Mol Gen Genet (1988) 213:191–201

MGG

Site-specific recombination promotes linkage between trimethoprim- and sulfonamide resistance genes. Sequence characterization of *dhfrV* and *sulI* and a recombination active locus of Tn*21*

Lars Sundström, Peter Rådström, Göte Swedberg, and Ola Sköld
Department of Pharmaceutical Microbiology, Biomedical Center, Uppsala University, P.O. Box 581, S-751 23 Uppsala, Sweden

Summary. A new gene for trimethoprim resistance, *dhfrV*, found in several plasmid isolates with different characteristics, was sequenced and found to correspond to a peptide of 157 amino acids showing 75% similarity with the previously characterized, drug resistant dihydrofolate reductase of type I. The sequenced surroundings of *dhfrV* in plasmid pLMO20, were found to be almost identical with genetic areas surrounding resistance genes in transposon Tn*21* and in R plasmid R388. The trimethoprim resistance genes of pLMO20 and R388 and the spectinomycin resistance gene of Tn*21* could be regarded as having been inserted, by recombination, into an evolutionary older structure containing the sulfonamide resistance gene, *sulI*. The latter gene was sequenced and found to correspond to a peptide of 279 amino acids and with a molecular weight of 30126 daltons. The inserted genes were found to be governed by a promoter situated in the highly conserved structure and also controlling expression of *sulI*. The insertion points of the different resistance genes were precisely defined, and at the 3′ ends of the inserted genes inverted repeats allowing the formation of stem and loop structures were found. Similar structures were found at the 3′ ends of the antibiotic resistance genes in Tn*7*, which could indicate similar recombination mechanisms to be effective in the evolutionary construction of all these different resistance elements.

Key words: Trimethoprim resistance – Sulfonamide resistance – Tn*21* – Tn*7* – Site-specific recombination

Introduction

The wide and rapid spread of antibiotic resistance among different pathogenic bacteria seems to involve a continuing recombinational incorporation of resistance genes into plasmids and transposons, which are efficient vehicles of transfer. In previous work a remarkable similarity was observed between the restriction enzyme digestion maps of parts of plasmids pLMO20, R388 and R100 (Tn*21*), carrying antibiotic resistance genes (Sundström et al. 1987). The different resistance genes seemed to have been inserted by a recombinational event.

In plasmid pLMO20 a 3.51 kb *Bam*HI fragment was characterized and found to contain a new gene (*dhfrV*) for trimethoprim resistance. This fragment, carrying *dhfrV*, was found in several, otherwise different plasmids and was in turn demonstrated to be very similar to a 3.79 kb *Bam*HI fragment in Tn*21* carrying the spectinomycin resistance gene *aadA* and a 3.82 kb fragment in R388 mediating type II trimethoprim resistance (*dhfrII*). The three different resistance genes *dhfrV*, *aadA* and *dhfrII* could be regarded as being inserted into a common structure carrying the gene *sulI*, which codes for a sulfonamide resistant dihydropteroate synthase.

In order to analyze the genetic organization of the resistance genes in pLMO20, R388 and Tn*21*, and to study the previously uncharacterized *dhfrV* and *sulI* genes, these were subcloned and investigated regarding expression and nucleotide sequences. Plasmid R6-5 was used in this work as source of the 3.79 kb *Bam*HI fragment representing Tn*21* (de la Cruz and Grinsted 1982), since R100 and R6-5 were shown to be identical in this area (Sharp et al. 1973; Schmidt and Klopfer-Kaul 1984).

Drug resistant dihydrofolate reductase of type V was found to be related to the previously characterized enzyme of type I mediated by transposon Tn*7*. The expression of *dhfrV* seemed to be governed by the same promoter as that controlling the dihydropteroate synthase of *sulI*. The sequence for the latter, given here, is the first known for an enterobacterial dihydropteroate synthase gene.

The sequence data also define a locus of insertion, flanked by GTTA sequences, for the *dhfrV*, *aadA* and *dhfrII* genes. The regions surrounding the insertions in pLMO20, R388 and Tn*21* are almost identical. The regions downstream of the insertions including the *sulI* gene are completely identical. At the 3′ ends of the inserted genes are potential stem and loop structures, which could be involved in recombination.

Materials and methods

Materials. Trimethoprim lactate was a gift from Wellcome Research Laboratories, Beckenham, England. Tetracycline and sulfathiazole were obtained from ACO Läkemedel, Solna, Sweden, and ampicillin was from Astra Läkemedel AB, Södertälje, Sweden. Spectinomycin was obtained from Upjohn S.A., Puurs, Belgium.

Plasmids, bacterial strains and growth conditions. The plasmid vectors used were pBR322 (Bolivar et al. 1977) and

Offprint requests to: L. Sundström

Table 1

Plasmids	Relevant characteristics	Reference or derivation
pLMO20	Ap Hg Km Su Tc Tp	Sundström et al. 1987
R388	Su Tp	Ward and Grinsted 1982
R6-5	Cm Hg Km Spc Su	Timmis et al. 1978
pLKO1B	Ap Su Tp	Sundström et al. 1987
pLKO2	Ap Tp	Sundström et al. 1987
pLKO22A	Ap Tp	0.48 kb *Hpa*I fragment from pLKO1B in *Sma*I of pUC18
pLKO22B	Ap	Like pLKO22A, but the *Hpa*I fragment is inverted
pLKO24	Ap Su	1.3 kb *Hin*dIII-*Bam*HI fragment from pLKO1B in pUC19
pGS72	Ap Su	Swedberg and Sköld 1983
pGS74	Ap Su Tp	Swedberg 1987
pGS105	Ap Spc Su	Swedberg 1987
pSUL101	Ap Su	1.23 kb *Hin*dIII-*Sac*I fragment from pGS105 in pUC19
pSUL102	Ap	Like pSUL101, but in pUC18

Abbreviations: Ap, ampicillin resistance; Cm, chloramphenicol resistance; Hg, mercuric chloride resistance; Km, kanamycin resistance; Nal, nalidixic acid resistance; Spc, spectinomycin resistance; Su, sulfonamide resistance; Tc, tetracycline resistance; Tp, trimethoprim resistance

pUC18/19 (Yanisch-Perron et al. 1985). Other plasmids used and their derivations are listed in Table 1. *Escherichia coli* strain C600 [F^- *thi thr leu lac tonA supE*] (Bachmann 1972) was used as host for pBR322 and its derivatives, and JM83 [*ara Δ (lac-proAB) rpsL φ80lacZ'*M15] (Yanisch-Perron et al. 1985) was used as host for pUC18/19. *E. coli* strain JM105 [*thi rpsL endA sbcB15 hspR4 Δ(lac-proAB) F' traD36 proAB lacI*^{q}*ZΔ*M15] was used for the propagation of M13 bacteriophage derivations mp18 and mp19 (Yanisch-Perron et al. 1985), used for cloning of DNA fragments at sequence analysis. To effect inhibition by trimethoprim and sulfonamide, Iso Sensitest medium (Oxoid, Basingstoke, UK) was used. Otherwise cultures were grown in either the rich LB medium (Miller 1972) or in the mineral salts medium M9 (Maniatis et al. 1982).

Construction of recombinant plasmids and analysis of them by digestion with restriction enzymes and gel electrophoresis. Recombinant derivatives of pBR322, pUC18/19 and of the replicative form of M13mp18/19 were constructed according to Davis et al. (1980). For screening purposes plasmid DNA was prepared from fresh bacterial colonies by the boiling method of Holmes and Quigley (1981). The conditions for restriction nuclease digestion, ligation and transformation, and for gel electrophoresis and purification of DNA fragments were as described previously (Sundström et al. 1987).

