© 1991 Oxford University Press
Nucleic Acids Research, Vol. 19, No. 21 5831–5838

Gene organization in the multiple DNA inversion region Min of plasmid p15B of *E.coli* 15T⁻: assemblage of a variable gene

Heinrich Sandmeier[1+], Shigeru Iida[1,2], Philipp Hübner[1], Rosemarie Hiestand-Nauer[1] and Werner Arber[1*]
[1]Abteilung Mikrobiologie, Biozentrum der Universität Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland and [2]Department of Biological Sciences & Technology, Science University of Tokyo, Yamazaki, Noda-Shi, Chiba 278, Japan

Received September 19, 1991; Accepted October 14, 1991

EMBL accession no. X62121

ABSTRACT

The bacteriophage P1-related plasmid p15B of *E. coli* 15T$^-$ contains a 3.5 kb long region which frequently undergoes complex rearrangements by DNA inversion. Site-specific recombination mediated by the Min DNA invertase occurs at six crossover sites and it eventually results in a population of 240 isomeric configurations of this region. We have determined 8.3-kb sequences of the invertible DNA and its flanking regions. The result explains how DNA inversion fuses variable 3′ parts to a constant 5′ part, thereby alternatively assembling one out of six different open reading frames (ORF). The resulting variable gene has a coding capacity of between 739 and 762 amino acids. A large portion of its constant part is composed of repeated sequences. The p15B sequences in front of the variable fusion gene encode a small ORF and a phage-specific late promoter and are highly homologous to P1 DNA. Adjacent to the DNA invertase gene *min*, we have found a truncated 5′ region of a DNA invertase gene termed *ψcin* which is highly homologous to the phage P1 *cin* gene. Its recombinational enhancer segment is inactive, but it can be activated by the substitution of two nucleotides.

INTRODUCTION

Recently, the prophage P1-related plasmid p15B of *E. coli* 15T$^-$ was shown to carry a complex DNA inversion system called Min (1). Its Min site-specific recombinase alternatively acts at any two out of a total of six crossover sites (*mix*) provided that the two considered sites are carried in inverted orientation. Successive recombinations can lead to 240 isomeric forms of the Min region (2). This p15B Min system belongs to the Din family of functionally related prokaryotic DNA inversion systems (1, 2), which are found in the genomes of phages P1 and Mu to alter their host range, in *Salmonella typhimurium* to alternatively express two different flagellin genes, and in the e14 element of some *Escherichia coli* strains (for review see Ref. 3). A characteristic property of these site-specific recombination systems is that efficient DNA inversion requires not only the system encoded Din recombinase and two *dix* crossover sites in inverted orientation but also the presence of a recombinational enhancer element and the FIS protein. The approximately 70 bp long enhancer sequence to which host protein FIS binds resides within the recombinase gene and it acts *in cis*, independent of its orientation, and widely independent of its distance from the crossover sites (for review see Ref. 4). It has been shown that Min also carries a recombinational enhancer which stimulates DNA inversion 300-fold (1).

The well-studied Din systems carry one invertible DNA segment between two crossover sites and they act as two-state genetic switches to alternatively express one of two sets of genes or gene variants. Since the Min system carries several invertible segments defined by six crossover sites and can generate 240 isomeric forms of its 3.5-kb DNA inversion region we wondered whether Min can switch between several genes or gene variants. Therefore, we determined the DNA sequence of a 8.3 kb long fragment which includes the Min invertible DNA. The obtained sequencing data met our expectations and revealed that DNA inversion assembles a variable ORF with six alternative 3′ ends. The data exhibited also that the variable gene and two adjacent genes are controlled by the phage P1-specific late promoter P_S. In addition to the *min* gene p15B carries a truncated DNA invertase gene, termed *ψcin*. It was thus likely that the Min region contained two recombinational enhancer sequences, an active one within *min* gene (1) and another one within *ψcin*. Functional analysis of this second putative enhancer sequence did not reveal any stimulation of DNA inversion. However, it becomes activated by altering two nucleotides within the *ψcin* fragment by means of site-directed mutagenesis. As was expected from earlier

* To whom correspondence should be addressed

+ Present address: Abteilung für Präventivzahnmedizin und orale Mikrobiologie, Zahnärztliches Institut der Universität Basel, Petersplatz 14, CH-4051 Basel, Switzerland

electron microscopic studies (5, 6) the DNA sequences which flank the invertible segments were found to be highly homologous to P1 DNA. This homology includes on the one side the P_S promoter, a small ORF and a 5′ end of the constant part of the variable gene and on the other side the *ψcin* gene. Based on these data an evolutionary implication of the gene organization of the Min region is discussed.

MATERIALS AND METHODS

Bacterial strains

E. coli K-12 strain WA921 *pin*, hsd_K, *lac*, *leu*, *met*, *thr*, *supE* and strain WA3782 *recA*, *pin*, *lac*, *met*, *supE*, hsd_S were described before (7, 8).

Plasmids

The pUC19-based plasmid pAW800 contains the Min DNA inversion system on the 8.3 kb long p15B *Bam*HI-5 fragment (1). DNA inversion tester plasmids pPHU9 and pPHU81 were described before (9). Plasmid pPHU133 was constructed by the insertion of a 200 bp long *Ban*I/*Dra*III fragment containing the p15B *ψcin* recombinational enhancer into the *Sma*I site of pPHU9. To mutagenize the *ψcin* enhancer a *Sph*I/*Sac*I fragment of pPHU133 was subcloned into the *Sph*I/*Sac*I sites of M13mp19. Site-directed mutagenesis (10) was performed with the synthetic oligodeoxyribonucleotide dGTTCATTTGTTGATATGCG. The mutated *Sph*I/*Sac*I fragment was then cloned into the *Sph*I/*Sac*I sites of pPHU9 yielding the plasmid pPHU134.

