

Two Type I restriction enzymes from Salmonella species:
purification and DNA recognition sequences

Valakunja Nagaraja, John C.W. Shepherd, Therese Pripfl and Thomas
A. Bickle*

Department of Microbiology
Biozentrum, University of Basel
Klingelbergstr. 70
CH-4056 Basel/Switzerland
061/25 38 80

*To whom correspondence should be addressed

SUMMARY

We have purified the type I restriction enzymes SB and SP from Salmonella typhimurium and S. potsdam, respectively, and determined the DNA sequences that they recognise. These sequences resemble those previously determined for the type I enzymes EcoB, EcoK and EcoA in that the specific part of the sequence is divided into two domains by a spacer of non-specific sequence that has a fixed length for each enzyme. Two main differences from the previously determined sequences are seen. Both of the new sequences are degenerate and one of them, SB, has one tri- and one pentanucleotide specific domains rather than the tri- and tetranucleotide domains seen for all of the other enzymes. The only conserved features of the recognition sequences are the adenosyl residues that are methylated in the modification reaction. For all of the enzymes these are situated ten or eleven base pairs apart, one on each strand of the DNA. This suggests that the enzymes bind to DNA along one face of the double helix making protein-DNA interaction in two successive major grooves with most of the non-specific spacer sequence in the intervening minor groove.

I. INTRODUCTION

The type I restriction and modification enzymes were first investigated in vivo more than twenty years ago (Arber & Dussoix, 1962; Dussoix & Arber, 1962) and they were also the first restriction enzymes to be purified and studied in vitro (Meselson & Yuan, 1968). Only two of the type I systems have been well studied although it is known that these enzymes are quite widely distributed, at least in the Enterobacteriaceae. The well studied examples are the enzymes EcoK and EcoB from the Escherichia coli strains K12 and B respectively. The enzymes are multisubunit and multifunctional proteins exhibiting DNA cleavage, ATP hydrolysis and DNA methylation at different stages of the enzymatic reaction. Both ATP and S-adenosyl methionine (AdoMet) are required for the restriction reaction in contrast to the simpler type II enzymes which require no cofactor other than Mg^{+2} (reviewed in Modrich, 1979; Endlich & Linn, 1981; Yuan, 1981; Bickle, 1982).

Genetic analysis has led to the identification of three genes coding for the subunits of the type I restriction enzymes (Boyer & Roulland-Dussoix, 1969; Hubacek & Glover, 1970; Glover & Colson, 1969). The genes are contiguous and located counterclockwise to serB at 98.5 minutes on the E. coli chromosome (Bachman & Low, 1983). The genes are designated hsdR, hsdM and hsdS. The hsdS gene product is responsible for the DNA sequence specificity of the enzymes; the hsdS and hsdM gene products are necessary for modification, and the products of all three genes are required for restriction (Sain and Murray, 1980; reviewed in Bickle, 1982; Suri et al., 1983).

The genetics of some type I restriction systems from various Salmonella species has also been investigated. The genes map at the same position in the chromosome as those from E. coli (Colson & Van Pel, 1974) and it was shown that they could complement genetic defects in the E. coli enzymes (Bullas et al., 1976, 1980; Bullas & Colson, 1975; Van Pel & Colson, 1974). More recently, Southern hybridisation experiments using probes derived from the E. coli K12 hsd region have shown a high degree of homology between the hsdM and hsdR genes of various species of Salmonella and different strains of E. coli (Sain & Murray, 1980; Murray et al., 1982). Furthermore, antibodies raised against the hsdR and hsdM subunits of EcoK show cross reactivity with EcoB, EcoD and the SP, SB and SQ systems of Salmonella, confirming their allelic nature (Murray et al., 1982). There is less homology among the different hsdS genes. Only two short regions within the various genes are sufficiently conserved to show homology in Southern hybridisation experiments (Murray et al., 1982) and DNA sequence analysis of the hsdS genes from different E. coli strains showed that these short regions were almost completely conserved while the rest of the genes were totally divergent (Gough & Murray, 1983). It was suggested that the two conserved regions of the gene most likely code for the parts of the protein that are involved in interactions with the other two subunits of the enzyme and with DNA (Gough & Murray, 1983).

