J. Mol. Biol. (1987) **198**, 159–170

Organization and Sequence of the *hsd* Genes of *Escherichia coli* K-12

W. A. M. Loenen†, Anne S. Daniel, H. D. Braymer‡ and Noreen E. Murray§

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

(Received 22 May 1987)

The nucleotide sequence of the *hsdR* and *M* genes, together with that for *hsdS* comprises an 8400 base segment spanning the entire *hsd* region of *Escherichia coli* K-12. The three *hsd* genes are transcribed in the same direction, but from two promoters. *hsdR* and *hsdM* are separated by 492 base-pairs, whereas the termination codon of *hsdM* overlaps the initiation codon of *hsdS*. p_{res} precedes *hsdR*, and our data indicate a transcription termination signal in the interval between *hsdR* and p_{mod}, as expected if transcription of *hsdM* and *S* is dependent on p_{mod}. Transcription from p_{res} is not influenced by the products of the *hsdM* and *S* genes, and the mechanism whereby restriction is prevented when the *hsd* region is transferred to a modification-deficient cell remains to be elucidated.

A segment of the predicted amino acid sequence of the M polypeptide shares homology with a variety of adenine methylases and may identify part of the active site for methylation of specific adenine residues. The R polypeptide shows homology with a variety of ATPases, and pronounced regions of α-helical structure are predicted, one of which is amphipathic.

1. Introduction

Type I restriction and modification enzymes are complex, multifunctional systems (for a review, see Bickle, 1982). They require ATP, *S*-adenosyl methionine and Mg^{2+} as co-factors and, in addition to their endonucleolytic and methylase activities, are potent DNA-dependent ATPases. Type I enzymes contain three different subunits, R, M and S; the products of the *hsdR*, *M* and *S* genes. The S polypeptide dictates DNA sequences specificity, and together with M is sufficient for modification, while the products of all three genes are essential for both nuclease and ATPase activities (Boyer & Roulland-Dussoix, 1969; Glover & Colson, 1969; Hubacek & Glover, 1970). The type I systems of *Escherichia coli* K-12 and B are allelic and the subunits of *Eco*K and *Eco*B are interchangeable (Boyer & Roulland-Dussoix, 1969; Hubacek & Glover, 1970). The approximate molecular weights of the subunits are: R, 135,000; M, 62,000; and S, 50,000 (Eskin & Linn, 1972; Sain & Murray, 1980).

The three closely linked *hsd* genes are transcribed in the same direction from two promoters; the first upstream from *hsdR*, the second between *hsdR* and *hsdM* (Sain & Murray, 1980). The separate promoters could permit differential control of the levels of endonuclease and methylase.

The steps in the restriction pathway have been analysed in detail *in vitro* for *Eco*K (for a review, see Yuan, 1981). Activation by *S*-adenosyl methione to give *Eco*K* allows specific binding at recognition sequences (s_K). If the site is fully methylated, the complex dissociates; if hemimethylated, it modifies the second strand; and only if the site is unmodified does restriction occur. In the latter case, *Eco*K* undergoes an ATP-dependent conformational change to *Eco*K$^+$, but remains bound to the target site while DNA is translocated past the enzyme by an ATP-dependent process prior to DNA cleavage.

The sequences of the *hsdS* genes of *Eco*K and the related *Eco*B and *Eco*D have been determined (Gough & Murray, 1983). We report here the DNA sequences of the *hsdM* and *R* genes of *Eco*K, and discuss the organization of the *hsd* genes and features of the polypeptides.

† Present address: Department of Biochemistry, Universiteit van Amsterdam, Postbus 20151, 1000 HD. Amsterdam, Nederland.

‡ Present address: Department of Microbiology, Louisiana State University, Baton Rouge, La. 70893, U.S.A.

§ Author to whom all correspondence should be addressed.

0022–2836/87/220159–12 $03.00/0

© 1987 Academic Press Limited

2. Materials and Methods

(a) *Bacteria, phages and plasmids*

The bacterial and phage strains are listed in Table 1. The *hsd* genotypes of some bacterial strains were changed by transferring mutations, or an alternative *hsdS* gene, from a *λhsd* phage to the *E. coli* chromosome (Gough & Murray, 1983). Plasmid pUC13 (Messing, 1983) was used as a general cloning vector and as a source of multiple cloning sites in the construction of some *λ* vectors. Plasmid pBg3 (Sain & Murray, 1980) was used as a source of *hsd* DNA (see Fig. 1).

(b) *Media and microbial methods*

Media were made and used as described by Murray *et al.* (1977). *β*-Galactosidase was assayed as described by Miller (1972).

(c) *Enzymes and chemicals*

DNA polymerase (Klenow fragment) was purchased from Boehringer-Mannheim; DNA polymerase I from NBL Enzymes Ltd., Cramlington, U.K.; restriction enzymes from either NBL or from New England Biolabs Inc. Beverly, MA. Phage T4 DNA ligase was made by Sandra Bruce. M13 17-mer primer and hybridization probe primer were from New England Biolabs Inc., deoxynucleoside triphosphates (dNTPs) and dideoxynucleoside triphosphates (ddNTPs) from PL Biochemicals Inc. Milwaukee WI, and deoxyinosine triphosphate from Sigma Ltd., Poole, U.K.

Deoxycytidine 5′-[α-^{32}P]triphosphate (110 TBq/nmol) and deoxyadenosine 5′-[α-^{35}S]thiotriphosphate (15 TBq/nmol) were from Amersham International.

(d) *DNA preparation, analysis and ligation reactions*

Phage *λ* and plasmid DNA were prepared as described by Sain & Murray (1980); M13 replicative form DNA was prepared by the method of Messing (1983). DNA was digested with restriction enzymes using the buffers recommended by the suppliers. Fragments were separated in agarose gels (0·7 to 1·0%, w/v) in TBE buffer, pH 8·2 (Biggin *et al.*, 1983) and, where necessary, isolated by electroelution in dialysis tubing (Maniatis *et al.*, 1982). Sonicated fragments were generated from circular plasmid DNA (Deininger, 1983), those in the size range of 300 to 600 bp† were incubated with Klenow polymerase and dNTPs to repair ends and were ligated to M13 vector cut at the *Sma*I site. DNA fragments with specific ends were ligated to appropriately cut M13 vectors.

(e) *Recovery of recombinants*

Phage were recovered by transfection (Mandel & Higa, 1970), but the efficiency of recovery of M13 derivatives

† Abbreviations used: bp, base-pair(s); kb, 10^3 bases or base-pairs; AdoMet, *S*-adenosyl methionine; r^- and m^-, restriction and modification deficient phenotypes, respectively; $r_K{}^+$ and $m_K{}^+$ identifies the specificity of the r and m system as that of *E. coli* K-12; *orf*, open reading frame; u.v., ultraviolet light.

Table 1
Bacterial and phage strains

A. *Bacteria* Strain	Relevant features	Source or reference
C600	$r_K{}^+$ $m_K{}^+$	Appleyard (1954)
ED8654	$r_K{}^-$ $m_K{}^+$	Murray *et al.* (1977)
WA960	$r_B{}^+$ $m_B{}^+$	Wood (1966)
C3000 *hsdR*	$r_B{}^-$ $m_B{}^+$	Zabeau *et al.* (1980)
NM522	*hsd*Δ/F′	Gough & Murray (1983)
psu0	*lacZ trpR*	Korn & Yanofsky (1976)
psu3	Rho^- *psu0*	Korn & Yanofsky (1976)
NM609	(*hsdRM*)Δ *psu0*	This laboratory
NM610	(*hsdMS*)Δ *psu0*	This laboratory
NM629	$hsdS_B$ *psu0*	This laboratory
ED8799	*lacZ*ΔM15 *hsdS*	Murray *et al.* (1977)
ED8514	*trp*(*OE*)Δ	Murray & Brammar (1973)
B. *Bacteriophages*		
λ NM1048	*hsd R* M S*	Sain & Murray (1980)
λ NM1050	*hsd R M S*	Sain & Murray (1980)
λ NM642	*hsd* R^{*-} *M S*	Sain & Murray (1980)
λ NM1106	*hsd R* (M S)*Δ	Sain & Murray (1980)
λ NM1119	*hsd R**Δ *M S*	Sain & Murray (1980)
λ WL263	*hsd M S* att^+	Loenen & Murray (1986)
λ TL25	*lacZ* imm^{21}	Linn & Ralling (1985)
λ NM459	*srI* (1–2)Δ$3^0 4^0 5^0$cI857*nin*5	This laboratory
λ JW19	*lacZ*/*trp* W205 fusion	Windass & Brammar (1979)
λ NM732	*trpE* imm^{21}	Hopkins *et al.* (1976)
M13 mp10	Vector for sequencing	Messing (1983)
M13 mp11	Vector for sequencing	Messing (1983)

All the bacteria are derivatives of *E. coli* K-12.