Assays for DNA inversion *in vivo*

Tester plasmids were transformed into WA3782(pPHU78) cells. Plasmid pPHU78 confers kanamycin (Km) resistance and overexpresses the Cin recombinase from the *lac*UV5 promoter (9). DNA inversion tester plasmids carry in addition to the *bla* gene for ampicillin (Ap) resistance a promoterless *cat* gene for resistance to chloramphenicol (Cm) which is activated by an operon fusion upon DNA inversion (11). The frequency of DNA inversion is then scored as described earlier (1).

Laboratory procedures

DNA sequencing used the chain termination method (12). DNA and protein sequences were analyzed with programs of the Genetics Computer Group (GCG) of the University of Wisconsin (13).

RESULTS AND DISCUSSION

Determination of the DNA sequence of the p15B Min region

Plasmid pAW800 contains the complete Min DNA inversion region within a 8.3-kb *Bam*HI fragment originating from p15B (1). The DNA sequence of this fragment was determined as sketched in Fig. 1. Since the Min region can adopt 240 isomeric forms through DNA inversion, we employed pAW800 derivatives carrying the invertible segments in a frozen configuration for sequencing the invertible region. Such plasmids were obtained by either deleting the *min* recombinase gene or by deleting all inversely oriented crossover sites as described before (2). The DNA sequence of the 8294 bp long p15B *Bam*HI-5 fragment shown in Fig. 2 includes the already published sequences of the *min* recombinase gene (1) and the six *mix* crossover sites (2).

Min inversion assembles a fusion gene with a constant 5′ part and one out of six alternative 3′ parts

A striking feature of the sequence is a long ORF termed *S*, which codes for 747 amino acids (aa). The major part of ORF *S* is encoded in segment R (*Sc*, 594 aa), but a 153 aa long 3′ end (*SvMl*) extends beyond the leftmost crossover site in Fig. 1 into the invertible segment M. DNA inversion of segment M substitutes the 3′ end of gene *S* (*SvMl*) by an alternative 166 aa

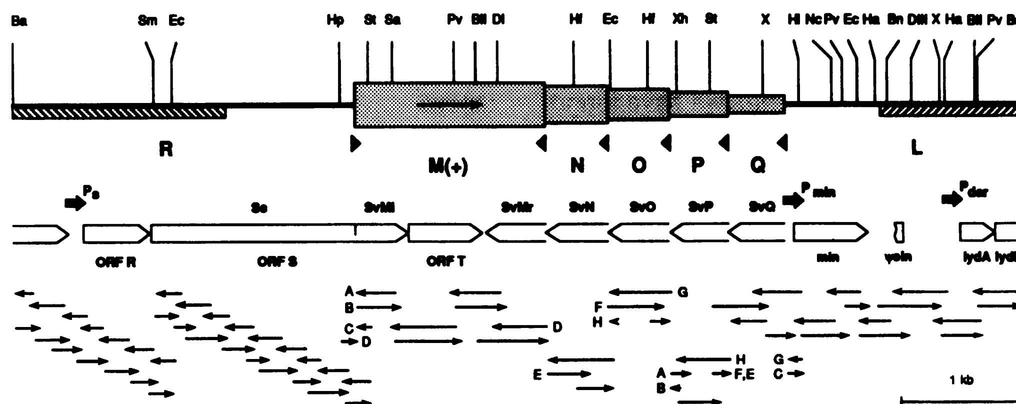

Figure 1. Map of the p15B *Bam*HI-5 fragment of recombinant plasmid pAW800, strategy for DNA sequencing, and gene organization in the Min region. Note that the configuration of invertible segments M, N, O, P and Q (stippled boxes) and the flanking segments R and L (thick lines) is chosen arbitrarily from 240 possible isomeric forms. The orientation of the invertible segment M in which ORF *T* is in the same direction as genes *R* and *S* is arbitrarily designated as M(+). The orientation of *mix* crossover sites is indicated by a filled-in triangle. The sequence of the R segment was deduced from *Exo*III generated deletion derivatives of two overlapping subfragments. DNA sequences of the invertible segments M, N, O, P and Q have been analyzed in plasmids with frozen configuration (see text). The configuration of some of these plasmids does not correspond to the one shown in the map. Hence, the length and the direction of DNA sequence determination could in those cases not be represented by a contiguous arrow and capital letters are used to identify the contiguous segments represented by split arrows. The sequence of segment L was derived from several subcloned fragments. ORFs deduced from the DNA sequence are represented by wide open arrows. Small, filled-in arrows indicate the location of promoter elements. For detailed description of ORFs *R*, *S*, *T*, *min*, *ψcin*, *lydA* and *lydB* and promoters P_S, P_{min} and P_{dar} see text. The hatched boxes represent sequences which are homologous to corresponding parts of the phage P1 genome (see text). Restriction sites are: Ba=*Bam*HI, BII=*Bgl*II, DI=*Dra*I, DIII=*Dra*III, Ec=*Eco*RV, Ha=*Hae*II, Sa=*Sal*I, Sm=*Sma*I, St=*Stu*I, Xh=*Xho*I. Selected sites relevant for sequencing or cloning are: Bn=*Ban*I, Hi=*Hinc*II, Hf=*Hin*fI, Hp=*Hpa*I, Nc=*Nco*I, Pv=*Pvu*II, X=*Xmn*I.

