Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 3659–3663, June 1986
Biochemistry

Homology between DNA polymerases of poxviruses, herpesviruses, and adenoviruses: Nucleotide sequence of the vaccinia virus DNA polymerase gene

(protein sequence/phosphonoacetic acid/Epstein–Barr virus)

PATRICIA L. EARL, ELAINE V. JONES†, AND BERNARD MOSS

Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

Communicated by James E. Darnell, Jr., January 21, 1986

ABSTRACT **A 5400-base-pair segment of the vaccinia virus genome was sequenced and an open reading frame of 938 codons was found precisely where the DNA polymerase had been mapped by transfer of a phosphonoacetate-resistance marker. A single nucleotide substitution changing glycine at position 347 to aspartic acid accounts for the drug resistance of the mutant vaccinia virus. The 5′ end of the DNA polymerase mRNA was located 80 base pairs before the methionine codon initiating the open reading frame. Correspondence between the predicted M_r 108,577 polypeptide and the 110,000 purified enzyme indicates that little or no proteolytic processing occurs. Extensive homology, extending over 435 amino acids, was found upon comparing the DNA polymerase of vaccinia virus and DNA polymerase of Epstein–Barr virus. A highly conserved sequence of 14 amino acids in the carboxyl-terminal regions of the above DNA polymerases is also present at a similar location in adenovirus DNA polymerase. This structure, which is predicted to form a turn flanked by β-pleated sheets, may form part of an essential binding or catalytic site that accounts for its presence in DNA polymerases of poxviruses, herpesviruses, and adenoviruses.**

DNA polymerases are encoded by members of at least three different virus families—e.g., poxviruses, herpesviruses, and adenoviruses. In the case of poxviruses, the induction of a new DNA polymerase may be correlated with the cytoplasmic site of replication. Within a few hours after infection with vaccinia virus, the prototypal member of the poxvirus family (1), a viral DNA polymerase composed of a single M_r 110,000 polypeptide with intrinsic 3′- to 5′-exonuclease activity appears (2). Evidence that this enzyme is specified by the virus has come from the isolation of phosphonoacetate (PAA)-resistant (PAA^R) and temperature-sensitive forms of DNA polymerase from cells infected with vaccinia virus mutants (3, 4). Marker transfer and rescue experiments served to localize the gene (5, 6), and a protein of correct size and peptide composition was formed by cell-free translation of viral mRNA that hybridized to the corresponding DNA fragment (5, 6). Two overlapping transcripts of about 3400 and 3800 nucleotides, with a common 5′ end, were mapped to the DNA polymerase locus (6, 7). To further study this DNA polymerase and compare it to others, we have sequenced a 5.4-kilobase pair (kbp) segment of the vaccinia virus genome. An uninterrupted open reading frame of predicted size was found, and a single base substitution that confers a PAA^R phenotype was identified. A computer-assisted database search revealed extensive amino acid sequence homology with an open reading frame of Epstein–Barr virus (EBV) that was predicted to encode DNA polymerase. A short, highly conserved region of identity between the DNA polymerases of vaccinia virus and EBV was also shared by adenovirus. To our knowledge, no homologies of any kind have previously been reported between members of these large and diverse virus families.