The DNA sequences recognised by EcoK and EcoB have been determined and also that recognised by a non-allelic type I enzyme, EcoA (Kan et al., 1979; Kröger & Hobom, 1984; Lautenberger et al., 1978; Ravetch et al., 1978; Sommer & Schaller, 1979; Suri et al., 1984). These are unusual recognition

sequences in that they are all bipartite, consisting of a specific tetranucleotide and a specific trinucleotide separated by a non-specific spacer sequence, six base pairs long for EcoK, seven base pairs long for EcoA and eight base pairs long for EcoB.

In order to learn more about the type I restriction enzymes and the kinds of DNA sequences that they can recognise, we have investigated two of the type I systems from Salmonella species. These are the SB system from S. typhimurium and the SP system from S. potsdam. In this paper we describe the purification of the enzymes and the determination of their DNA recognition sequences. A comparison of these recognition sequences with those of EcoK and EcoB shows that the only positions that are conserved between all four sequences are the two adenosyl residues that become methylated during the modification reaction. These two residues are in all cases ten base pairs apart suggesting that, during modification, the enzymes are interacting with one face of the double helix only, making protein-DNA contacts in two successive major grooves.

2. MATERIALS AND METHODS

Biological materials

The strains used for the purification of restriction enzymes were L4001 for SB and L4002 for SP. These are E. coli K12 derivatives to which the hsd genes from the Salmonella strains were transferred by transduction (Bullas et al., 1976). Bacteriophage T7 was a gift from F.W. Studier and was the strain that was used to determine the genome sequence (Dunn & Studier, 1983). λcI85757 and the plasmids used were from the collection of this laboratory. Plasmid DNA was purified from cleared lysates by CsCl-ethidium bromide equilibrium centrifugation (Clewell & Helinski, 1969).

Nucleic acid methods

Type II restriction enzymes were from New England Biolabs or from Boehringer. DNA fragments were separated on either 1 % agarose gels or on 4 % polyacrylamide gels depending on the expected fragment sizes. The gels were stained with ethidium bromide and photographed. To detect tritiated DNA bands the gels were fluorographed according to Chamberlain, (1979). They were soaked for 30 min in 1 M sodium salicylate, dried and exposed to preflashed Fuji RX X ray film for one to several days.

Assays for type I restriction enzymes

Enzyme purifications were monitored using the non-modified DNA and AdoMet-dependent ATPase activity of the enzymes as described in Hadi et al. (1975). In most cases enzyme was also detected by immunoblotting essentially according to Towbin et al. (1979) using antibodies against the EcoK hsdM and hsdR subunits and detection with ¹²⁵I-labeled protein A. DNA methylation assays

were as described in Suri et al. (1984) with some minor modifications. Large scale reaction mixtures had a volume of 600 μ l and contained 100 mM N-hydroxyethyl piperazine-N'-2-ethane sulphonic acid pH 7.2 or 7.4, 6.5 mM $MgCl_2$, 14 mM 2-mercaptoethanol, 3 μ M 3H -methyl AdoMet (Amersham, 62 to 78 Ci/mmol), 30-40 μ g of DNA and 10 μ g of enzyme. Reactions were generally incubated for 12-14 h at 37°C. The reactions were terminated by two extractions with an equal volume of buffer saturated phenol and the DNA was separated from low molecular weight radioactive material by gel filtration on a Biogel A-0.5 M column. Smaller scale assays had a volume of 25 μ l and contained about a twentieth of the amount of DNA and enzyme.