R

```
GGATCCCATATATGCTCCGGTGGACTTTGAACGTCATCCCTATGGCACGGTCCTGATTCCAGAAAGCACACTGGAAACCACCGGCGGCACATTCGGCGAG   100
BamHI
ATGTTTCTGACTTCCAGAGGAATGATCAGTATTCCCATAAACGACCTGGCCCGGACAATGGGTATTACTGGAACGATAGATCAGTCCGCAATTACAGAAG   200
AAATTCTCAGAAAGTTTAATCAGTTCGTAAAGCCTCTACTGCCACTGCATATAGTGTTTGATGGGCTTACGCTCTATTTGTCGGTTGTTGTAAATGAACA   300
GGCCGACATGATTACTTTGAATGAGATTTCTGATACCGAAAAAGCATTCTGCTGGTTTGAAACTTCGGATACAACTTCGCTTACTGAAGTTACGTCGATT   400
AACGCCCCGATCACTGCAACGCCGGGCGGCACTATTGTGAAAGCAACGCCTACGTTTGATCGCACCCGCGCAGATGATTTGCTGTTGGATAGCGATGCGT   500
                                                                        stop unidentified ORF ->|
GACAATCACCCGTCCGCAGGCGGTGACAAGTTACTTATCTTACAATGAGGCTTCACAACATTGATTAGGGAAAATCATGTCTGACGTCTCAACAAACCTC   600
                           Ps -22      .Ps -10                                 |-> start R
TATAAGAGTCAGTTGTTGGACTATTACTATCAGCGGCGCGCTGAATCGTCCATTAACAAAGGCTCTCGATTTTTAATCAGCAAGGCCGTTTTTGGTACCA   700
GTTCACTGGTTACTAAGAAAGGAGATGGCACTTATGAGATTGGAGAACTGCCTAAGGCTTTCGATCTGGCAGAACTGACCAGTCAATTTTGCACCATCAA   800
CCTCGTCCCAACCTACTCAGGCGGGATAATTACTGTCCGAATGGACCTTGATCAAAGTCAGTTGCAGGAAGGGAAAAACTACCCATTCAACACTCTGGTT   900
GTTCTGGATAACGAGAATAAGCCAATCGCCATTATTTGTGTCCAGGAAGACTCGCTGTATGTGGGCAAAACATATACCGCAGTTATGGCCATAAACTCGA   1000
CTACAGCATAAGGATATGCTTGATGAATGACGTTACAGTTGTTACATCGGTTACTTACCCATCACCCGAGTCGTTGGCTCTGGTGGCTGATGTGCAATAC   1100
stop R ->|      |-> start S
CACGAACCATATCTGTCAGCCGCTCTAAACCGAAAATTCAGGGGAATTGTTGACCCGGGATTTTATGCTGGTTTCTTGCCTAAGCCTGGCGGTGGGATGA   1200
                                                      SmaI
ACCTGTTAATCACCTCAGTGGATGGAGATAAAACCGCTGGCGCGGCGTCAGTGGATATTGGTGAATTCTACCAGGTAACTATTCAGCATCGTAAGGATAT   1300
                                                                                                EcoRV
CTCTCTTGCACTTAACGCAGGCAAGAAATATGCAATTGTGCTGAAGGGAAGATACCTTCTTGGAGAAGATACCTATCAGGTGAATACCGCGTCACATATT   1400
CATGCAGCTGAATTTGTTGCCAGAACCTATACCGATTCATATCAGTTAGGTGATGGGGAACTGCTGGTTTGTACGGTGAATATCCCTGCTGGCGTATCTA   1500
CCATTACTCAAGAGATGATTGATACATCCGAGCGTATCAACCGCACGATCGGCATTGATATTTCAGACTCTGTAACCAGTACCAGAAGTGATGTTGCTGC   1600
GAGTTCGCTGGCAGTTAAAAAAGCCTACGATCTGGCGAAAAGCAAGTATACGGCGCAGGATGCAAGCACAACGCAAAAGGGATTAGTTCAGCTCAGTAGT   1700
          |           re 0               | re 1
GCCACTAACAGCACGTCTGAAGTGCTGGCCGCCACACCGAAAGCTGTCAAGGCTGCATATGACCTGGCTAACGGGAAGTATACAGCCCAGGATGCAACCA   1800
                                                                          re 1 | re 2
CGACACAAAAAGGGATAGTTCAGCTCAGTAGCGACACCAACAGCACTTCTGAAACATTAGCTGCAACTCCAAAAGCGGTTAAAGCTGCATACGATCTAGC   1900
AGCCGGAAAGGCACCATCCAGTCATACACATCCCTGGAATCAGATCACTGGTGTGCCAACAGCTTCATTGACAGCGAAAGGCATCACTCAGCTCAGTAGT   2000
    re 2 | re 3                                                     :
GCCACTAACAGCACGTCTGAAGTGCTGGCCGCCACACCGAAAGCTGTCAAGGCTGCATATGATTTGGCTAACGGGAAGTATACAGCCCAGGATGCAACCA   2100
                                                                          re 3 | re 4