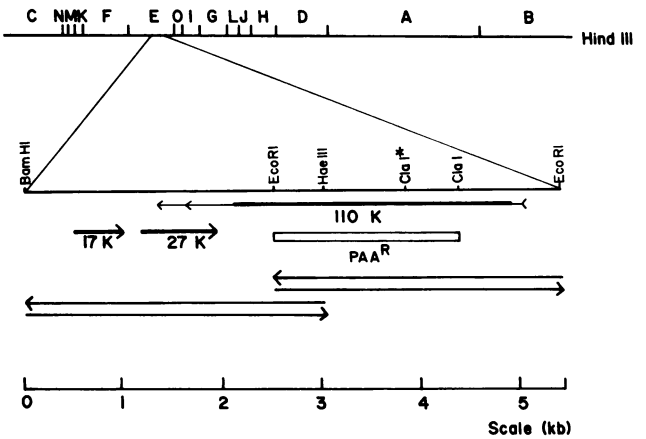

FIG. 1. Location of the DNA polymerase gene and sequencing strategy. The upper line represents a *Hin*dIII map of the entire vaccinia virus 180-kbp genome. The expanded portion represents the 5.4-kbp region that was sequenced (Fig. 2). The transcript for the M_r 110,000 (110K) DNA polymerase gene is shown by a thin line, the open reading frame by a heavy line. Open reading frames for the M_r 17,000 (17K) and the 27,000 (27K) putative peptides are also shown. The 2-kbp *Eco*RI–*Cla* I fragment that was reported to transfer PAA-resistance (5) is depicted by an open box. The *Cla* I site present in the PAA^R mutant but absent from wild-type virus is marked (*). The parallel arrows running in opposite directions indicate the fragments that were subcloned into M13 and completely sequenced.

MATERIALS AND METHODS

DNA Sequencing. Cloned DNA was derived from wild-type vaccinia virus (strain WR) or a PAA^R mutant (3). An *Eco*RI fragment from nucleotide 1 to 2867 and a *Hae* III fragment from nucleotide 2372 to the *Hae* III site in the plasmid vector were excised from pEJ4 and PEJ1, respectively. pEJ4 contains the 2.9-kbp *Eco*RI fragment in which most of the DNA polymerase gene is located. pEJ1 contains a 7.5-kbp *Bam*HI–*Hin*dIII fragment that extends from the middle to the right end of the *Hin*dIII E fragment (Fig. 1). The DNA fragments were cloned in both orientations into M13mp18 (8) and were propagated in *Escherichia coli* strains JM105 or

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: PAA, phosphonoacetic acid; PAA^R, PAA-resistant; EBV, Epstein–Barr virus; bp, base pair(s).

†Present address: Department of Molecular Genetics, Smith, Klein and French Laboratories, Philadelphia, PA 19201.

FIG. 2. Nucleotide sequence of a 5.4-kbp segment of the *Hin*dIII E fragment of a PAA^R mutant of vaccinia virus and deduced amino acid sequence of the DNA polymerase gene. The sites used for subcloning into M13mp18 are shown in italics. The entire sequence was determined in both orientations using the dideoxynucleotide chain termination technique (10). The long open reading frame from the methionine codon at nucleotide 498 to the termination codon at nucleotide 3311 codes for a M_r 108,577 protein and represents the DNA polymerase gene. The asterisk above nucleotide 418 indicates the start of transcription as determined by nuclease S1 mapping (Fig. 3). Two additional open reading frames, coding for putative proteins of M_r 17,617 and 27,109 are on the strand opposite of that coding for DNA polymerase. Initiation and termination codons for the three open reading frames are boxed. The wild-type nucleotides at positions 1435 and 1537, with their resulting amino acid changes, are shown above and below the nucleotide and amino acid sequences of the PAA^R mutant, respectively. Arrows indicate the direction of the leftward open reading frames.

JM109. Serial deletions, approximately 180 bp apart, were made with exonuclease III according to the method of Henikoff (9). The entire region was sequenced in both directions by the dideoxynucleotide chain termination meth-

od (10) using 5′-[α[^{35}S]thio]dATP and buffer-gradient sequencing gels (11). Sequences were aligned with the aid of SEQ (12) and NUCALN (13) computer programs. The sequence of a 2-kbp fragment of DNA from wild-type vaccinia virus was carried out as above except that synthetic oligonucleotide primers were used. Protein database searches were made by implementing FASTP (14) on an IBM-XT computer. Amino acid secondary structure determinations were made using the MicroGenie program from Beckman.

Nuclease S1 Mapping. DNA from pEJ4 was cleaved with *Hae* III, 5′ end-labeled with [γ-^{32}P]dATP and T4 polynucleotide kinase, and then cleaved with *Eco*RI using standard methods (42). After agarose gel purification, the 595-bp DNA fragment was hybridized to RNA isolated from the cytoplasm of cycloheximide-treated vaccinia virus-infected HeLa cells (7). Nuclease S1-resistant DNA was resolved on a 6% denaturing polyacrylamide gel. A parallel sequence ladder was prepared from the same labeled DNA fragment chemically cleaved at guanosine and adenosine residues (15).

