clarity.

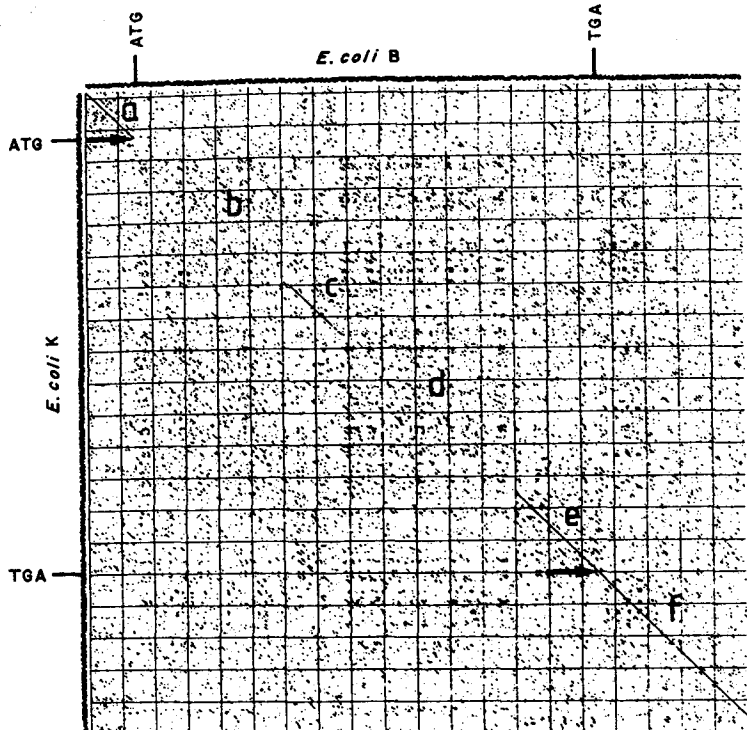


FIG. 9. Dot matrix comparisons between the *hsdK* and *B* regions. Each dot indicates conservation of at least 6 bp within a region of 9 bp. The letters K and B indicate the sequence plots on the corresponding axis. One square corresponds to 100 bp. The start and end of the coding region for *hsdS* are indicated. Additional sequence upstream of the *hsdS* gene, not included in the comparison of K with B, is highly conserved.

(d) Conservation and divergence of sequence in the *hsdS* genes

In order to assess variation in regions not essential to the maintenance of sequence specificity, the sequence of the *hsdB* region was determined from two different strains of *E. coli*. However, despite changes in the pattern of the neighbouring restriction targets, the sequence of the *hsdS* gene of *E. coli B* was identical to that from 629. Our comparisons are therefore confined to sequences representing three specificity systems, K, B and D.

The sequences of the three different *hsdS* genes show a surprising lack of homology. Indeed, there are essentially only two short conserved regions, the first of approximately 100 bp is in the middle of the gene, while the second is in the terminal 250 bp. These findings are illustrated by dot matrix comparisons (Hiebert *et al.*, 1980) between the sequences of the different *hsd* regions in which diagonal lines correspond to regions of homology (Figs 9, 10 and 11). For each pairwise combination, the two gaps in the diagonal line (b and d) indicate non-homologous regions of approximately 600 and 700 bp, respectively. Whilst

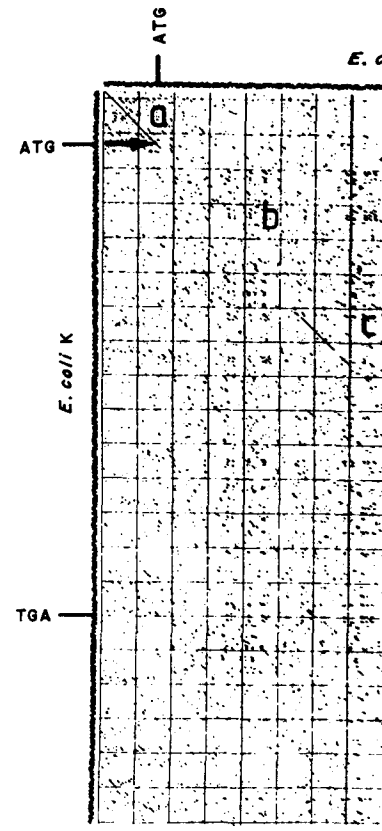
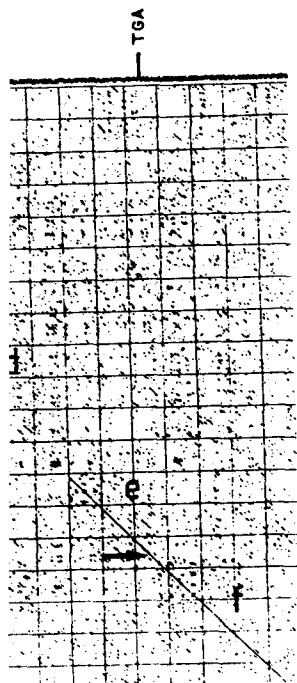