CGGCTCAAAAAGGGATAGTCCAACTCAGCAGTGCTACCAATAGCACGTCCGAAGTGCTGGCCGCCACACCGAAAGCTGTCAAGGCTGCATATGACCTGGC   2200
TAACGGGAAGTATACAGCCCAGGATGCAACCACGACACAAAAAGGGATAGTTCAGCTCAGTAGCGACACCAACAGCACTTCTGAAACATTAGCTGCAACT   2300
    re 4 | re 5
CCAAAAGCGGTTAAAGCTGCATACGATCTAGCAGCCGGAAAGGCACCATCCAGTCATACACATCCCTGGAATCAGATCACTGTTGTGCCAACAGCTTCAT   2400
                                    re 5 | re 6
TGACAGCGAAAGGCATCACTCAGCTCAGTAGTGCCACTAACAGCACGTCTGAAGTGCTGGCCGCCACACCGAAAGCTGTCAAGGCTGCATATGATTTGGC   2500
             :
TAACGGGAAGTATACAGCCCAGGATGCAACCACGGCTCAAAAAGGGATAGTCCAACTCAGCAGTGCTACCAATAGCACGTCCGAAGTGCTGGCCGCCACA   2600
    re 6 | re 7
CCGAAAGCTGTCAAGGCTGCATATGATTTGGCTAATGGAAAGCAAGCGGCAGACGCTACGCTCACTGCTCTAGCGGCACTAGCTACAGCAGCAGATAAAC   2700
                                    re 7 |
TCCCTTATTTCACAGGTGTTGATCGTGCCGCGTTAACTGCATTGACAAGTGTTGGACGCGCCATTCTTGGTAAGACCAGTATTCAGAGCCTTCTTGATTA   2800
                                  HpaI
CCTTGGTTTGGGGGAAGGCTCTGCACTGCCTGTTGGTGTGCCCGTTCCGTGGCCCTTAGAAACACCACCAACGGGCTGGCTAAAATGCAATGGTGCAGCA   2900
  mixR'M1"
TTTTCTTCTGAAATGTATCCCAAACTGGCAAAGGCCTACCCCACCAATAAATTACCGGATTTACGCGGTGAGTTTATCCGTGGTTGGGATGATGGGCGAG   3000
                                  StuI
GTGTGGATGCGGGAAGGGTCATCTTAAGCATACAGGGGTGGTTAACAGGAAGTCATTATCATAATATTCGGTCATGGGACGCGTGGGATAACACGGTATT   3100
GGTGCCAAATGACAGAGGGGGGGATAGTCTGTTGTCGACAGATAACGCCGTCCGGCAAGGAGCGATTAATGGTAAATTTACCAGTCAATACAGAACGGAG   3200
                                    SalI
TTATCTGGGGGGAATGAAACCCGCCCACGTAACATTGCCTTCAATTATATTGTGAGAGCAGCATAATGGATAATGCGATATTAAATAGCGAACTTATAGC   3300
                                                       stop SvM1 ->| |-> start T
CATACAGGCAGGAAACATTATCGTTTATAACTATGATGGTGGTAATCGGGAATATATTTCTGCATCAACTGAATATCTTGCTGTTGGCGTTGGTATTCCG   3400
GCAAATTCTTGTTTGGATGCTCCAGGCTCACATAAAGCAGGTTATGCGATTCTCCGTTCAGAGGATTTAAGTTCATGGGAGTATGTGCCAGATCATCGTG   3500
GCGAAACTGTCTATAGCATTGACACAGGGAATCCCGAAGAAATCACGGTGTTGGGTGACTATCCGGAAAATACAACCACTATCGCCCCGCTAACACCATA   3600
CGACAAATGGGATGGAGAGAAATGGGTGGTTGATACTGAGGCTCAACATAGTGCAGCTGTAGAGGCAGCAGAAACAAAACGTCAGTCATTGATTGATACT   3700
                                                         PvuII
GCGATGGATTCCATTAGTCTGATTCAGTTGAAATTACGGGCTGGACGGAAGTTGACGCAGGCAGAAACCACGCAGCTTAACTCCGTGCTAGATTATATAG   3800
ACGAGCTGAACGCGATGGATTTAACCACGGCACCAGATCTCAACTGGCCTGAAAAACAACTTTCTACAGCCAGTTGATATAATTAATCCGCCCTCACAAT   3900
                                   BglII.                               stop T ->|    |<- stop SvMr
ATAATTAAATGCAATATTACGAGGGCGGGTTTCTAAGCCAGCAGTACCTAAATTAGCTACGGACTGTTTATATGTTTTAAAAGTTCCATAATTAGGGGCT   4000
                                                                             DraI
GGCAATCCTGCATCGCTTGTGTTCCCCCATCTGATAACTCCCGTATTACCACTAGTAGCCCATGCCTCGTCAAAGTAGAAATTAATTGAGCGGTCAGTTG   4100
CTACGGTTGATTTTGACGGTAATCCGTGAGCATGATCCTCCGTTGCATATCCTTGCTCGCTTAAAATAGAGCGACCAGAGTCAATCCCTCGCCCATCATC   4200
CCAGCCACGGATAAACTCACCGCGTAAATCAGGTAATTTTAACGTTGGGTAAACCTTTGCCAGATTTGGGTACTTTTCAGAAGAAAATGCTGCACCATTG   4300
```

