Vol. 97, No. 3, 1980 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
December 16, 1980 Pages 1116-1123

CONSTRUCTION OF RECOMBINANT cDNA VIA A RIBOSUBSTITUTED HAIRPIN

Gary V. Paddock* and Jim Gaubatz+

Department of Basic and Clinical Immunology and Microbiology,
Medical University of South Carolina, Charleston, South Carolina 29403

Received October 30,1980

SUMMARY

A new technique for recombinant cDNA construction is presented in which ribonucleotides are added to the 3' end of cDNA via ribosubstitution prior to synthesis of the second DNA strand. The hairpin floppy loop is then opened by alkaline hydrolysis. This method eliminates the requirement for S1 nuclease and thus shows promise as a means for preservation of mRNA 5'-end sequences in recombinant cDNAs and for eliminating a potential source of error in those sequences.

INTRODUCTION

One of the major areas of research in molecular biology today concerns gene organization and expression in eukaryotic cells. Much effort has been spent on studies of RNA transcription and its subsequent processing to mRNA. It is currently thought that the genome sequences surrounding the cap site contain the signal for initiation of mRNA transcription or, alternatively, that very rapid processing cleaves away the first few nucleotides followed by capping at the 5' end (1-5). In either case, the nucleotide sequences contained in the 5' untranslated regions of mRNA, especially those near the cap site, are of prime importance to proper gene regulation, as illustrated by the extensive conservation of sequences found in this region for alpha and beta globin and other mRNA species (1,2,6). In addition to transcription and processing of mRNA, these sequences undoubtedly play an important role in the translation of protein from mRNA. Indeed, the importance of the nucleotides contained in the 5' untranslated regions of mRNA is emphasized by the variety of methods designed to sequence them (6-12). Yet none of these methods permits sequencing of the 5' end of an impure mRNA obtained in low yield (which is the case for most mRNAs). Furthermore, none of the cloning techniques developed thus far (13-19) have been successful in preserving these important terminal sequences. In fact, the most popular of these techniques

*To whom correspondence should be addressed.
+Present address: Department of Biochemistry, MSB 2170, University of South Alabama, Mobile, Alabama 36688.

0006-291X/80/231116-08$01.00/0

is destined to destroy these sequences in part because it relies on the use of S1 nuclease (13). We have sought to preserve these important 5'-end signals by developing methodology which avoids the need for S1 nuclease (20,21). The floppy loop method described here employs a ribosubstitution step so that cleavage of the hairpin loop can be carried out by alkali or ribonuclease. A second technique also developed in this laboratory, involving the synthesis of a "third" DNA strand and no cleavage step (21), will be further described elsewhere.

MATERIALS AND METHODS

Construction of recombinant cDNAs via the floppy loop technique was carried out as follows. Complementary DNA was synthesized in 0.1 ml to 0.4 ml 50 mM Tris-HCl, pH 8.3; 10 mM $MgCl_2$; 20 mM 2-mercaptoethanol; 30 μg/ml actinomycin D; 20 μg/ml oligo$(dT)_{12-18}$, obtained from Collaborative Research; 40-60 μg/ml rabbit globin mRNA; 1 mM each dATP, dGTP, dTTP, and dCTP [dCTP containing 50-250 μCi/umol ^3H or ^{32}P]; and 150 U/ml AMV reverse transcriptase, supplied by J. Beard through the Biological Carcinogenesis Branch, National Cancer Institute, NIH. Incubations in early experiments were carried out at 37°C for 1 hr, but we have found incubation at 45°C for 15-20 min to be far superior. The reaction mixture was then extracted with an equal volume of water-saturated phenol, and residual phenol was removed by extraction with an equal volume of ether. The cDNA was precipitated with 40 μg/ml yeast tRNA, 0.1 vol 3 M NaAc (pH 5.5) and 2.25 vol ethanol. The cDNA was resuspended in 0.2 ml H_2O, brought to 0.3 M in NaOH, and incubated at 90°C for 30 min to hydrolyze the mRNA. The cDNA was then precipitated with 0.1 vol 3 M NaAc (pH 5.5), 0.1 vol 3 N HCl, 40 μg/ml yeast tRNA, and 2.5 vol ethanol. The cDNA was then chromatographed through a 0.6x15 cm column of Sephadex G-100 in 0.01 M triethylammonuim bicarbonate, pH 8.5. The columns were prepared in silanized glass or plastic pipettes. We have also used 1-ml tuberculin syringes, but in this case the oligo(dT) may not be removed. Yeast tRNA (40 μg) was chromatographed on a column prior to cDNA samples to fill nonspecific binding sites. The peak fractions were pooled and precipitated with 0.1 vol 3 M NaAc (pH 5.5), 40 μg/ml tRNA, and 2.25 vol EtOH.