FIG. 10. Dot matrix comparisons between the *hsdK* and *B* regions.

and e represent the conserved sequences available for the *hsdM* gene is (10 or 11) as is the non-coding region downstream (10 and 11). No obvious direct repeats are present along non-diagonal lines. Recent sequence analyses suggest that the sequences share structural similarities (10 appear to be shared by the *hsdS* polypeptides analysed by Sauer and colleagues, probably arrangements of their subunits interacting to display rotational symmetry. In contrast to the *hsdSK* and *hsdSB* polypeptides are homologous consonant with this, the functional enzyme is responsible for target recognition, it is sufficient for the binding of DNA.



and D regions. See Fig. 9 for details.

Glu
 Asp Thr Leu Leu Ala Glu Val Asp Ser
 Phe Arg Glu Ala Val Leu
 Tyr Ala Asp Tyr Ile Glu Lys Glu Val
 Leu Ala Lys Ala Phe Arg Gly Glu Leu
 Glu Asn Ser Ala Ala Ala Leu Leu Glu
 Ser Arg Lys Lys Ser
 Phe

ions of the *hsdS* polypeptides. D, K and
 sequence and deviations from it are indicated
 35-663 for D (Fig. 7), 701-829 for K (Fig. 7),
 1216-1482 for D (Fig. 7), 1355-1621 for K (Fig. 7).
 asterisks indicate the termination of translation.

SEQUENCE DIVERSITY AMONG RELATED GENES

17

```

EcoB 5'-T-G-A-N-N-N-N-N-N-N-T-G-C-T
      A-C-T-N'-N'-N'-N'-N'-N'-A-C-G-A-5'
      :           :           :
EcoK 5' N-A-A-C-N-N-N-N-N-N-G-T-G-C-N
      N'-T-T-G-N'-N'-N'-N'-C-A-C-G-N'-5'
  
```

FIG. 13. The DNA sequences recognized by *EcoB* and *EcoK*. The recognition sequence for *EcoB* is interrupted by any sequence of 8 nucleotides, that for *EcoK* by any sequence of 6 nucleotides.

The regions of the *hsdS* gene that share homology are so highly conserved (Fig. 12) that a common ancestry is indicated, while the two interspersed stretches of non-homology could imply that diversification results from the insertion of modules of DNA sequences. An alternative interpretation is that the common ancestry of apparently non-homologous regions has simply been obscured by many changes. We prefer this second interpretation, in part, because particularly for *hsdK* and *hsdD*, we find occasional short stretches (e.g. 10 bp) whose sequence and position have both been conserved. However, the significance of these apparent conserved sequences is challenged, since analyses indicate that the variable regions of different *hsdS* genes are no more related to each other than they are to the DNA sequence of bacteriophage ϕ X174 (T. Bickle & J. Shepherd, personal communication). If diversification had arisen by the insertion of modules of DNA sequence, we might expect to find evidence for related DNA sequences in some *E. coli* strains. Experiments so far have not detected such sequences in the DNAs of various *E. coli* strains.

The roles of the various domains of the specificity polypeptides remain to be demonstrated. However, it seems reasonable to suggest that the highly conserved carboxy terminus has been maintained for interaction with the *hsdM* subunit. The less well-conserved central region could maintain a common configuration, as expected for the active site, while retaining enough variability to confer different target specificities. Divergence in the amino acid sequences flanking these regions would be allowed, providing that the appropriate tertiary configuration is retained. A quite different interpretation would correlate the two variable domains of the *hsdS* polypeptides with the two domains of the DNA recognition sequences (Fig. 13). This second possibility predicts that *hsdS* genes in which the variable regions are derived from different specificity genes could confer novel specificities. Some support for this notion might be implied from the observation that a new specificity system has been associated with recombination between the *S. typhimurium* (SB) and *S. potsdam* (SP) *hsd* genes (Bullas *et al.*, 1976).