M(+)

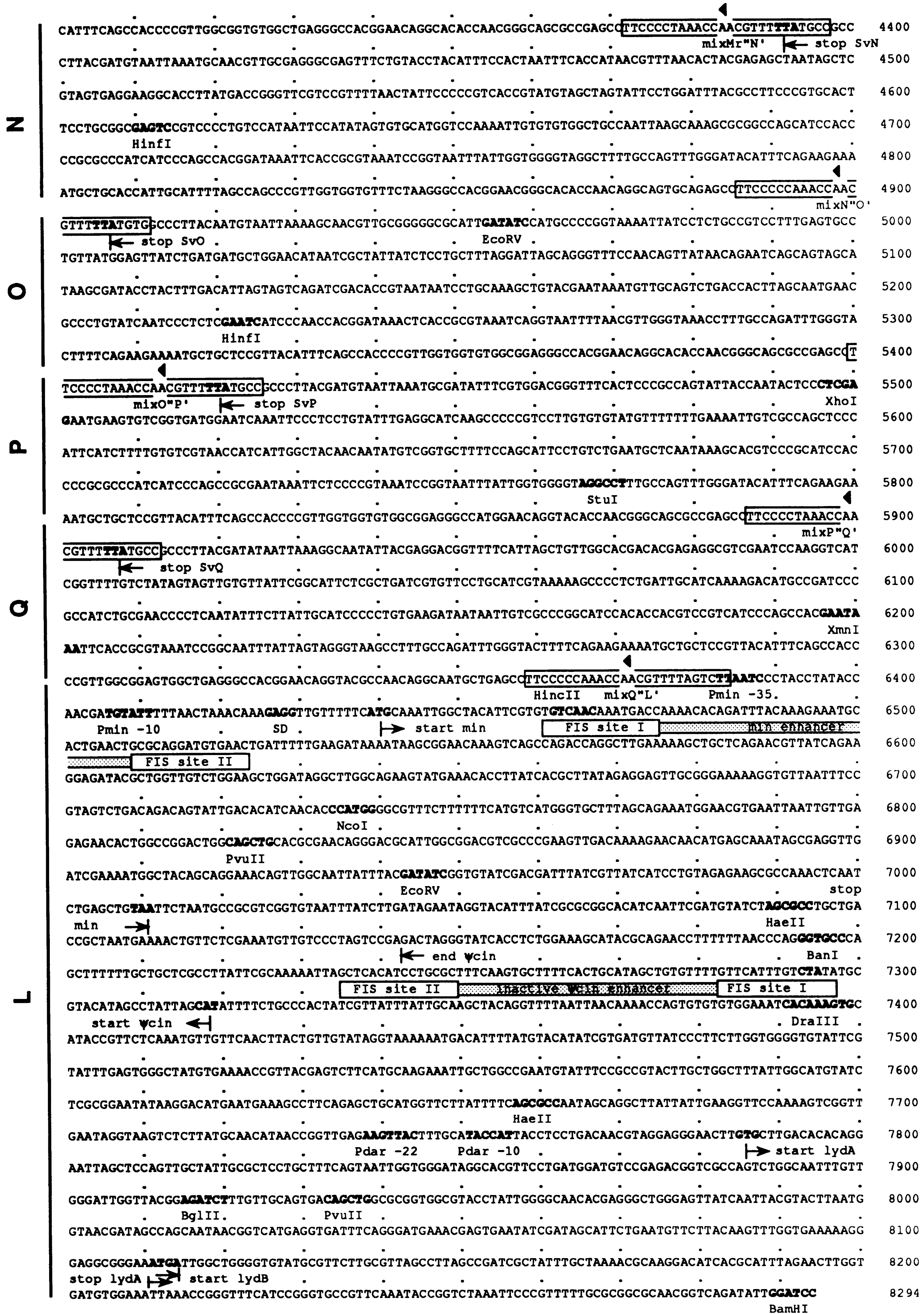

Figure 2. DNA sequence of the p15B *Bam*HI-5 fragment including the Min DNA inversion system in configuration RM(+)NOPQL. The sequence of the imperfect palindromic *mix* crossover sites are boxed and their orientation, which is determined by the asymmetrical central dinucleotides (AA or in inverse orientation TT), is indicated by filled-in triangles. Recognition sites of restriction enzymes, promoter elements, potential Shine-Dalgarno (SD) sequences and start and stop codons for translation are printed in bold. The location of repeated elements in the R segment (re0 to re7) and the location of recombinational enhancer sequences are indicated. Vertical stippled lines within re3 and re6 mark the boundaries of small sequences which are not homologous to the other repeated elements (see also Fig. 5).

long 3′ end termed *SvMr* (Fig. 3A and 3B). If Min recombination inverts segment M in conjunction with one or several of the segments N, O, P and Q (as exemplified in Fig. 3B and 3C) the 3′ end of gene *S* is substituted by either *SvN* (168 aa), *SvO* (168 aa), *SvP* (159 aa), or *SvQ* (145 aa). Therefore, *S* is a variable gene and Min inversion assembles one constant 5′ part and one out of six different 3′ parts. In addition Min recombination inverts also *T*, a 203-aa ORF encoded within segment M and hence provides an on and off switch for its expression in the M(+) and M(−) orientation, respectively (see also below).

The precise fusion of the variable parts (*Sv*) to the constant part (*Sc*) is the consequence of the particular gene organization in the Min region and the fidelity of the DNA inversion reaction.

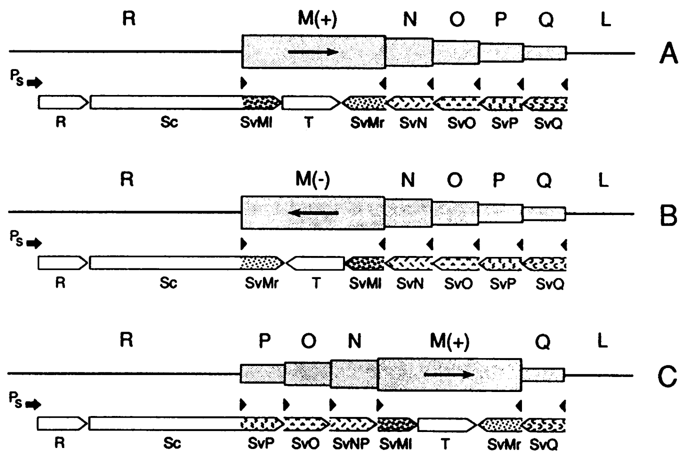

Figure 3. Representative drawing of 3 successive configurations of the multiple invertible segments of the Min system. Upon site-specific DNA inversion of segment M the *SvMl* portion of gene *S* as drawn in **A** becomes replaced by *SvMr* as drawn in **B**. A consecutive recombination at the *mix* sites flanking the invertible segments MNOP replaces *SvMr* by *SvP* (see C). By analogous DNA inversions each of the six 3′ variable parts can be fused to the constant 5′ part of gene *S*. Hence, the diversity of 240 isomeric forms (2×5!) of this 3.5-kb DNA region serves to produce 6 different alternatively fused forms of gene *S*. The black arrow indicates the location of the P_S promoter, which transcribes genes *R*, *S* and if in the right configuration, *T*. For other symbols see Fig. 1.

Min inversion and the related DNA inversions are conservative site-specific recombination reactions. DNA is cut at the central dinucleotides (AA in all of the *mix* sites) of the two recombining crossover sites carried in inverted orientation and the resulting 2 bp long 3′ overhangs are subsequently religated in absence of DNA synthesis (14−16). Since all of the *Sv* ORFs are in the same phase relative to the crossover point, each recombinatiòn event at the crossover site flanking *Sc* assembles with high fidelity a variant of gene *S*. The detached variable parts are put in reserve without sequence alteration and can be used for successive gene fusions

The alternative fusion of a constant part with one of several variable parts of a gene by Min inversion is reminiscent of the consequences of DNA inversion occurring in a Shufflon system on IncI plasmid R64 (17), in which the Shufflon site-specific DNA recombinase inverts seven DNA segments independently or in groups and thereby fuses one of seven variable 3′ parts of ORFs to one constant 5′ ORF. Although both the Shufflon and the Min systems contain similar complex invertible DNA regions, their recombinase systems are not related (1, 2).

Repeated sequences within the Min region

Dotplot comparison of the inversion region (Fig. 4) revealed that the six variable parts, *Sv*, share two domains with approximately 80% homologous sequences interrupted by completely dissimilar sequences (see also Fig. 5).

