Complete sequence of IS3

Kurt P. Timmerman and Chen-Pei D. Tu*

Department of Molecular and Cell Biology, The Pennsylvania State University, University Park, PA 16802, USA

Received 27 December 1984; Revised and Accepted 31 January 1985

ABSTRACT
We have determined the nucleotide sequence of IS3. Our IS3 isolate has 39 bp inverted repeats (IR's) with 6 mismatches, and is 1258 bp long. IS3 contains a large open reading frame (ORF) of 288 codons with a smaller, partially overlapping ORF of 91 codons on the opposite strand in codon-codon register. The large ORF is preceded by and has a 4 bp overlap with a 99 codon ORF that has potential transcriptional and translational start signals. Thus, IS3 could encode a bicistronic mRNA. The Shine-Dalgarno sequence for this 99 codon ORF could be sequestered in a stem-loop structure, but only if the transcript began outside IS3, as was first seen with IS10. This could be a means for preventing fortuitous activation of IS3 by outside promoters. No DNA sequence homology was found between IS3 and other prokaryotic IS elements, but there is slight amino acid sequence homology and significant conservation of hydrophaticity patterns between the putative transposases of IS3 and IS2.

INTRODUCTION
Bacterial insertion sequences (IS elements) are small transposable elements that contain no phenotypic markers. Presumably, they carry only the genetic information necessary for transposition (see 1 and 2 for reviews). IS elements range in size from about 200 bp to 1500 bp or more. Short IR's at the ends of the elements are characteristic of each IS element. Upon insertion, IS elements also generate short direct repeats of host sequence, 2 to 20 bp long. The size, although not necessarily the sequence, of the direct repeats is also characteristic of each element.

IS3, which has been found to exist naturally in E. coli, is reported to contain imperfect 38 bp IR's and to generate 3 bp direct repeats upon insertion (3,4). Two copies of IS3 are located on the F plasmid and four to five copies are found on the E. coli chromosome (5,6).

IS3 was first detected by Malamy in 1966 as a spontaneous lactose metabolism mutant in E. coli (7). The mutant was not suppressible by nonsense suppressors, but otherwise, its nature was unknown. In 1970 Malamy determined that this mutant, MS505, contained an insertion of DNA in the omega region of the lacZ gene (8). This insertion was designated IS3 in 1972, and its size...
was determined by EM measurements to be 1170 ± 100bp (9).

In addition to inactivating genes, IS3 is also capable of activating them. Glansdorff et al. found 3 separate mutations in the control region of the divergently transcribed argECBH operon that activated the weakly expressed argE gene. The argE gene had been largely inactivated by a deletion that removed its promoter. These three mutants were found to contain insertions of IS3 in orientation II (10). The same group later reported that the activation of argE by IS3 in orientation II is due to an outward reading promoter contained in IS3, interior to the right IR (4). IS2 has also been shown to activate genes (11,13,14). Jaurin and Normark reported that IS2 mediated gene activation is caused by the juxtaposition of a -35 region in the right IR of IS2 (in orientation II only) and a -10 region in the host sequence (15).

We have shown earlier that activation by IS elements of a promoterless chloramphenicol acetyltransferase (CAT) gene on a pBR322 derivative can be selected for in the presence of chloramphenicol (Cm). Insertions of IS2 and IS3 5 bp (IS2), 71 bp (IS3), and 767 bp (IS2) upstream from the ATG codon of the pDH5110 CAT gene are able to activate gene expression, though to different extents (16). In this communication we report the complete nucleotide sequence of IS3 and a comparison of its gene organization with that of other known IS elements.

MATERIALS AND METHODS

Antibiotics and Nucleotides

Chloramphenicol (Cm) and ampicillin were purchased from Sigma Chemicals (St. Louis, MO). Spectinomycin sulfate was a generous gift from The Upjohn Co. (Kalamazoo, MI). The four dNTP's were purchased from Pharmacia P-L Biochemicals Inc. (Milwaukee, WI). The [γ-32P] dATP (specific activity >2000 Ci/mmole) was obtained from ICN (Irvine, CA). All four [α-32P] dNTP's were purchased from ICN or NEN (Boston, MA). [α-32P] Cordycepin 5'-triphosphate (specific activity about 3000 Ci/mmole) was a generous gift from NEN. [α-32P] dideoxyadenosine 5'-triphosphate (specific activity about 3000 Ci/mmole) was purchased from Amersham (Arlington Heights, IL).