Purification of SB and SP

The procedure was essentially the same for both enzymes. Cells (L4001 for SB or L4002 for SP) were grown at 37°C in 100 l of a yeast extract-tryptone medium until late log phase and harvested by centrifugation. The yield of cells was about 2.5 g/l of culture. A purification starting with 160 g of cells will be described. The cells were suspended in 20 mM Tris.HCl, pH 8.0, 10 mM $MgCl_2$, 7 mM 2-mercaptoethanol and 5 % glycerol. All further operations were at 0-4°C. The cells were opened by sonication and the extract was clarified by centrifugation, first at 18,000 rev/min for 20 min in a Sorvall SS34 rotor and then at 38,500 rev/min for 3 h in an IEC A170 rotor. The supernatant was made 0.2 M in NaCl and nucleic acids were precipitated by the addition of a 10 % neutralised solution of polyethyleneimine to a final concentration of 0.5 %. After stirring for 1 h, the precipitate was removed by centrifugation. The supernatant was fractionated by the addition of solid $(NH_4)_2SO_4$ and the protein precipitating

between 35 and 70 % of saturation was collected by centrifugation and dissolved in 20 mM Tris-HCl, pH 8.0, 75 mM NaCl, 7 mM 2-mercaptoethanol, 0.1 mM EDTA and 10 % glycerol (Buffer A). This solution was dialysed twice against Buffer A and a protein precipitate that appeared was removed by centrifugation.

The dialysate was loaded on a DEAE-Sephacel column (bed volume 200 ml) equilibrated with Buffer A and after washing with the same buffer the column was eluted with a linear gradient between 75 mM and 600 mM NaCl. The enzymes eluted at a molarity of approximately 250 mM. The enzyme containing fractions were pooled and the proteins were concentrated by precipitation with $(NH_4)_2SO_4$ (70 % of saturation). The precipitate was dissolved in a small volume of Buffer A, dialysed against Buffer A overnight, and then applied to a column of Sephacryl S-200 (column width 2.6 cm, bed height 100 cm) equilibrated with Buffer A. The column was eluted at 13.6 m/h with buffer A and the active fractions were pooled and loaded on to a heparin agarose column (20 ml bed volume) equilibrated with Buffer A. The column was eluted with a linear gradient between 75 and 800 mM NaCl and enzymatic activity appeared at about 340 mM NaCl. The active fractions were pooled and dialysed against Buffer B (20 mM Tris.HCl, pH 7.0, 25 mM NaCl, 7 mM 2-mercaptoethanol, 0.1 mM EDTA and 10 % glycerol). The dialysate was applied to a column of ATP-agarose (bed volume 6 ml; Sigma A-6888) equilibrated with Buffer B and the column was eluted with a linear gradient of NaCl between 25 mM and 1 M. The enzyme eluted at about 300 mM. At this stage, the enzymes were more than 90 % pure as judged by SDS-polyacrylamide gel electrophoresis and contained no contaminating nucleases or DNA methylases.

modification methylase, by definition, methylates the DNA within the recognition sequence. The method depends on the availability of fully sequenced DNA molecules and the possibility of manipulating these DNA sequences in a computer. DNA of known sequence is methylated by the modification methylase activity that the type I enzymes show in the absence of ATP using ^3H -methyl AdoMet as the methyl group donor. The labeled DNA is then digested with different type II restriction enzymes and the resulting fragments are separated by gel electrophoresis. The fragments which contain a recognition site for the type I enzyme are radioactive and can be identified after fluorography. In this way, using different combinations of type II enzymes, the type I recognition sequences can be precisely mapped to well defined regions of the DNA molecule. A computer is then used to search for sequences present within the labeled regions and absent from the rest of the DNA molecule. This procedure normally produces a short list of candidates for the recognition sequence and these are tested using other sequenced DNA molecules. The approach works well for the determination of simple, short recognition sequences such as those recognised by type III restriction enzymes but requires some modification before it will find the longer, interrupted sequences recognised by type I restriction enzymes.

Determination of the recognition sequence of SP

We began the determination of the SP site using phage λ DNA (Sanger et al., 1982). The DNA was methylated with the SP enzyme and the methylated DNA was digested with various type II restriction enzymes and analysed by gel electrophoresis and fluorography. The SP sites were first mapped roughly with hexamer

3. RESULTS

The SB and SP restriction enzymes

The purification scheme for the SB and SP restriction enzymes is described in Materials and Methods. Both enzymes are very similar to the well known enzymes EcoK and EcoB with respect to their enzyme structure and function, as expected from their allelic nature. Both enzymes contain three subunits with the same molecular weights as those from EcoK and present in similar ratios as judged from SDS-polyacrylamide gel electrophoresis and they have very similar reaction properties: they require both ATP and AdoMet for restriction which is followed by massive ATP hydrolysis and, in the absence of ATP they act as modification methylases.