The ribosubstitution step, modified from Whitcome et al. (22), was carried out in 0.1 ml 67 mM Tris-HCl (pH 7.4), 0.67 mM $MnCl_2$; 1.0 mM 2-mercaptoethanol; 330 μM each rATP, rGTP, rUTP, and rCTP; 100 U/ml DNA polymerase I (Klenow large fragment from Boehringer-Mannheim); and 10-40 μg/ml cDNA for 10 min at 37°C. Recently we have also added 5'-rAMP at 0.3 mM to prevent any possible degradation by the 3'-exonuclease activity contained in DNA polymerase I (23) during the slow ribosubstitution reaction. The cDNA reaction mixture was then extracted with water-saturated phenol and the residual phenol removed by ether extraction. Following the extraction, cDNA was chromatographed through Sephadex G-100, and the excluded material was precipitated with EtOH as above.

Second-strand synthesis was carried out as described by Higuchi et al. (13). The cDNA was resuspended in 0.12 ml 67 mM potassium phosphate, pH 7.3; 6.7 mM $MgCl_2$; 1 mM 2-mercaptoethanol; 33 μM each of dTTP, dCTP, dATP, and dGTP (dCTP containing 10 to 20 mCi/μmol ^3H or ^{32}P) and incubated with 40 U/ml DNA polymerase I (Klenow large fragment) for 30 min at 37°C. The double-stranded hairpin cDNA reaction mixture was extracted with water-saturated phenol and ether and precipitated with ethanol as above. The precipitate was resuspended in 0.1 ml H_2O, brought to 0.3 N in NaOH, and incubated at 90°C for 30 min to hydrolyze the ribonucleotide link. The DNA was then precipitated with 0.1 vol 3 M NaAc, pH 5.5, 0.1 vol 3 N HCl, 40 μg/ml yeast tRNA, and 2.25 vol ethanol. The DNA was then resuspended in 0.1 ml 0.1 M Tris-HCl, pH 8.0, and treated

with bacterial alkaline phosphatase (Worthington BAPF) as described by Shinagawa and Padmanabhan (24). The reaction mixture was supplemented with 0.1% SDS and incubated with 1.5-2.0 units of phosphatase which had been resuspended in 0.1 M Tris-HCl (pH 8), 10 μM $ZnSO_4$ for 1 hr at 37°C. [The removal of phosphate can be followed by including 1 μCi ^{32}P-dNTP and at various times spotting 1 μl for cellulose thin-layer chromatography (methanol:HCl:H_2O, 7:2:1). The reaction mixture can be frozen until the degree of reaction is determined by autoradiography of the chromatograms.] The reaction mixture was then extracted four times with equal volumes of water-saturated phenol and one time with ether and then chromatographed through Sephadex G-100, and the peak fractions were precipitated with ethanol as above.

The procedure for tailing the globin DNA with poly(dC) and Pst-cleaved pBR322 plasmid DNA with poly(dG) was modified from Chang et al. (25), using terminal transferase [prepared in collaboration with M.S. Coleman at the University of Kentucky by the method of Chang and Bollum (26)] in 140 mM cacodylic acid, 30 mM Tris-base, 0.1 mM dithiothreitol, 1 mM $CoCl_2$, final pH 7.6, so that approximately 30 C residues were added to the globin DNA or 15-20 G residues to the pBR322 DNA. Incorporation was followed by acid precipitation of ^{3}H-dNTP. The DNAs were then subjected to extraction with an equal volume of water-saturated phenol and removal of residual phenol with ether. The pBR322 DNA was concentrated by precipitation with ethanol. The DNAs were then purified of precursors by chromatography through an 0.6x15 cm A5m (Bio-Rad) column (previously treated with 40 μg yeast tRNA) in 0.1 M NaCl, 0.01 M Tris-HCl, pH 7.5. The peak fractions were then pooled and precipitated with ethanol as for the Sephadex G-100 columns above. More recently, we have used Sephadex G-200 columns for purification of the globin DNAs at this step in order to obtain better recovery of the DNA. Annealing of approximately equimolar amounts of globin and pBR322 DNA was carried out as described by Chang et al. (25) in 10 mM Tris-HCl, pH 7.5, 0.25 mM EDTA, 0.1 M NaCl.

