Volume 5 Number 11 November 1978 Nucleic Acids Research

Recognition sequence of restriction endonuclease KpnI from Klebsiella pneumoniae

Joanne Tomassini, Ranajit Roychoudhury, Ray Wu* and Richard J.Roberts†

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, and
†Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, NY 11724, USA

Received 24 August 1978

ABSTRACT

We have determined the recognition sequence of the restriction endonuclease *KpnI*, previously isolated from *Klebsiella pneumoniae*[1]. The enzyme cleaves the twofold rotationally symmetric sequence $\begin{matrix} 5'\text{-G-G-T-A-C}\overset{\downarrow}{-}\text{C-3'} \\ 3'\text{-C}\underset{\uparrow}{-}\text{C-A-T-G-G-5'} \end{matrix}$ at the positions indicated by the arrows, producing 3' protruding cohesive ends, four nucleotides in length. The specific cleavage site was unambiguously deduced using both 3' and 5' end analyses of *KpnI* generated restriction fragments of simian-virus 40 (SV40) DNA (1 site), adenovirus-2 (Ad-2) DNA (8 sites), and a plasmid (pCRI) DNA (2 sites).

INTRODUCTION

The discovery of site-specific endonucleases nearly a decade ago has led to significant advances in DNA sequence analysis, the physical mapping of genes, recombinant DNA research, and gene isolation, providing new approaches to the study of gene structure and function. The usefulness of these endonucleases is attributed to the characteristic recognition sequences which differ for each enzyme in both nucleotide sequence and position of cleavage within the recognition site[2]. Therefore, knowledge of the recognition sequence of a restriction endonuclease is necessary.

A direct and rapid method of sequence analysis is desirable for determining the short recognition sequence of a restriction endonuclease. Efficient end-labeling techniques are essential to such analysis. Techniques used in the sequence analysis of cleavage sites which possess 5' terminal extensions or flush ends are well developed. DNA with 5' protruding ends can be labeled either by polynucleotide kinase or by DNA polymerase in repair synthesis of the complementary strand[3]. Flush-end cleavage sites can also be end labeled using polynucleotide kinase. On the other hand, DNA with 3' protruding ends can be readily sequenced only after the addition of a [^{32}P]rNMP to the 3'OH end using terminal deoxynucleotidyl transferase. The restriction endonuclease *KpnI* was found to produce 3' protruding ends. The recognition sequence of this enzyme has been determined using two-dimensional mobility-shift analysis on 3' terminally labeled KpnI restriction fragments, in conjunction with 3' and 5' terminal mononucleotide and dinucleotide identification. The methods employed here can be generally extended to elucidate the recognition sequences of other restriction endonucleases bearing 3' protruding ends.

EXPERIMENTAL PROCEDURE

Materials

DNA -- Form I superhelical SV40 DNA[3], pCRI DNA[4], and linear Ad-2 DNA[5] were purified according to published procedures.

Enzymes -- The restriction endonucleases KpnI (specific activity 50,000 units/mg) and AluI (specific activity 20,000 units/mg) were purchased from New England Biolabs. Calf thymus terminal deoxynucleotidyl transferase (specific activity 5,500 units/mg) was obtained from P.L. Biochemical Incorporated. Polynucleotide kinase (specific activity 20,000 units/mg) was obtained from Biogenics Research Corporation. AMV reverse transcriptase (specific activity 39,000 units/mg) was supplied by the Office of Program Resources and Logistics, Viral Cancer Program, Viral Oncology, Division of Cancer Cause and Prevention, National Cancer Institute, Bethesda, MD 20014. Micrococcal nuclease and spleen phosphodiesterase from Worthington Biochemical Corporation were further purified as previously described[3]. Venom phosphodiesterase, also from Worthington Biochemical Corporation, was further purified[6] before use. Pancreatic DNase (specific activity 3,750 units/mg) and bacterial alkaline phosphatase (BAPF grade, specific activity 36 units/mg) were purchased from Worthington Biochemical Corporation. Exonuclease I was purified according to Lehman and Nussbaum[7].