Strategy for the determination of the SB and SP enzyme DNA recognition sequences

The determination of type I restriction enzyme recognition sequences is complicated by the fact that these enzymes cleave the DNA randomly at distances that can be as much as several thousand base pairs from their recognition sites (Horiuchi & Zinder, 1972; Adler & Nathans, 1973; Bickle et al., 1978). The recognition sequences of EcoK and EcoB were determined by sequencing the DNA at mutated sites (Kan et al., 1979; Lautenberger et al., 1978, 1979; Ravetch et al., 1978; Sommer & Schaller, 1979) but this approach has severe limitations.

We have applied an approach that we originally developed for the determination of type III restriction enzyme recognition sequences (Hadi et al., 1979; Piekarowicz et al., 1981) and more recently adapted to elucidate the EcoA recognition sequence (Suri et al., 1984). This method relies on the fact that the

cutting enzymes and then more precisely with fine cutting enzymes. Figure 1 shows an experiment where the DNA has been cut with enzymes that have many cleavage sites on the λ genome. The figure demonstrates that λ DNA has eight sites for the SP enzyme, all of which could be mapped to within 50-300 base pair regions of the 48.5 kb λ genome. When this data was analysed as described in the last section, no sequence was found that was common to the labeled regions and absent from the rest of the genome even allowing for interruptions at different positions by non-specific sequences of lengths between 2 and 20 nucleotides.

We therefore turned to other sequenced DNA molecules and screened a number of them for the presence of SP sites by testing whether they were a substrate for the *in vivo* methylation reaction. We found some substrates for the enzyme, pBR322 and some of its derivatives. ϕ X174 DNA on the other hand was not a substrate. We methylated pBR322 DNA with the SP enzyme and mapped the sites (Fig. 2). There are two sites for SP in pBR322 DNA since two fragments are labeled in most of the digests, the exception being the HqiAI digest in which three labeled bands are found (Fig. 2, lane 5), two of them of lower intensity compared to the third one. This result indicates that one of the SP recognition sites contains a HqiAI site and that the two tritiated methyl groups in that SP site are found on two different HqiAI fragments. The two sites in pBR322 DNA were mapped and this time the analysis produced a single sequence that could be a candidate for the SP site:

5'-A-A-C-N-N-N-N-N-N-N-G-T-G-C-3'

This is the EcoK recognition sequence (Kan *et al.*, 1979) and it occurs five times in the λ genome in regions where we had mapped SP sites. We then reasoned that perhaps the SP recognition site was degenerate and that the other site recognised by the enzyme might differ from the EcoK site by a single nucleotide. The λ genome was then searched to see whether the EcoK site together with a sequence differing from it by one nucleotide could account for the SP methylation data. This search was successful and predicted the SP site to be:

5'-A-A-C-N-N-N-N-N-N-N-G-T-R-C-3'

where R is either purine.

Support for the idea that the SP site was degenerate and might include the EcoK site came from earlier experiments where it was shown that SP modified λ DNA was not restricted in cells expressing EcoK (Bullas *et al.*, 1980). We have tested this biochemically. First, we have shown that SP modified λ DNA is not a substrate for EcoK using the DNA cleavage dependent ATPase activity of the enzyme as an assay (Table 1). Second, we have methylated EcoK modified DNA with the SP enzyme and ³H-methyl-AdoMet. As shown in Figure 3, only three of the eight SP sites are now methylatable, precisely the ones in those regions predicted to contain the variant of the EcoK site.

