Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 3141–3145, May 1986
Biochemistry

Homology between RNA polymerases of poxviruses, prokaryotes, and eukaryotes: Nucleotide sequence and transcriptional analysis of vaccinia virus genes encoding 147-kDa and 22-kDa subunits

(*Escherichia coli*/*Saccharomyces cerevisiae*/*Drosophila melanogaster*/evolution)

STEVEN S. BROYLES AND BERNARD MOSS

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

Communicated by Aaron J. Shatkin, February 7, 1986

ABSTRACT **We have determined the nucleotide sequence of a region of the vaccinia virus genome encoding RNA polymerase subunits of 22 and 147 kDa and have mapped the 5′ and 3′ ends of the two mRNAs. The predicted amino acid sequence of the vaccinia 147-kDa subunit shows extensive homology with the largest subunit of *Escherichia coli* RNA polymerase, yeast RNA polymerases II and III, and *Drosophila* RNA polymerase II. The regions of homology between the five RNA polymerases are subdivided into five separate domains that span most of the length of each. A sixth domain shared by the vaccinia and the eukaryotic polymerases is absent from the *E. coli* sequence. In all specified regions, the vaccinia large subunit has greater homology with eukaryotic RNA polymerases II and III than with the *E. coli* polymerase. Vaccinia virus and eukaryotic RNA polymerases may therefore have evolved from a common ancestral gene after the latter diverged from prokaryotes.**

The presence of a transcription and RNA-modification system within the core of vaccinia virus and other poxviruses contributes to their remarkable ability to replicate in the cytoplasm of animal cells (1). Because of its large size and multisubunit composition (2–4), the vaccinia virus RNA polymerase more closely resembles RNA polymerases of eukaryotic and prokaryotic organisms than those of viruses such as T7 (5). Nevertheless, the vaccinia virus RNA polymerase can be distinguished from that of its host with regard to the precise size of the subunits (2–4), α-amanitin insensitivity (3, 4), and specificity for vaccinia virus promoters (6, 7). That soluble vaccinia virus extracts can initiate (8, 9) and terminate (G. Rohrmann, L. Yuen, and B.M., unpublished data) transcription suggests that this system may have significant value as a research tool. Determination of the nature of the RNA polymerase is central both for our understanding of the biology and evolution of this extensive virus family and for evaluating its significance as an experimental model. We report here the nucleotide sequences of two vaccinia virus genes encoding RNA polymerase subunits and a computer analysis that reveals extensive similarities between the amino acid sequences of the large RNA polymerase subunits of vaccinia virus, *Escherichia coli*, yeast, and *Drosophila*.

MATERIALS AND METHODS

Vaccinia Virus DNA and RNA. The vaccinia virus DNA used for sequencing was derived from subclones of the *Hin*dIII J and H genome fragments of the WR strain (10). The DNA sequence across the *Hin*dIII junction was confirmed with an *Eco*RI clone, pV7.1 (11). Vaccinia virus RNA was purified from infected HeLa cells by centrifugation in CsCl gradients (12). "Late" RNA was extracted from cells 6 hr after infection, and "early" RNA, from cells infected for 4 hr in the presence of cycloheximide.

DNA Sequence Analysis. Overlapping fragments of vaccinia virus DNA were inserted into M13 mp18 or mp19 phage vectors (13). Inserts in replicative-form DNA were deleted unidirectionally with exonuclease III (14), and the single-stranded phage DNA was sequenced by the dideoxynucleotide chain-termination method (15). In some places, synthetic oligonucleotide primers were used in order to completely sequence both DNA strands. Computer analysis of overlapping DNA sequences was performed with the SEQ (16) and NUCALN (17) programs. Protein-homology searches were made with the FASTP program (18).

RESULTS

Nucleotide Sequence of the Vaccinia Virus Genes Encoding the 147-kDa and 22-kDa RNA Polymerase Subunits. Eight putative RNA polymerase subunit genes have been mapped by immunoprecipitation of cell-free translation products of mRNAs that hybridized to cloned vaccinia virus DNA fragments (E. V. Jones, C. Puckett, and B.M., unpublished work), and two by screening of a λgt11 library with a monoclonal antibody (19). We determined the sequence of a 6067-base-pair region spanning the map positions of the largest subunit and one of the small subunits (Fig. 1). Two open reading frames (ORFs) had the size, location, and orientation expected for the RNA polymerase subunits. Molecular masses of 22,008 and 146,967 daltons are predicted from the sequence. A third ORF, located between the two RNA polymerase genes and oriented in the opposite direction, is predicted to code for a protein of 15,159 daltons. Another ORF, to the right of the gene coding for the 147-kDa subunit, predicts a protein of 19,700 daltons corresponding to a late protein previously mapped to this region (20). The TAA termination codons of these oppositely oriented ORFs are overlapping and share two nucleotides.

The sequence of a segment of vaccinia DNA that overlapped nucleotides 1–1711 of Fig. 1 has recently been reported (21). The two sequence determinations are in agreement except at nucleotides 904, 1324, 1623, and 1625. Plucienniczak *et al.* (21) also noted that an ORF for a 41-kDa polypeptide (which ends with a TAG at nucleotide 336 in Fig. 1) overlaps that of the 22-kDa-subunit gene, but in a different frame.

Nuclease S1 Mapping of RNA 5′ and 3′ Ends. The general locations of RNAs mapping in the region now known to contain the genes for the 22-kDa and 147-kDa RNA polymerase subunits were determined by the nuclease S1 mapping technique (11, 12). For a more precise localization of

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.

Abbreviation: ORF, open reading frame.