Transformation of *E. coli* X1776 to tetracycline resistance was performed by the method of Norgard et al. (27). Clones were then screened for absence of ampicillin resistance and for presence of globin gene sequences using the colony hybridization technique (28) as modified by Sippel et al. (29).

Recombinant cDNAs were prepared for sequence analysis by cleavage with Bst N1 (N.E. Biolabs) and labeled with ^{32}P using T4 polynucleotide kinase (Boehringer-Mannheim) and [gamma-^{32}P]rATP (ICN). The labeled DNA was then cleaved with Pst I (N.E. Biolabs), and the fragments were separated by electrophoresis through a 6% polyacrylamide gel. The fragments containing the insert ends were easily determined by comparison with non-Pst I cleaved DNA loaded in an adjacent position.

Sequence analysis was performed by the method of Maxam and Gilbert (30) with the G+A reaction as described by Rogers et al. (31), except that all reactions at 90°C were performed under mineral oil in polypropylene tubes instead of in sealed glass capillaries.

RESULTS AND DISCUSSION

The floppy loop technique deviates from the S1 nuclease cDNA cloning method as follows (see Fig. 1): First, the 3' end of the cDNA acts as the primer for the addition of a short stretch of ribonucleotides by *E. coli* DNA polymerase I. Secondly, the ribonucleotide-terminated cDNA is used as a primer-template for synthesis of the second DNA strand by DNA polymerase, which is followed by treatment with alkali and phosphatase to open the hairpin loops and to regenerate a 3'-OH at this position, respectively. Obviously, treatment with alkali results in denaturation of double-stranded cDNA. Since