We now searched the DNA sequence data bank for DNA molecules that would contain the 5'AAC(N₆)GTAC-3' sequence but not the EcoK site. SV40 DNA (Reddy *et al.*, 1978) satisfied this requirement. We methylated SV40 DNA *in vitro* with the SP enzyme and processed it as described above for λ or pBR322 DNA. Here, however, we took advantage of the fact that the GTAC part of the sequence is a site for the type II restriction enzyme, RsaI. Figure 4 shows

firstly that SV40 does contain a single site for the SP enzyme which maps at the expected place. Secondly, a new, slowly moving band is seen in the RsaI digest of SP methylated DNA compared to non-methylated DNA due to non-digestion by RsaI at the methylated site and this band contains most of the radioactivity.

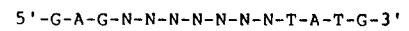
Figure 5 compiles all of the SP recognition sequences that we have tested. An examination of all of these sequences shows that the indicated seven base pair sequence represents the SP site. All four possible nucleotides can be found flanking the sequence and at every position within the 6 base pair non-specific spacer sequence.

Determination of the SB recognition sequence

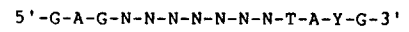
To determine the SB recognition sequence we did similar experiments to those described in the last section for SP. An experiment to map the SB sites in λ DNA is shown in Figure 6. There are five SB sites in the λ genome which could be precisely mapped but, as with SP, a computer analysis of this data produced no sequences as candidates for the SB site, leading us to suspect that again as for SP, the SB recognition sequence might be degenerate. We, therefore, again began to screen various plasmids by in vitro methylation for the presence of SB sites in the hope that we could find one that had only one of the possible recognition sequences.

Figure 7 shows an analysis of pBR322 DNA. This plasmid contains two sites for SB since two fragments are labeled in most of the digests. These sites could be accurately mapped, especially because one of them contains within it a sequence recognised by HgiAI and thus this digest shows three labeled

bands in the fluorogram, two having lower intensities (Fig. 7, lane a). Analysis of this data showed that a single, unique sequence could be the SB site:



A computer search of the λ DNA sequence showed that it is present once at position 45,955, a region where one of the SB sites maps. We again searched the λ sequence using the putative SB site derived from the pBR322 analysis but systematically allowing a "wobble" at each position in the sequence. This search was fruitful in that it gave a unique solution that was compatible with all of the data from both λ and pBR322:



where Y is any pyrimidine.

The next step in the analysis was to verify these sequences. We searched the DNA data bank for other molecules containing both variants of the sequence so that, if the sequences were correct, we could predict the labeling patterns that we should obtain after in vitro methylation. Figure 8 demonstrates an experiment based on this search with two pBR322 derivatives both containing cloned cat genes. pBR325 (Bolivar et al., 1977; Prentki et al., 1981) was predicted to contain three sites, all with T in the wobble position, while pSHI44 DNA (Marcoli et al., 1980) should have four sites, three with T (the same as in pBR325) and one with C in the wobble position. The results show that these sites are present in the predicted restriction fragments of both plasmids.

A close examination of all of the SB sites that we found in λ DNA and the plasmids did not allow us to conclude that the sequence discussed above is the SB recognition sequence. In all

of these sequences the nucleotide in the non-specific spacer adjacent to the specific tetranucleotide is a purine and we could not exclude that the enzyme requires a purine in this position. To settle this point we went back to the DNA sequence data bank to find DNA molecules containing sequences that were otherwise SB sites but with a pyrimidine at the position in question. We found two convenient DNA molecules satisfying these criteria. ϕ X174 DNA (Sanger et al., 1977) which has two sites, both with a T in this position, and phage T7 DNA which has 15 of the sequences, 7 with a purine and 8 with a pyrimidine at the suspect position.