Enzymes

The restriction endonucleases were purchased from BRL, Inc. (Gaithersburg, MD), New England Biolabs (Beverly, MA), or Boehringer Mannheim (Indianapolis, IN). AMV reverse transcriptase was obtained from Life Sciences, Inc. (St. Petersburg, FL). The terminal deoxynucleotidyl transferase was purchased from Pharmacia P-L Biochemicals Inc. (Milwaukee, WI).
**Bacterial Strains and Plasmids**

_E. coli_ strain C600 (17) was used in all experiments. The plasmid pDH5110 contains a 784 bp TaqI fragment from the pTU2 (18) CAT structural gene inserted by Dr. Shing Chang into the pBR322 ClaI site in the same orientation as the tetracycline resistance gene. Since cutting the CAT gene at the TaqI site destroys the promoter, cells harboring pDH5110 are Cm8.

**Isolation of pABV5**

After prolonged incubation at 37°C of pDH5110-containing cells in liquid media with 100 ug/ml Cm (which is a lethal concentration for Cm8 cells), CmR cells can be isolated at a very low frequency. One of these isolates containing a DNA insertion in pDH5110, designated as pABV5, was further studied to determine the cause of the activation (16).

**Plasmid DNA Purification**

For the initial analysis of the CmR isolates, plasmid DNA was purified by the rapid alkaline extraction procedure of Birnboim and Doly (19). The cleared lysate, CsCl-ethidium bromide gradient method (20) was used for large scale plasmid preparations from spectinomycin amplified cultures as previously described (18).

**Restriction Mapping**

The site of insertion of IS3 was localized and the IS element itself was mapped with the following restriction enzymes: AluI, AvaII, BamHI, DdeI, HaeIII, HhaI, HindIII, HinfI, HpaII, PstI, PvuII, Sau3A, and TaqI. All digests were run under the conditions specified by the supplier except that BSA was omitted. Gel fragments were run on 3.5% polyacrylamide gels and visualized after ethidium bromide staining or autoradiography (21,22). The order of the IS3 DdeI fragments was partially elucidated by reverse transcriptase labelling with cold dTTP and either [α-32P] dCTP, dATP or dGTP. The fragments were separated on a 5%/20% polyacrylamide step gel and visualized by autoradiography (22).

**Preparation of Labelled DNA Fragments**

All sites (with the exception of the PstI sites) were labelled with [α-32P] dNTP's using reverse transcriptase (22). After running through 3.5% polyacrylamide gels, the fragments were purified by electroelution (23) and DE52 minicolumns (22). The HpaII fragment (bp 536-657) was run on a 8% strand separating gel after labelling (24), and the top strand was sequenced. The fragment from the IS3 PstI site leftward (as seen in orientation II) was labelled with [α-32P] ddATP using terminal deoxynucleotidyl transferase (TdT) (25). The IS3 PstI rightward fragment was labelled with [α-32P]
cordycepin 5'-triphosphate (26) using TdT. The TdT-labelled fragments were run on Sephadex G-50 columns to remove unreacted triphosphates prior to the secondary TaqI digests and purification on polyacrylamide gels, as described above.

DNA Sequencing

All fragments were sequenced using the G, A>C, G+A, C>T, and C reactions as described by Maxam and Gilbert (24). The sequences were determined using either 40cm thin polyacrylamide gels (6%, 8%, or 20%) or 80cm gels (5% or 6%) (27).

Computer Analysis: Computer analyses of the sequences were accomplished with programs by Conrad and Mount (28), the SEQ-Sequence Analysis Programs of the Molgen Project (29), the PR1 program by Harr et al. (30), the Matrix program by Zweig (31), and The Hydropathicity Program of the BionetTM Resource (32).