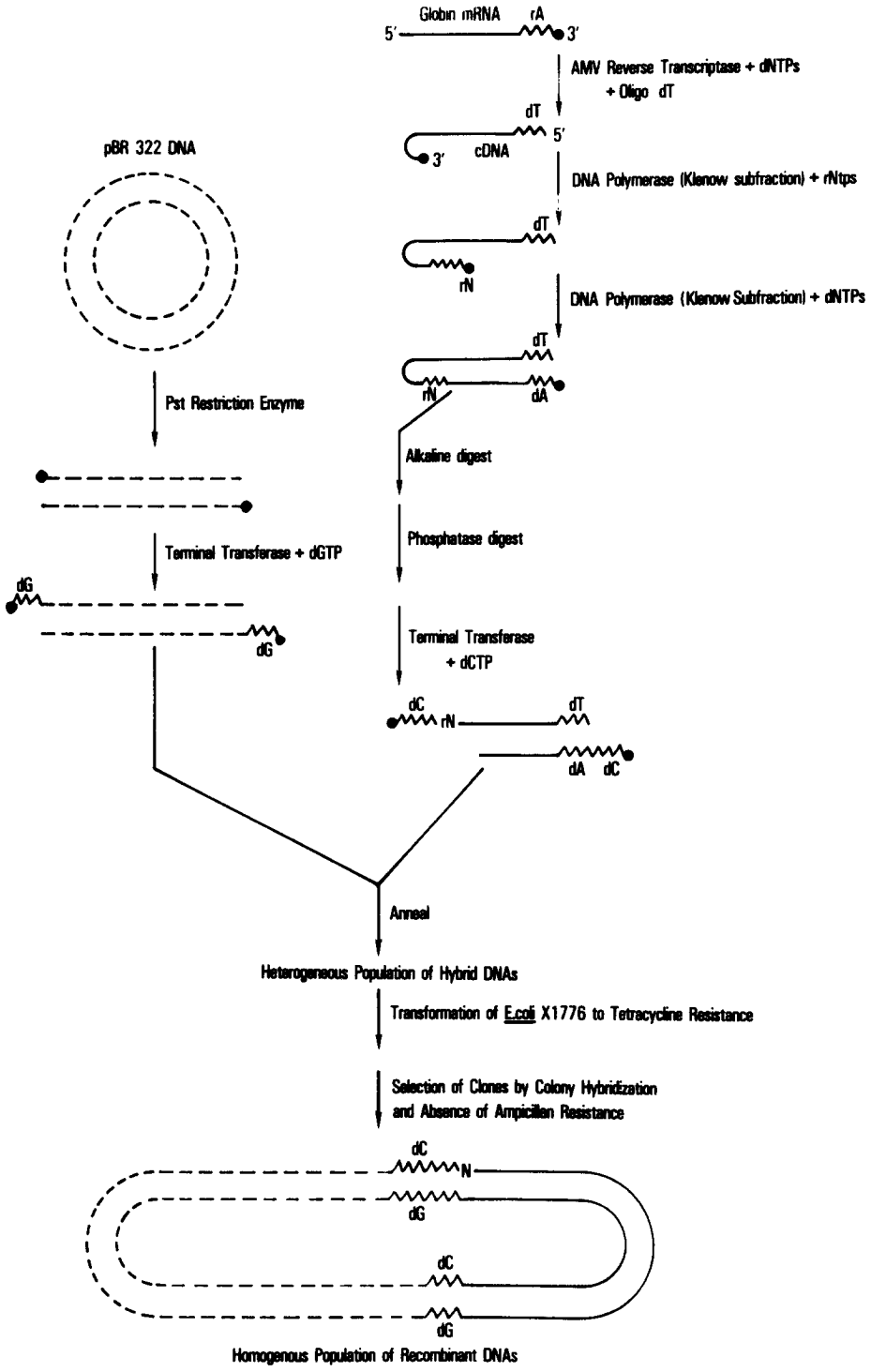

Figure 1. Schematic illustration of the ribosubstitution floppy loop recombinant cDNA technique.

the cDNA must be hybridized to the plasmid, however, reannealing of complementary sequences should take place without any additional steps. The resulting double-stranded DNA is then tailed with poly(dC) and annealed to the poly(dG)-tailed plasmid, pBR322, previously cleaved at the Pst site.

In preliminary experiments, we employed several options for incorporating the ribonucleotide desired for linking the first and second strands, because we were uncertain as to the best conditions for generating a reasonable length of ribonucleotides. In these experiments, ribonucleotides were added either by terminal transferase or by ribosubstitution with DNA polymerase. All but one of the ribonucleotides were removed prior to second-strand synthesis by treatment with alkali, and the 3'-OH end was regenerated by treatment with phosphatase. The terminal transferase method has the advantage of marking the 5' end with a known residue, but has the disadvantage that the nucleotide added may not base-pair with the first strand. It is known that the purpose of the 3'-exonuclease activity contained in the DNA polymerase is to remove mismatched nucleotides; thus the ribonucleotide linker could be lost. This may be prevented, however, by adding a 5'-nucleoside monophosphate to the reaction mixture during second-strand synthesis (23). On the other hand, the ribosubstitution methods have the advantage of further stabilizing the cDNA 3' hairpin but have the disadvantage of uncertainty with regard to the identity of the nucleotide at the cDNA 3' terminus, i.e., the nucleotide complementary to the 5'-mRNA terminal nucleotide, disregarding the inverted cap nucleotide, 7 meG. This is because it cannot be determined whether this nucleotide or the ribonucleotide added to it is the same (i.e, cytosine) as the poly(dC) tail added later. Of course, after plasmid replication in E. coli, the ribonucleotide will be converted to a deoxynucleotide. Thus, the following possibilities must always be considered when a sequence is eventually determined: (a) the real sequence is one nucleotide longer (i.e., both the 3' nucleotide and the ribonucleotide are cytosine); (b) the real sequence is as determined (i.e., the 3' nucleotide is not cytosine but the ribonucleotide is cytosine); or (c) the real sequence is one nucleotide shorter (i.e., neither the 3' nucleotide nor the ribonucleotide is cytosine).

Of the various options tried, priming of the second-strand synthesis directly after synthesis of the linker was found to be much simpler, and the elimination of the extra steps required for removal of all but a single ribonucleotide resulted in smaller losses from handling. However, all of the various options for ribosubstitution were successful in yielding recombinant globin cDNA/pBR322 DNA molecules as determined by colony hybridization and antibiotic resistance. The efficiency of transformation obtained with these

cDNAs was comparable to that obtained with the S1 method, which we performed concurrently. Of 48 clones analyzed, 30 had lost ampicillin resistance, indicating insertion of a rabbit globin cDNA segment into the original Pst site. Colony hybridization indicated that all 30 had globin gene inserts. Analysis of 27 of the recombinant plasmids revealed that we had recreated Pst sites at both ends of the insert in 21 recombinants (80%). The remainder had at least one of the possible Pst sites recreated.