The constant part, *Sc*, contains an array of directly repeated sequences, termed re0 to re7 (Fig. 5). Detailed analysis revealed that this 1032 bp/344 aa long region (Fig. 2 positions 1611−2642) is composed entirely of repeated elements which show between 64% and 100% homology. Five of these elements (re1, re2, re4, re5 and re7) are 132 bp/44 aa long, another two (re3 and re6) are 168 bp/56 aa in length and the smallest (re0) is 36 bp/12 aa long. Thereby, the elements re2 to re4 and re5 to re7 form continuous stretches of 432 bp/144 aa, the composition of which differs only by 4 nucleotides or by 1 amino acid.

In proximity of the *min* gene is a 176 bp long DNA sequence in inverted orientation relative to *min*. This segment, termed *ψcin*, is 67% homologous to the 5′ end of *min* and will be further characterized below.

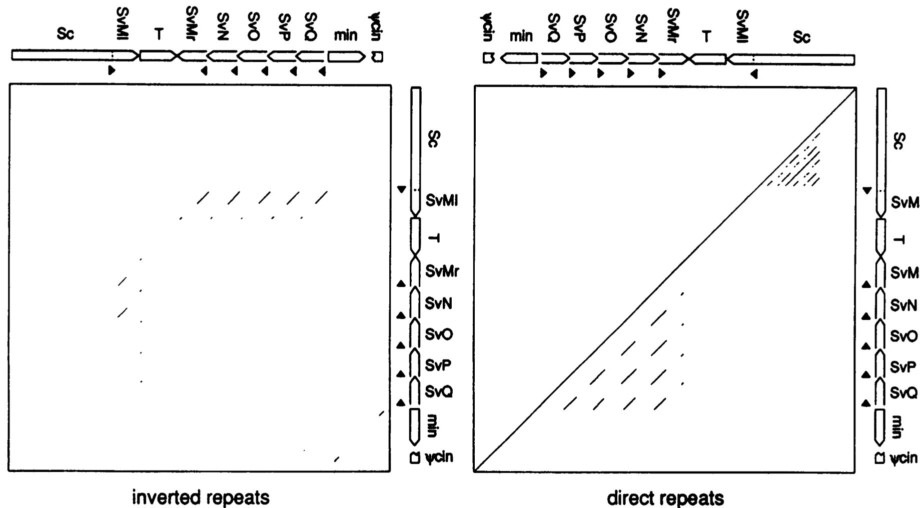

Figure 4. Dotplots of the p15B *Bam*HI-5 fragment from positions 1000 to 7500 reveal directly and invertedly repeated sequences. Two identical sequences in direct or indirect orientation were gradually displaced and matches of 21 nucleotides within a window of 30 nucleotides were marked as dots. Crossover sites and ORFs are indicated as filled-in triangles and open arrows, respectively.

The variable gene *S* and genes *R* and *T* are controlled by a phage specific late promoter

Upstream of *Sc* are an ORF coding for 144 aa termed *R* and the 3′ end of an unidentified ORF (Figs. 1 and 2). A 1.7-kb region including these ORFs and the 5′ end of *Sc* (see Fig. 1) was found to be 97% homologous to the corresponding part of the P1 genome. It has been shown earlier that a *Bam*HI/*Sma*I subfragment of this p15B region and the corresponding fragment of the P1 genome carry promoters (P_S), which are activated relatively late in the cycle of vegetative growth of P1 phage (18). The precise location of the P1 late promoter P_S and the mode of expression of the P1 genes *R* and *S* from P_S have been studied (18). The P_S promoter consists of the −10 box common to *E. coli* promoters and a palindromic element centered around position −22. The latter is believed to be the site of action of the late promoter activating protein, the product of P1 gene *10* (19). The sequences of the two p15B P_S promoter elements −10 (Fig. 2 positions 541−546) and −22 (Fig. 2 positions 528−536) are identical to the corresponding elements of P1 P_S.

An analysis of the P1 tail fiber operon under control of the P_S promoter revealed that the expression of genes *R* and *S* are partially coupled when translated (20). This is likely to be also the case for p15B. As to the expression of ORF *T*, one can expect that p15B P_S transcribes *T* if the M segment is in the M(+) orientation. Neither a terminator nor a promoter-like sequence is found in front of *T* and the start codon of gene *T* overlaps with the stop codon of *SvMl* (Fig. 2 positions 3264−3268). Therefore genes *S* and *T* are likely to be translationally coupled.

The functions of genes *R*, *S* and *T* have not yet been elucidated. Since P_S, *R* and part of *S* of p15B are almost identical to the corresponding sequences in the tail fiber operon of phage P1, it is tempting to speculate, that the gene products encoded by the Min region are analogous to tail fiber proteins. This is in line with the earlier expressed hypothesis (21, 22) that plasmid p15B is a defective prophage.

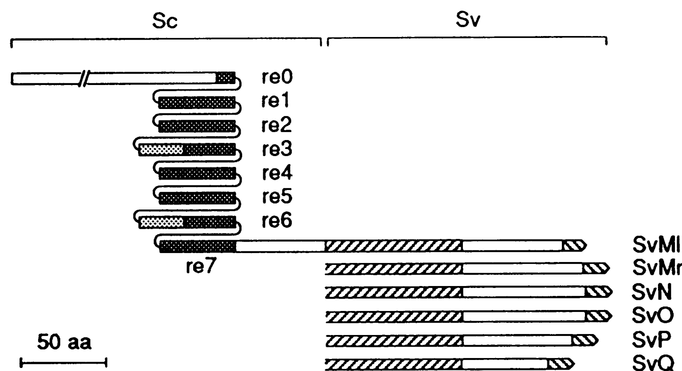

Figure 5. Linear structure of the variable gene *S* with repeated elements and six possible carboxyterminal ends. The aligned repeated elements re0 to re7 are shown as shaded boxes. Homologous domains among the six variable 3′ ends *SvMl*, *SvMr*, *SvN*, *SvO*, *SvP* and *SvQ* are hatched.

Table 1. Activity of the p15B *ψcin* recombinational enhancer and its mutant derivative.