An asterisk indicates that the major part of *hsdR* is present but fused in phase to the *redA* gene of λ (⁻ and Δ denote point or deletion mutations, respectively, within *hsdR**).

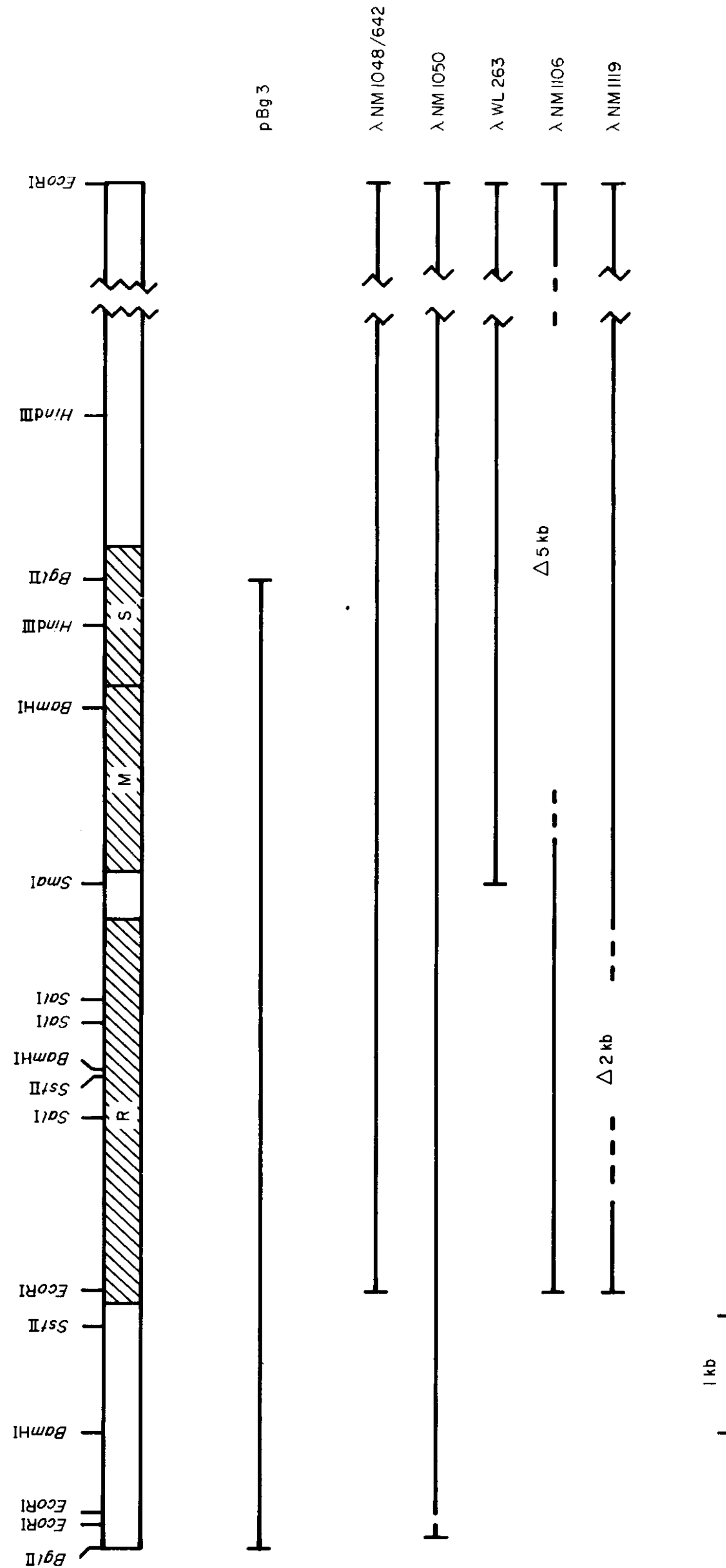

Figure 1. Map of the *hsd* region of *E. coli* K-12. Shaded areas identify the 3 genes. The extent of the region carried in plasmid pBg3 and λ*hsd* phages is indicated.

```
SstII  10                       30                       50                       70                       90                       110
CCGCGGCAGCCTCATGAACATCACGCGCGGCTGCACCTTCCGGTTTGTTGCCAGATAACGCAGAACAGGTTCAATAAATTTGTCATAGGTAGGAACCGTCATAGTACATCCTTGCAGAA
                                   EcoK            -35                          -10

    130                      150                      170                      190                      210                      230
TCAGGTAGATGTTTTTCGGCTACTATAGCACTACAAAAATAGACGAACACGTTAGAAATGAGTCAGTTGCTGTGACCGTGGTCATTGCCCGGAAAGGTACAGAAAGCTAAGATGAGATGT
                                                                                                                     MetL

    250                      270                      290                      310  EcoR1               330                      350
TATGGGCCTTAAATATTTGGACAGGCCCGCACAGCAATGGATTAATAACAATGATGAATAAATCCAATTTTGAATTCCTGAAGGGCGTCAACGACTTCACTTATGCCATCGCCTGTGCGG
euTrpAlaLeuAsnIleTrpThrGlyProHisSerAsnGlyLeuIleThrMetMetAsnLysSerAsnPheGluPheLeuLysGlyValAsnAspPheThrTyrAlaIleAlaCysAlaA

    370                      390                      410                      430                      450                      470
CGGAAAATAACTACCCGGATGATCCCAACACGACGCTGATTAAAATGCGTATGTTTGGCGAAGCCACAGCGAAACATCTTGGTCTGTTACTCAACATCCCCCCTTGTGAGAATCAACACG
laGluAsnAsnTyrProAspAspProAsnThrThrLeuIleLysMetArgMetPheGlyGluAlaThrAlaLysHisLeuGlyLeuLeuLeuAsnIleProProCysGluAsnGlnHisA

    490                      510                      530                      550                      570                      590
ATCTCCTGCGTGAACTCGGCAAAATCGCCTTTGTTGATGACAACATCCTCTCTGTATTTCACAAATTACGCCGCATTGGTAACCAGGCGGTGCACGAATATCATAACGATCTCAACGATG
spLeuLeuArgGluLeuGlyLysIleAlaPheValAspAspAsnIleLeuSerValPheHisLysLeuArgArgIleGlyAsnGlnAlaValHisGluTyrHisAsnAspLeuAsnAspA

    610                      630                      650                      670                      690                      710
CCCAGATGTGCCTGCGACTCGGGTTCCGCCTGGCTGTCTGGTACTACCGTCTGGTCACTAAAGATTATGACTTCCCGGTGCCGGTGTTTGTGTTGCCGGAACGTGGTGAAAACCTCTATC
laGlnMetCysLeuArgLeuGlyPheArgLeuAlaValTrpTyrTyrArgLeuValThrLysAspTyrAspPheProValProValPheValLeuProGluArgGlyGluAsnLeuTyrH

    730                      750                      770                      790                      810                      830
ACCAGGAAGTGCTGACGCTAAAACAACAGCTTGAACAGCAGGTGCGAGAAAAAGCGCAGACTCAGGCAGAAGTCGAAGCGCAACAGCAGAAGCTGGTTGCCCTGAACGGCTATATCGCCA
isGlnGluValLeuThrLeuLysGlnGlnLeuGluGlnGlnValArgGluLysAlaGlnThrGlnAlaGluValGluAlaGlnGlnGlnLysLeuValAlaLeuAsnGlyTyrIleAlaI

    850                      870                      890                      910                      930                      950
TTCTGGAAGGCAAACAGCAGGAAACCGAAGCGCAAACCCAGGCTCGCCTTGCGGCACTGGAAGCACAGCTCGCCGAGAAGAACGCGGAACTGGCAAAACAGACCGAACAGGAACGTAAGG
leLeuGluGlyLysGlnGlnGluThrGluAlaGlnThrGlnAlaArgLeuAlaAlaLeuGluAlaGlnLeuAlaGluLysAsnAlaGluLeuAlaLysGlnThrGluGlnGluArgLysA

    970                      990                      1010                     1030                     1050                     1070
CTTACCACAAAGAAATTACCGATCAGGCCATCAAGCGCACACTCAACCTTAGCGAAGAAGAGAGTCGCTTCCTGATTGATGCGCAACTGCGTAAAGCAGGCTGGCAGGCCGACAGCAAAA
laTyrHisLysGluIleThrAspGlnAlaIleLysArgThrLeuAsnLeuSerGluGluGluSerArgPheLeuIleAspAlaGlnLeuArgLysAlaGlyTrpGlnAlaAspSerLysT

    1090                     1110                     1130                     1150                     1170                     1190
CCCTGCGCTTCTCCAAAGGCGCACGTCCGGAACCCGGCGTCAATAAAGCCATTGCCGAATGGCCGACCGGAAAAGATGAAACGGGTAATCAGGGCTTTGCGGATTATGTGCTGTTTGTCG
hrLeuArgPheSerLysGlyAlaArgProGluProGlyValAsnLysAlaIleAlaGluTrpProThrGlyLysAspGluThrGlyAsnGlnGlyPheAlaAspTyrValLeuPheValG

    1210                     1230                     1250                     1270                     1290                     1310
GCCTCAAACCCATCGCGGTGGTAGAGGCGAAACGTAACAATATCGACGTTCCCGCCAGGCTCAATGAGTCGTATCGCTACAGTAAATGTTTCGATAATGGCTTCCTGCGGGAAACCTTGC
lyLeuLysProIleAlaValValGluAlaLysArgAsnAsnIleAspValProAlaArgLeuAsnGluSerTyrArgTyrSerLysCysPheAspAsnGlyPheLeuArgGluThrLeuL

    1330                     1350                     1370                     1390                     1410                     1430
TTGAGCACTACTCACCGGATGAAGTGCATGAAGCAGTGCCAGAGTATGAAACCAGCTGGCAGGACACCAGCGGCAAACAACGGTTTAAAATCCCCTTCTGCTACTCGACCAACGGGCGCG
euGluHisTyrSerProAspGluValHisGluAlaValProGluTyrGluThrSerTrpGlnAspThrSerGlyLysGlnArgPheLysIleProPheCysTyrSerThrAsnGlyArgG

    1450                     1470                     1490                     1510                     1530                     1550
AATACCGCGCAACAATGAAGACCAAAAGCGGCATCTGGTATCGCGACGTGCGTGATACCCGCAATATGTCGAAAGCCTTACCCGAGTGGCACCGCCCGGAAGAGCTGCTGGAAATGCTCG
luTyrArgAlaThrMetLysThrLysSerGlyIleTrpTyrArgAspValArgAspThrArgAsnMetSerLysAlaLeuProGluTrpHisArgProGluGluLeuLeuGluMetLeuG

    1570                     1590                     1610                     1630                     1650                     1670
GCAGCGAACCGCAAAAACAGAATCAGTGGTTTGCCGATAACCCTGGCATGAGCGAGCTGGGCCTGCGTTATTATCAGGAAGATGCCGTCCGCGCGGTTGAAAAGGCAATCGTCAAGGGGC
lySerGluProGlnLysGlnAsnGlnTrpPheAlaAspAsnProGlyMetSerGluLeuGlyLeuArgTyrTyrGlnGluAspAlaValArgAlaValGluLysAlaIleValLysGlyG

    1690                     1710                     1730                     1750                     1770                     1790 SalI
AACAAGAGATCCTGCTGGCGATGGCGACCGGTACCGGTAAAACCCGTACGGCAATCGCCATGATGTTCCGCCTGATCCAGTCCCAGCGTTTTAAACGCATTCTCTTCCTTGTCGACCGCC
lnGlnGluIleLeuLeuAlaMetAlaThrGlyThrGlyLysThrArgThrAlaIleAlaMetMetPheArgLeuIleGlnSerGlnArgPheLysArgIleLeuPheLeuValAspArgA