Nucleotides -- [α-^{32}P]rCTP and [α-^{32}P]rATP (specific activities 200 Ci/mmole) were purchased from ICN Parmaceuticals. [α-^{32}P]dCTP (350 Ci/mmole) and [γ-^{32}P]ATP (1,000-3,000 Ci/mmole) were from Amersham Corporation. The four dinucleoside monophosphates used were obtained from Collaborative Research Incorporated.

Chromatography -- Dry cellulose acetate strips were obtained from Schleicher and Schuell Incorporated, New Hampshire. The DEAE-cellulose thin layer plates (20 x 20 cm) and Homo-mix VI were prepared as described previously[8].

Methods

Repair synthesis using AMV reverse transcriptase -- The reaction was carried out as described[9], and an aliquot of the reaction mixture was filtered on a Whatman GF/A glass filter to measure the incorporation[3].

3' and 5' end labeling -- KpnI restricted fragments of SV40 and pCRI DNA were labeled at the 5' ends using polynucleotide kinase and [γ-^{32}P]rATP. The addition of a single [^{32}P] ribomononucleotide to the 3'OH ends using terminal deoxynucleotidyl transferase and [α-^{32}P]rCTP or [α-^{32}P]rATP was performed as described[10,11].

Identification of the 5' terminal mono- and dinucleotides -- Ad-2 DNA (23 μg) was incubated in a reaction mix (100 μl) containing 6 mM Tris-HCl (pH 7.9), 6 mM $MgCl_2$, 6 mM $SHCH_2CH_2OH$, 25 units KpnI for 1 h at 37°C. Alkaline phosphatase (10 μg) was added and the volume increased to 200 μl by adjusting to 50 mM Tris-HCl (pH 8.9), 10 mM $MgCl_2$ and incubation continued for an additional hour at 37°C. Following extraction with phenol (4 x 200 μl), the DNA fragments were precipitated with ethanol (2 volumes) and recovered by centrifugation (100,000 g for 20 min). The DNA was phosphorylated in a reaction mix (100 μl) containing 50 mM Tris-HCl (pH 9.5), 10 mM $MgCl_2$, 5 mM dithiothreitol, 5% glycerol, 10 μM [γ-^{32}P]ATP (specific activity 1,000 Ci/mmole), 10 units polynucleotide kinase, and incubated for 1 h at 37°C. Unreacted [γ-^{32}P]ATP was removed by passage through a Sephadex G-50 column, run in 0.1 M Tris-HCl (pH 7.9), 0.001 M Na_2EDTA. Labeled KpnI fragments were recovered from the void volume by precipitation with ethanol and centrifugation (as above).

The DNA was then incubated in a reaction mix (10 μl) containing 0.1 M sodium acetate (pH 5.0), 0.005 M $MgCl_2$, 5 μg pancreatic DNase for 30 min at 37°C. An aliquot of this mixture (1 μl) was removed and incubated in a fresh reaction mix (10 μl) containing 0.1 M Tris-HCl (pH 8.9), 0.05 M $MgCl_2$, 2 μg venom phosphodiesterase for 30 min at 37°C. A second aliquot (1 μl) was incubated in a reaction mix (10 μl) containing 66 mM glycine-NaOH (pH 9.6), 6.6 mM $MgCl_2$, 3.3 mM dithiothreitol, 5 units exonuclease I for 30 min at 37°C. The products from both reactions were analyzed by electrophoresis on Whatman 540 paper[12] at pH 3.5. The four standard dinucleotides were prepared by phosporylating the corresponding dinucleoside monophosphates with polynucleotide kinase and [γ-^{32}P]ATP, as described above, and the products purified by electrophoresis on DEAE-cellulose paper in 7% (v/v) formic acid.