Nucleotide sequencing. All the studied nucleotide sequences were determined by the dideoxynucleotide chain termination method of Sanger et al. (1977). Single-stranded sequencing templates from recombinant bacteriophage M13mp18/19 were prepared from 1.5 ml cultures. Overnight cultures of JM105 in rich LB medium were diluted 1:100, and infected with recombinant M13 phage freshly picked from plaques. After 5 h of culture at 37° C, cells were carefully removed by centrifugation, and phages were precipitated by the addition of 200 µl of a 20% w/v polyethylene glycol (PEG-6000) solution in 2.5 M NaCl. Precipitates collected by centrifugation were resuspended in 100 µl of 10 mM Tris-HCl, 1 mM disodium EDTA, pH 8.0; DNA was extracted with phenol and chloroform and finally precipitated with ethanol. Estimation of the size of recombinant, single-stranded DNA in mp18/19 was performed by sampling 20 µl from cell-free phage suspensions, mixing with 2 µl of 2% sodium dodecyl sulfate solution, and gel electrophoresis.

Chain elongation was performed with *E. coli* DNA polymerase (Klenow fragment) at 30° C. For labelling, adenosine 5'-[^{35}S] thiotriphosphate (New England Nuclear, Dreieich, W. Germany) was used. The pLMO20 and R388 sequences were determined from both strands. About half of the Tn*21* sequence, including the regions around the homology branch points at each side of the *aadA* gene, was determined from both strands. The rest of the sequence from Tn*21* was determined from one strand only. Fragments from the recombinant plasmids listed in Table 1 were obtained by digestion with one or several of the restriction nucleases *Ava*I, *Bam*HI, *Bgl*I, *Bgl*II, *Eco*RI, *Eco*RV, *Hin*dIII, *Hpa*I, *Pst*I, *Pvu*II, *Rsa*I, *Sac*I and *Sph*I. As sources for DNA fragments of pLMO20, the cloned derivatives of it, pLKO1B, pLKO2, pLKO24 were used. Similarly the derivatives pGS72 and pGS74, cloned from R388, were used. The derivative pGS105 from R6-5, was used as a representation of Tn*21*. The fragments obtained were either cloned directly into M13mp18/19 or they were further digested with *Taq*I and/or *Hpa*II. The small fragments obtained in this way were randomly cloned by ligation into the *Acc*I site of M13mp18/19. Alternatively, small fragments were generated by *Sau*3A digestion and ligated into the *Bam*HI site of the vector phage. By such stepwise random cloning a good and even representation of the analysed DNA was obtained.

Results

Related sequences around the antibiotic resistance genes in plasmids pLMO20, R388 and R6-5 (Tn21)

In previous work (Sundström et al. 1987), detailed restriction enzyme digestion maps of parts of pLMO20, R388 and Tn*21* demonstrated that the trimethoprim resistance genes *dhfrII* and *dhfrV* and the spectinomycin resistance gene *aadA* could be regarded as inserted in a common genetic structure containing the sulfonamide resistance gene *sulI*. In order to analyse further the genetic organization of the resistance traits in the transferable resistance elements mentioned, and to characterize the new trimethoprim resistance gene *dhfrV* and also the previously known gene for sulfonamide resistance, *sulI*, nucleotide sequence determinations were performed according to the plan shown in Fig. 1. This figure shows the precise locations of *dhfrV* in pLMO20,

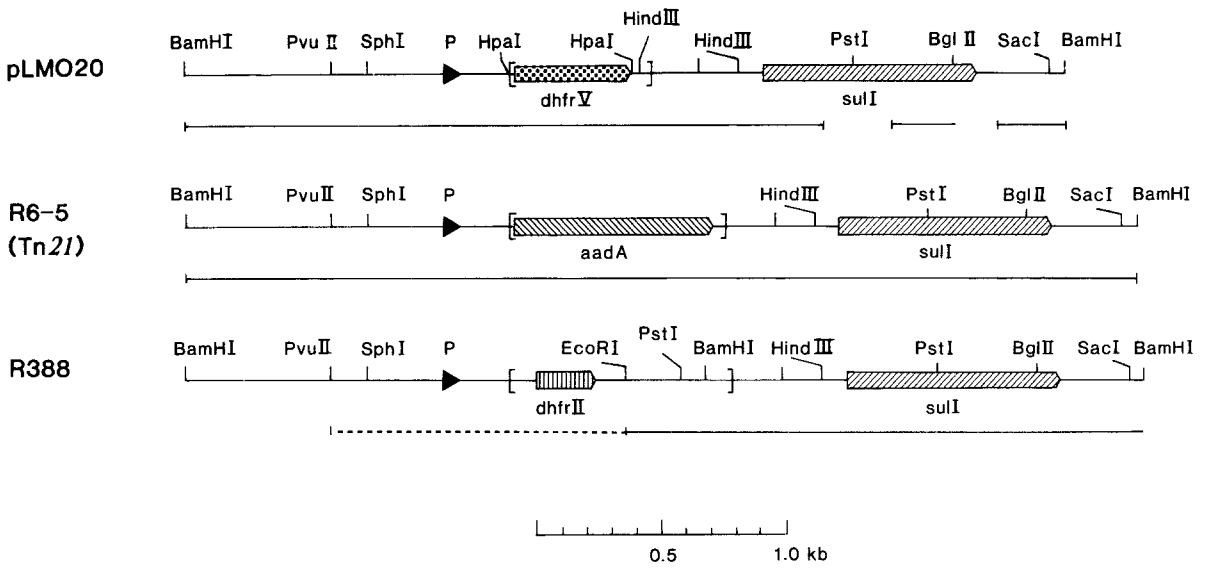

Fig. 1. Physical maps of the interrelated *Bam*HI fragments from plasmids pLMO20, R388 and R6-5 (Tn*21*). The sequenced parts are continuously *underlined*, while the *dotted line* marks the sequence quoted from other published work (Swift et al. 1981; Zolg and Hänggi 1981). Nonhomologous regions are within *brackets*. The common promoter is indicated by *P* and the arrows indicate the transcription direction

of *aadA* in Tn*21* and of *dhfrII* in R388, and also of *sulI* in all the three elements. The sequence data for pLMO20, shown in Fig. 2 demonstrate almost exact identity between the leftmost *Bam*HI-*Hpa*I fragment (Fig. 1) and a corresponding part (*Pvu*II to bracket, Fig. 1) of R388 (Swift et al. 1981; Zolg and Hänggi 1981). There were only four nucleotide discrepancies, which are listed in Table 2. In the comparison with the corresponding part (*Bam*HI to bracket, Fig. 1) of Tn*21* from R6-5 the DNA sequences were identical except for three discrepancies, one of which was an insertion of GGG at position 1173 in the Tn*21* sequence (Fig. 2, Table 2). At the leftmost *Hpa*I site in pLMO20 (left bracket in Fig. 1, box in Fig. 2) the homology ceases abruptly. At a position about 200 nucleotides to the left of the conserved double *Hin*dIII site (Fig. 1) the homology resumes (right bracket in Fig. 1, box in Fig. 2), and from this well-defined point the corresponding sequences are identical all the way to the right *Bam*HI site (Fig. 1) in pLMO20, R388 and Tn*21*. The similarity between these three genetic elements will be discussed in more detail below.

Identification of the dhfrV gene coding for type V dihydrofolate reductase

The drug resistant dihydrofolate reductase expressed from plasmid pLMO20 is distinct from the other previously known plasmid mediated, trimethoprim resistant enzymes of type I, II, III and IV (Sundström et al. 1987; Young and Amyes 1986), and has been named type V. In the pLMO20-specific portion, mentioned above, an open reading frame in the same direction as the *sulI* gene (see below) and corresponding to 157 amino acids was found (Fig. 2). This reading frame is flanked by *Hpa*I cleavage sites, and on digestion with *Hpa*I a 0.48 kb fragment was obtained, which included the assumed translation start but had lost the termination codon TAA (Fig. 2). This *Hpa*I fragment was ligated into the *Sma*I site of plasmid pUC18 (see the Materials and methods) and thus transcribed from the *lac* promoter of the plasmid. About 50% of the obtained recombinant plasmids, pLKO22A, mediated resistance to trimethoprim, while 50%, pLKO22B, were sensitive to this drug. By restriction enzyme digestion analysis it was demonstrated that expression of resistance was dependent on the fragment orientation. Only those recombinant plasmids, where the reading frame found was governed by the *lac*-promoted transcription of the plasmid, made resistant transformants. Because of the lost termination codon in the fragment, the translated product becomes fused, at its carboxy-terminal end, to a peptide of 23 amino acids coded for by the pUC18 linker region and out of frame with the *lacZ'* sequence. The fused peptide thus does not seem to interfere with the function of the drug resistant enzyme.