RESULTS

The nucleotide sequence of IS3 was determined from the IS3 insertion in the plasmid pABV5. We have numbered the sequence in orientation II, and therefore, the large ORF reads from left to right (Figure 1). The direct repeats, ATT, and the location of insertion, position 5087 or 5090 of pDH5110, were determined by comparing the sequence of pABV5 with that of pDH5110. As can be seen by our sequencing strategy shown in Figure 2, the majority of IS3 has been sequenced on both strands. Some of the restriction sites are also listed, and a complete list of sites is available upon request.

The exact length of IS3 is 1258 bp. Upon examining this sequence, several significant ORF's were found (Figure 2). The largest ORF, ORF I, which could encode a protein of 288 amino acids, begins at position 362 and ends at position 1226, which is inside the right IR. ORF II, 99 codons long, goes from position 66 to a termination codon ending at position 365. The first two bases of this stop codon are the last two bases of the start codon for ORF I. ORF III is found on the complementary strand, beginning at position 1240 in the IR, and ending at position 967. This ORF uses the same triplets as ORF I, but in the opposite direction, and it starts outside of the sequence used for ORF I. All three reading frames on both strands contain stop codons within 51 bp of the ends of IS3. This may account for the

Figure 1. The complete nucleotide sequence of insertion element IS3. The IS3 sequence is 1258 bp long and is printed with both strands. The 3 bp direct repeats are underlined and the arrows designate the 39 bp IR's. Two stars, at positions 1190 and 1237, mark the nucleotides that differ from the partial sequences of Sommer et al. and Charlier et al., respectively (3,4).
Figure 2. The sequencing strategy and major open reading frames of IS3. IS3's location in pABV5, relative to the CAT structural gene, is shown at the top. The open arrows represent the IR's. The arrows above and below the map of IS3 indicate the fragments sequenced on the top and bottom strands, respectively. ORF's I, II, and III are shown as open boxes. Several of the restriction sites used for sequencing are labelled. Restriction sites: A, AluI; B, BstEII; D, HindIII; H, HinfI; P, PstI; and T, TaqI.

polarity sometimes seen with IS3 insertions (see 1 for review).

By analyzing the sequence of IS3 with the computer program of Harr et al. (30), we located a putative promoter sequence that could be responsible for transcription of the two longest ORF's. The promoter, as shown in Figure 3, has a −35 region at position 17 and a −10 region at position 40 (33). The −10 region, along with a possible Shine-Dalgarno sequence (34), is contained in a stem-loop structure. No likely promoters were found just preceding the longest ORF, ORF I, but the promoter just described might also transcribe this ORF to generate a bicistronic message. The third ORF, which begins at position 1240 on the complementary strand, is too close to the end of IS3 to allow for a complete promoter within the IS element. We also failed to find any promoters at or near the IS3-pDH5110 junction. Our search for transcription termination signals, either rho-dependent or independent, failed to produce any likely structures.

Some transposable elements, such as Tn3, transpose to regions that are AT rich and contain sequences homologous to the ends of the element (18). A computer comparison (29) of the sequence of IS3 with the sequence near the site of insertion revealed a region located 82 bp 5' to the IS3 insertion in
Figure 3. Possible stem-loop structure for a transcript beginning outside of IS3. By sequestering the Shine-Dalgarno sequence (SD) this structure could prevent translation of ORF II. If the IS3 promoter shown were used, the transcript would begin in the loop and no stem-loop structure could form. A -35 region closer to the consensus sequence TTGACA (32) is found 3 bp to the right of the marked -35 region. However, the better -35 region would cause the spacer sequence to be reduced from the optimal 17 bp to 14 bp. The first base is base 1 in IS3, and IR marks the end of the left IR. The boxed ATG is the start codon for ORF II.

pDH5110 that shares some homology with IS3 at positions 1139-1158. However, with a limited number of IS3 insertion site sequences available, it is difficult to propose any consensus sequence at or near the insertion sites.