Identification of the 3' terminal mononucleotide (nearest-neighbor analysis) -- DNA fragments of KpnI digested pCRI and SV40 DNA were labeled at the 3' terminal ends using the terminal transferase. The DNA was completely digested to 3' mononucleotides with micrococcal nuclease and spleen phosphodiesterase, then chromatographed on orange ribbon filter paper with 50 nmoles each of dAp, dGp, dTp and dCp as carrier and internal markers[3]. The separated mononucleotides were visualized with UV light, cut from the paper and quantitated by liquid scintillation counting[13].

In experiment I, pCRI DNA labeled with [α-^{32}P]rCMP at the 3' ends was used and analyzed by one-dimensional paper chromatography[3]. In experiment II, SV40 DNA labeled with [α-^{32}P]rAMP was used and analyzed by two-dimensional paper chromatography[3].

Two-dimensional mapping of 3' terminally labeled SV40 DNA and pCRI DNA -- Superhelical SV40 and pCRI DNA (50 μg) were incubated in 100 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 70 mM NaCl, 1 mM dithiothreitol, and 200 units of *KpnI* for 2-4 h at 37°C in a final volume of 100 μl. Following two phenol extractions, the DNA was EtOH precipitated and labeled at the two 3'OH ends with terminal transferase and [α-^{32}P]rATP (for SV40 DNA) or [α-^{32}P] rCTP (for pCRI DNA) as described[3]. After labeling, the reaction mixture was adjusted to 500 mM Tris-HCl (pH 8.0) and the DNA was precipitated with EtOH twice. The precipitated DNA was incubated in 200 μl of 6.6 mM Tris-HCl (pH 7.8), 6.6 mM $MgCl_2$, 1 mM dithiothreitol, 0.5 mM 5'-dAMP and 40 units of *AluI* for 8 h at 37°C. The DNA was treated with alkali and alkaline phosphatase[10], phenol extracted four times and passed over a Sephadex G-50 column. The two labeled SV40 DNA fragments (112 and 26 nucleotides) were partially separated on the column and the DNA in the peak fractions (eluted with 0.01 M triethylamine bicarbonate, pH 8.5) could be sequenced directly. SV40 DNA with uniquely labeled ends were also isolated by polyacrylamide slab gel electrophoresis[3]. DNA fragments isolated by either method were digested with pancreatic DNase for 3, 15 and 60 min. One-third of the combined digest was treated with venom phosphodiesterase[3] for 5 min, and heated to 90° for 10 min to inactivate the enzyme. This step yields labeled mononucleotide. Both digests were combined and an aliquot (20,000 cpm) was electrophoresed on cellulose acetate (at pH 3.5) for the first dimension, followed by homochromatography on DEAE-cellulose (using Homo-mix VI[8]) for the second dimension[14]. The sequence was deduced by calculating the mobility shifts[15] of the homologous series of oligonucleotides obtained from each unique 3' labeled end.

Unfractionated 3'-labeled pCRI DNA fragments were treated with alkali and alkaline phosphatase[10], passed through a Sephadex G-50 column and then EtOH precipitated. The sequence of labeled DNA fragments were analyzed as described above.

RESULTS

Before a detailed analysis of the KpnI recognition sequence was begun, preliminary information on the type of cleavage was obtained. Inefficient labeling of the single SV40 site at its 5' ends with polynucleotide kinase[3] suggested the break to be either flush ended or 3' protruding. The failure of reverse transcriptase or DNA polymerase I to incorporate radioactive deoxynucleotides[9] was also consistent with a flush ended or 3' protruding sequence at the KpnI cleavage site. The addition of a [^{32}P]rCMP or [^{32}P] rAMP to the 3'OH ends[11] using terminal transferase proceeded efficiently, indicating the presence of 3' protruding ends.