The relation of type V to other dihydrofolate reductases

The reading frame of 157 amino acids shown in Fig. 2 and corresponding to a molecular weight of 17531 daltons, seems to represent the dihydrofolate reductase, type V. There is a possibility, however, that translation could start at the ATG, 5 codons distal to the GTG given as the start codon, resulting in a molecule that would be shorter by 5 amino acids. Comparison with other dihydrofolate reductases, on the other hand, makes GTG the most likely initiation codon. A reasonable ribosome binding sequence (GGAAC) precedes the putative GTG start codon. The *dhfrI* gene borne on transposon Tn7 has been sequenced and studied in detail (Fling and Richards 1983; Simonsen et al. 1983). The corresponding type I enzyme also comprises 157 amino acids, and a comparison of the amino acid sequences shows that 75% (118/157) of the amino acids are identical in the two enzymes I and V (Fig. 3). At the nucleotide level the similarity is only 68%, however. This discrepancy is explained by neutral substitutions in 67 codons, of which 63 were third base changes. It was not established whether the functional type V enzyme is also dimerical, as is type I. In Fig. 3 a further comparison is made between the enzyme of type V and the *E. coli* K-12 chromo-

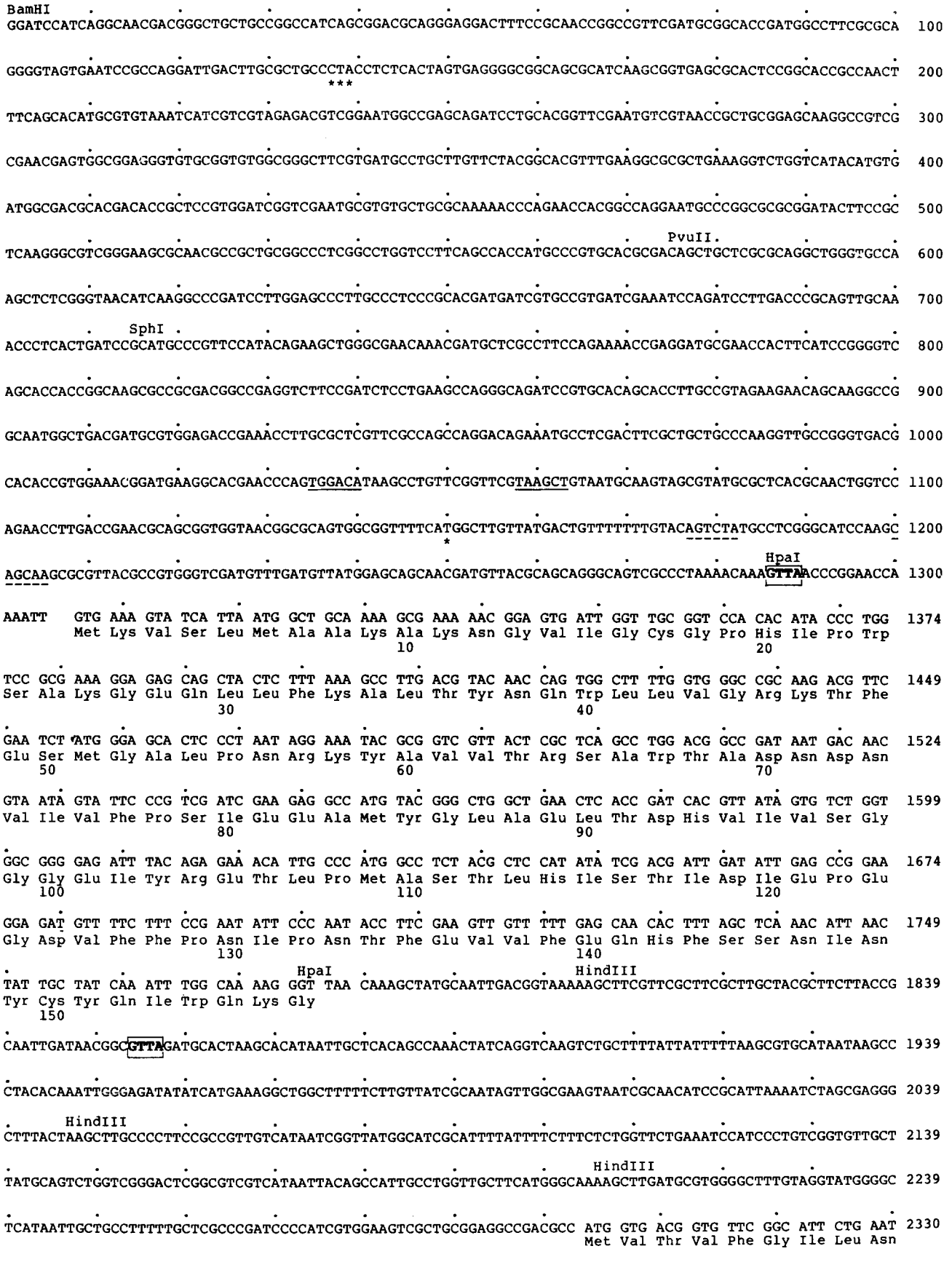

Fig. 2. The nucleotide sequence of *dhfrV* and its surroundings in pLMO20. Some of the restriction nuclease target sites are indicated to allow alignment with the survey of the sequenced areas shown in Fig. 1. The translated *dhfr* sequence comprises nucleotides 1306–1776. The amino-terminal of the translated *sulI* gene (cf. Fig. 4) starts at nucleotide 2304. The GTTA sequences indicating where homology with R388 and Tn*21* ends and starts again, are *boxed*. The common promoter sequences are *underlined*. The start and stop codons of orf2 which runs in the opposite direction, are marked by *asterisks*, and the suggested orf2 promoter hexamers by *interrupted lines*. The sequence given here is identical to that found for the corresponding area (nucleotides 1–1289) of Tn*21* with the exceptions mentioned in Table 2

Table 2. Observed sequence aberrations

Position in pLMO20 (Fig. 2)	pLMO20	Tn*21*	R388
901	G	C	C
907	G	C	C
1036	G	G	T
1061	G	G	A
1173	G	GGGG	G
1290–1852		no similarity	

Data for R388 are from this work, from Swift et al. (1981), and from Zolg and Hänggi (1981)

somal enzyme, the gene of which was sequenced and studied in detail by Smith and Calvo (1980). The amino acid similarity is 37% (58/157) and somewhat more pronounced than the corresponding similarity of 31% (49/157) seen with the type I enzyme. The unaltered amino acid sequences shared by the three enzymes are in those regions observed in a broader comparison of dihydrofolate reductases from various sources (Simonsen et al. 1983) to be most conserved.

The dhfrV *and* sulI *genes in pLMO20 have a common promoter*

The *dhfrII* promoter in R388 was localized in mutagenesis experiments (Swift et al. 1981). This localization is within the region to the left (Fig. 1) of the antibiotic resistance genes of R388, that is similar to the corresponding parts of pLMO20 and Tn*21*. The nucleotide sequence found in pLMO20 (Fig. 2) is identical to that defined as the *dhfrII* promoter by Swift et al. (1981), with two exceptions: the nucleotides at positions 1036 and 1061 in both the pLMO20 and Tn*21* sequences are G instead of a T and an A, respectively, in R388. The rest of the DNA sequences are almost identical (cf. Table 2) down to the point of divergence between pLMO20, R388 and Tn*21* and seem to be devoid of *E. coli* promoters (Harley and Reynolds 1987). The *dhfrV* gene then begins 16 nucleotides beyond the divergence point (Fig. 2), and for the *aadA* gene in Tn*21* this distance is only 6 nucleotides (Fig. 5). These observations indicate that the three inserted resistance genes are governed by the same promoter located in the conserved area upstream of the divergence point (Fig. 1). It was recently reported from this laboratory (Swedberg 1987) that the expression of sulfonamide resistance from *sulI* in Tn*21* was dependent on transcription from a promoter located upstream of *aadA*. Also in pLMO20 and in R388, there is no promoter structure that could be connected to *sulI* (Figs. 2, 5). This indicates that the promoter described by Swift et al. (1981) and also in this study for pLMO20 would be the common promoter for *dhfrV-sulI* in pLMO20, for *dhfrII-sulI* in R388 and for *aadA-sulI* in Tn*21*. The inserted antibiotic resistance genes would thus be expressed in an operon-like way together with *sulI*. The distance between the promoter and the start of the resistance gene is rather long in all three instances; in pLMO20 for example, the distance is 233 nucleotides. The function of these long leader sequences remains to be explained. However, the leaders contain open reading frames, the translation of which could be related to the final expression of antibiotic resistance. It could be mentioned that the ATG, at position 1080 of Fig. 2, is the start