We compared the DNA sequence of IS3 with that of IS1, IS2, IS4, IS5,

IS3
MKYVFIEKHAESIKAMGVLRVASGWYTWGERRTRISTQGGFQHCDGVRNLAFTRSKQGARYPRLTD
IS2
MNSSKPLMDQGSKCARIAGWGVSLVSCRLVSPAGLHLVILRRTQMODGRRSHTDIDLVRHRAVGE

IS3
ELRAOGYPFNVKTVAYSLRPQ6RRAKASRY
IS2
PTYGRFVWALLPPPDELGMPAINAKPRVRPRNONALLFFRPVPPSKPHFVAVKESWHRCSERFFCGM

IS3
LYLAVVIDLWSRAVIGWSMSPRMTAQLACDALOMALWRKRPRIVHVTTPGDDYKSADYQAOQLKMRHLNRGSMSAK
IS2
ERLRFALDDCDEALHWAFTGGFNTQVDVMLGAVERRFGNDLPSSPEWNTINGSGYRANTQFARMLGEPKNTAV

IS3
GCCYDACKVESFFHLSKVECHIGEHFSREIMATVFNYIECDNNWRHRKSWGGLSPEGENKNLA
IS2
RSPESGLAESFKTVKRDYISMPKPDGLTAANLAEAEFHYNEWHPSALGYSFREYLRQRACHGLSLIMRCLEI

Figure 4. The amino acid sequences of the putative transposases of IS3 and IS2. The sequences are written N- to C-terminal and are aligned to show matches produced by the dot matrix program of Zweig (31). Regions containing at least three matches with greater than 40% homology are boxed. Only matches in the boxes are designated with dashes.
Figure 5. Hydrophaticity plots for the putative transposases of IS3, IS2, and IS102. The plots, aligned to show possible similarities, were generated by the Hydrophaticity Program of BioNet\textsuperscript{TM} (32) using the values of Kyte and Doolittle (57). All other parameters were set to their default values. Positive values indicate hydrophobic regions, which are often located on the interior of proteins. IS2's putative transposase shows some amino acid homology to IS3's while neither show significant homology to IS102's, which is included as a negative control.

IS10, IS26, IS30, IS50, IS101, IS102, ISH1, ISH2, and ISH50 (35-46) using the dot matrix program of Zweig and a condition of 12 of 14 bp match (31). No significant sequence homologies were found. Using a setting of 4 of 6 amino acid match, the same program was also used to compare the amino acid sequences of the putative transposases (ie: the longest ORF) of each of these IS elements, with the exception of IS101 which is probably too small to encode a transposase (47). The only noticeable homology was found between the putative transposases of IS2 and IS3 (Figure 4). The hydrophaticity plots for these putative transposases also show some similarity (Figure 5).

DISCUSSION

An insertion of IS3 that activates the CAT gene in pDH5110 has been isolated. We have sequenced this IS3, compared our results to those of other groups, and have examined the sequence for signals involved in transcription...
and translation.

IS3 has been reported to have 38 bp IR's with 32 bases matching (actually one IR is claimed to be 38 bp long and the other 39 bp) (3). However, our data shows that the IR's of our isolate are both 39 bp long with 33 bp matching. The difference is position 1237 in the right IR (Figure 1): a "C" that is not present in the sequence published by Sommer et al. (3). This "C" pairs with a "G" which is present on the other IR of both ours and their sequences. Charlier et al. published the sequence of 188 bp at the right end of IS3, and their IR sequence matches ours (4). There is, however, one difference between our sequence and that of Charlier et al.: the base at our position 1190 is a "C" while their's is a "G" (Figure 1). Since this base is located in the spacer region of the outward-reading promoter of IS3, the difference should have no effect on the ability of IS3 to activate genes in orientation II (4). Naturally, these differences could be due to different isolates of IS3.

A more substantial difference between our data and that of Sommer et al. involves the size of IS3. Our sequence data gives a length for IS3 of 1258 bp. This is within the range given by Malamy et al. of 1170 ± 100 bp as determined by EM mapping (9). But, Sommer et al. show a length of about 1400 bp (3). Their restriction map agrees with ours in approximately the first 300 bp and the last 500 bp. In the middle region, where the extra 150 bp presumably are located, Sommer et al. did not list any restriction sites. So, we cannot tell whether the size given by Sommer et al. is off by 150 bp, or if we have a different isolate of IS3.

Klaer et al. report that IS4 has potential rho-independant transcription termination signals near both termini (37). They state that these potential termination signals could prevent transcripts that begin outside IS4 from continuing through the IS, and they could also cause the polarity that is seen with many IS elements. If outside transcripts were allowed to read through an IS element, the resulting activation of the element could lead to a higher frequency of transposition which could be deleterious to the host, and therefore, the element itself.