    1810                     1830                     1850                     1870                     1890                     1910
GTTCTCTTGGCGAACAGGCGCTGGGCGCGTTTGAAGATACGCGTATTAACGGCGACACCTTCAACAGCATTTTCGACATTAAAGGGCTGACGGATAAATTCCCGGAAGACAGCACCAAAA
rgSerLeuGlyGluGlnAlaLeuGlyAlaPheGluAspThrArgIleAsnGlyAspThrPheAsnSerIlePheAspIleLysGlyLeuThrAspLysPheProGluAspSerThrLysI

    1930                     1950                     1970                     1990                     2010                     2030
TTCACGTTGCCACCGTACAGTCGCTGGTGAAACGCACCCTGCAATCAGATGAACCGATGCCGGTGGCCCGTTACGACTGTATCGTCGTTGACGAAGCGCATCGCGGCTATATTCTCGATA
leHisValAlaThrValGlnSerLeuValLysArgThrLeuGlnSerAspGluProMetProValAlaArgTyrAspCysIleValValAspGluAlaHisArgGlyTyrIleLeuAspL

    2050                     2070                     2090                     2110                     2130                     2150
AAGAGCAGACCGAAGGCGAACTGCAGTTCCGCAGCCAGCTGGATTACGTCTCTGCCTACCGTCGCATTCTCGATCACTTCGAATGCGGTAAAAATCGCTCTCACCGCCACCCCGGCGCTA
ysGluGlnThrGluGlyGluLeuGlnPheArgSerGlnLeuAspTyrValSerAlaTyrArgArgIleLeuAspHisPheGluCysGlyLysAsnArgSerHisArgHisProGlyAlaT

    2170                     2190                     SstII                    2230                     BamH1                    2270
CATACTGTGCAGATTTTCGGCGAGCCGGTTTACCGTTATACCTACCGTACCGCGGTTATCGACGGTTTTCTGATCGACCAGGATCCGCCTATTCAGATCATCACCCGCAACGCGCAGGAG
hrTyrCysAlaAspPheArgArgAlaGlyLeuProLeuTyrLeuProTyrArgGlyTyrArgArgPheSerAspArgProGlySerAlaTyrSerAspHisHisProGlnArgAlaGlyG

    2290                     2310                     2330                     2350                     2370                     2390
GGGGTTTATCTCTCCAAAGGCGAGCAGGTAGAGCGCATCAGCCCGCAGGAGAAGTGATCAATGACACCCTGGAAGACGATCAGGATTTTGAAGTCGCCGACTTTAACCGTGGCCTGGTGA
lyGlyLeuSerLeuGlnArgArgAlaGlyArgAlaHisGlnProAlaGlyGluValIleAsnAspThrLeuGluAspAspGlnAspPheGluValAlaAspPheAsnArgGlyLeuValI

    2410                     2430                     2450                     2470                     2490                     2510
TCCCGGCGTTTAACCGCGCCGTCTGTAACGAACTCACCAATTATCTTGACCCGACCGGATCGCAAAAAACGCTGGTCTTCTGCGTCACCAATGCCCATGCCGATATGGTGGTGGAAGAGC
leProAlaPheAsnArgAlaValCysAsnGluLeuThrAsnTyrLeuAspProThrGlySerGlnLysThrLeuValPheCysValThrAsnAlaHisAlaAspMetValValGluGluL

    2530                     2550                     2570                     2590                     2610                     2630
TGCGTGCCGCGTTCAAGAAAAAGTATCCGCAACTGGAGCACGACGCGATCATCAAGATCACCGGTGATGCCGATAAAGACGCGCGCAAAGTGCAGACCATGATCACCCGCTTCAATAAAG
euArgAlaAlaPheLysLysLysTyrProGlnLeuGluHisAspAlaIleIleLysIleThrGlyAspAlaAspLysAspAlaArgLysValGlnThrMetIleThrArgPheAsnLysG

    2650                     SalI                     2690                     2710                     2730                     2750
AGCGGCTGCCCAATATCGTGGTAACCGTCGACCTGCTGACGACCGGCGTCGATATTCCGTCGATCTGTAATATCGTGTTCCTGCGTAAAGTACGCAGCCGCATTCTGTACGAACAGATGA
luArgLeuProAsnIleValValThrValAspLeuLeuThrThrGlyValAspIleProSerIleCysAsnIleValPheLeuArgLysValArgSerArgIleLeuTyrGluGlnMetL

    2770                     2790                     2810                     2830                     2850 SalI                2870
AAGGCCGCGCCACGCGCTTATGCCCGGAGGTGAATAAAACCAGCTTTAAGATTTTTGACTGTGTCGATATCTACAGCACGCTGGAGAGCGTCGACACCATGCGTCCGGTGGTGGTGCGCC
ysGlyArgAlaThrArgLeuCysProGluValAsnLysThrSerPheLysIlePheAspCysValAspIleTyrSerThrLeuGluSerValAspThrMetArgProValValValArgP

    2890                     2910                     2930                     2950                     2970                     2990
CGAAGGTGGAACTGCAAACGCTGGTCAATGAAATTACCGATTCAGAAACCTATAAAATCACCGAAGCGGATGGCCGCAGTTTTGCCGAGCACAGCCATGAACAACTGGTGGCGAAGCTCC
roLysValGluLeuGlnThrLeuValAsnGluIleThrAspSerGluThrTyrLysIleThrGluAlaAspGlyArgSerPheAlaGluHisSerHisGluGlnLeuValAlaLysLeuG

    3010                     3030                     3050                     3070                     3090                     3110
AGCGTATCATCGGTCTGGCCACGTTTAACCGTGACCGCAGCGAAACGATAGATAAACAGGTGCGTCGTCTGGATGAGCTATGCCAGGACGCGGCGGGCGTGAACTTTAACGGCTTCGCCT
lnArgIleIleGlyLeuAlaThrPheAsnArgAspArgSerGluThrIleAspLysGlnValArgArgLeuAspGluLeuCysGlnAspAlaAlaGlyValAsnPheAsnGlyPheAlaS

    3130                     3150                     3170                     3190                     3210                     3230
CGCGCCTGCGGGAAAAAGGGCCGCACTGGAGCGCCGAAGTCTTTAACAAACTGCCTGGCTTTATCGCCCGTCTGGAAAAGCTGAAAACGGACATCAACAACCTGAATGATGCGCCGATCT
erArgLeuArgGluLysGlyProHisTrpSerAlaGluValPheAsnLysLeuProGlyPheIleAlaArgLeuGluLysLeuLysThrAspIleAsnAsnLeuAsnAspAlaProIleP

    3250                     3270                     3290                     3310                     3330                     3350
TCCTCGATATCGACGATGAAGTGGTGAGTGTAAAATCGCTGTACGGTGATTACGACACGCCGCAGGATTTCCTCGAAGCCTTTGACTCGCTGGTGCAACGTTCCCCGAACGCGCAACCGG
heLeuAspIleAspAspGluValValSerValLysSerLeuTyrGlyAspTyrAspThrProGlnAspPheLeuGluAlaPheAspSerLeuValGlnArgSerProAsnAlaGlnProA

    3370                     3390                     3410                     3430                     3450                     3470
CATTGCAGGCAGTTATTAATCGCCCGCGCGATCTCACCCGTAAAGGGCTGGTCGAGCTACAGGAGTGGTTTGACCGCCAGCACTTTGAGGAATCTTCCCTGCGCAAAGCATGGAAAGAGA
laLeuGlnAlaValIleAsnArgProArgAspLeuThrArgLysGlyLeuValGluLeuGlnGluTrpPheAspArgGlnHisPheGluGluSerSerLeuArgLysAlaTrpLysGluT

    3490                     3510                     3530                     3550                     3570                     3590
CGCGCATGAAGATATCGCCGCCCGGCTGATTGGTCATATTCGCCGCGCTGCGGTGGGCGATGCGCTGAAACCGTTTGAGGAACGTGTCGATCACGCGCTGACGCGCATTAAGGGCGAAAA
hrArgMetLysIleSerProProGlyEnd

    3610                     3630                     3650                     3670                     3690                     3710
CGACTGGAGCAGCGAGCAATTAAGCTGGCTCGATCGTTTAGCGCAGGCGCTGAAAGAGAAAGTGGTGCTCGACGACGATGTCTTCAAAACCGGCAACTTCCACCGTCGCGGCGGGAAGGC

    3730                     3750                     3770                     3790                     3810                     3830
GATGCTGCAAAGAACCTTTGACGATAATCTCGATACCTGCTGGGCAAATTCAGCGATTATATCTGGGACGAGCTGGCCTGACACGTATACACTTCATCCTTCAGGCTGCCTCTGCGTTGG
```