Plasmid	Enhancer source	Relative enhancer activity
pPHU9	−	1
pPHU81	*cin*	30
pPHU133	*ψcin*	0.25
pPHU134	mutated *ψcin*	8.5

DNA inversion frequencies of the indicated plasmids were determined as described in Materials and Methods and were normalized with the value obtained with pPHU9 carrying no enhancer. DNA inversion frequency on pPHU9 was 5.1×10^{-4} Cm^R per Ap^R colonies. The relevant sequences of *cin*, *ψcin* and its mutant carried in pPHU134 are given in Fig. 6.

Expression of *min* might be autoregulated

Since the p15B recombinase is expressed constitutively (1), transcription of the *min* gene is not dependent on the P_S late promoter. The putative P_{min} promoter sequence and a putative ribosomal binding site (SD) were identified on the basis of the structural homologies to the consensus sequences of the corresponding *E. coli* regulatory elements (23−25). The −35 box of P_{min} overlaps a *mix* crossover site (Fig. 2 positions 6383−6388) as is the case for the promoters of the related *din* genes. Since the binding of the Din recombinases to the crossover sites is likely to interfere with the initiation of the *din* gene transcript, it has been speculated that this could provide the basis for an autoregulation of expression of the Din family of recombinases (26, 27).

Homology of the Min region with the P1 genome includes a truncated *cin* gene

Both ends of the p15B fragment present in pAW800 are highly homologous to corresponding DNA of the P1 genome (Fig. 1). A high sequence conservation of 96% identical nucleotides is

```
                        10        20        30        40        50        60        70        80        90
                         .         .  SITE I      .         .         .         .   SITE II         .
cin             ATGCTAATAGGCTATGTACGCGTATCAACAAATGAACAAAACACTGCTTTACAACGAAACGCTCTTGAAAGCGCAGGATGTGAGCTAATT
FIS consensus                         K  YR  W  YR  M                                 K  YR  W  YR  M
min                 A   T     CA T  T  G           C        A A       A    T  A  GA CT             A  G
ψcin                                 A   AG                 A   A G  GT    A  A
mutated ψcin                         A                      A   A G  GT    A  A
cin mutants                              AG                 A   A G  GA    T  C
                                         down                      neutral
```

Figure 6. Comparison of recombinational enhancer sequences and their mutants. The top line shows the *cin* enhancer (26). Its two regions protected by FIS protein against DNAseI digestion (32) are underlined and labelled with site I and site II. The consensus sequence for FIS binding site derived by mutational analysis of the *cin* enhancer sequence (9) is given below. Blanks in the *min* sequence (cf. Fig. 2, site I positions 6461−6475 and site II positions 6509−6523), in the inactive *ψcin* element (cf. Fig. 2, site I positions 7297−7283 and site II positions 7249−7235) and in the newly produced mutant *ψcin* with regained recombinational enhancer activity denote bases identical to the P1 *cin* enhancer sequence. The last line pools the sequence data of nine independent single base substitutions of the *cin* enhancer (9). Two mutant *cin* enhancers altered within site I had been classified as down (0−20% of the activity of the wild type *cin* enhancer on pPHU81) and seven sequence alterations carried between site I and site II as neutral (61−150% of control activity).

observed for the 560-bp sequenced parts of the P1 and p15B *dar* operons including the P_{dar} promoter and the lysis controlling genes *lydA* and part of *lydB* (28, 29, SI unpublished results). It has been shown earlier that the p15B P_{dar} promoter is structurally related to P_S and that it functions as a phage-specific late promoter (18). Upstream of P_{dar} a 330-bp region is less conserved (75%) and adjacent to the latter a 270-bp sequence resumes higher homology (89%) to P1 DNA (26, SI, RHN unpublished results). This region lies 132 bp downstream of *min* and includes a 176 bp long sequence (Fig. 2 positions 7143–7319) which is in inverted orientation to *min* (see above) and which corresponds to the 5′ part of the P1 *cin* gene for DNA invertase. This p15B sequence, termed pseudo *cin* gene (*ψcin*), is non-functional (the ninth codon is an amber codon, Fig. 2 positions 7295–7293).

The *ψcin* gene contains a silent recombinational enhancer sequence which can be activated by site-directed mutagenesis

Since the recombinational enhancer of P1 is encoded within the first 100 nucleotides of the *cin* gene (11), we suspected that plasmid p15B carries two recombinational enhancer sequences, one residing within the *min* gene (Fig. 2 positions 6461–6523) and one within the inactive *ψcin* (Fig. 2 positions 7297–7235). The sequence within the *min* gene has already been demonstrated to be an active recombinational enhancer (1). Since it has been shown that an increasing number of recombinational enhancer elements present on the same replicon can elevate the stimulatory effect on DNA inversion (11), we wondered whether *ψcin* DNA contributes to Min inversion. To test this, the putative second enhancer of p15B was subcloned into a DNA inversion tester plasmid and its activity was determined *in vivo* as described in the Materials and Methods section. As Table 1 shows, this sequence does not stimulate DNA inversion compared to a tester plasmid containing no enhancer sequence. Therefore, the *ψcin* gene contains no active recombinational enhancer.

DNA sequence comparison between the P1 *cin* enhancer and its *ψcin* homolog on p15B reveals ten basepair changes (Fig. 6). Since the P1 *cin* enhancer was analyzed by many point mutations (9) we can presume the effect of most of these ten basepair changes: only two of these substitutions which lie within FIS site I would be detrimental, while another six single changes would have no or only slight effect. Incidentally, the substitute nucleotides within FIS site I concern positions 2 and 3 of the amber codon of *ψcin* (cf. Fig. 2 positions 7293 and 7294, Fig. 6 *ψcin* positions 26 and 27). To investigate if back-mutation of these two basepairs would restore the enhancer activity, the *ψcin* enhancer of p15B was mutated by oligodeoxyribonucleotide-directed site-specific mutagenesis and subsequently tested for its activity by DNA inversion assay (see Materials and Methods). The results summarized in Table 1 reveal that the mutated p15B *ψcin* enhancer is partly active compared to the P1 *cin* enhancer. Since the sequences spanning the FIS consensus of the mutated p15B *ψcin* enhancer are identical with those of the P1 *cin* enhancer (Fig. 6), the difference in activity between these two enhancer sequences is most likely due to the sequence differences of the DNA segment between the two FIS sites. This central DNA segment between the two FIS sites could allow for a FIS protein induced preference to a bent DNA conformation (9, 30). Hence, the data presented here are consistent with the notion that this central segment is able to modulate the recombinational enhancer activity.