Sequence analysis of the 3'OH ends

Sequence analysis of the 3'OH ends was carried out as follows. DNA fragments with unique labeled ends were isolated by digestion of terminally labeled DNA with a second restriction endonuclease, followed by Sephadex G-50 column chromatography (unpublished results) or polyacrylamide gel electrophoresis[3]. Each isolated single-end labeled DNA fragment was digested with pancreatic DNase, and one-third of the partial digest was treated with venom phosphodiesterase to liberate labeled mononucleotides. The mixture of digestion products was fractionated by two-dimensional electrophoresis-homochromatography as shown in Figure 1. The sequence was determined by calculating the quantitative mobility shifts[15] between adjacent oligonucleotides (Table 1). For example, in Figure 1a, by visual inspection it is apparent that the shift from spot 2 to 3 could be due to the addition of either an A or a G. However, by mobility-shift calculation, the addition of an A gives a S^{calc} value of 0.12 while the addition of a G gives a value of 0.36. Since the observed shift was 0.16, an A must have been added to the nucleotide in spot 2 to give nucleotide 3. Using similar calculations, the sequence at the KpnI cleavage site of the SV40 DNA fragment (shown in Figure 1a) was deduced to be 5'-T-A-G-G-T-A-C-OH. A parallel analysis using unfractionated pCRI DNA fragments as substrate gave the sequence, 5'-(T or C)-G-G-T-A-C-OH (Figure 1b). Since the sixth nucleotide from the 3' end is no longer unique, the KpnI endonuclease recognition sequence from the 3' end is deduced to be 5'-G-G-T-A-C-OH.

Determination of the cleavage position in the recognition sequence

The exact cleavage position of KpnI endonuclease was determined by nearest-neighbor analysis of KpnI-digested 3' end-labeled SV40 and pCRI DNA[3], and by analysis of mono- and dinucleotides present in 5' labeled Ad-2 DNA. The 3' terminal mononucleotide was identified as Cp (Table 2) after