Fig. 2.

Figure 2. DNA sequence of *hsdR* and *hsdM* with predicted amino acid sequence. The −35 and −10 regions of the putative promoters are identified (approximate positions 50 and 3910). Other features shown are an *Eco*K site (position 35); the repeated "Ala-Gln-Thr-Gln-Ala' sequence within regions of α-helix (positions 775 and 870); and 2 possible stem-loop structures between *hsdR* and *hsdM* (positions 3700 and 3790).

was increased by the modification described by Hanahan (1983).

(f) *Detection and sequencing of M13 recombinants*

A library of plasmid pBg3 sonicated DNA fragments was made in M13mp10 and screened for plaques that hybridized to an *hsd*-specific probe (Benton & Davis, 1977; Rigby *et al.*, 1977). Single-stranded template DNA, prepared by the method of Sanger *et al.* (1980), was sequenced by the dideoxy chain termination method using [α-^{35}S]thio-dATP, and analysed on buffer gradient gels (Biggin *et al.*, 1983). Compressed sequences were resolved by the use of dITP in place of dGTP. The sequences of DNA fragments were compiled by computer programs (Staden, 1982). Sections of sequence, initially obtained on only 1 strand, were selected from the library of recombinants using strand-specific probes (Hu & Messing, 1982). A few gaps were completed using defined fragments and, in 2 instances, synthetic oligonucleotides as primers.

3. Results

(a) *The nucleotide sequence of the* hsd *region of* E. coli *K-12*

The *hsd* genes of *E. coli* K-12 have been cloned as a contiguous segment of DNA in bacteriophage λ. A single 11·3 kb *Eco*RI fragment cloned in λ(λNM642 or λNM1048; Fig. 1) lacked the beginning of the *hsdR* gene, but a derivative (λNM1050) was isolated in which the segment of bacterial DNA was extended to include the start of the *hsdR* gene and the promoter p_{res} (Sain & Murray, 1980).

The sequence of the *hsdS* gene has been determined (Gough & Murray, (1983). Plasmid pBg3, which includes *hsdR* and *hsdM* (see Fig. 1), was used as a source of DNA to sequence the remainder of the *hsd* region (Fig. 2). A contiguous sequence of 8·4 kb, from the *Sst*II site upstream from *hsdR* to the *Hin*dIII site downstream from *hsdS* (see Fig. 1) includes the entire region. Two coding sequences separated by 492 bp correlate with *hsdR* and *hsdM*. Their codon usage is as expected for *E. coli* genes that are poorly expressed (Grantham *et al.*, 1981).

(i) *The* hsdR *gene*

The only open reading frame within the 8·4 kb of DNA long enough to encode the R polypeptide is mainly within the *Eco*RI fragment originally cloned in λNM1048 but extends into the adjacent *Eco*RI fragment included in λNM1050. Immediately upstream from the first methionine codon (Fig. 2) is a purine-rich region that could serve as a ribosome binding site (Shine & Dalgarno, 1974). Using the

first methionine as initiation codon, a polypeptide with a molecular weight of 124,802 is predicted. This is in excellent agreement with estimates deduced from the properties of the *hsdR* polypeptide observed in bacterial extracts (Eskin & Linn, 1972; Sain & Murray, 1980). One end of the fragment cloned in λNM1048 would be the *Eco*RI site in the *hsdR* gene (bp 312, Fig. 2). The reading frame of this coding sequence and its orientation within the λ vector predict a fusion polypeptide having the major portion of the R polypeptide and the N terminus of λ exonuclease. A fusion polypeptide of appropriate mobility has been reported (Sain & Murray, 1980). Alternative upstream GTG initiation codons that maintain the reading frame are not precluded.

(ii) *The* hsdM *gene*

λ*hsd* phages carrying only the 7·7 kb *Sma*I–*Eco*RI fragment (see λWL263, Fig. 1) include *hsdM* and *hsdS* (Loenen & Murray, 1986). A coding sequence of 1590 bases starts 120 bases downstream from this *Sma*I site and has a stop codon overlapping the start codon of the *hsdS* gene. This coding sequence is preceded by a potential ribosome-binding site and encodes a polypeptide with a molecular weight of 59,289, in good agreement with the mobility of the M polypeptide (Sain & Murray, 1980).

(iii) *Other* orfs

Additional *orfs* may not be significant since no other polypeptide encoded by this region was detected following infection of u.v. irradiated cells with λ*hsd* phages (Sain & Murray, 1980). However, small polypeptides would not readily be detected. One *orf* on the antisense strand of the *hsdM* gene could encode a polypeptide with a molecular weight of 35,000, but it lacks obvious signals for either the initiation of transcription or translation, and its codon usage and amino acid composition are uncharacteristic of *E. coli*.

(b) *The polypeptides*

*Eco*K (Fig. 2; and see Gough & Murray, 1983) is the only type I restriction enzyme for which the complete amino acid sequence has been predicted. Our analyses of these polypeptides, therefore, rely on comparisons with other proteins that interact with the same substrates or co-factors. DNA, ATP and AdoMet all serve as both allosteric co-factors and substrates for the activities of *Eco*K (Bickle, 1982; Yuan, 1981).

Many lines of evidence implicate the S polypeptide in the determination of sequence specificity (see Bickle, 1982; Fuller-Pace & Murray, 1986), but it is not necessarily the only *hsd* polypeptide to interact with DNA. A second DNA binding site becomes available once the activated *Eco*K has bound to its recognition sequence. As a consequence of this second interaction, looped structures are formed that become supercoiled as the DNA is translocated in an ATP-dependent process past the bound *Eco*K complex (Yuan *et al.*, 1980). ATP effects the conformational change that is a prerequisite to the formation of the looped structures, and its subsequent hydrolysis is essential for both DNA translocation and nucleolytic activity (Bickle *et al.*, 1978). Furthermore, while ATP stimulates methylation of hemimethylated DNA by the *Eco*K complex, it does not affect the methylase activity of the enzyme comprising only the M and S subunits (Suri *et al.*, 1984). All the known ATP-dependent activities require the R subunit and the most likely location of an ATP-binding site is, therefore, in the R polypeptide. In contrast, methylation occurs in the absence of the R polypeptide and a relevant active centre must be made by S, or M, or a complex of both polypeptides. It has been shown, however, that *Eco*K binds as many as five molecules of AdoMet and, since AdoMet serves as both allosteric effector and substrate, binding sites other than the active centre are predicted (Hadi *et al.*, 1975). The characterization of the complex encoded by each of three *hsd* mutants supports the correlation of the S polypeptide with specific interaction with DNA, M with methylation and R with the binding of ATP (Hadi & Yuan, 1974).