In vitro methylation experiments showed that ϕ X174 DNA was not a substrate for the enzyme (not shown) and that only the seven sites in T7 DNA with a purine were labeled (Figure 9). The sequences around all of the SB sites and SB-related sites that we have investigated are shown in Figure 10. It is apparent that, apart from the sequences that we have discussed, all other positions can contain any of the four nucleotides and the SB recognition sequence is, therefore,

5'-G-A-G-N-N-N-N-N-N-R-T-A-Y-G-3'

4. DISCUSSION

The purified SP and SB restriction enzymes from S. potsdam and S. typhimurium are very similar to the classical type I enzymes, EcoK and EcoB, with respect to their subunit structure and enzymatic activities. This substantiates the earlier findings (reviewed in Bickle, 1982) that all of these enzymes belong to a homogeneous, genetically related group of probably ancient origin. Moreover, the common ancestry of these enzymes is well demonstrated by the extensive homology between their structural genes (Murray et al., 1982). This study also showed that the homology is much less well preserved among the different hds genes whose product is responsible for DNA sequence recognition and this result has been confirmed by sequencing the genes (Gough & Murray, 1983). The sequencing showed that, apart from two short regions that were highly conserved between the different genes, the rest of the genes had completely divergent sequences. Because the hdsS gene product must interact with the hdsM and hdsR gene products, which are both highly conserved, it was thought likely that the conserved regions of the hdsS gene product are involved in these interactions. The non-homologous regions of the hds gene would then code for the portion of the polypeptide that recognises the specific DNA sequence, which is different for each enzyme.

A consideration of the five type I recognition sequences published to date (EcoA, EcoB, EcoK, SB and SP) allows a certain number of generalities to be made. These sequences are shown in Figure 11. The general organisation of all of them is very similar in that they all consist of two specific nucleotide domains separated by a spacer of non-specific sequence which is

of fixed length for any one enzyme but which varies in length between quite strict limits for the different enzymes. For the sequences shown in Figure 11 the lengths are between six and eight base pairs although we have determined the recognition sequence for another type I enzyme where the length is five (V. Nagaraja, unpublished data).

Kan et al. (1979) saw homology at four out of the seven positions in the EcoK and EcoB sites. This homology is no longer apparent when the larger set of sequences in Figure 11 are examined. With the exception of EcoA, which is not allelic to the other four (Suri et al., 1984), the only conserved positions are the A + T base pairs that are the most likely targets for methylation during the modification reaction. All of the enzymes place one methyl group in each strand, one in the tetrameric part and one in the trimeric part. For EcoB there is only one A in the top strand of the sequence as shown in Figure 11 and this is in the trinucleotide part of the sequence. In the lower strand, analysis of methylated oligonucleotides suggested that the first A in the tetranucleotide domain was the methylation site (Van Ormondt et al., 1973). For EcoK there is only one A in the lower strand of the sequence, in the tetranucleotide, and in the upper strand the second A in the trinucleotide is the methylation site because, when the site overlaps with a HindII site, EcoK modification prevents HindII cleavage (Roy & Smith, 1973; Bickle, 1982). The fact that SP methylation protects the DNA from EcoK (this paper) suggests that the pattern of methylation is the same for both enzymes. For SB, the top strand has only one A, in the trinucleotide and, while the bottom strand can have two As, only one is invariable and this therefore is the one that is

methylated. All of these methylated nucleotides are exactly ten base pairs apart. This together with the constraints on the length of the non-specific spacer strongly suggests that during the modification reaction the enzyme interacts along one face of the DNA helix, making contacts in two successive major grooves with most of the non-specific sequence in the minor groove separating them. A similar kind of interaction is also possible for the non-allelic EcoA because in this case the methylated residues are eleven base pairs apart (Suri et al., 1984).

Another interesting feature that emerges from our results is the degeneracy of the sequences of the enzymes from Salmonella, not found for those previously published. The situation is especially interesting for the enzyme SB where instead of the trinucleotide and tetranucleotide domains found for the other enzymes we find trinucleotide and pentanucleotide domains with a degeneracy at two positions within the pentanucleotide.