Our examination of the sequence of IS3 did not yield any good potential rho-independent or rho-dependent transcription termination signals. However, a structure that could fulfill the purpose of preventing fortuitous activation (39) of the element was found (Figure 3). The stem-loop structure which begins inside the interior end of the left IR would sequester the ORF II putative promoter's -10 region and the Shine-Dalgarno sequence (34). If a
transcript from outside IS3 were to continue into the left end of IS3, the transcript could form a stem-loop structure and prevent or reduce the translation of the IS3 encoded functions. But, a transcript beginning at the IS3 promoter would not be able to form this stem-loop structure.

Dalrymple et al. and Kroger and Hobom found very similar structures in IS30 and IS5, respectively (41, 48). As in IS3, these stem-loop structures begin inside the interior end of the IR and would sequester the -10 region and part of the Shine-Dalgarno sequence. In the cases of IS10, ISH1, and ISH2, a stem-loop could form just before the major ORF, and it would contain the initiation codon in the stem (39, 45, 44). Presumably, these stem-loop structures would act in the same manner as the structures in IS3 and IS30. It is possible to extend the stem-loop structure in IS3 so that it would also sequester the initiation codon, but such a structure has less thermodynamic stability in the absence of specific binding proteins, as determined by the SEQ program (29), than the one shown in Figure 3.

IS1, IS2, IS4, IS5, IS102, and IS903 have one large ORF with a smaller ORF on the opposite strand, in codon-codon register, and contained within the larger ORF (49). IS3, however, is similar to ISH1 and ISH2 in that the strand opposite to the one with the major ORF encodes a smaller ORF in codon-codon register that is not totally contained within the major ORF (44, 45).

It is believed, but not yet proven, that these ORF's encode transposition functions. There are a number of reports that support these beliefs. Bernardi and Bernardi reported that mutations in the C-terminal portion of IS102's large ORF affected the frequency of deletion formation (50). More significantly, IS5 has been shown to produce, in E. coli minicells, two proteins corresponding to its two overlapping genes (49). Trinks et al. have also found a protein synthesized by IS4 in E. coli minicells that they believe is a product of the large ORF (51).

IS3 contains an ORF, ORF I, comparable in size to that of other IS elements. But, our computer search turned up no obvious promoter sequences preceding ORF I. Due to the low transposition frequency observed with IS3 and other IS elements, the transcription and translation signals for their transposition functions may be inefficient and bear no resemblance to the canonical ones. However, the smaller ORF on the same strand as ORF I, but in a different reading frame, does appear to have the proper transcription and translation start signals. The first two nucleotides of the smaller ORF's (ORF II's) stop codon are the second and third nucleotides of ORF I's start codon (ATGA). The transcript beginning from ORF II's promoter might be a
bicistronic message, containing ORF II and ORF I. The Protein C and Protein A genes in φX174, the hisD and hisC genes in Salmonella typhimurium, and the toxA and toxB genes in E. coli, have all been reported to have an analagous overlap using the sequence ATGA (52-54). If ORF II and ORF I are contained in a bicistronic message, they may have coupled translation, in which translation of ORF I would be dependent on translation of ORF II (52-56). We have examined the possible ORF's of IS1, IS2, IS4, IS5, IS10, IS26, IS30, IS50, IS102, ISH1, ISH2, and ISH50, and have not found any similar arrangements of their large ORF's (34-45).

The significance of ORF III on the "bottom" strand of IS3 is not clear. The pattern of a large and small ORF on opposite strands in codon-codon register seems to be conserved among many IS elements. But, ORF III begins too close to the end of IS3 to allow for a complete promoter in IS3. The next possible start codon, a GUG at position 1174, would lead to a protein with 69 amino acids, but this ORF also lacks a promoter. Therefore, this ORF would probably be under control of outside promoters. If ORF III encodes a protein involved in transposition, the regulation of IS3 could vary greatly, depending on the presence of external promoters at the location of insertion.