(i) *ATP binding domains*

Walker *et al.* (1982) compared the sequence of the α and β-subunits of ATP synthase of *E. coli* with those of adenylate kinase, phosphofructokinase and several other ATP binding proteins, and identified conserved sequences in two domains. For adenylate kinase, crystallographic evidence correlates these conserved domains with a pocket-like structure identified as the adenine nucleotide binding site (Pai *et al.*, 1977). One conserved sequence (domain A) includes a glycine-rich flexible loop followed by a lysine residue, whilst the second (B) includes a hydrophobic β-sheet structure, located at the back of the ATP binding site, and an aspartic acid residue that could bind magnesium ions.

Our search of both the R and M polypeptide sequences for regions sharing homology with those identified as common to various ATP-binding proteins (Walker *et al.*, 1982; Finch & Emmerson, 1984) identified one sequence in R resembling that of domain A of *E. coli* ATP synthase (Table 2). The sequence in R, however, lacks the first conserved glycine residue, or the alternative alanine residue identified in some proteins. A putative B domain is not easy to find. Potential hydrophobic β-sheets followed by an aspartic acid residue are located around positions 133 and 515 (Table 2) but only the first has the arginine and glycine residues upstream (Walker *et al.*, 1982).

(ii) *Adenine methylase domains*

X-ray crystallographic analysis has not been reported for any modification methylases and AdoMet binding domains have not been characterized.

Table 2

Alignment of homologous sequences in adenine nucleotide-binding proteins

Protein	1st residue	Sequence	Reference
Domain A			
Adenylate kinase	008	S K I I F V V G G P G S G K G Y Q C E K I V	a
E. coli ATPase α	163	G Q R E L I I G D R Q T G K T A L A I D A I	b
E. coli ATPase β	144	G G K V G L F G G A G V G K T V N M M E L I	c
DnaB	223	S D L I I V A A R P S M G K T T F A M N L V	d
UvrD	022	R S N L L V L A G A G S G K T R V L V H R I	e
*Eco*K HsdR	482	Q Q E I L L A M A T G T G K T R T A I A M M	
Domain B			
Adenylate kinase	106	F E R K I G Q P T − L L L Y V D	a
E. coli ATPase α	269	Y F R D R G E D A − L I I Y D D	b
E. coli ATPase β	131	K F R D E G R D V − L L F V D N	c
*Eco*K HsdR	125	C L R L − G F R L A V W Y Y R L V T K D	
*Eco*K HsdR	503	M F R L I Q S R F K R I L F L V D	

In domain A, the boxed sequences are identical; in domain B, the 3rd box contains conserved hydrophobic residues. References: a, Pai *et al.* (1977); b, Gay & Walker (1981); c, Saraste *et al.* (1981); d, Nakayama *et al.* (1984); e, Finch & Emmerson (1984).

We have looked for features of amino acid sequences of M and S that are common to DNA methylases. A comparison of the M polypeptide with the *Eco*RI methylase identified an extensive region of homology. This region shows some homology with a number of methylases, including a dam methylase and both type II and type III modification enzymes. S. Slatko and colleagues (personal communication from New England Biolabs) have identified part of this sequence (Asp/Asn, Pro, Pro, Tyr/Phe) as characteristic of enzymes that methylate adenine. The sequence in Table 4 is not present in the cytosine methylases we have examined, and so could identify part of the active site for the specific methylation of adenine.

(iii) *DNA binding sites*

Proteins interact with DNA in a diversity of roles and forms. Some DNA-binding proteins share a helix-turn-helix motif (for a review, see Pabo & Sauer, 1984) in which one of these two helices dictates the sequence specificity (Wharton & Ptashne, 1985), whilst others align recognition domains within the major groove of the DNA helix in different ways (Miller *et al.*, 1985; McClarin *et al.*, 1986). The S polypeptide imparts sequence specificity to the *Eco*K complex, yet no DNA recognition motif has been detected (Gough & Murray, 1983; and this work), although regions of α-helical content have been predicted (Argos, 1985).

Table 3

Alignment of homologous sequences in proteins that bind DNA

Protein	1st residue	Sequence	Reference
λ CI	033	Q E S V A D K M G M G Q S G V G A L F N	a
*Eco*K HsdM	483	D V L A A E A M G E L V Q A L S E L D A	
DnaA	154	A R Q V A D N P G G A Y N P L F L Y G G	b
*Eco*K HsdR	449	N Q W F A D N P G M S E L G L R Y Y Q E	

The boxed sequences are the conserved amino acids known to be involved in DNA binding. References: a, Sauer (1978); b, Hansen *et al.* (1982).

Table 4

Alignment of homologous sequences in E. coli *adenine methylases*

Protein	1st residue	Sequence	Reference
HsdM	251	G S D G E N L P K A H I V A T N P P F G S A A	
*Eco*RI	124	S E S I D L L K K S D I V V T N P P F S L F R	a
*Eco*RV	178	E K T I G M V N R D D V V Y C D P P Y I G R H	b
Dam	230	A D S M A R A D D A S V V Y C D P P Y A P L S	c

The boxed sequences contain conserved amino acid residues. References: a, Greene *et al.* (1981), Newman *et al.* (1981); b, Bougueleret *et al.* (1984); c, Brooks *et al.* (1983).

We have also found the sequence KVNMIYIDPPY in the type III enzyme P1 (T. A. Bickle, personal communication); but as this sequence is incomplete, we have not included it.

The *Eco*K complex recognizes the methylated state of the target sequence, and some domain, or domains, of the complex encoded by *hsdM*, or possibly *S*, must distinguish the three possible states of the recognition sequence (see Burkhardt *et al.*, 1981; Loenen & Murray, 1986). The R polypeptide may include a DNA binding domain involved in the translocation process or associated with nucleolytic activity. We therefore searched the amino acid sequences of the M and R polypeptides for potential DNA binding domains, although we have no evidence to implicate sequence specificity in these processes.

The consensus amino acids characteristic of the helix-turn-helix motif were identified in both the M and R polypeptides. No evidence indicates whether these sequences are relevant to interaction with DNA, and the two identified in Table 3 are included after consideration of both the secondary structure predictions and the nature of the amino acids within the postulated helices; most particularly, the conservation of hydrophobic residues at positions 4, 8 and 10 (Drummond *et al.*, 1986; but see Table 3) and homology with predicted helical domains in other polypeptides. The putative motif in the R polypeptide has hydrophobic residues at positions 4 and 10 (see Table 3) and a neutral one at position 8, but this latter residue is part of an identical sequence within the DnaA polypeptide of *E. coli* (Hansen *et al.*, 1982).

(iv) *Other features*

We have compared the sequences of the R, M and S polypeptides with themselves and each other using UWGCG software (Devereux *et al.*, 1984) to search for repeated sequences, as well as conserved features. We find no evidence for conservation of sequences when R and M are compared, although it has been suggested that there might be some cross-reactivity to their antisera (Bickle, 1982).

The amino acid sequence Ala-Gln-Thr-Gln-Ala is repeated in the R polypeptide at positions 180 and 212. Each of these repeats (see Fig. 2) is within, or more likely close to, extensive regions of predicted α-helix. These α-helical regions are also prominent because of the high proportion of glutamine (20%) and positively charged amino acids in comparison to the whole of the molecule. In the first, residues 160 to 177 make a reasonable amphipathic helix with a hydrophobic patch (residues 161, 165, 168, 172 and 176). The basic residues (162, 169, 177) are mainly on one side of the helix, and there is one mainly neutral but polar surface. After a predicted turn (residues 197, 198) and β-sheet, residues 217 to 240 give a mainly hydrophobic region on one side of the next helix, but otherwise the amino acid distribution does not indicate well-defined characteristics with respect to the helical organization.

These α-helical regions are prominent features of the R polypeptide, and an exhaustive search (Lyall *et al.*, 1986) of the NBRF protein database found remarkably similar segments in some filamentous proteins, for example the mouse glial fibrillary acid protein (Lewis *et al.*, 1984). In these proteins, the structural relevance of the α-helical regions is the formation of interpolypeptide coiled coils. Crystallographic information is necessary before we can tell whether such an organization is relevant to the *Eco*K complex.