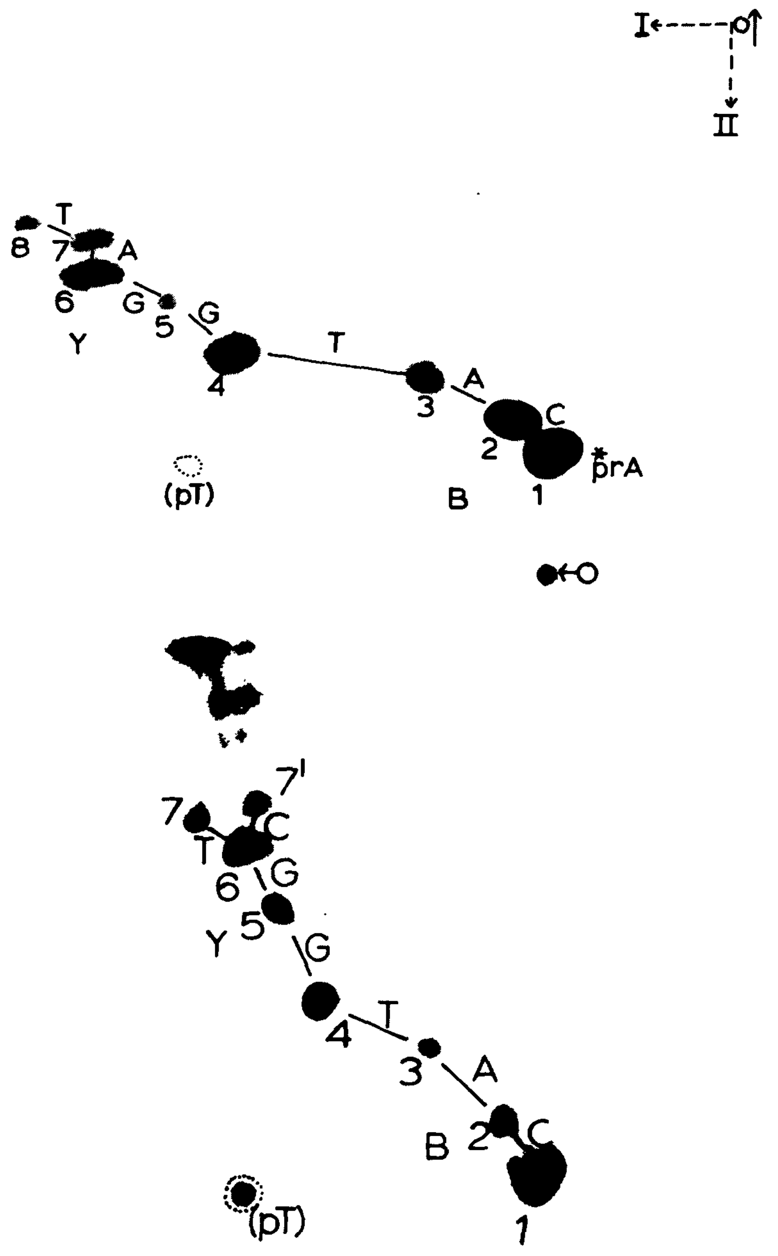

Fig. I Two-dimensional maps of the partial pancreatic DNase and venom phosphodiesterase digestion products of 3' terminally labeled SV40 DNA (a) and pCRI DNA (b).

Table 1. Experimental and calculated mobility shifts used in determining the KpnI recognition sequence.[†]

Figure	Oligonucleotide (N)	Distance from origin (in mm)	‡U_T^{obs}	U_T^{calc}	S^{obs}	S^{calc}	d Value (in mm)	Sequence deduced
Ia	1	41	0.32	0.40				$\overset{*}{p}$rA
					0.09	0.02(+C)	7.0(py)	
(SV40 DNA)	2	52	0.41	0.42				pC$\overset{*}{p}$rA
					0.16	0.12(+A)	9.0(pu)	
	3	73	0.57	0.54				pApC$\overset{*}{p}$rA
					0.36	0.39(+T)	3.0(py)	
	4	118	0.93	0.93				pTpApC$\overset{*}{p}$rA
					0.11	0.16(+G)	9.0(pu)	
	5	132	1.04	1.09				pGpTpApC$\overset{*}{p}$rA
					0.14	0.15(+G)	9.0(pu)	
	6	150	1.18	1.24				pGpGpTpApC$\overset{*}{p}$rA
					0.00	-0.02(+A)	9.0(pu)	
	7	150	1.18	1.22				pApGpGpTpApC$\overset{*}{p}$rA
					0.12	0.17(+T)	2.0(py)	
	8	166	1.30	1.39				pTpApGpGpTpApC$\overset{*}{p}$rA
	(pdT)	(127)	(1.0)	(1.0)				
Ib	1	8	0.09	0.07				$\overset{*}{p}$rC
					0.10	0.14(+C)	14.0(py)	
(pCRI DNA)	2	16	0.19	0.21				pC$\overset{*}{p}$rC
					0.21	0.18(+A)	21.0(pu)	
	3	34	0.40	0.39				pApC$\overset{*}{p}$rC
					0.35	0.40(+T)	10.0(py)	
	4	64	0.75	0.79				pTpApC$\overset{*}{p}$rC
					0.16	0.20(+G)	22.0(pu)	
	5	77	0.91	0.99				pGpTpApC$\overset{*}{p}$rC
					0.13	0.15(+G)	16.0(pu)	
	6	88	1.04	1.14				pGpGpTpApC$\overset{*}{p}$rC
					0.14	0.18(+T)	7.0(py)	
	7	102	1.20	1.32				pTpGpGpTpApC$\overset{*}{p}$rC
					-0.05	-0.08(+C)	11.0(py)	
	7'	85	1.00	1.06				pCpGpGpTpApC$\overset{*}{p}$rC
	(pdT)	(85)	(1.0)	(1.0)				

[†] Experimental data presented is in reference to the figures Ia and Ib, from which the KpnI recognition sequence was deduced.