Oligonucleotide Mutagenesis and Marker Transfer. A 2.9-kbp *Eco*RI fragment of the wild-type DNA polymerase gene was inserted into the *Eco*RI site of replicative form M13mp18. The single-stranded phage DNA was isolated and mutagenized using oligonucleotide TATCGTCATCGATGAAC or TTTCAAATTTTTGAATG or both (16). Marker transfer of PAA-resistance was carried out as described (5).

RESULTS

Nucleotide Sequence. We sequenced a 5400-bp segment of the vaccinia virus genome that encompassed the region in which the DNA polymerase gene had been previously mapped. An open reading frame of 2814 nucleotides was located within the boundaries of the DNA polymerase message and included a DNA segment shown to transfer PAA-resistance to wild-type vaccinia virus (Fig. 1). The deduced amino acid sequence (Fig. 2) predicts a M_r 108,577 polypeptide in agreement with previous estimates of 110,000 for the purified enzyme (2). Two additional open reading frames encoding putative polypeptides of 17,617 and 27,109 were found downstream of and in opposite orientation to the DNA polymerase gene (Figs. 1 and 2). The vaccinia virus proteins corresponding to these open reading frames have not yet been identified.

Identification of a Nucleotide Conferring Resistance to PAA. A 2-kbp segment of DNA from wild-type vaccinia virus that includes the site containing the PAA-resistance mutation was sequenced. Only two nucleotide differences, approximately 100 bp apart were found (Fig. 2). To determine whether one or both of these mutations caused the PAA-resistance phenotype, synthetic oligonucleotides containing single-base changes were used to mutagenize a 2.9-kbp *Eco*RI segment of wild-type DNA that had been cloned in a single-stranded M13 vector. Marker transfer experiments (Table 1) indicated that the substitution at nucleotide 1537 was sufficient to confer PAA-resistance. This nucleotide change creates a *Cla* I site that is absent from wild-type DNA. A *Cla* I site difference between wild-type and mutant DNA as well as the loss of the ability to transfer the PAA-resistance marker upon *Cla* I digestion has been noted (5). Thus the drug-resistance phenotype of this mutant was caused by a change of guanosine to adenosine.

Location of the 5′ End of the DNA Polymerase mRNA. RNA was isolated from vaccinia virus-infected cells to locate precisely the 5′ end of the DNA polymerase message. Since DNA polymerase is expressed at early times, an inhibitor of protein synthesis was added prior to and during infection to amplify early transcripts. Purified cytoplasmic RNA was then hybridized to a DNA fragment that was asymmetrically labeled at the *Hae* III site at nucleotide 595 (Fig. 2). After digestion of residual single-stranded DNA with nuclease S1, the lengths of the protected DNA fragments were compared by polyacrylamide gel electrophoresis to an A+G sequence ladder prepared with the same end-labeled DNA used for nuclease protection. Although a cluster of bands was detected (Fig. 3), only one of these corresponded to a purine. Since direct analysis of the 5′ ends of total early RNA has revealed only purines (17, 18), the adenosine residue at position 418 is likely to be the site of transcription initiation. The first three ATG triplets are followed by termination codons within a short distance. The fourth ATG, starting the open reading frame, occurs 80 bp downstream of the predicted RNA start site. Inspection of these ATG triplets indicates that only the

Table 1. Transfer of PAA resistance

DNA	PAA^R virus titer, pfu/ml
None	5.3
PAA^R	100
PAA^{S*}	7.1
PAA^S, C → T at 1435	1.5
PAA^S, G → A at 1537	128
PAA^S, C → T at 1435 and G → A at 1537	175

Synthetic oligonucleotides containing single-base changes were used to mutagenize a 2.9-kbp *Eco*RI segment of wild-type DNA inserted into a single-stranded M13 vector. CV-1 cells were infected with wild-type PAA^S vaccinia virus and transfected with mutagenized double-stranded replicative form M13 DNA. The yield of PAA^R recombinant vaccinia virus was determined by plaque assay in the presence of PAA at 300 μg/ml.
*PAA^S is DNA from wild-type PAA-sensitive virus.