The sizes for the inserted sequences in the 27 rabbit globin recombinants tested ranged up to 585 nucleotides, with most being 200 to 400 nucleotides including the poly(dC)-poly(dG) tails. A similar size range was observed for a duck globin floppy loop preparation, except that at least one recombinant contained an insert of approximately 800 nucleotides. Thus, most of our recombinants did not have inserted sequences representing the full length of the globin mRNA. It must be kept in mind, however, that our first goal was to develop the technique. In order to try a variety of approaches and options, we conserved our cDNA by not subjecting it to a preparative sizing step in an acrylamide gel. Analysis of this particular cDNA preparation revealed that a good portion was less than full length. On the other hand, restriction enzyme analysis suggested, on the basis of the size of the inserts and comparison with cleavage sites expected for the rabbit globin genes (16,32), that a few might contain the important 5'-end mRNA sequences. Thus far, we have determined that the methodology does not introduce erroneous sequences, in that the first two rabbit globin recombinants we have sequenced (Fig. 2) matches perfectly with the known sequence for rabbit beta globin (16). This is an important consideration because the S1 nuclease technique has a known potential for introducing errors into the nucleotide sequence (33) through a mechanism that is avoided by the floppy loop technique.

We believe that, while not yet perfected, the floppy loop cDNA cloning method offers the promise of obtaining the 5' end sequences of mRNA and thus full length clones as well. We have simplified the technique to the point that second-strand synthesis directly after the ribosubstitution step would seem to be the method of choice. This also allows us alternatives to opening the loop with alkali. We have recently obtained clones of duck globin cDNA recombinants where the loop was opened using RNase H (R. Frankis and G. Paddock, manuscript in preparation). This eliminates the need to create a 3'-OH end with phosphatase, because a 3'-OH end results from cleavage with RNase H. In addition to simplicity, opening the loop by treatment with ribonuclease instead of alkali will probably prove to be more suitable for cloning of recombinant cDNAs representing minor mRNA species, because this

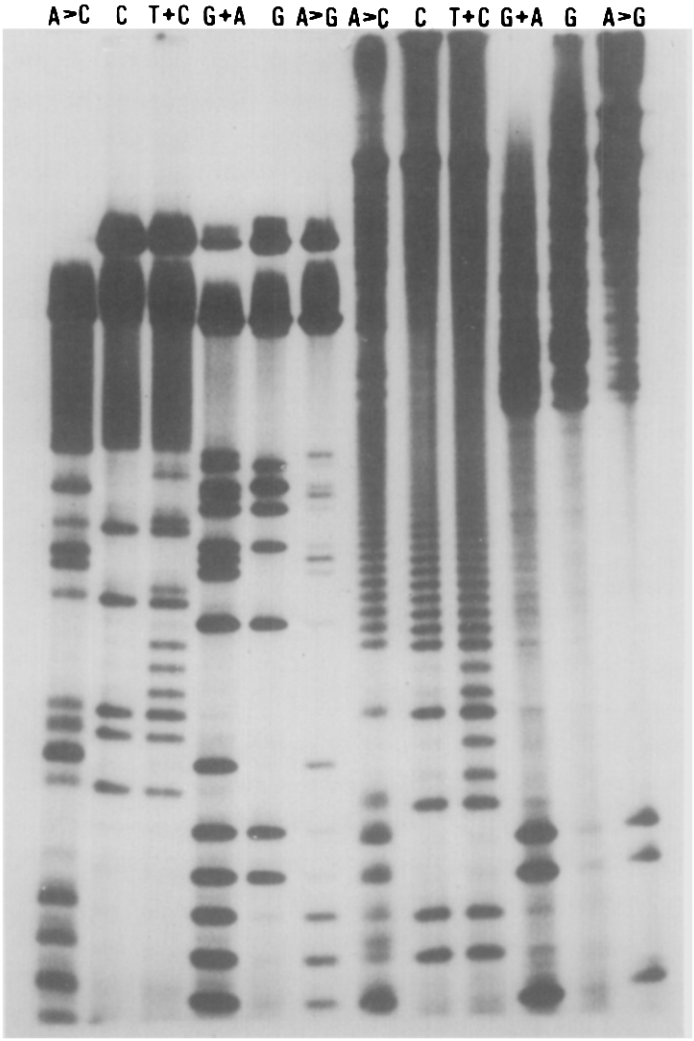

Figure 2. Sequence analysis of rabbit beta globin gene sequences cloned by the floppy loop method. The right-hand set of six lanes represent the complement to rabbit globin mRNA beginning at nucleotide 62. The left-hand set of six lanes represent the mRNA sequence ending at nucleotide 269 (16).

eliminates the risk of denaturation and thus the need for subsequent reannealing. This denaturation, however, has not seemed to be a problem with the abundant globin mRNA sequences.