Finally, how does EcoK recognise one of the two sequences recognised by SP? It has been shown by DNA hybridisation and heteroduplex analysis that the hsdS genes of these two enzymes are no more closely related to each other than they are to the rest of the class (Murray et al., 1982; Fuller-Pace et al., 1984). The genes are about 1.5 kb long and have two homologous regions, one about 100 bp and the other about 250 bp long, the rest of the DNA sequences are totally dissimilar. If one assumes that the regions of the protein that interact with the SP and EcoK recognition sequences ought to have a similar amino acid sequence, this would imply that one of the conserved regions of the hsdS gene codes for the DNA binding region of the protein. The 250 bp conserved region shows too little variation between

EcoK, EcoB and EcoD to allow it to recognise different DNA sequences (Gough & Murray, 1983); it is most probably involved in interactions with the other subunits of the enzyme. The 100 bp region shows more variation (Gough & Murray, 1983) and would be the most likely candidate for the DNA binding domain. Alternatively, the DNA binding domains of the hdsS gene product are situated in the variable regions of the gene, in which case we would have a situation analogous to the immunoglobins, where hypervariable regions allow the recognition of different epitopes.

ACKNOWLEDGEMENTS

We thank F.W. Studier for a gift of T7 phage, Bruno Suri for many helpful discussions and Noreen Murray for communicating results prior to publication. This work would not have been possible without access to the EMBL DNA sequence data bank. Supported by grants from the Swiss National Science Foundation.

DNA	coordinate	sequence	methylated
pSHI44	2288	AGAGTGCACCATATGC	+
	4252	TGAGCTGGTGATATGG	+
	5530	TGAGAATAGTGATGC	+
	5031	GGAGCCTGTCATACGC	+
λ	45955	AGAGATTGATGTATGA	+
	1308	CGAGCCTTCAATACGC	+
	8115	TGAGACAAAGGTACGC	+
	10287	TGAGCACGGTGTACGT	+
	31679	AGAGCAGGCGGTACGC	+
T7	1491	GGAGTAATGTATATGC	+
	8278	GGAGCTAACGATATGA	+
	14342	GGAGAAATCAATATGA	+
	22955	GGAGATATACATATGG	+
	36541	GGAGTGATATGTATGG	+
	10527	TGAGAAGACAATACGG	+
	34785	AGAGATAGTAGTACGT	+
	1456	CGAGTGGCCTTTATGA	-
	3783	AGAGTCTTCCTTATGC	-
	5875	AGAGTGGCCTTTATGA	-
	13573	GGAGAAACCATTATGT	-
	16840	GGAGAAATTATTATGG	-
	18811	CGAGATTGAGTTACGC	-
	34715	AGAGTTACCACTACGA	-
	36874	CGAGATGCTCCTACGT	-
Øx174	53	GGAGGCTTTTTTATGG	-
	2459	TGAGGATAAATTATGT	-
Consensus:		GAGNNNNNR ₁ AYG	

DNA	coordinate	sequence	
pBR322	1652	GAACATAATGGTGCA	
	4026	TAACCCACTCGTGCA	
SV40	1385	TAACACTTCCGTACA	
λ	6942	CAACACGATGGTGCT	
	14981	TAACCCGGACGTGCT	
	16370	AAACGTCTTGGTGCG	
	34764	AAACGATAAAGTGCG	
	47001	CAACGCCCCGGTGCT	
	430	GAACTATTGAGTACG	
T7	18724	GAACGAGGCGGTACA	
	32803	AAACGAATGAGTACT	
	15161	GAACATCTCAGTGCC	
Consensus:	26603	GAACCC ₁ TATCGTGCC	
	32632	CAACGCTACGGTGCC	
	37459	AAACCTACGGTGCC	
	6345	CAACCAATCAGTACA	
	8202	AAACCCGCCAGTACA	
	28109	CAACTGGTGGTACT	
	31460	CAACGTTGATGTACG	
	35465	GAACCAATGTGTACT	
	Consensus:		AACNNNNNGTRC