(v) *The normal amino-terminal sequence of the R polypeptide is not essential for restriction*

A λ*hsd* phage in which the 11·3 kb *Eco*RI fragment is in the orientation in which the *hsd* genes are transcribed from p_L (λNM1048 in Fig. 1) encodes a fusion polypeptide with the N terminus of the λ exonuclease and the major portion of the R polypeptide (Sain & Murray, 1980). Complementation tests (Table 5) indicate that this fusion polypeptide is biologically active. Thus, infection of an *hsdR* strain conferring B-modification specificity ($r_B^- m_B^+$) with the λ*hsd* phage described above produces a B-restriction-proficient cell, and the incoming (K-modified) phage has a reduced efficiency of plating, because it is sensitive to B-specific restriction. The efficacy of the R fusion polypeptide in the active restriction complex is confirmed by complementation tests using λ*hsd* phages carrying mutant *hsd* genes (see Table 5). Deletion of the *S* and *M* genes of the incoming phage is without effect, whilst mutations in the *hsdR* gene prevent restriction, clearly implicating the fusion polypeptide. The DNA sequence (see Fig. 2) indicates that the fusion polypeptide lacks the first 26 amino acid residues, yet it forms an active *Eco*K complex.

(c) *Transcription signals*

(i) *Promoter signals for* hsdR *and* hsdM

Genetic evidence indicates a promoter, p_{res}, for *hsdR* in the *hsdRMS* phage λNM1050 (Sain & Murray, 1980) and one, p_{mod}, for *hsdM* and *S* downstream from the *Sma*I site (Loenen & Murray,

Table 5

An R polypeptide lacking 26 N-terminal amino acid residues retains activity

λ*hsd* phages		Efficiency of plating on	
Stock no.	Genotype	r_B^- m_B^+	r_B^+ m^+_B
NM1050	R^+ M^+ $S^+ \cdot$K	10^{-1}	10^{-5}
NM1048	R^* M^+ $S^+ \cdot$K	10^{-1}	10^{-5}
NM1106	R^* $(M\ S)$ $\Delta\cdot$K	10^{-1}	10^{-5}
NM642	R^{*-} M^+ $S^+ \cdot$K	1	10^{-5}
NM1119	$R^*\Delta$ M^+ $S^+ \cdot$K	1	10^{-5}

*R** denotes that the beginning of the *hsdR* gene in the λ*hsd* phage is replaced by part of the *redA*(exo) gene of the vector and a fusion polypeptide results (Sain & Murray, 1980). R^{*-} indicates a point mutation and R*Δ a deletion within the *hsdR* portion of the gene fusion.

The λ*hsd* phages (see Fig. 1) were derived from *E. coli* K-12 and were grown on an m_K^+ host (hence ·K). Their efficiency of plating (e.o.p.) on $r_B^-m_B^+$ and $r_B^+m_B^+$ is estimated relative to that on an $r_K^+m_K^+$ strain. λNM1050 has a wild-type *hsdR* gene and this phage is *ral*⁻, the other phages are *ral*⁺, but *ral*⁻ derivatives behave in the same way. Phages grown on $r_B^-m_B^+$ hosts (λ*hsd*.B) plate on an $r_B^+m_B^+$ strain with an e.o.p. of 1.

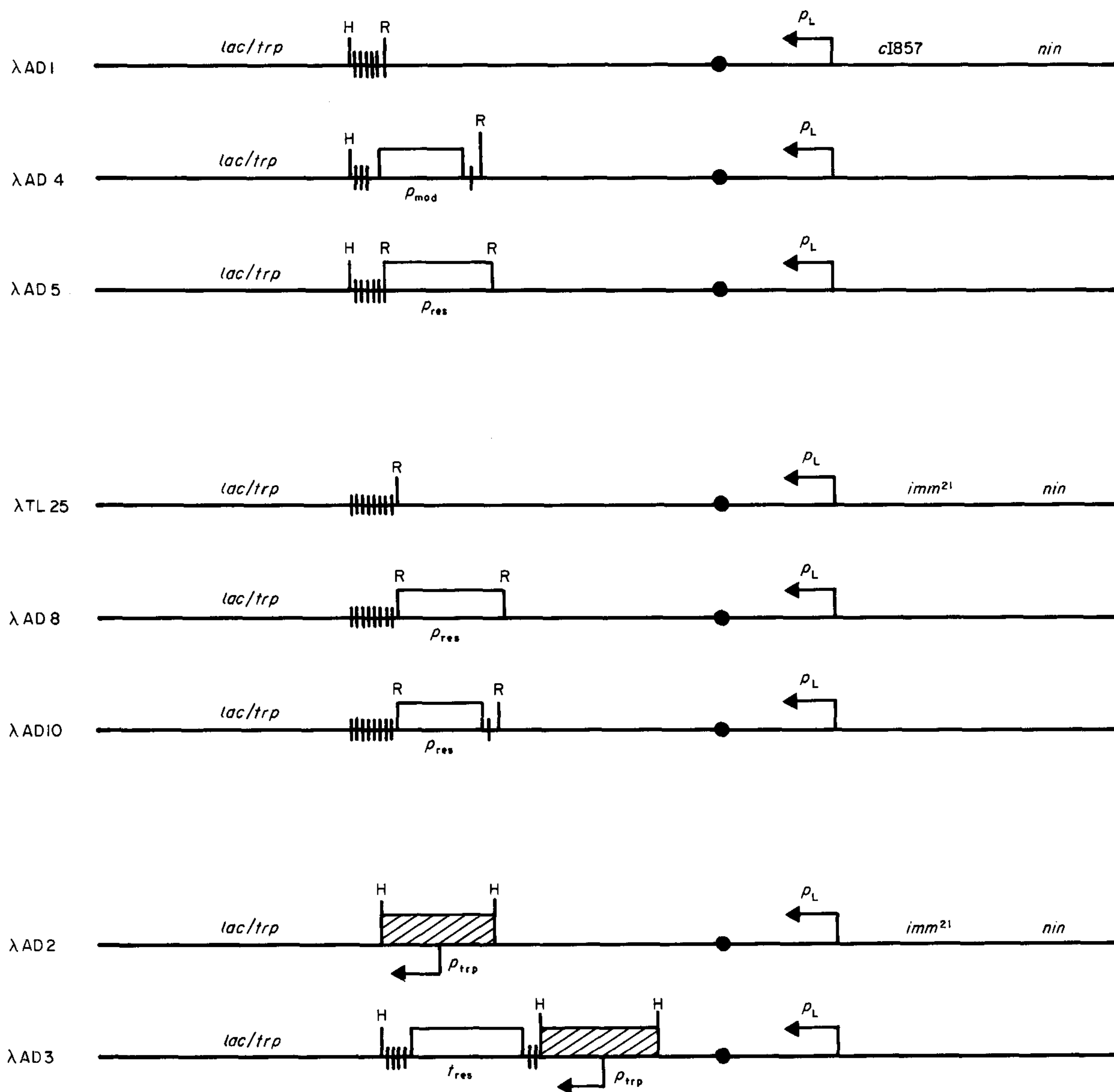

Figure 3. The λ*lacZ* fusion phages used to assay promoter and terminator activities (Table 6). t_{res}, p_{mod} and the smaller p_{res} fragment, were all transferred to the polylinker region of pUC13. This enabled the use of flanking restriction sites in the construction of the above phages. λAD1 was constructed by joining the left arm of λJW19 and the right arm of λNM459, *via* the polylinker region from pUC13. λAD4 was made in the same way, using the pUC13 plasmid containing p_{mod}. For λAD5, the 2 kb *Eco*RI fragment was inserted directly into λAD1. λAD8 is a simple insertion of the 2 kb p_{res} fragment at the unique *Eco*RI target in λTL25. λAD10 uses the same *Eco*RI target to bring in the smaller p_{res} fragment previously cloned in pUC13. The junction between both p_{res} fragments and the right-hand side of the phages is identical. λAD2 has the left arm of λAD1 ligated to a partial *Hin*dIII digest of the λ*trpE* phage λNM732. λAD3 was made in a similar way introducing between p_{trp} and *lacZ* a *Sma*I–*Hin*dIII fragment containing the putative terminator.

1986) in the interval between *hsdR* and *M* (see λWL263 in Fig. 1). Computer analyses identify potential promoters at positions consistent with these data, 157 bp upstream from *hsdR* and 68 bp upstream from *hsdM* (see Fig. 2), although both sequences predict relatively weak binding sites for RNA polymerase. Preliminary experiments using appropriate DNA fragments cloned in the expression vector pK06 (McKenney *et al.*, 1981) provided support, *in vivo* and *in vitro*, for each of these presumptive promoters (W. Loenen & J. Pratt, unpublished results).