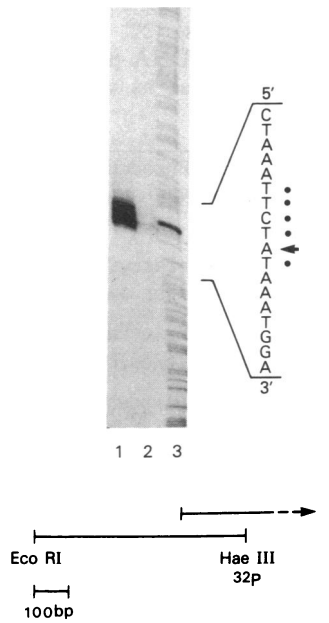

FIG. 3. Nuclease S1 mapping of the 5′ end of the mRNA for vaccinia virus DNA polymerase. Plasmid DNA was cleaved with *Hae* III (nucleotide 595 in Fig. 2), 5′ end-labeled and then cleaved with *Eco*RI (nucleotide 1). After gel purification, the 595-bp fragment was hybridized with RNA isolated from cycloheximide-treated vaccinia virus-infected HeLa cells, treated with nuclease S1, and resolved on a 6% denaturing polyacrylamide gel. Lanes 1 and 2, samples from different preparations of early RNA. Lane 3, same labeled DNA fragment chemically cleaved at guanosine and adenosine residues (15). The sequence to the right of the autoradiogram represents the complement of the actual sequence shown. The arrow points to the nucleotide most strongly protected, dots indicated adjacent nucleotides that are also protected.

fourth has the favored consensus for a codon that initiates translation (19).

Protein Database Search. The deduced amino acid sequence of the DNA polymerase and of the two other open reading frames were used to search the National Biomedical Research Foundation protein database.‡ For this purpose we implemented the program FASTP (14), which assigns scores based on amino acid identities and conservative replacements. No significant similarities were uncovered when the two open reading frames downstream of the DNA polymerase were used as queries. However, a single highly significant match was obtained when the 938-amino acid sequence of the vaccinia DNA polymerase was tested. The protein identified in this manner was the open reading frame predicted to encode the DNA polymerase of EBV (20, 21). The relevance of this finding was enhanced by the identical biological function of the two proteins and the similar locations of the matched regions within them. For the initial query, FASTP picked out the region from amino acids 430–544 and 490–604 of the vaccinia virus and EBV DNA polymerases, respectively (Fig. 4). Since FASTP locates only the best match between two proteins, the sequences on both sides of the selected region of vaccinia virus DNA polymerase were searched for additional similarities to EBV DNA polymerase. In this manner, another three regions extending over an additional 320 amino acids were matched (Fig. 4). The number of identical and conserved amino acid matches are shown in Table 2.

A marked degree of homology was found between a 14-amino acid sequence starting at position 720 of the vaccinia virus DNA polymerase and position 749 of EBV DNA polymerase (Fig. 4). Because this region was so highly conserved, we used it as a query to again search the protein databank with FASTP. The highest scoring match was to adenovirus DNA polymerase (22, 23). The 14-amino acid region from vaccinia virus DNA polymerase matched the adenovirus DNA polymerase as well as the EBV DNA polymerase (Fig. 5). All three DNA polymerases contain the same six consecutive amino acids in the center of the homology region. For EBV and adenovirus, the strong homology was evident even at the nucleotide level, whereas there was usually an adenosine or thymidine substitution in the wobble position of the vaccinia virus sequence.

Although the second and third highest scores obtained with the 14-amino acid query were to the polymerase 3 protein of influenza virus (24–26) and to the open reading frame predicted to encode the DNA polymerase of hepatitis B virus (27–29), the degree of similarity was less striking. Neither *E. coli* DNA polymerase I nor phage T7 DNA polymerase had evident homology.