This unexpected finding is supported by recent molecular studies. A probe made from the *hsdS* gene of *E. coli* K has been used to show relatedness of the *hsdS* genes of *E. coli* K with those of *S. typhimurium* (SB) and *S. potsdam* (SP) (Murray *et al.*, 1982). The DNA fragment from which this probe was made (fragment b in Fig. 5) is now known to include the short conserved sequence from the centre of the specificity gene. Heteroduplex analyses of the *hsdS* genes of these two *Salmonella* strains identify a short (~100 bp) region of homology flanked by two non-homologous regions (unpublished results from this laboratory). It seems probable that the same short region is conserved in *Salmonella* species as well as

in *E. coli*. This conserved region could provide sufficient homology for recombination between the *hdsS* genes of the SB and SP systems and, of course, between heterologous *hdsS* genes of *E. coli*. We are particularly indebted to Hans Lehrach for his constructive comments on the sequencing strategies, to Hans Lehrach for the two-dimensional comparisons of the DNA sequences, and John Shepherd for searching the DNA sequences for homology. We thank T. Bickle, G. Cesareni, N. Gough and K. Murray for their constructive comments on this manuscript.

REFERENCES

- Anderson, S., Gait, M. J., Mayol, L. & Young, I. G. (1980). *Nucl. Acids Res.* **8**, 1731-1742.
- Appleyard, R. K. (1954). *Genetics*, **39**, 440-459.
- Arber, W. & Dussoix, D. (1962). *J. Mol. Biol.* **5**, 18-36.
- Arber, W. & Linn, S. (1969). *Annu. Rev. Biochem.* **38**, 467-500.
- Benton, W. D. & Davis, R. W. (1977). *Science*, **196**, 180-182.
- Bertani, G. & Weigle, J. (1953). *J. Bacteriol.* **65**, 113-121.
- Bolivar, R., Rodriguez, R., Green, P., Betlach, M., Heynaker, M. & Boyer, H. (1977). *Gene* **2**, 85-113.
- Boyer, H. W. & Roulland-Dussoix, D. (1969). *J. Mol. Biol.* **41**, 459-472.
- Bullas, L. R. & Colson, C. (1975). *Mol. Gen. Genet.* **139**, 177-188.
- Bullas, L. R., Colson, C. & van Pel, A. (1976). *J. Gen. Microbiol.* **95**, 166-172.
- Clewell, D. & Helinski, D. R. (1969). *Proc. Nat. Acad. Sci., U.S.A.* **62**, 1159-1166.
- Dretzen, G., Bellard, M., Sassone-Corsi, P. & Chambon, P. (1981). *Anal. Biochem.* **112**, 295-298.
- Endlich, B. & Linn, S. (1981). In *The Enzymes* (Boyer, P. D., ed.), vol. 14, pp. 137-150. Academic Press, New York.
- Eskin, B. & Linn, S. (1972). *J. Biol. Chem.* **247**, 6183-6191.
- Garoff, H. & Ansoerge, W. (1981). *Anal. Biochem.* **115**, 450-457.
- Glover, S. W. & Colson, C. (1969). *Genet. Res.* **13**, 227-240.
- Gough, N. M. & Murray, K. (1982). *J. Mol. Biol.* **162**, 43-68.
- Hieter, P. A., Max, E. E., Seidman, J. G., Maizel, J. V. & Leder, P. (1980). *Cell*, **22**, 197-207.
- Hu, N.-T. & Messing, J. (1982). *Gene*, **17**, 271-277.
- Hubacek, J. & Glover, S. W. (1970). *J. Mol. Biol.* **50**, 111-127.
- Kan, K. C., Lautenberger, J. A., Edgell, M. H. & Hutchison, C. A. III (1979). *J. Mol. Biol.* **130**, 191-210.
- Lautenberger, J. A., Kan, N. C., Lackey, D., Linn, S., Edgell, M. H. & Hutchison, C. A. III (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 2271-2275.
- Maxam, A. M. & Gilbert, W. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 560-564.
- Maxam, A. M. & Gilbert, W. (1980). *Methods Enzymol.* **65**, 499-560.
- Meselson, M., Yuan, R. & Heywood, J. (1972). *Annu. Rev. Biochem.* **41**, 447-466.
- Messing, J. & Vieira, J. (1982). *Gene*, **19**, 269-276.
- Messing, J., Gronenborn, B., Muller-Hill, B. & Hofschneider, P. H. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 3642-3646.
- Messing, J., Crea, R. & Seeburg, P. H. (1981). *Nucl. Acids Res.* **9**, 309-321.
- Mills, D. R. & Kramer, F. R. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 2233-2235.
- Modrich, P. (1979). *Quart. Rev. Biophys.* **12**, 315-369.
- Murray, N. E., Gough, J. A., Suri, B. & Bickle, T. A. (1982). *EMBO J.* **1**, 535-539.
- Oppenheim, D. S. & Yanofsky, C. (1980). *Genetics*, **95**, 785-795.
- Platt, T. & Yanofsky, C. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 2399-2403.
- Ravetch, J. V., Horiuchi, K. & Zinder, N. D. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 2266-2270.
- Sain, B. & Murray, N. E. (1980). *Mol. Gen. Genet.* **188**, 1-10.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 5467-5471.
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Stent, G. B. (1978). *Nature (London)*, **278**, 447-451.
- Shine, I. & Dalgarno, L. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 1342-1346.
- Sommer, R. & Schaller, H. (1979). *Mol. Gen. Genet.* **188**, 1-10.
- van Pel, A. & Colson, C. (1974). *Mol. Gen. Genet.* **139**, 177-188.
- Wilson, G. G., Tanyashin, V. I. & Murray, N. E. (1980). *J. Mol. Biol.* **131**, 1-10.
- Wood, W. B. (1966). *J. Mol. Biol.* **16**, 118-131.
- Yuan, R. (1981). *Annu. Rev. Biochem.* **50**, 281-302.