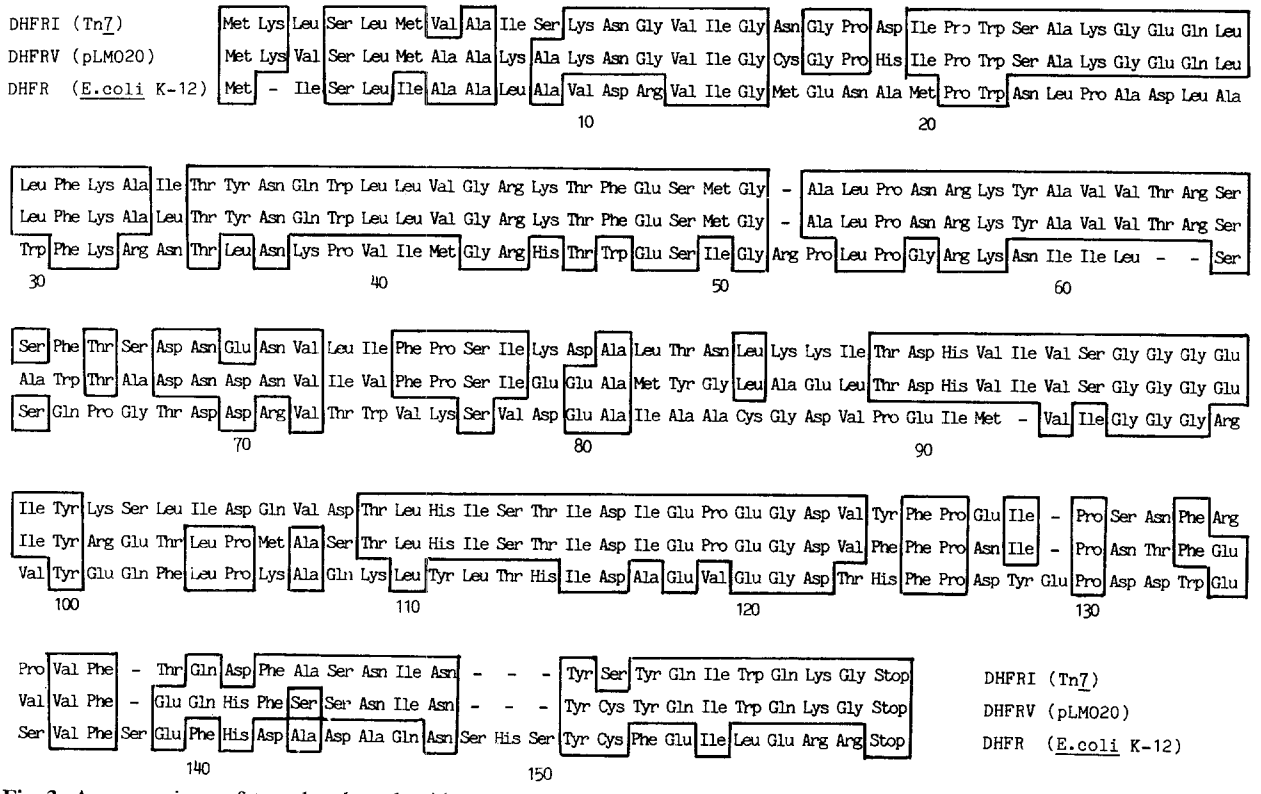

Fig. 3. A comparison of translated nucleotide sequences for *dhfrV* from pLMO20 (Fig. 2), *dhfrI* from Tn*7* (Fling and Richards 1983) and the chromosomal *fol* gene of *Escherichia coli* K-12 (Smith and Calvo 1980). Regions of amino acid homology are *boxed*. The numbering is according to the *E. coli* K-12 enzyme

```
EcoRI
GAATTCCTCTCGAACCGTTTGGGGTTATGTACAGGACTCAGAGAGGACAATCTGCGCATACTTTGTTCAATGGACTGTGGGCGCTCCAGATCACTTTCCC  100

AACTTTGACTTTCTCATCGGAACATGGGGCGACGATGCAGTAAGTGACAAAGTTCTTTCCTCGTGGCTATTCAATCCCGAAGCAGGTTCATTCATGATCA  200

          PstI
TTGATGCTTCTAGTCGTCCTGCAGCGGCATCGGATATGTGCAACCACGCTCTTAGTCGTGCCGAGGTTCTTGCCGCACCGGGTATGAAGTCTTTGGCGTC  300

          .BamHI
CCAATGCCTAGATGCGGTCTGGATCCAAGACGAGCGCGTTGCGGAAATTCGGGGTTGGAAGAATGACGCCTAACCGGTCGTTCGAGCGGACTGCCCTCGG  400

CAAGCCTCGGTCAGCCGCTCAACTTCAACGTTAGATGCACTAAGCACATAATTGCTCACAGCCAAACTATCAGGTCAAGTCTGCTTTTATTATTTTTAAG  500

CGTGCATAATAAGCCCTACACAAATTGGGAGATATATCATGAAAGGCTGGCTTTTTCTTGTTATCGCAATAGTTGGCGAAGTAATCGCAACATCCGCATT  600

          . HindIII.
AAAATCTAGCGAGGGCTTTACTAAGCTTGCCCCTTCCGCCGTTGTCATAATCGGTTATGGCATCGCATTTTATTTTCTTTCTCTGGTTCTGAAATCCATC  700

                                                  . HindIII .
CCTGTCGGTGTTGCTTATGCAGTCTGGTCGGGACTCGGCGTCGTCATAATTACAGCCATTGCCTGGTTGCTTCATGGGCAAAAGCTTGATGCGTGGGGCT  800

TTGTAGGTATGGGGCTCATAATTGCTGCCTTTTTGCTCGCCCGATCCCCATCGTGGAAGTCGCTGCGGAGGCCGACGCC ATG GTG ACG GTG TTC   894
                                                                                    Met Val Thr Val Phe

GGC ATT CTG AAT CTC ACC GAG GAC TCC TTC TTC GAT GAG AGC CGG CGG CTA GAC CCC GCC GGC GCT GTC ACC GCG   969
Gly Ile Leu Asn Leu Thr Glu Asp Ser Phe Phe Asp Glu Ser Arg Arg Leu Asp Pro Ala Gly Ala Val Thr Ala
                10                                  20                                  30

GCG ATC GAA ATG CTG CGA GTC GGA TCA GAC GTC GTG GAT GTC GGA CCG GCC GCC AGC CAT CCG GAC GCG AGG CCT  1044
Ala Ile Glu Met Leu Arg Val Gly Ser Asp Val Val Asp Val Gly Pro Ala Ala Ser His Pro Asp Ala Arg Pro
                            40                                  50

GTA TCG CCG GCC GAT GAG ATC AGA CGT ATT GCG CCG CTC TTA GAC GCC CTG TCC GAT CAG ATG CAC CGT GTT TCA  1119
Val Ser Pro Ala Asp Glu Ile Arg Arg Ile Ala Pro Leu Leu Asp Ala Leu Ser Asp Gln Met His Arg Val Ser
                60                                  70                                  80

ATC GAC AGC TTC CAA CCG GAA ACC CAG CGC TAT GCG CTC AAG CGC GGC GTG GGC TAC CTG AAC GAT ATC CAA GGA  1194
Ile Asp Ser Phe Gln Pro Glu Thr Gln Arg Tyr Ala Leu Lys Arg Gly Val Gly Tyr Leu Asn Asp Ile Gln Gly
                            90                                  100
                                        PstI
TTT CCT GAC CCT GCG CTC TAT CCC GAT ATT GCT GAG GCG GAC TGC AGG CTG GTG GTT ATG CAC TCA GCG CAG CGG  1269
Phe Pro Asp Pro Ala Leu Tyr Pro Asp Ile Ala Glu Ala Asp Cys Arg Leu Val Val Met His Ser Ala Gln Arg
                110                                 120                                 130

GAT GGC ATC GCC ACC CGC ACC GGT CAC CTT CGA CCC GAA GAC GCG CTC GAC GAG ATT GTG CGG TTC TTC GAG GCG  1344
Asp Gly Ile Ala Thr Arg Thr Gly His Leu Arg Pro Glu Asp Ala Leu Asp Glu Ile Val Arg Phe Phe Glu Ala
                            140                                 150

CGG GTT TCC GCC TTG CGA CGG AGC GGG GTC GCT GCC GAC CGG CTC ATC CTC GAT CCG GGG ATG GGA TTT TTC TTG  1419
Arg Val Ser Ala Leu Arg Arg Ser Gly Val Ala Ala Asp Arg Leu Ile Leu Asp Pro Gly Met Gly Phe Phe Leu
                160                                 170                                 180

AGC CCC GCA CCG GAA ACA TCG CTG CAC GTG CTG TCG AAC CTT CAA AAG CTG AAG TCG GCG TTG GGG CTT CCG CTA  1494
Ser Pro Ala Pro Glu Thr Ser Leu His Val Leu Ser Asn Leu Gln Lys Leu Lys Ser Ala Leu Gly Leu Pro Leu
                            190                                 200

TTG GTC TCG GTG TCG CGG AAA TCC TTC TTG GGC GCC ACC GTT GGC CTT CCT GTA AAG GAT CTG GGT CCA GCG AGC  1569
Leu Val Ser Val Ser Arg Lys Ser Phe Leu Gly Ala Thr Val Gly Leu Pro Val Lys Asp Leu Gly Pro Ala Ser
                210                                 220                                 230
                                                                    . BglII .
CTT GCG GCG GAA CTT CAC GCG ATC GGC AAT GGC GCT GAC TAC GTC CGC ACC CAC GCG CCT GGA GAT CTG CGA AGC  1644
Leu Ala Ala Glu Leu His Ala Ile Gly Asn Gly Ala Asp Tyr Val Arg Thr His Ala Pro Gly Asp Leu Arg Ser
                            240                                 250

GCA ATC ACC TTC TCG GAA ACC CTC GCG AAA TTT CGC AGT CGC GAC GCC AGA GAC CGA GGG TTA GAT CAT GCC TAG  1719
Ala Ile Thr Phe Ser Glu Thr Leu Ala Lys Phe Arg Ser Arg Asp Ala Arg Asp Arg Gly Leu Asp His Ala
                260                                 270

CATTCACCTTCCGGCCGCCCGCTAGCGGACCCTGGTCAGGTTCCGCGAAGGTGGGCGCAGACATGCTGGGCTCGTCAGGATCAAACTGCACTATGAGGCG 1819

GCGGTTCATACCGCGCCAGGGGAGCGAATGGACAGCGAGGAGCCTCCGAACGTTCGGGTCGCCTGCTCGGGTGATATCGACGAGGTTGTGCGGCTGATGC 1919

                                                                                    SacI
ACGACGCTGCGGCGTGGATGTCCGCCAAGGGAACGCCCGCCTGGGACGTCGCGCGGATCGACCGGACATTCGCGGAGACCTTCGTCCTGAGATCCGAGCT 2019

                                        BamHI
CCTAGTCGCGAGTTGCAGCGACGGCATCGTCGGCTGTTGCACCTTGTCGGCCGAGGATCC                                          2079
```