The number of regular IS elements in E. coli that have been sequenced is now eleven (IS1, IS2, IS3, IS4, IS5, IS10, IS26, IS30, IS50, IS102, and IS903) (35-43). Dalrymple et al. compared IS30 to all of these elements (except for IS3) plus ISH1, ISH2, and ISH50 (41). They found no significant homologies between their nucleotide sequences or the amino acid sequences of their putative transposases. We also found no significant homologies between the sequences of IS3 and IS1, IS2, IS4, IS5, IS10, IS26, IS30, IS50, IS102, ISH1, ISH2 and ISH50. However, when we compared the putative transposase amino acid sequences of these elements, we found several regions of slight homology (Figure 4) between IS3 and IS2. The hydropathicity plots of these two putative transposases (Figure 5), when shown in an alignment similar to that used in Figure 4, also suggest some similarity. The regions showing marked hydrophobicity often correlate to regions located on a protein's interior (57). As negative controls, the hydropathicity plots of the similarly-sized putative transposases from IS5, IS26, and IS102 were compared to those of IS2 and IS3. Of the three negative controls, IS102 appeared to be most similar to IS2 and IS3, and so is included in Figure 5. We feel that the hydropathicity plots show much more similarity between IS2 and IS3 than between either of these and IS102. IS3 and IS2 are of similar size (1258 and 1327 bp, respectively), have IR's of similar size (39 and 41 bp) and are both capable
of activating genes. But, they generate different size direct repeats.

We have observed that IS3 shows some similarity to other IS elements in its gene organization, but not in its nucleotide sequence. The organization of ORF's I and II, in which they overlap by 4 bp and share a putative promoter, is also unique to IS3. The functional significance of these findings remain to be investigated.

ACKNOWLEDGMENTS

We would like to thank Anne B. Vojtek for isolating pABV5, Kathy Rockey and Eileen McConnell for typing the manuscript and figures, and the various authors for making their computer programs available. This work was supported by NSF (PCM8022650).