REFERENCES

- Adler, S.P. & Nathans, D. (1973). Biochim. Biophys. Acta, **299**, 177-188.
- Arber, W. & Dussoix, D. (1962). J. Mol. Biol. **5**, 18-36.
- Bachman, B.J. & Low, K.B. (1983). Microbiol. Rev. **47**, 180-230.
- Bickle, T.A. (1982) In The Nucleases (Linn, S. & Roberts, R.J. eds.) pp 85-108, Cold Spring Harbor Press, New York.
- Bickle, T.A., Brack, C. & Yuan, R. (1978). Proc. Natl. Acad. Sci. U.S.A. **75**, 3099-3103.
- Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.L., Heyneker, A.L., Boyer, H.W., Crosa, J.H. & Falkow, S. (1977). Gene **2**, 95-113.
- Boyer, H.W. & Roulland-Dussoix, D. (1969). J. Mol. Biol. **41**, 459-472.
- Bullas, L.R. & Colson, C. (1975). Mol. Gen. Genet. **139**, 177-188.
- Bullas, L.R., Colson, C. & Van Pel, A. (1976). J. Gen. Microbiol. **95**, 166-172.
- Bullas, L.R., Colson, C. & Neufeld, B. (1980). J. Bacteriol. **141**, 275-292.
- Chamberlain, J.P. (1979). Anal. Biochem. **98**, 132-135.
- Clewell, D. & Helinski, D.R. (1969). Proc. Natl. Acad. Sci. U.S.A. **62**, 1159-1166.
- Colson, C. & Van Pel, A. (1974). Mol. Gen. Genet. **129**, 325-337.
- Dunn, J.J. & Studier, F.W. (1983). J. Mol. Biol. **166**, 477-535.
- Dussoix, D. & Arber, W. (1962). J. Mol. Biol. **5**, 37-49.
- Endlich, B. & Linn, S. (1981) In The Enzymes, vol. **14** part B (Boyer, P.D. ed.) pp 137-156, Academic Press, New York.
- Fuller-Pace, F.V., Bullas, L.R., Delius, H. & Murray, N.E. (1984). Proc. Natl. Acad. Sci. U.S.A., in press.
- Glover, S.W. & Colson, C. (1969). Genet. Res. **13**, 227-240.
- Gough, J.A. & Murray, N.E. (1983). J. Mol. Biol. **166**, 1-19.
- Hadi, S.M., Bickle, T.A. & Yuan, R. (1975). J. Biol. Chem. **250**, 4159-4164.
- Hadi, S.M., Bächli, B., Shepherd, J.C.W., Yuan, R., Ineichen, K. & Bickle, T.A. (1979). J. Mol. Biol. **134**, 655-666.
- Horiuchi, K. & Zinder, N.D. (1972). Proc. Natl. Acad. Sci. U.S.A. **69**, 3220-3224.
- Hubacek, J. & Glover, S.W. (1970). J. Mol. Biol. **50**, 111-127.
- Kan, K.C., Lautenberger, J.A., Edgell, M.H. & Hutchison, C.A. III (1979) J. Mol. Biol. **130**, 191-210.
- Kröger, M. & Hobom, G. (1984). Nucl. Acids Res. **12**, 887-899.
- Lautenberger, J.A., Kan, N.C., Lackey, D., Linn, S., Edgell, M.H. & Hutchison, C.A. III (1978). Proc. Natl. Acad. Sci. U.S.A. **75**, 2271-2275.
- Lautenberger, J.A., Edgell, M.H., Hutchison, C.A. III, & Godson, G.N. (1979). J. Mol. Biol. **131**, 871-875.
- Marcoli, R., Iida, S. & Bickle, T.A. (1980). FEBS Lett. **110**, 11-14.
- Meselson, M. & Yuan, R. (1968). Nature (Lond.) **217**, 1110-1114.
- Modrich, P. (1979). Quart. Rev. Biophys. **12**, 315-369.
- Murray, N.E., Gough, J.A., Suri, B. & Bickle, T.A. (1982). EMBO J. **1**, 535-539.
- Piekarowicz, A., Bickle, T.A., Shepherd, J.C.W. & Ineichen, K. (1981). J. Mol. Biol. **146**, 167-172.
- Prentki, P., Karch, F., Iida, S. & Meyer, J. (1981). Gene **14**, 289-299.
- Ravetch, J.V., Horiuchi, K. & Zinder, N.D. (1978). Proc. Natl. Acad. Sci. U.S.A. **75**, 2266-2270.