Fig. 4. The nucleotide sequence of the *sulI* gene and its surroundings in R388. Some of the restriction nuclease target sites are indicated to allow alignment with the survey of the sequenced areas shown in Fig. 1. The translated *sulI* sequence comprises nucleotides 880–1716. The GTTA sequence indicating the start point of homology (see text) is *boxed*. The nucleotide sequence from Tn*21* corresponding to the area shown here was also determined and found to be identical to that given here from nucleotide 429 onwards

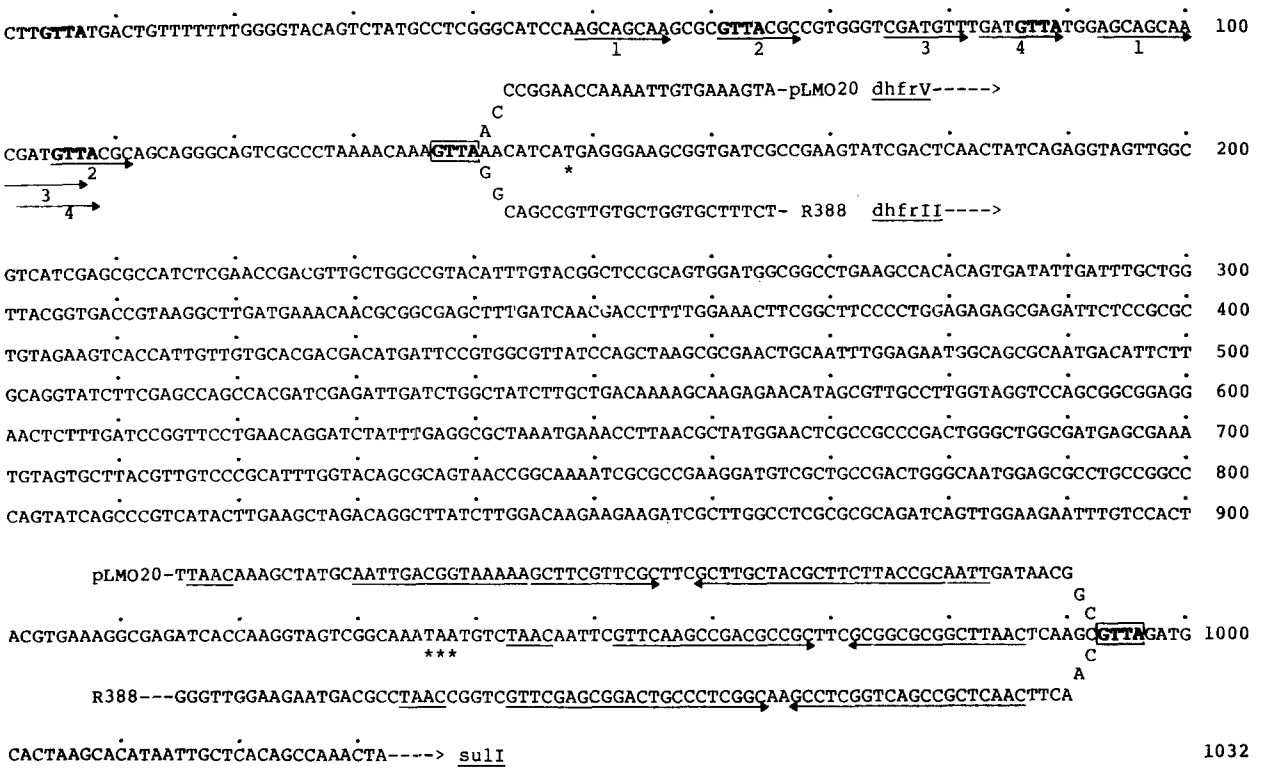

Fig. 5. The nucleotide sequence of the area within the branch points (*boxed*, see text) of Tn*21* including the *aadA* gene and its surroundings. The sequence between the start codon (marked by the *asterisk*) and stop codon (marked by the *three asterisks*) is identical to the corresponding region of the *aadA* sequence given by Hollingshead and Vapnek (1985) for plasmid R538-1. Parts of the sequences for pLMO20 and R388 close to the branch points are also shown. The numbered short *arrows* indicate direct repeats and the longer ones indicate inverted repeats

of a reading frame that is open across the upper branch point. In pLMO20 this potentially translated sequence is out of frame with *dhfrV*, but in Tn*21* an in-frame fusion is possible with *aadA*.