[‡] Terms used above are briefly described as follows (for details, refer to reference 15):

$$U_T^{obs} = \frac{\text{distance from origin to oligonucleotide (N) in the first dimension}}{\text{distance from origin to pdT in the first dimension}}$$

U_T^{calc} = calculated electrophoretic mobilities (first dimension) of oligonucleotides relative to pdT.

S_{obs} = $(U_{T,\, n+1})^{obs} - (U_{T,\, n})^{obs}$ = observed mobility shift between oligomer n and n+1.

S^{calc} = $(U_{T,\, n+1})^{calc} - (U_{T,\, n})^{calc}$ = calculated mobility shift between oligomer n and n+1. The S value is the calculated mobility shift due to the addition of the mononucleotide shown in the parenthesis.

d value = difference of mobility between oligonucleotides in the second dimension (homochromatography). Small d value indicates a pyrimidine (py) and large d value indicates a purine (pu) addition.

$\overset{*}{p}$ = ^{32}P

complete digestion of 3' labeled SV40 DNA with micrococcal nuclease and spleen phosphodiesterase followed by nearest-neighbor analysis. The 5' terminal mononucleotide was identified as pC (Table 3) after complete digestion of 5' labeled Ad-2 DNA fragments with pancreatic DNase and venom phosphodiesterase. Direct comparison with the four standard dinucleotides, pC-N, showed the 5'-dinucleotide (resistant to exonuclease I action[7]) to

Table 2. Nearest-neighbor analysis of *KpnI* digested 3'-end labeled pCRI and SV-40 DNA.

Digestion Products	% of total[δ]	
	Expt. I (pCI DNA)	Expt. II (SV40 DNA)
Ap	3.5	6.0
Gp	6.6	6.4
Tp	10.5	23.0
Cp	79.4	64.6

[δ] The ^{32}P counts of the four mononucleotide digestion products are expressed as a percentage of the total mononucleotides detected. The total counts recovered for the four products are 20,000 cpm in Experiment I, and 3,400 cpm in Experiment II. Whatman #1 paper can be used in place of orange ribbon paper[3,13] (Schleicher and Schuell, Inc., Keene, N.H.).

Table 3. Identification of the 5' mono- and dinucleotides present after cleavage with *KpnI*.

Mononucleotide	Dinucleotide	% of total
pC		92.4
pA		2.4
pG		2.0
pT		3.3
	pC-C	25.6
	pC-A	33.1
	pC-G	17.1
	pC-T	24.2

be 5' pCpN where N is dA, dG, dT or dC (Table 3). Thus, 5' pC is the only unique nucleotide at the 5' terminus of the *Kpn*I endonuclease cleavage site. When SV40 DNA was used in a parallel experiment, a 5' terminal pC was also found (data not shown). In the case of a 3' protruding recognition site, the 5' end analysis is complicated by the inefficiency of labeling the DNA at the 5' end. We found that the efficiency of 5' end labeling was improved when the restriction fragments were heat denatured or the strands were separated prior to the kinase reaction.

From these analyses, we conclude that *Kpn*I recognizes the hexanucleotide depicted in Structure I. Cleavage at the sites indicated by the arrows produces a 3'-protruding tetranucleotide as shown in Structure II.

$$\begin{array}{lcl} 5'\text{-G-G-T-A-C}\overset{\downarrow}{-}\text{C-3}' & & 5'\text{-G-G-T-A-C-OH }3' \\ 3'\text{-C}\underset{\uparrow}{-}\text{C-A-T-G-G-5}' & \longrightarrow & 3'\text{-Cp}\qquad\quad 5' \\ \underline{\text{Structure I}} & & \underline{\text{Structure II}} \end{array}$$