An evolutionary implication of the Min inversion region on the p15B plasmid

Although the regions flanking the Min system are highly homologous to the corresponding regions of P1, the p15B invertible segments spanning 3.5 kb are completely dissimilar to the 4.2-kb inversion segment of P1 (26, 31, SI, RHN unpublished results, see also Fig. 1). It appears that an ancestral inversion region of P1 had at one time been substituted by a related DNA inversion system, which might or might not had already contained multiple inversion segments. The *ψcin* gene on the one side and the sequence for the phage P1-specific late promoter P_S as well as the ORF *R* region on the other side must be remnants of such a process. Comparative studies to be described in detail elsewhere showed that the Min invertible segments are related to part of the invertible segment of the Pin inversion system of the e14 element of *E. coli*.

ACKNOWLEDGEMENTS

We would like to thank Reinhard Doelz for maintaining an excellent computer software environment. This work was supported by grant 3.624.087 of the Swiss National Science Foundation.

REFERENCES

1. Iida, S., Sandmeier, H., Hübner, P., Hiestand-Nauer, R., Schneitz, K. and Arber, W. (1990) *Molec. Microbiol.*, **4**, 991–997.
2. Sandmeier, H., Iida, S., Meyer, J., Hiestand-Nauer, R. and Arber, W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1109–1113.
3. Glasgow, A.C., Hughes, K.T. and Simon, M.I. (1989) In Berg, D.E. and Howe, M.M. (ed.), Mobile DNA. Amer. Soc. Microbiol., Washington, D. C., pp. 637–660.
4. Johnson, R.C. and Simon, M.I. (1987) *Trends Genet*, **3**, 262–267.
5. Iida, S., Huber, H., Hiestand-Nauer, R., Meyer, J., Bickle, T.A. and Arber, W. (1984) *Cold Spring Harbor Symp. Quant. Biol.*, **49**, 769–777.
6. Meyer, J., Stalhammar-Carlemalm, M., Streiff, M., Iida, S. and Arber, W. (1986) *Plasmid*, **16**, 81–89.
7. Wood, W.B. (1966) *J. Molec. Biol.*, **16**, 118–133.
8. Iida, S., Schrickel, S. and Arber, W. (1982) *FEMS Lett.*, **15**, 269–273.
9. Hübner, P. and Arber, W. (1989) *EMBO J.*, **8**, 577–585.
10. Vandeyar, M.A., Weiner, M.P., Hutton, C.J. and Batt, C.A. (1988) *Gene*, **65**, 129–133.
11. Huber, H.E., Iida, S., Arber, W. and Bickle, T.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3776–3780.
12. Sanger, F., Nicklen, S. and Coulson, R. (1977) *Proc. Nat. Acad. Sci. USA*, **74**, 5463–5467.
13. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucl. Acids Res.*, **12**, 387–395.
14. Iida, S. and Hiestand-Nauer, R. (1986) *Cell*, **45**, 71–79.
15. Klippel, A., Cloppenburg, K. and Kahmann, R. (1988) *EMBO J.*, **7**, 3983–3989.
16. Johnson, R.C. and Bruist, M.F. (1989) *EMBO J.*, **8**, 1581–1590.
17. Komano, T., Kubo, A. and Nisioka, T. (1987) *Nucl. Acids Res.*, **15**, 1165–1172.
18. Guidolin, A., Zingg, J.-M., Lehnherr, H. and Arber, W. (1989) *J. Molec. Biol.*, **208**, 615–622.
19. Lehnherr, H., Guidolin, A. and Arber, W. (1991) *J. Bacteriol.*, **173**, (in press).
20. Guidolin, A., Zingg, J.-M. and Arber, W. (1989) *Gene*, **76**, 239–243.
21. Ikeda, H., Inuzuka, M. and Tomizawa, J. (1970) *J. Molec. Biol.*, **50**, 457–470.
22. Meyer, J., Iida, S. and Arber, W. (1983) *J. Molec. Biol.*, **165**, 191–195.
23. Gold, L. and Stormo, G. (1987) In Neidhard, F.C. (ed.), *Escherichia coli* and *Salmonella typhimurium*. Amer. Soc. Microbiol., Washington, D. C., pp. 1302–1307.
24. Harley, C.B. and Reynolds, R.P. (1987) *Nucl. Acids Res.*, **15**, 2343–2361.

25. Oliphant, A.R. and Struhl, K. (1988) *Nucl. Acids Res.*, **16**, 7673–7683.
26. Hiestand-Nauer, R. and Iida, S. (1983) *EMBO J.*, **2**, 1733–1740.
27. Yarmolinsky, M.B. and Sternberg, N. (1988) In Calendar, R. (ed.), The Bacteriophages. Plenum Press, N. Y., pp. 291–438.
28. Iida, S., Streiff, M.B., Bickle, T.A. and Arber, W. (1987) *Virology*, **157**, 156–166.
29. Streiff, M.B., Iida, S. and Bickle, T.A. (1987) *Virology*, **157**, 167–171.
30. Hübner, P., Haffter, P., Iida, S. and Arber, W. (1989) *J. Molec. Biol.*, **205**, 493–500.
31. Kamp, D., Kardas, E., Ritthaler, W., Sandulache, R., Schmucker, R. and Stern, B. (1984) *Cold Spring Harbor Symp. Quant. Biol.*, **49**, 301–311.
32. Haffter, P. and Bickle, T.A. (1987) *J. Molec. Biol.*, **198**, 579–587.