Table 2. Sequence homology between DNA polymerases of vaccinia virus and EBV

Region	Amino acids, no.	Identical amino acids, %	Related amino acids,* %
I	135	25.2	58.5
II	115	22.6	70.4
III	51	33.3	72.5
IV	134	19.4	71.6

Data are derived from Fig. 4.
*Sum of identical and conservative substitutions.

DISCUSSION

An open reading frame of 938 amino acids was found precisely where the DNA polymerase gene of vaccinia virus had been mapped. The correspondence between the predicted M_r 108,577 polypeptide and the 110,000 purified DNA polymerase (2) suggests that little or no proteolytic processing occurs. The overall amino acid composition appears unremarkable with 12.5% acidic, 12.7% basic, 11.6% aromatic, and 27.1% hydrophobic residues. Inspection of the codon usage indicated an adenosine and thymidine preference for the wobble position consistent with the low G+C content of the vaccinia virus genome.

The DNA polymerase of vaccinia virus is sensitive to PAA, making isolation of resistant mutants possible (3, 4). We demonstrated that the drug resistance of one PAA^R mutant

‡National Biomedical Research Foundation (1985) *Protein Sequence Data Base of the Protein Identification Resource* (Washington, DC), Release No. 6.0.

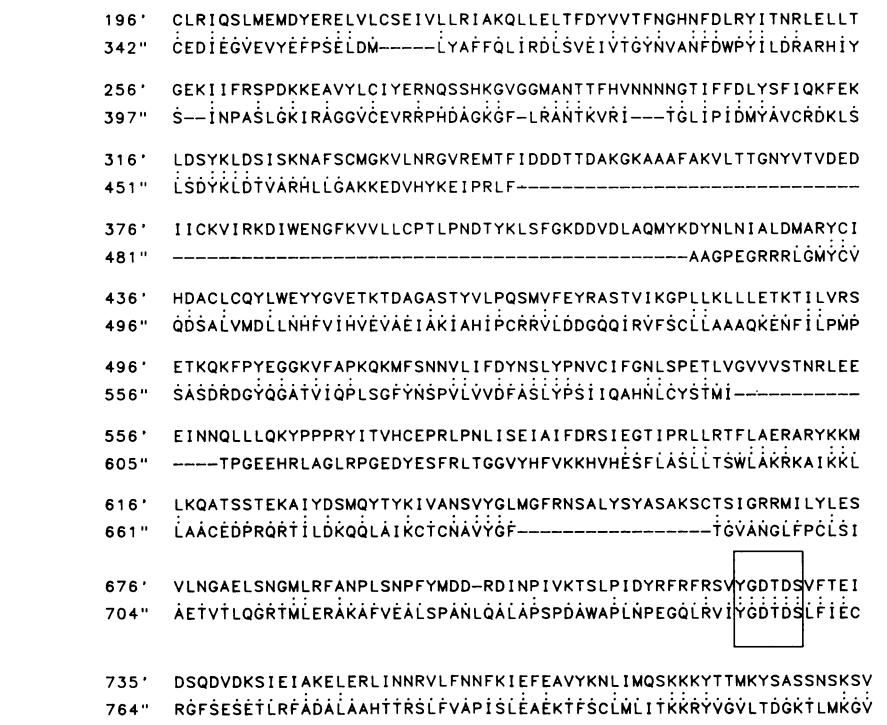

FIG. 4. Homology between the amino acid sequences of DNA polymerases from vaccinia virus and EBV. Numbers on the left represent the positions of amino acid residues in the two viral polymerases. Sequences are denoted for vaccinia virus (') and for EBV ("). Identical amino acids are indicated (:) and conservative substitutions (.). The four regions of homology between vaccinia and EBV are compared in Table 2.