Nucleotide sequence analysis of the DNA region harbouring the sulI *gene*

There are two plasmid-borne sulfonamide resistance genes known, *sulI* and *sulII*. Both express drug resistant dihydropteroate synthases and can be distinguished by DNA hybridization methods (Swedberg and Sköld 1983).

The new sequence data for R388 shown in Fig. 4, comprising 2079 nucleotides of the *Eco*RI-*Bam*HI segment, could be connected to published sequences (Swift et al. 1981; Zolg and Hänggi 1981) to cover the entire 3.25 kb of antibiotic resistance in R388. A comparison of the sequence shown in Fig. 4 with the sequence for Tn*21* demonstrated complete identity from the point of convergence at nucleotide 429 (Fig. 4) to the *Bam*HI end of the presented sequence. Earlier work demonstrated that the *sulI* gene is located within a 1.75 kb fragment from R388 (Swedberg and Sköld 1983). Further subcloning localized the *sulI* gene to a 1.23 kb *Hin*dIII-*Sac*I fragment (cf. Figs. 1, 4) from which, when it was ligated into vector pUC19, sulfonamide resistance was expressed, that was under the control of the *lac* promoter in this plasmid, pSUL101. When the same fragment was ligated into plasmid pUC18 and thus in the reverse direction in relation to the promoter, (pSUL102) no expression of sulfonamide resistance was observed. The *sulI* gene of the cloned fragment is thus expressed in the *Hin*dIII-*Sac*I direction and has no promoter of its own. In this orientation two open reading frames coding for peptides longer than 50 amino acids were found in the conserved region of R388 and Tn*21*. The longer of these, which starts entirely within the *Hin*dIII-*Sac*I fragment, starts with a GTG at nucleotide 793 (Fig. 4). However, there is no reasonable ribosome-binding site in front of this codon. At position 880, however, there is an ATG preceded by a GGAGG sequence, which could serve as the locus for ribosome binding. With the ATG at position 880 as the start codon a peptide of 279 amino acids would be produced. The molecular weight deduced for the putative protein is 30126. This is suggested to represent the dihydropteroate synthase mediating sulfonamide resistance. The complete nucleotide sequence of the *sulI* region in pLMO20 was not obtained, but for those 70% that were analysed (cf. Fig. 1) there was identity to R388 and Tn*21*, so it is reasonable to assume identity between the three genetic elements in this vicinity. The other, shorter open reading frame (orf1) mentioned above to occur near *sulI* starts at nucleotide 539 (Fig. 4), comprises 115 codons and overlaps *sulI* with its stop codon at nucleotide 884. Its function is unexplained.

Observations indicating a mechanism for the recombinational insertion of dhfrV, dhfrII *and* aadA

As described above, the antibiotic resistance genes *dhfrV*, *dhfrII* and *aadA* could be regarded as inserted into a nucleo-

tide sequence that was found to be almost identical in pLMO20, R388 and Tn*21*. This is summarized in Fig. 5, which shows the common points of divergence and convergence, and also the complete sequence of *aadA* in Tn*21* from R6-5. Parts of the sequences from pLMO20 and R388 are also shown (cf. Figs. 2, 4). It can be seen that the short sequence GTTA is directly repeated at the two branch points. This tetranucleotide sequence also occurs four times in the 100 nucleotides of conserved sequence upstream of the divergence point, although it would be distributed by chance only once every 256 nucleotides. Several of these GTTAs are also part of several 7- to 8-nucleotide-long direct repeats shown by the arrows in Fig. 5.

In each of pLMO20 and Tn*21*, only one gene seems to be situated between the two branch points. In R388, however, there is a second open reading frame (orf3) downstream of *dhfrII*, with its translational stop at nucleotide 371, well above the point of convergence at nucleotide 429 (Fig. 4). The function of this open reading frame remains to be explained.

The sequence for the *aadA* gene of R6-5 (Tn*21*), including surrounding sequences with homology branch points, is shown in Fig. 5. The longest reading frame between the branch points translates into 263 amino acids and is in perfect agreement with the corresponding part of the *aadA* of plasmid R538-1 (Hollingshead and Vapnek 1985). The start and stop codons of this reading frame are indicated in Fig. 5. The *aadA* gene of Tn*7*, with the same start and stop codons, is very similar but shows one lost codon (849–851) and two base substitutions at positions 830 and 897 (Fling et al. 1985). This reading of *aadA* in Tn*21* is in frame with the translatable leader sequence. Several possible start sites occur along this conserved sequence, which could extend the protein at its amino-terminal end. In the case of Tn*7* the sequence following the *aadA* gene was identical to that shown in Fig. 5 down to GTTA.

This trailing sequence showed a feature that was found to be common to all the genes within the branch points, including orf3. This was the occurrence of sequences allowing stem and loop structures to be formed with TAAC at one end and the branch point GTTA at the other. This is indicated by the arrows at the end of the *aadA* gene of Fig. 5 and at the trailing parts of the *dhfrV* of pLMO20 and at the orf3 of R388, also indicated in Fig. 5. A similar structure could also be discerned in the sequence associated with the *dhfrII* of R388 (Swift et al. 1981; Cameron et al. 1986).

Further analysis of the sequence upstream of the *dhfrV* gene in pLMO20 also revealed another open reading frame of 337 codons, orf2. It reads in the opposite direction compared to the antibiotic resistance genes, and extends from an ATG at nucleotide 1150 to a TAG at nucleotide 139 (Fig. 2). The considerable length of orf2 and the lack of open reading frames of significant length in the other direction, suggest that orf2 could be a gene. This interpretation is supported by the occurrence of potential promoter sequences at nucleotides 1182–1177 (–10, TAGACT) and at nucleotides 1205–1200 (–35, TTGCTG). Furthermore, the carboxy-terminal part of the translated nucleotide sequence showed a marked similarity to the proposed DNA-binding region comprising about 40 amino acids in a group of phage integrase proteins (Argos et al. 1986; Eisenstein et al. 1987). The translated orf2 sequence is shown in Fig. 6, where a consensus sequence for the proposed DNA-binding region

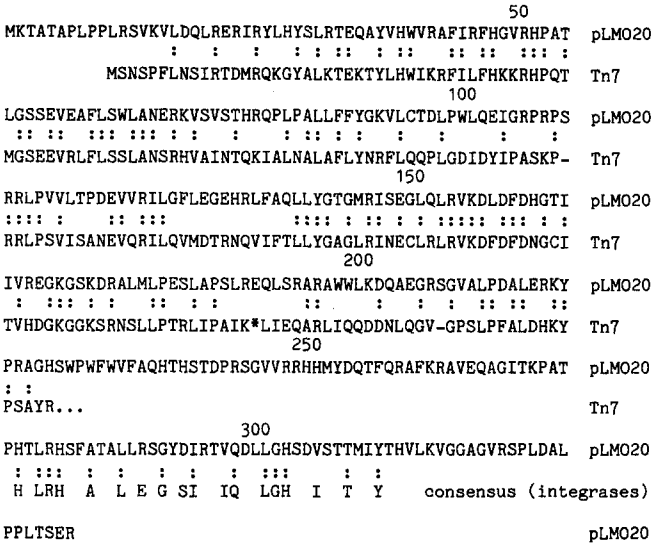

Fig. 6. Comparison of the predicted amino acid sequence of orf2 in pLMO20 with a phage integrase consensus sequence (see text) and with the sequence from an open reading frame to the left of *dhfrI* in Tn*7* (Simonsen et al. 1983)

close to the carboxy-terminals of the integrase proteins of bacteriophages P2, 186, P22, λ, ϕ80 and P4, of the bacteriophage P1 Cre recombinase and also of the *fimB* and *fimE* inversion proteins of *E. coli* is shown. This consensus sequence consists of those 18 amino acids which were identical in 4 out of 9 of the mentioned proteins (Eisenstein et al. 1987). It can be seen that the putative orf2 protein has 14 amino acids in common with this consensus sequence. The orf2 sequence of pLMO20 showed, furthermore, a remarkable similarity to the putative protein translated from an open reading frame to the left of *dhfrI* in Tn*7* (Simonsen et al. 1983; cf. also Fig. 7). A similar observation was made previously by Hall and Vockler (1987); part of the plasmid pKM101 (R46) which is very similar to the resistance region of Tn*21*, was found to contain an orf differing from our orf2 by only 2 amino acids (Ala instead of Pro at positions 82 and 84, respectively).