ACKNOWLEDGMENTS

This is publication no. 375 from the Department of Basic and Clinical Immunology and Microbiology, Medical University of South Carolina.
We acknowledge the technical support of Louise Just and Linda Paddock and the

editorial assistance of Charles L. Smith. This research was supported in part by grants to Gary Paddock from NIH (GM-24783) and from State of South Carolina Institutional Support Funds.

REFERENCES

1. Konkel, D.A., Tilghman, S.M., and Leder, P. (1978). Cell 15, 1125-1132.
2. Konkel, D.A., Maizel, J.V., Jr. and Leder, P. (1979). Cell 18, 865-873.
3. Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979). Nature 278, 428-434.
4. Nishioka, Y. and Leder, P. (1979). Cell 18, 875-882.
5. Kinniburgh, A.J. and Ross, J. (1979). Cell 17, 915-921.
6. Lockard, R.E. and RajBhandary, U.L. (1976). Cell 9, 747-760.
7. Baralle, F.E. (1977). Cell 10, 549-558.
8. Baralle, F.E. (1977). Nature 267, 279-281.
9. Baralle, F.E. (1977). Cell 12, 1085-1095.
10. Legon, S. (1976). J. Mol. Biol. 106, 37-53.
11. Chang, J.C., Temple, G.F., Poon, R., Neumann, K.H. and Kan, Y.W. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 5145-5149.
12. Chang, J.C., Poon, R., Neumann, K.H. and Kan, Y.W. (1978). Nucl. Acids Res. 5, 3515-3522.
13. Higuchi, R., Paddock, G.V., Wall, R. and Salser, W. (1976). Proc. Natl. Acad. Sci. U.S.A. 73, 3146-3150.
14. Maniatis, T., Kee, S.G., Efstratiadis, A. and Kafatos, F.C. (1976). Cell 8, 163-182.
15. Rougeon, F., Kourilski, P. and Mach, B. (1975). Nucl. Acids Res. 2, 2365-2378.
16. Efstratiadis, A., Kafatos, F.C. and Maniatis, T. (1977). Cell 10, 571-585.
17. Rabbits, T.H. (1976). Nature 260, 221-225.
18. Rougeon, F. and Mach, B. (1976). Proc. Natl. Acad. Sci. U.S.A. 73, 3418-3422.
19. Wood, K.O. and Lee, J.C. (1976). Nucl. Acids Res. 3, 1961-1971.
20. Frankis, R., Gaubatz, J., Lin, F-K. and Paddock, G.V. The Twelfth Miami Winter Symposium (ed. Whelan, W.J., and Schultz, J., Academic Press, New York), vol. 17, in press (1980).
21. Gaubatz, J. and Paddock, G.V. (1980). Fed. Proc. 39, 1782.
22. Whitcome, P., Fry, K. and Salser, W. (1974). Methods Enzymol. 29, 295-321.
23. Byrnes, J.J., Downey, K.M., Que, B.G., Lee, M.Y.W., Black, V.L. and So, A.G. (1977). Biochemistry 16, 3740-3746.
24. Shinagawa, M. and Padmanabhan, R. (1979). Anal. Biochem. 95, 458-464.
25. Chang, A.C.Y., Nunberg, J., Kaufman, R.J., Erlich, H.A., Schimke, R.T. and Cohen, S.N. (1978). Nature 275, 617-624.
26. Chang, L.M.S. and Bollum, F.J. (1971). J. Biol. Chem. 246, 909-916.
27. Norgard, M.V., Keem, K. and Monahan, J.J. (1978). Gene 3, 279-292.
28. Grunstein, M. and Hogness, D.S. (1975). Proc. Natl. Acad. Sci. U.S.A. 72, 3691-3695.
29. Sippel, A.E., Land, H., Lindenmaier, W., Nguyen-Huu, M.C., Wurtz, T., Timmis, K.N., Giesecke, K. and Schutz, G. (1978). Nucl. Acids Res. 5, 3275-3294.
30. Maxam, A.M. and Gilbert, W. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 560-564.
31. Rogers, J., Clarke, P. and Salser, W. (1979). Nucl. Acids Res. 6, 3305-3321.
32. Heindell, H.C., Lin, A., Paddock, G.V., Studnicka, G.M., and Salser, W.A. (1978). Cell 15, 43-54.
33. Richards, R.I., Shine, J., Ulbrich, A., Wells, J.R.E., and Goodman, H.M. (1979). Nucl. Acids Res. 7, 1137-1146.