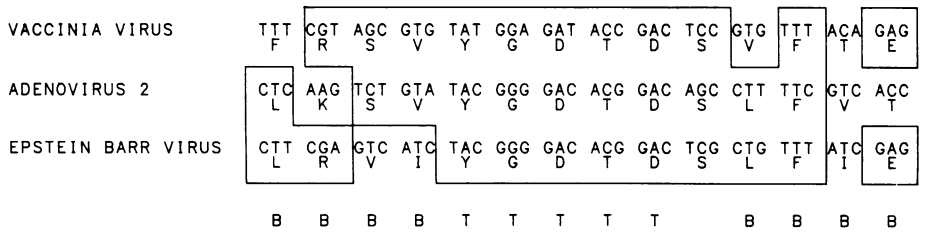

FIG. 5. Similar 14-amino acid sequence near the carboxyl ends of DNA polymerases of vaccinia virus, EBV, and adenovirus. Amino acids of vaccinia virus, adenovirus type 2 and EBV are shown starting at positions 720, 864, and 749, respectively. The predicted secondary structure for the region is shown. B, β-pleated sheet. T, reverse turn.

was caused by the substitution of a guanosine for an adenosine at amino acid number 347. Since PAA is an analog of pyrophosphate, the mutation probably occurs within or adjacent to the pyrophosphate binding site. By precisely mapping additional PAA^R mutants, it may be possible to determine the boundaries of the binding domain. The availability of the nucleotide sequence of the DNA polymerase also will facilitate the identification of the sites of other drug-resistance and temperature-sensitive mutations (4, 30).

Three out-of-frame ATG triplets were found within an 80-bp leader region upstream of the ATG codon initiating the long open reading frame. The fourth codon is the only one, however, that has a typical eukaryotic consensus sequence surrounding the ATG (19). This arrangement allows a reasonable level of translation, since the DNA polymerase is made in large amounts in vaccinia virus-infected cells (2). It might be of interest, nevertheless, to measure the effects produced by deleting the upstream ATG residues on the expression of DNA polymerase.

One important result of the nucleotide sequence determination was the finding of homology between the DNA polymerases of vaccinia virus, EBV and adenovirus types 2 and 5. The similarity of vaccinia virus and EBV polymerase is extensive, extending over a 435-amino acid region. In one 14-amino acid segment, near the carboxyl end of the two DNA polymerases, there are nine residues that are identical and four additional residues that are conservative substitutions. Furthermore, six of the identical amino acids occur consecutively. The significance of this finding was enhanced by the discovery of a similar 14-amino acid segment near the carboxyl end of adenovirus DNA polymerase. In this region of the vaccinia virus and adenovirus enzyme, there are nine identities (eight of which are consecutive) and five additional conservative substitutions. Remarkably, all three DNA polymerases have a common hexapeptide. A similar amino acid segment also is present in the DNA polymerase of herpes simplex virus type 1 (39, 40). Secondary structure analysis (31) suggests that the 14 amino acids form a reverse turn flanked by β-pleated sheet structures. It is reasonable to speculate that this region forms part of a binding or catalytic site that may be present in many DNA polymerases.

Three other genes of vaccinia virus—e.g., thymidine kinase (32, 33), a growth factor (34), and the large subunit of RNA polymerase (41) have been shown to share significant homology with their eukaryotic cellular counterparts (35–38, 43). Based on these results and the present finding of similarities between DNA polymerases of unrelated virus families, we predict that cellular DNA polymerase α will have considerable homology with vaccinia virus DNA polymerase.

We thank Steven Broyles and George Rohrmann for many helpful discussions, David Lipman for advice on computer programs, Donald Coen and John Quinn for communicating results prior to publication, and Mary Rust for typing the manuscript.