*To whom correspondence should be addressed

REFERENCES

1. Iida S., Meyer, J., and Arber, W. (1983) in Mobile Genetic Elements, Shapiro, J.A. Ed., pp. 159-221, Academic Press, New York.
2. Kleckner, N. (1981) Annu. Rev. Genet. 15, 341-404.
3. Sommer, H., Cullum, J., and Saedler, H. (1979) Molec. Gen. Genet. 177, 85-89.
4. Charlier, D., Piette, J., and Glansdorff, N. (1982) Nucleic Acids Res. 10, 5935-5948.
5. Hu, S., Ptashne, K., Cohen, S.N., and Davidson, N. (1975) J. Bacteriol. 123, 687-692.
6. Deonier, R.C., Hadley, R.G., and Hu, M. (1979) J. Bacteriol. 137, 1421-1424.
7. Malamy, M.H. (1966) Cold Spring Harbor Symp. Quant. Biol. 31, 189-201.
8. Malamy, M.H. (1970) in The Lactose Operon, Beckwith, J.R. and Zipser, D. Eds., pp. 359-373, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
9. Malamy, M.H., Fandt, M., and Szybalski, W. (1972) Molec. Gen. Genet. 119, 207-222.
10. Glansdorff, N., Charlier, D., and Zafarullah, M. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, 153-156.
11. Brunel, F., Heusterspreute, M., Merches, M., Thi, V.H., Pilacte, M.-F., and Davison, H. (1983) Plasmid 9, 201-214.
12. Jund, R. and Loison, G. (1982) Nature 296, 680-681.
13. Wals, A., Ratziun, B., and Carbon, J. (1978) Proc. Natl. Acad. Sci. USA 75, 6172-6176.
14. Saedler, H., Reif, H.J., Hu, S., and Davidson, N. (1974) Molec. Gen. Genet. 132, 265-289.
15. Jaurin, B. and Normark, S. (1983) Cell 32, 809-816.
16. Timmerman, K.P., Vojtek, A.B., and Tu, C.-P. (1983) Fed. Proc. 42, 2160 (Abstract).
17. Appleyard, R.K. (1954) Genetics 39, 440.
18. Tu, C.-P.D. and Cohen, S.N. (1980) Cell 19, 151-160.
19. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
20. Kupersztchok, Y.M. and Helinski, D.R. (1973) Biochem. Biophys. Res. Commun. 54, 1451-1459.
21. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
22. Wu, R., Jay, E., and Roychoudhury, R. (1976) Methods Cancer Res. 12, 87-176.
23. McDonell, M.W., Simon, M.N., and Studier, F.W. (1977) J. Mol. Biol. 110, 119-146.
24. Maxam, A.M. and Gilbert, W. (1980) Meth. Enzymol. 65, 499-560.
25. Yousaf, S.I., Carroll, A.R., and Clarke, B.E. (1984) Gene 27, 309-313.
26. Tu, C.-P.D. and Cohen S.N. (1980) Gene 10, 177-183.
27. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110.
28. Conrad, B. and Mount, D.W. (1982) Nucleic Acids Res. 10, 31-38.
29. Clayton, J., Friedland, P., Kedes, L., and Brutlag, D. Copyright 1981, Board of Trustees, Stanford University.
30. Harr, T., Haggstrom, M., and Gustafsson, P. (1983) Nucleic Acids Res. 11, 2943-2957.
31. Zweig, S.E. (1984) Nucleic Acids Res. 12, 767-776.
32. The Bionet™ Resource, IntelliGenetics, Palo Alto, CA.
33. Siebenlist, U., Simpson, R.B., and Gilbert, W. (1980) Cell 20, 269-281.
34. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342-1346.
35. Ohtsubo, H. and Ohtsubo, E. (1978) Proc. Natl. Acad. Sci. USA 75, 615-619.
36. Ghosal, D., Sommer, H., and Saedler, H. (1979) Nucleic Acids Res. 6, 1111-1121.
37. Klaer, R., Kuhn, S., Tillmann, E., Frits, H.-J., and Starlinger, P. (1981) Molec. Gen. Genet. 181, 169-175.
38. Engler, J.A. and vanBree, M.P. (1981) Gene 14, 155-163.
39. Halling, S.M., Simons, R.W., Way, J.C., Walsh, R.B., and Kleckner, N. (1982) Proc. Natl. Acad. Sci. USA 79, 2608-2612.
40. Mollet, B., Itda, S., Shepherd, J., and Arber, W. (1983) Nucleic Acids Res. 11, 6319-6330.
41. Dalrymple, B., Caspers, P., and Arber, W. (1984) EMBO J, 2145-2149.
42. Auerswald, E.-A., Ludwig, G., and Schaller, H. (1980) Cold Spring Harbor Symp. Quant. Biol. 45, 107-113.
43. Bernardi, A. and Bernardi, F. (1981) Nucleic Acids Res. 9, 2905-2911.
44. Simsek, M., DasSarma, S., RajBhandary, U.L., and Khorana, H.G. (1982) Proc. Natl. Acad. Sci. USA 79, 7268-7272.
45. DasSarma, S., RajBhandary, U.L., and Khorana, H.G. (1983) Proc. Natl. Acad. Sci. USA 80, 2201-2205.
46. Xu, W.-L. and Doolittle, W.F. (1983) Nucl. Acids Res. 11, 4195-4199.
47. Fischhoff, D.A., Vovis, G.F., and Zinder, N.D. (1980) J. Mol. Biol. 144, 247-265.
48. Kroger, M. and Hobom, G. (1982) Nature 297, 159-162.
49. Rak, B., Lusky, M., and Hable, M. (1982) Nature, 297, 124-128.
50. Bernardi, F. and Bernardi, A. (1984) Molec. Gen. Genet. 195, 452-458.
51. Trinks, K., Habermann, P., Beyreuther, K., Starlinger, P., and Ehring, T. (1981) Molec. Genet. 182, 183-188.
52. Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison III, C. A., Slocombe, P. M., and Smith, M. (1977) Nature 265, 687-695.
53. Riggs, D. and Artz, S. (1984) Molec. Gen. Genet. 196, 526-529.
54. Yamamoto, T., Tamura, T.-a., Yokota, T., and Takamo, T. (1982) Molec. Gen. Genet. 188, 356-359.
55. Oppenheim, D. S. and Yanofsky, C. (1980) Genetics 95, 785-795.
56. Baughman, G. and Nomura, M. (1983) Cell 34, 979-988.
57. Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132.