Our particular concern is whether transcription of *hsdR* is dependent on, or affected by, the expression of *hsdM* and *S*. Such a pattern of transcriptional control would favour modification of DNA, before the synthesis of restriction enzyme, following transfer to a new environment. The putative promoters were cloned in a λ*lacZ* expression vector to permit sensitive assays from operon fusions present as single chromosomal copies. In the lysogenic state, the synthesis of β-galactosidase from this *lacZ*/*trpA* fusion is dependent on the inclusion of a promoter between the *trpA* gene and

Table 6
β-Galactosidase assays of promoter and terminator activities

Prophage	Host			
A. Promoters	*psu0*	NM609 *(hsdRM)*Δ	NM610 *(hsdMS)*Δ	NM629 *hsdS*$_B$
λ AD1 (vector)	72±16	45±19	76±16	n.t.
λ AD4 (p_{mod})	115±19	126±29	128±14	n.t.
λ TL25 (vector)	123±39	182±60	120±6	159±38
λ AD8 (p_{res})	506±79	496±50	533±70	512±78
λ AD10 (p_{res})	495±41	456±109	473±115	610±91

	Host	
B. Terminator	*psu0* (Rho$^+$)	*psu3* (Rho$^-$)
λ AD2 (vector)	3140±439	4304±1019
λ AD3 (t_{res})	41±7 (99% termination)	535±124 (88% termination)

The vectors and their derivative fusions are shown in Fig. 3. NM609, NM610 and NM629 were derived from the $r_K^+m_K^+$ host *psu0*. n.t. not tested.

Mean LacZ units (Miller, 1972) are shown with their 95% confidence limits. Between 4 and 14 independent lysogens contributed to each figure. Lysogens that gave significantly higher activities (2, 3 or 4 times the levels shown) were assumed to be multiple lysogens and were discounted. This conclusion was justified by the demonstration that lysogens re-made using phage recovered from both presumed mono and multiple lysogens, gave a similar spectrum of values.

the phage attachment site (see Fig. 3).

The 160 bp *Sma*I–*Sau*3A fragment presumed to include p_{mod} was cloned in λAD1 and lysogens of the resulting phage (λAD4) isolated in the *lacZ* host, *psuO*, and two *hsd* deletion derivatives. The levels of *β*-galactosidase support the presence of p_{mod} in λAD4 and indicate that its transcriptional activity is unaffected by the K-specific modification enzyme (Table 6). The 2 kb *Eco*RI fragment including the promoter for *hsdR*, p_{res}, was also cloned in λAD1 (λAD5), but the levels of *β*-galactosidase (data not shown) elicited by the λAD5 prophage were too low to be meaningful. Fragments including the p_{res} sequence were therefore recloned in λTL25, a *lacZ* expression vector especially designed to prevent translational polarity from interfering with the assay based on the level of *β*-galactosidase (Linn & Ralling, 1985). λAD8 includes the 2 kb *Eco*RI fragment cloned in λTL25 and an analogous construct (λAD10) differs only by the brevity of the sequence upstream from the promoter (approx. 300 bp upstream rather than approx. 1800 bp). As prophage, both elicit effective expression of *β*-galactosidase, and neither shows any response to changes in the *hsd* genotype of the host (Table 6).

These data indicate that transcription of *hsdR* is not affected by either K-specific methylation or the presence of the K-modification enzyme.

The relatively high background of *β*-galactosidase elicited by the λ*lacZ* expression vector limits their sensitivity in detecting the apparently weak promoters p_{mod} and p_{res}. In vector λAD1, the activity of p_{res} was probably obscured by translational polarity, whilst p_{mod} was detected in spite of an out-of-phase fusion. In vector λTL25, we have failed to detect p_{mod}. This may reflect inefficient translation of the mRNA elicited from this particular fusion. In two alternative fusions, using multicopy expression vectors, p_{mod} is detected readily (unpublished observations of Ross & Braymer, and our laboratory).

(ii) *Transcription termination signals between* hsdR *and* hsdM

hsdM and *S* overlap and consequently translation of their messenger RNA may be coupled (Oppenheim & Yanofsky, 1980). In contrast, *hsdR* and *hsdM* are separated by 492 bp, and this intergenic sequence includes, in addition to p_{mod}, inverted repeat sequences capable of forming stem-loop structures (see Fig. 2). The 1 kb *Sal*I–*Sma*I fragment that includes the end of *hsdR* and most of the intergenic sequence upstream from p_{mod} was transferred to the λ*lac*/*trp* vector λAD2. In the resulting phage (λAD3), transcription of the *lacZ* gene is dependent on either p_{trp} or the leftward promoter of λ (p_L) and, in either case, must traverse the intergenic segment of *hsd* DNA (see Fig. 3). A functional *lacZ* gene is indicated in lytic infection when plaques are assayed on indicator medium (Miller, 1972). In a lysogen of the *trpR* host (*psu*0), however, when transcription from p_L is repressed, and only p_{trp} is active, little *β*-galactosidase is detected (Table 6). This suggests a transcription termination signal in λAD3 between p_{trp} and *lacZ*. Analysis of the DNA sequences flanking the new junctions of λAD3 showed no sequence homology, hence the only stem-loop structures predicted are within the intergenic DNA. The failure to express *β*-galactosidase does not appear to be the result of translational polarity, since the level remains low even in *psu3*, a *rho* derivative of *psu0* (Table 6). The data suggest a termination signal in the DNA segment separating *hsdR* and *hsdM*. This would

make transcription of *hsdM* and *hsdS* dependent on p_{mod}.

4. Discussion

The nucleotide sequence of an 8·4 kb segment of DNA confirms the gross organization of the three genes of the *hsd* region of *E. coli* K-12, previously deduced by genetic analyses (Sain & Murray, 1980). Our data indicate two transcriptional units; the first includes only *hsdR* and the second both *hsdM* and *hsdS*. Since the products of the latter two genes are sufficient for the K-specific methylase, this organization could allow transcription from p_{mod} to provide methylase and consequent modification of DNA before the activation of p_{res} and the production of the R polypeptide, an essential component of the restriction enzyme complex. Furthermore, since there is a target for the K-methylase in the immediate vicinity of p_{res} (see Fig. 2), methylation might influence the efficiency of this promoter. Our data provide no evidence for transcriptional control of *hsdR* by the K-specific methylase, nor indeed by either one of its constituent polypeptides. Translational control remains a possibility, as does an explanation based on a difference between the kinetics of association of the modification and restriction complexes. If the latter were so, differential strengths of the two promoters could ensure that on transfer to a new cellular environment the production of methylase would precede that of the restriction enzyme. At present we have no comparative estimate of the strengths of the two promoters from a prophage system, although preliminary estimates using the plasmid pK06 indicate that p_{mod} may be twice as effective as p_{res} (our unpublished data).

We have no direct evidence to implicate any mechanism of post-transcriptional control. Nevertheless, biological experiments in which the possibility of co-ordinated control of the *hsd* transcripts is prevented can result in "self-destruction". This is true of host DNA when a phage encoding the K-specific methylase enters an unmodified cell in which a multicopy plasmid expresses the R polypeptide (J. E. Kelleher & A. Daniel, unpublished observations). Similarly, when a K-modified *λhsd* phage in which the *hsdR* gene is transcribed from p_L enters a restriction-deficient cell in which the host chromosome encodes a B-specific modification enzyme, the phage DNA is attacked by the resulting B-specific endonuclease (Table 5). In these two encounters the R polypeptide is assumed to be over-produced either as the result of increased gene dosage, or increased copy number aggravated by a strong promoter. Similar results are available for the *E. coli* A and E specificity systems, although the entire *hsd* region is readily transferred to an *hsdR* strain of *E. coli* by either transformation or conjugation (Fuller-Pace *et al.*, 1985; Suri & Bickle, 1985).

Close to *hsdS* is a gene previously referred to as gene X (Sain & Murray, 1980). The absence of this gene (strain NM610 in Table 6) has no effect on transcription from either p_{res} or p_{mod}. We conclude that this gene, which is immediately downstream of *S*, is not directly relevant to expression of the *hsd* genes. *mcrB*, a gene whose product is involved in the restriction of methylated DNA, has been located very close to *hsd* (Raleigh & Wilson, 1986), and gene X seems to be part of the *mcr* system (Ross & Braymer, 1987). The deletion in NM610, and one that begins in *hsdS* and extends less than 1000 bases downstream, are associated with the loss of both the *hsd* and *mcrB* restriction systems (our unpublished observations).

The overlap of the *hsdM* and *S* genes (see Fig. 2) is a feature consistent with translational coupling (Oppenheim & Yanofsky, 1980). The simplest expectation would be that the M and S polypeptides are required in equimolar amounts. Although it has been claimed that the ratio of M to S in the *Eco*K complex is 2 : 1 (Meselson *et al.*, 1972), a ratio of 1 : 1 in the K-methylase has been deduced by Suri *et al.* (1984). If the ratio in both complexes is the same, the results based on the methylase are preferable, since they depend on the analysis of a much smaller protein complex. The equimolar ratio is consistent with the simplest prediction of translational coupling, involving an efficient Shine–Dalgarno sequence upstream from *hsdS* (Shine & Dalgarno, 1974).

The interval between *hsdR* and p_{mod} is longer than required for a signal to terminate transcription. Its length may suggest some further functional significance. Alternatively, it could be a relic of the evolutionary origin of a restriction system from a sequence-specific methylase by the addition of an extra gene. In this context, we note that for the type I restriction system, R124, in contrast to the K and A systems, the *hsdR* gene is oriented in the opposite direction from that of *hsdM* and *hsdS* (C. Price & T. A. Bickle, personal communication).