provide sufficient homology for B and SP systems and, of course

for constant advice with the DN_A dimensional comparisons of the DN_A A sequences for homology. We thank their constructive comments on the

(1980). *Nucl. Acids Res.* **8**, 1731-1743

36.
3, 467-500.
180-182.
-121.

leynaker, M. & Boyer, H. (1977). *Gene*

J. Mol. Biol. **41**, 459-472.

39, 177-188.

J. Microbiol. **95**, 166-172.

Proc. Nat. Acad. Sci., U.S.A. **62**, 1159-1166.

on, P. (1981). *Anal. Biochem.* **112**, 298

oyer, P. D., ed.), vol. 14, pp. 137-156

3-6191.

5, 450-457.

77-240.

2, 43-68.

J. V. & Leder, P. (1980). *Cell*, **22**, 197

10, 111-127.

Hutchison, C. A. III (1979). *J. Mol. Biol.*

J., Edgell, M. H. & Hutchison, C. A. III

115, 275.

Proc. Nat. Acad. Sci., U.S.A. **74**, 560-564.

Mol. Biol. **65**, 499-560.

Annu. Rev. Biochem. **41**, 447-466.

ehneider, P. H. (1977). *Proc. Nat. Acad.*

Nucl. Acids Res. **9**, 309-321.

Proc. Nat. Acad. Sci., U.S.A. **76**, 2233-2235.

9.

A. (1982). *EMBO J.* **1**, 535-539.

95, 785-795.

Proc. Nat. Acad. Sci., U.S.A. **72**, 2399-2403.

(1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**

Sain, B. & Murray, N. E. (1980). *Mol. Gen. Genet.* **180**, 35-46.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463-5467.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980). *J. Mol. Biol.* **143**, 161-178.

Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M. & Paba, C. O. (1982). *Nature (London)*, **298**, 447-451.

Shine, I. & Dalgarno, L. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 1342-1346.

Sommer, R. & Schaller, H. (1979). *Mol. Gen. Genet.* **168**, 331-335.

van Pel, A. & Colson, C. (1974). *Mol. Gen. Genet.* **135**, 51-60.

Wilson, G. G., Tanyashin, V. I. & Murray, N. E. (1977). *Mol. Gen. Genet.* **156**, 203-214.

Wood, W. B. (1966). *J. Mol. Biol.* **16**, 118-133.

Yuan, R. (1981). *Annu. Rev. Biochem.* **50**, 285-315.

Edited by S. Brenner