Genetic organization of the right end of Tn21

When nucleotides 1–1289 of Fig. 2 (and Table 2), nucleotides 141–992 of Fig. 5, and nucleotides 430–2079 of Fig. 4 are linked together, a sequence of 3794 bp covering the antibiotic resistance region of Tn*21* is obtained (cf. Fig. 1). This segment could then be linked to previously published sequence data for the *tnpA* gene (Ward and Grinsted 1987), for the *tnpR* gene (Diver et al. 1983) and for the suggested modulator gene *tnpM* (Hyde and Tu 1985). Taken together these data comprise the continuous sequence of 7.89 kb of the right end of Tn*21*.

Discussion

The rapid spread of antibiotic resistance traits among pathogenic bacteria seems to depend on the recombinational insertion of resistance genes into plasmids and transposons. The highly conserved sequences surrounding *dhfrV* in pLMO20, *aadA* in Tn*21* and *dhfrII* in R388 described in this work, could be regarded as a recipient structure featur-

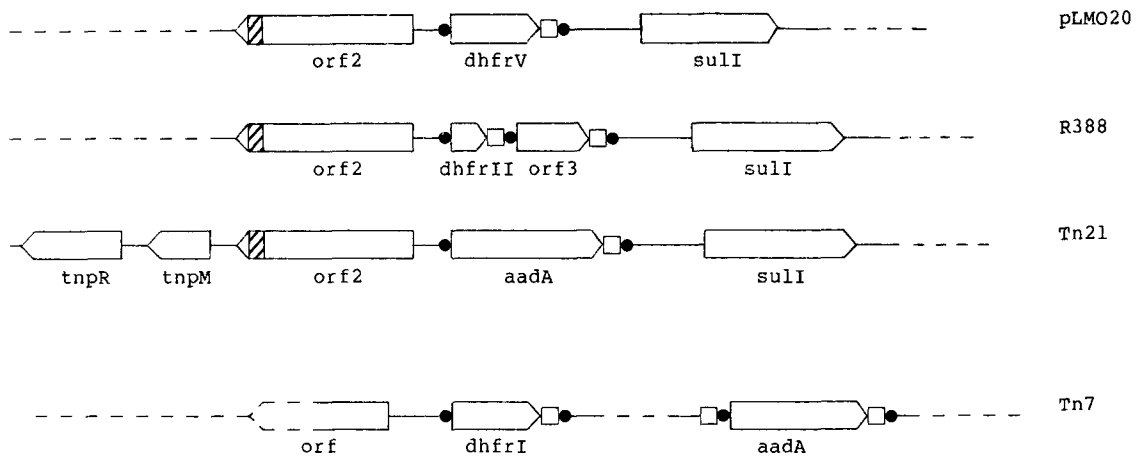

Fig. 7. A comparison of the genetic organization of the antibiotic resistance region of transposon Tn*21* with that of plasmids pLMO20 and R388 and of transposon Tn*7*. Data for Tn*7* are from Fling and Richards (1983), Simonsen et al. (1983) and Fling et al. (1985). The GTTA sequences at branch points (see text) are indicated by the *filled circles*. Sequences corresponding to the potential stem and loop structures described in the text are indicated by the *open squares*. Regions of unknown sequence are marked with an *interrupted line*. The *hatched areas* denote that part of orf2 representing the suggested DNA-binding region, the sequence of which is shown in Fig. 6

ing such a phenomenon. These sequences could then provide an evolutionary older background containing an efficient promoter and thus receive and express recombinational insertions of new resistance genes in response to changing patterns of antibiotics usage. In the cases of *dhfrV* in pLMO20 and *dhfrII* in R388 these recombinations result in a genetic structure endowing its host with resistance to both trimethoprim and sulfonamide, which could be seen as a response to the very frequent use of this drug combination. The occurrence of *sulI* in the presumed receptor structure could reflect an early distribution of this gene in response to the early and widespread use of sulfonamides.

Sulfonamide resistance as represented by the two genes *sulI* and *sulII*, is known to be one of the resistance traits most widely spread on resistance plasmids (Bukhari et al. 1977). Both of these genes express drug resistant variations of the target enzyme for sulfonamide, dihydropteroate synthase. The two plasmid-borne enzymes can sharply distinguish between sulfonamide and the normal substrate, para-aminobenzoic acid, in that they are virtually insensitive to the inhibiting effect of the drug, while showing an affinity for the normal substrate similar to that of the chromosomal enzyme (Swedberg and Sköld 1980). The sequence of the *sulI* gene was determined and found to correspond to a peptide of 279 amino acids with a molecular weight of 30,126. No other enterobacterial dihydropteroate synthase has previously been characterized at the molecular level.

The precise point of insertion indicated in Fig. 5 for the three genes *dhfrV*, *dhfrII* and *aadA* in the common sequence also containing *sulI*, is interpreted to be the target site of a site-specific recombination mechanism. The short GTTA sequence flanking the inserted genes could be the part of a target sequence that gets duplicated at recombination. Furthermore, the four direct repeats situated close to the suggested target site (Fig. 5) could indicate the occurrence of a DNA structure that is efficient as a substrate for an enzymic recombination system. From Fig. 1 it can be seen that the inserted sequences were not much longer than the inserted resistance genes with the exception of *dhfrII* in R388. In this case, another reading frame, orf3, was found downstream of *dhfrII*. Its function is not known. A recombination phenomenon similar to that inferred here was experimentally demonstrated by Tanaka et al. (1985), who observed a recombinational incorporation of the OXA-1 β-lactamase gene into Tn*21*. The site of recombination was not determined precisely, but restriction nuclease mapping indicated a location upstream of *aadA* which could coincide with the upper branch point described here.

At the 3′ ends of the inserted genes *dhfrV*, *dhfrII* and *aadA* (also orf3) there were weakly conserved inverted repeats allowing the formation of stem and loop structures, which included the GTTA sequence at the lower branch points. The constant occurrence of these structures as regions attached to the inserted genes suggests that they are involved in the putative recombination phenomenon. Similar structures, with the potential of hair pin formation, were observed in Tn*7* within an element comprising 54 nucleotides (Fling et al. 1985; Wiedemann et al. 1986). Similar 54-nucleotide-long elements were identified by Cameron et al. (1986) and inferred to be involved in the recombinational insertion of the *aadA* gene conferring streptomycin/spectinomycin resistance and the *aadB* gene conferring gentamicin resistance into R plasmids pSa and pDGO100, respectively.

Finally, the predicted protein from orf2, described here, with its remarkable similarity to site-specific recombinases (Fig. 6) could be the effector of antibiotic resistance gene insertion into pLMO20, Tn*21* and R388. A similar proposal has been made by Ouellette and Roy 1987.

In Fig. 7 a comparison of similarities and genetic organization of the antibiotic resistance regions is made between Tn*21* and pLMO20, R388 and Tn*7*. As described above, there is almost complete identity between Tn*21* and pLMO20 and R388 in the areas flanking the inserted genes. This similarity can be extended to include pSa, which also carries an *aadA* gene that shows 88% homology to that of Tn*21* and Tn*7* (Tait et al. 1985; Cameron et al. 1986). In Fig. 7 it is also shown that all the antibiotic resistance genes in the compared structures, including Tn*7* have, at their 3′ ends, potential stem and loop structures starting with TAAC and ending with GTTA (cf. also Fig. 5).

In Tn*7*, the two resistance genes *dhfrI* and *aadA* both

have such a structure at their 3′ ends. A potential stem and loop structure of the type described, also occurs at the 5′ end of *aadA* (Fig. 7). This could be interpreted to have been involved in the insertion of another part of Tn*7*, situated between the two resistance genes. This could be seen as a parallel to the situation in R388, where two reading frames, *dhfrII* and orf3 are inserted in tandem (Fig. 7). Furthermore, a similarity was also shown between the *dhfrV* gene in the Tn*21*-like structure of pLMO20 and the *dhfrI* gene of Tn*7*. Both genes code for peptides of 157 amino acids and the deduced sequences of these showed 75% homology. It is thus quite likely that both genes have a common ancestry and that they could have a common mechanism of recombinational insertion.