Our analyses of the predicted amino acid sequences of the *Eco*K polypeptides identify two pronounced α-helical regions in R, one of which is amphipathic and, within R and M, regions that may be relevant to the binding of ATP and DNA. A well-conserved amino acid sequence common to adenine methylases is identified in the M polypeptide. Definitive correlations await the analysis of mutants in combination with structural studies of the polypeptides.

We are indebted to Professor T. Blundell, Birkbeck College, London, for his analysis of the predicted α-helical regions in the R polypeptide, Dr John F, Collins, this Department, for computer searches on this region, and Mr J. Keyte, Department of Biochemistry, University of Leicester, for oligonucleotides. We also appreciate the help of our colleagues, in particular, John Collins, Gill Cowan and Alexander Gann for critical discussion of the manuscript, Betty McCready for preparing the manuscript, and Annie Wilson for drawing the Figures. The research was supported by the Medical Research Council.

References

Appleyard, R. K. (1954). *Genetics*, **39**, 440–459.
Argos, P. (1985). *EMBO. J.* **4**, 1351–1355.
Benton, W. D. & Davis, R. W. (1977). *Science*, **196**, 180–182.
Bickle, T. A. (1982). In *The Nucleases* (Linn, S. M. & Roberts, R. J., eds), pp. 85–108, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Bickle, T. A., Brack, C. & Yuan, R. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 3099–3103.
Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 3963–3965.
Bougueleret, L., Schwarzstein, M., Tsugita, A. & Zabeau, M. (1984). *Nucl. Acids Res.* **12**, 3659–3676.
Boyer, H. W. & Roulland-Dussoix, D. (1969). *J. Mol. Biol.* **41**, 459–472.
Brooks, J. E., Blumenthal, R. M. & Gingeras, T. R. (1983). *Nucl. Acids Res.* **11**, 837–851.
Burkhardt, J., Weissmann, J. & Yuan, R. (1981). *J. Biol. Chem.* **256**, 4024–4032.
Deininger, P. L. (1983). *Anal. Biochem.* **129**, 216–223.
Devereux, J., Haeberli, P. & Smithies, O. (1984). *Nucl. Acids Res.* **12**, 387–395.
Drummond, M., Whitty, P. & Wootton, J. (1986). *EMBO J.* **5**, 441–447.
Eskin, B. & Linn, S. (1972). *J. Biol. Chem.* **247**, 6183–6191.
Finch, P. W. & Emmerson, P. T. (1984). *Nucl. Acids Res.* **12**, 5789–5799.
Fuller-Pace, F. V. & Murray, N. E. (1986). *Proc. Nat. Acad. Sci., U.S.A.* **83**, 9368–9372.
Fuller-Pace, F. V., Cowan, G. M. & Murray, N. E. (1985). *J. Mol. Biol.* **186**, 65–75.
Gay, N. J. & Walker, J. E. (1981). *Nucl. Acids Res.* **9**, 2187–2194.
Glover, S. W. & Colson, C. (1969). *Genet. Res.* **13**, 227–240.
Gough, J. A. & Murray, N. E. (1983). *J. Mol. Biol.* **166**, 1–19.
Grantham, R., Gautier, C., Guoy, M., Jacobzone, M. & Mercier, R. (1981). *Nucl. Acids Res.* **9(1)**, r43–r74.
Greene, P. I., Gupta, M., Boyer, H. W., Brown, W. E. & Rosenburg, J. M. (1981). *J. Biol. Chem.* **256**, 2143–2153.
Hadi, S. M. & Yuan, R. (1974). *J. Biol. Chem.* **249**, 4580–4586.
Hadi, S. M., Bickle, T. A. & Yuan, R. (1975). *J. Biol. Chem.* **250**, 4159–4164.
Hanahan, D. (1983). *J. Mol. Biol.* **166**, 557–580.
Hansen, E. B., Hansen, F. G. & von Meyenburg, K. (1982). *Nucl. Acids Res.* **10**, 7373–7385.
Hopkins, A. S., Murray, N. E. & Brammar, W. J. (1976). *J. Mol. Biol.* **107**, 549–569.
Hu, N. & Messing, J. (1982). *Gene*, **17**, 271–277.
Hubacek, J. & Glover, S. W. (1970). *J. Mol. Biol.* **50**, 111–127.
Korn, L. J. & Yanofsky, C. (1976). *J. Mol. Biol.* **106**, 231–241.
Lewis, S. A., Balcarek, J. M., Krek, V., Shelanski, M. & Cowan, N. J. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 2743–2746.
Linn, T. & Ralling, G. (1985). *Plasmid*, **14**, 134–142.
Loenen, W. A. M. & Murray, N. E. (1986). *J. Mol. Biol.* **190**, 11–22.
Lyall, A., Hill, C., Collins, J. F. & Coulson, A. F. W. (1986). In *Parallel Computing 85* (Feilmeier, M., Joubert, G. & Schendel, U., eds), pp. 235–240, Elsevier Science Publishers B.V., North Holland, Amsterdam.
Mandel, M. & Higa, A. (1970). *J. Mol. Biol.* **53**, 159–162.
Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). Editors of *Molecular Cloning; A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
McClarin, J. A., Frederick, C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J. & Rosenberg, J. M. (1986). *Science*, **234**, 1526–1541.
McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C. & Rosenberg, M. (1981). In *Gene Amplification and Analysis* (Chirikjian, J. G. & Papas, T. S., eds), vol. 2, pp. 383–415, Elsevier North-Holland, Amsterdam.
Meselson, M., Yuan, R. & Heywood, J. (1972). *Annu. Rev. Biochem.* **41**, 447–466.
Messing, J. (1983). *Methods Enzymol.* **101**, 20–78.
Miller, J. M. (1972). In *Experiments in Molecular Genetics*, pp. 352–383, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Miller, J., McLachlan, A. D. & Klug, A. (1985). *EMBO J.* **4**, 1609–1614.
Murray, N. E. & Brammar, W. J. (1973). *J. Mol. Biol.* **77**, 615–624.
Murray, N. E., Brammer, W. J. & Murray, K. (1977). *Mol. Gen. Genet.* **150**, 53–61.
Nakayama, N., Arai, N., Bond, M. W., Kaziro, Y. & Arai, K. (1984). *J. Biol. Chem.* **259**, 97–101.
Newman, A. K., Rubin, R. A., Kim, S. H. & Modrich, P. (1981). *J. Biol. Chem.* **256**, 2131–2139.
Oppenheim, D. S. & Yanofsky, C. (1980). *Genetics*, **95**, 785–795.
Pabo, C. O. & Sauer, R. T. (1984). *Annu. Rev. Biochem.* **53**, 293–321.
Pai, E. G., Sachsenheimer, W., Schirmer, R. H. & Schultz, G. E. (1977). *J. Mol. Biol.* **114**, 37–45.
Raleigh, E. A. & Wilson, G. (1986). *Proc. Nat. Acad. Sci., U.S.A.* **83**, 9070–9074.
Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977). *J. Mol. Biol.* **113**, 237–251.
Ross, T. & Braymer, H. D. (1987). *J. Bacteriol.* **169**, 1757–1759.
Sain, B. & Murray, N. E. (1980). *Mol. Gen. Genet.* **180**, 35–46.
Sanger, F., Coulson, A. R., Barell, B. G., Smith, A. J. H. & Roe, B. A. (1980). *J. Mol. Biol.* **43**, 161–178.
Saraste, M., Eberle, A., Gay, N. J., Runswick, N. J. & Walker, J. E. (1981). *Nucl. Acids Res.* **9**, 5287–5296.
Sauer, R. T. (1978). *Nature (London)*, **276**, 301–302.
Shine, J. & Dalgarno, L. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 1342–1346.
Staden, R. (1982). *Nucl. Acids Res.* **10**, 4731–4751.
Suri, B. & Bickle, T. A. (1985). *J. Mol. Biol.* **186**, 77–85.
Suri, B., Nagaraja, V. & Bickle, T. A. (1984). *Curr. Top. Microbiol. Immunol.* **108**, 1–9.
Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). *EMBO J.* **1**, 945–951.
Wharton, R. & Ptashne, M. (1985). *Nature (London)*, **316**, 601–605.
Windass, J. D. & Brammar, W. J. (1979). *Mol. Gen. Genet.* **172**, 329–337.
Wood, W. B. (1966). *J. Mol. Biol.* **16**, 118–133.
Yuan, R. (1981). *Annu. Rev. Biochem.* **50**, 285–315.
Yuan, R. Hamilton, D. L. & Burckhardt, J. (1980). *Cell*, **20**, 237–244.
Zabeau, M., Friedman, S., van Montague, M. & Schell, J. (1980). *Mol. Gen. Genet.* **179**, 63–73.

Edited by S. Brenner