1. Moss, B. (1985) in *Virology*, eds. Feilds, B. N., Knipe, D. M., Chanock, R. M., Melnick, J., Roizman, B. & Shope, R. (Raven, New York), pp. 685–703.
2. Challberg, M. D. & Englund, P. T. (1979) *J. Biol. Chem.* **254**, 7812–7819.
3. Moss, B. & Cooper, N. (1982) *J. Virol.* **43**, 673–678.
4. Sridhar, P. & Condit, R. C. (1983) *Virology* **128**, 444–457.
5. Jones, E. V. & Moss, B. (1984) *J. Virol.* **49**, 72–77,
6. Traktman, P., Sridhar, P., Condit, R. C. & Roberts, B. E. (1984) *J. Virol.* **49**, 125–131.
7. Jones, E. V. & Moss, B. (1984) *J. Virol.* **53**, 312–315.
8. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9**, 309–321.
9. Henikoff, S. (1984) *Gene* **28**, 351–359.
10. Sanger, F., Coulson, A. R., Barrell, B. J., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178.
11. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
12. Stone, T. W. & Potter, K. N. (1984) *Nucleic Acids Res.* **12**, 367–378.
13. Wilbur, W. J. & Lipman, D. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 726–730.
14. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441.
15. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
16. Zoller, M. J. & Smith, M. (1984) *DNA* **3**, 479–488.
17. Boone, R. F. & Moss, B. (1977) *Virology* **79**, 67–80.
18. Keith, J., Gershowitz, A. & Moss, B. (1980) *J. Virol.* **36**, 601–605.
19. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
20. Bankier, A. T., Deininger, P. L., Farreu, P. J. & Barrell, B. G. (1983) *Mol. Biol. Med.* **1**, 21–45.
21. Baer, R., Bankeir, A. T., Biggin, M. D., Deininger, P. L., Farreu, P. J., Gibson, T. J., Hatfull, G., Hudson, G. S., Satchwill, S. C., Segium, C., Tuffney, P. S. & Barrell, B. G. (1984) *Nature (London)* **310**, 207–211.
22. Gingeras, T. R., Sciaky, D., Gelinas, R. E., Bing-Dong, J., Yen, C. E., Kelly, M. M., Bullock, P. A., Parsons, B. C., O'Neill, K. E. & Roberts, R. J. (1982) *J. Biol. Chem.* **257**, 13475–13491.
23. Alestrom, P., Akusjarui, G., Pettersson, M. & Pettersson, U. (1982) *J. Biol. Chem.* **257**, 13492–13498.
24. Kaptein, J. S. & Nayak, D. P. (1982) *J. Virol.* **42**, 55–63.
25. Fields, S. & Winter, G. (1982) *Cell* **28**, 303–313.
26. Jones, K. L., Huddleston, J. A. & Brownlee, G. G. (1983) *Nucleic Acids Res.* **11**, 1555–1566.
27. Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P. & Charnay, P. (1979) *Nature (London)* **281**, 646–650.
28. Ono, Y., Onda, H., Sasada, R., Igarashi, K., Sugino, Y. & Nishioka, K. (1983) *Nucleic Acids Res.* **11**, 1747–1757.
29. Pasek, M., Goto, T., Gilbert, W., Zink, B., Schaller, H., Mackay, P., Leadbetter, G. & Murray, K. (1979) *Nature (London)* **282**, 575–579.
30. DeFilippes, F. M. (1984) *J. Virol.* **52**, 474–482.
31. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
32. Weir, J. P. & Moss, B. (1983) *J. Virol.* **46**, 530–537.
33. Hruby, D. E., Maki, R. A., Miller, D. B. & Ball, L. A. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3411–3415.
34. Venkatesan, S., Gershowitz, A. & Moss, B. (1982) *J. Virol.* **44**, 637–646.
35. Bradshaw, H. D., Jr., & Deininger, P. L. (1984) *Mol. Cell. Biol.* **4**, 2316–2320.
36. Blomquist, M. C., Hunt, L. T. & Barker, W. C. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7363–7367.
37. Brown, J. P., Twardzik, D. R., Marquardt, H. & Todaro, G. J. (1985) *Nature (London)* **313**, 491–492.
38. Reisner, A. H. (1985) *Nature (London)* **313**, 801–803.
39. Gibbs, J. S., Chiou, H. C., Hall, J. D., Mount, D. W., Retondo, M. J., Weller, S. K. & Coen, D. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7969–7973.
40. Quinn, J. P. & McGeoch, J. M. (1985) *Nucleic Acids Res.* **13**, 8143–8163.
41. Broyles, S. & Moss, B. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3141–3145.
42. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), p. 122.
43. Kuoh, T. J. & Engler, J. A. (1984) *Nucleic Acids Res.* **12**, 3959–3971.