A succession of recombinations could be discerned as being responsible for the development of Tn*21*-like transposons, like the insertion of an 11.2 kb segment into an ancestral transposon, Tn*501* (Brown et al. 1986), and the subsequent site-specific addition of many different genes for antibiotics resistance.

Acknowledgements. This work was supported by a grant from the Swedish Medical Research Council to Ola Sköld and Göte Swedberg, by Fellowships from the Swedish Academy of Pharmaceutical Sciences and Lennander's stipendiary fund to Lars Sundström and from the I.F. Foundation for Pharmaceutical Research to Peter Rådström.

References

Argos P, Landy A, Abremski K, Egan JB, Haggard-Ljungquist E, Hoess RH, Kahn ML, Kalionis B, Narayana SVL, Pierson LS, Sternberg N, Leong JM (1986) The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J 5:433–440

Bachmann BJ (1972) Pedigrees of some mutant strains of *Escherichia coli* K-12. Bacteriol Rev 36:525–557

Bolivar F, Rodriguez RL, Green PJ, Betlach MC, Heyneker HL, Boyer HW (1977) Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 2:95–113

Brown NL, Misra TK, Winnie JN, Schmidt A, Seiff M, Silver S (1986) The nucleotide sequence of the mercuric resistance operons of plasmid R100 and transposon Tn*501*: further evidence for *mer* genes which enhance the activity of the mercuric ion detoxification system. Mol Gen Genet 202:143–151

Bukhari AI, Shapiro JA, Adhya SL (1977) DNA insertion elements, plsmids and episomes. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Cameron FH, Obbink DJG, Ackerman VP, Hall RM (1986) Nucleotide sequence of the AAD(2″) aminoglycoside adenylyltransferase determinant *aadB*. Evolutionary relationship of this region with those surrounding *aadA* in R538-1 and *dhfrII* in R388. Nucleic Acids Res 14:8625–8635

Davis RW, Botstein D, Roth JR (1980) A manual for genetic engineering: advanced bacterial genetics. Cold Spring Harbor Laboratories, Cold Spring Harbor, New York

de la Cruz F, Grinsted J (1982) Genetic and molecular characterization of Tn*21*, a multiple resistance transposon from R100.1. J Bacteriol 151:222–228

Diver WP, Grinsted J, Fritzinger DC, Brown N, Altenbuchner J, Rogowsky P, Schmitt R (1983) DNA sequences of and complementation by the *tnpR* genes of Tn*21*, Tn*501*, and Tn*1721*. Mol Gen Genet 191:189–193

Eisenstein BI, Sweet DF, Vaughn V, Friedman DI (1987) Integration host factor is required for the DNA inversion that controls phase variation in *Escherichia coli*. Proc Natl Acad Sci USA 84:6506–6510

Fling ME, Richards C (1983) The nucleotide sequence of the trimethoprim-resistant dihydrofolate reductase gene harbored by Tn*7*. Nucleic Acids Res 11:5147–5158

Fling ME, Kopf J, Richards C (1985) Nucleotide sequence of the transposon Tn*7* gene encoding an aminoglycoside-modifying enzyme, 3″(9)-*O*-nucleotidyltransferase. Nucleic Acids Res 13:7095–7106

Hall RM, Vockler C (1987) The region of the IncN plasmid R46 coding for resistance to β-lactam antibiotics, streptomycin/spectinomycin and sulfonamides is closely related to antibiotic resistance segments found in IncW plasmids and in Tn*21*-like transposons. Nucleic Acids Res 15:7491–7501

Harley CB, Reynolds RP (1987) Analysis of *E. coli* promoter sequences. Nucleic Acids Res 15:2343–2361

Hollingshead S, Vapnek D (1985) Nucleotide sequence analysis of a gene encoding a streptomycin/spectinomycin adenyltransferase. Plasmid 13:17–30

Holmes DS, Quigley M (1981) A rapid boiling method for the preparation of bacterial plasmids. Anal Biochem 114:193–197

Hyde DR, Tu C-PD (1985) *tnpM*: A novel regulatory gene that enhances Tn*21* transposition and suppresses cointegrate resolution. Cell 42:629–638

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratories, Cold Spring Harbor, New York

Ouellette M, Roy PH (1987) Homology of ORF:s from Tn*2603* and from R46 to site-specific recombinases. Nucleic Acids Res 15:10055

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schmidt F, Klopfer-Kaul I (1984) Evolutionary relationship between Tn*21*-like elements and pBP201, a plasmid from Klebsiella pneumoniae mediating resistance to gentamicin and eight other drugs. Mol Gen Genet 197:109–119

Sharp PA, Cohen SN, Davidson N (1973) Electron microscope heteroduplex studies of sequence relations among plasmids of *Escherichia coli*. II. Structure of drug resistance (R) factors and F factors. J Mol Biol 75:235–255

Simonsen CC, Chen EY, Levinson AD (1983) Identification of the type I trimethoprim-resistant dihydrofolate reductase specified by the *Escherichia coli* R-plasmid R483: Comparison with procaryotic and eucaryotic dihydrofolate reductases. J Bacteriol 155:1001–1008

Smith DR, Calvo JM (1980) Nucleotide sequence of the *E. coli* gene coding for dihydrofolate reductase. Nucleic Acids Res 8:2255–2274

Sundström L, Vinayagamoorthy V, Sköld O (1987) Novel type of plasmid-borne resistance to trimethoprim. Antimicrob Agents Chemother 31:60–61

Swedberg G (1987) Organization of two sulfonamide resistance genes on plasmids of gram-negative bacteria. Antimicrob Agents Chemother 31:306–311

Swedberg G, Sköld O (1980) Characterization of different plasmid-borne dihydropteroate synthases mediating bacterial resistance to sulfonamides. J Bacteriol 142:1–7

Swedberg G, Sköld O (1983) Plasmid-borne sulfonamide resistance determinants studied by restriction enzyme analysis. J Bacteriol 153:1228–1237

Swift G, McCarthy BJ, Heffron F (1981) DNA sequence of a plasmid-encoded dihydrofolate reductase. Mol Gen Genet 181:441–447

Tait RC, Rempel H, Rodriguez RL, Kado CI (1985) The aminoglycoside-resistance operon of the plasmid pSa: nucleotide sequence of the streptomycin-spectinomycin resistance gene. Gene 36:97–104

Tanaka M, Matsushita K, Yamamoto T (1985) Genesis of a complex transposon encoding the OXA-1 (type II) β-lactamase gene. Antimicrob Agents Chemother 28:227–234

Timmis K, Cabello F, Cohen SN (1978) Cloning and characterization of *Eco*RI and *Hin*dIII restriction-endonuclease-generated fragments of antibiotic resistance plasmids R6-5 and R6. Mol Gen Genet 162:121–137

Ward E, Grinsted J (1987) The nucleotide sequence of the *tnpA* gene of Tn*21*. Nucleic Acids Res 15:1799–1806

Ward JM, Grinsted J (1982) Physical and genetic analysis of the Inc-W group plasmids R388, Sa, and R7K. Plasmid 7:239–250

Wiedemann B, Meyer JF, Zuhlsdorf MT (1986) Insertions of resistance genes into Tn*21*-like transposons. J Antimicrob Chemother [Suppl C] 18:85–92

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119

Young H-K, Amyes SGB (1986) A new mechanism of plasmid trimethoprim resistance. Characterization of an inducible dihydrofolate reductase. J Biol Chem 261:2503–2505

Zolg JW, Hänggi UJ (1981) Characterization of an R plasmid-associated, trimethoprim-resistant dihydrofolate reductase and determination of the nucleotide sequence of the reductase gene. Nucleic Acids Res 9:697–710

Communicated by W. Goebel

Received September 21, 1987 / March 17, 1988

Note added in proof

The gene for the recently identified (Sundström et al. 1987) trimethoprim resistance trait borne on plasmid pLMO20 was named *dhfrV*. Similarly, the two known genes for sulfonamide resistance of type I and II (Swedberg and Sköld 1983; Swedberg 1987) have been named *sulI* and *sulII*, respectively