DISCUSSION

*Kpn*I, like the restriction endonuclease *Hae*II[16] and *Pst*I[17], is of particular interest as it generates fragments bearing 3' cohesive ends. To date, relatively few enzyme recognition sequences containing 3' terminal tetranucleotide extensions have been determined due to difficulties encountered in labeling the 3' ends. The advantage of labeling the protruding 3' ends with terminal transferase for sequence analysis of the cleavage site has been clearly demonstrated here. The methods used in determining the *Kpn*I restriction site sequence and the cleavage position within the site by 3' end labeling of DNA with terminal transferase, quantitative mobility-shift analysis, and the 3' end and 5' end analyses, are especially useful for studying short terminal sequences such as the restriction enzyme recognition sequences.

The *Kpn*I recognition sequence is deduced from one sensitive site in SV40 DNA and two sites in pCRI DNA. It is possible that this hexanucleotide recognition sequence may show some variation if more sites are analyzed. So far, variation has not been found since the structural gene for rat growth hormone also include a *Kpn*I site with the same sequence G-G-T-A-C-C, even though Seeburg et al[18] placed the cleavage site between the two G's. Furthermore, a *Kpn*I site with the same sequence is found in BK virus DNA (R. Yang and R. Wu, unpublished observation).

ACKNOWLEDGEMENTS

This work was supported by a research grant, 77-20313, from the National Science Foundation and in part by funds provided by the Cornell University Agricultural Experiment Station.

[*] Individual to whom correspondence should be sent.

REFERENCES

1 Smith, D.L., Blattner, F.R. and Davies, J. (1976) Nuc. Acids Res. 3, 343-353.

2 Roberts, R.J. (1976) CRC Crit. Rev. Biochem. 4, 123-164.

3 Wu, R., Jay, E. and Roychoudhury, R. (1976) Methods in Cancer Res. 12, 87-176.

4 Brutlag, D., Fry, K., Nelson, T., and Hung, P. (1977) Cell 10, 509-519.

5 Ginsberg, H.S., Bello, H.J. and Levine, A.J. (1967) in The Molecular Biology of Viruses, Colter, J.S. and Paranchyck, W. Eds. pp. 547-572 Academic Press, New York.

6 Sulkowski, E. and Laskowski, M.Sr. (1971) Biochim. Biophys. Acta 240, 463-468.

7 Lehman, I.R. and Nussbaum, A.L. (1964) J. Biol. Chem. 239, 2628-2636.

8 Jay, E., Bambara, R., Padmanabhan, R. and Wu. R. (1974) Nuc. Acids Res. 1, 331-353.

9 Bahl, C.P., Wu, R., Stawinski, J. and Narang, S.A. (1977) Proc. Nat. Acad. Sci. U.S.A. 74, 966-970.

10 Kössel, H. and Roychoudhury, R. (1971) Eur. J. Biochem. 22, 271-276.

11 Roychoudhury, R., Jay, E. and Wu, R. (1976) Nuc. Acids Res. 3, 863-878.

12 Barrell, B.G. (1971) in Procedures in Nucleic Acid Research, Cantoni, G.L., and Davies, D.R., Eds., Vol. 2, pp 751-779, Harper and Row, New York.

13 Wu, R. (1970) J. Mol. Biol. 51, 331-339.

14 Brownlee, G.G. and Sanger, F. (1969) Eur. J. Biochem. 11, 395-399.

15 Tu, C.D., Jay, E., Bahl, C.P. and Wu, R. (1976) Anal. Biochem. 74, 73-93.

16 Tu, C.D., Roychoudhury, R. and Wu, R. (1976) Biochem. Biophys. Res. Commun. 72, 355-362.

17 Brown, N.L. and Smith, M. (1976) FEBS Letters 65, 284-287.

18 Seeburg, P.H., Shine, J., Martial, J.A., Baxter, J.D. and Goodman, H.M. (1977) Nature 270, 486-494.