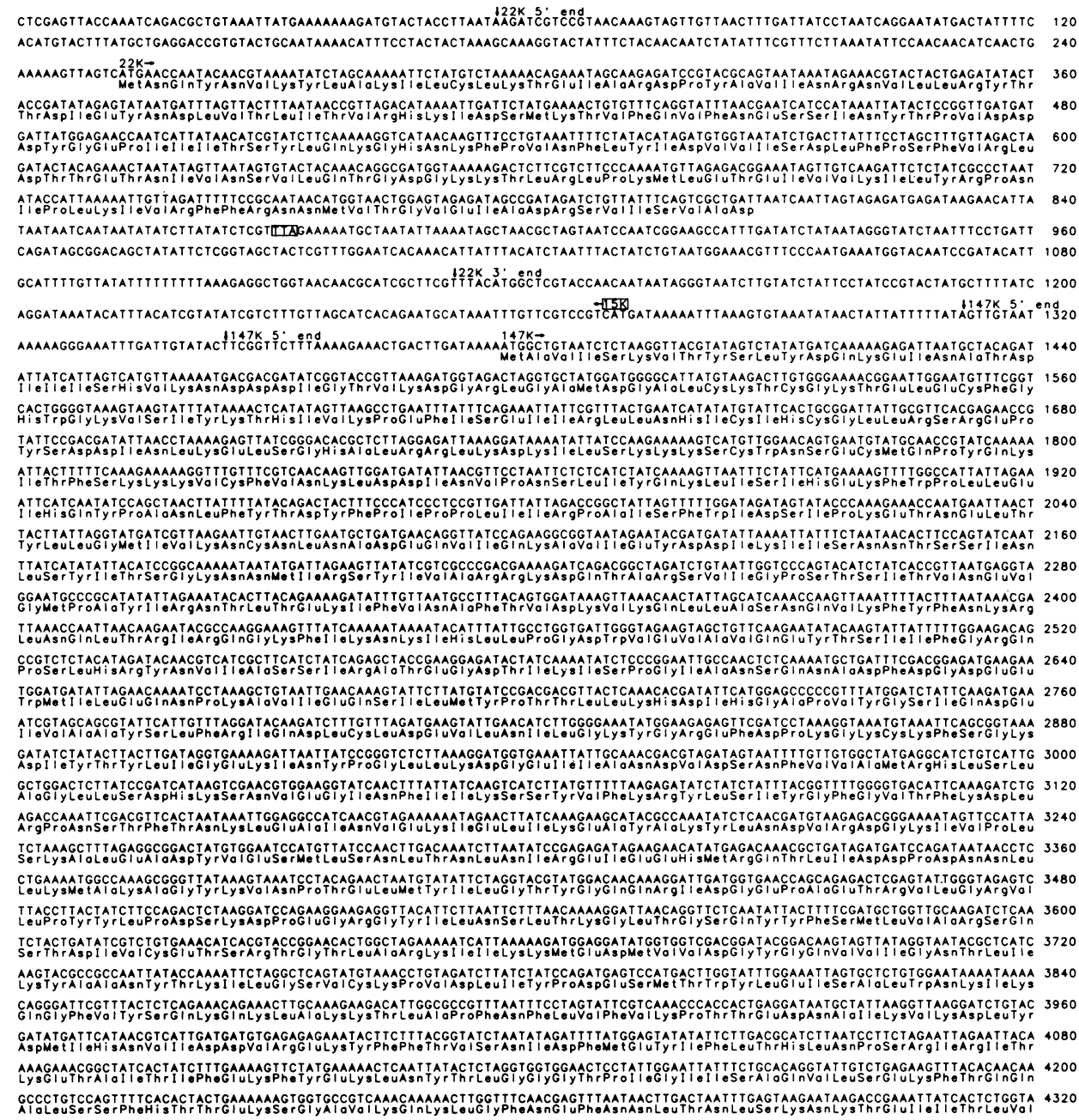

FIG. 1. Nucleotide sequence of the genes for vaccinia RNA polymerase 22-kDa and 147-kDa subunits. The amino acid sequences of the 22-kDa (22K) and 147-kDa (147K) polypeptides are shown below the coding sequences. The locations of mRNA ends as determined by nuclease S1 analysis are indicated. The initiation and termination codons of the ORF for the putative 15-kDa (15K) polypeptide and the 19-kDa (19K) late polypeptide, which are oriented in the opposite direction, are also shown.

these sites, nuclease S1 cleavage products were displayed alongside sequence ladders on high-resolution polyacrylamide gels (Fig. 2). The 5′ end of an early 22-kDa-subunit mRNA mapped to nucleotide 57 of Fig. 1, indicating that the 5′ noncoding leader sequence for the 22-kDa-subunit mRNA is 196 nucleotides long. No 5′ end was detected in the vicinity of this region when late RNA was analyzed in this manner. Nuclease S1 analysis of the 3′ end of the 22-kDa-subunit message indicated some heterogeneity. The S1 signal was centered around position 1132, which is 325 nucleotides downstream from the termination codon for the 22-kDa-subunit gene. The total length for this mRNA is predicted to be 1075 nucleotides, almost twice that of the coding sequence, but in excellent agreement with the experimentally determined length of 1070 nucleotides (12). Nuclease S1 analysis of the 5′ end of the 147-kDa-subunit mRNA revealed signals at nucleotides 1310–1312 and 1345–1346, just upstream of the ORF initiation codon at 1378. Primer-extension experiments support the existence of two populations of 147-kDa-subunit message with 5′ termini at nucleotides 1313 and 1345 (data not shown). When late vaccinia RNA was hybridized to the probe, an intense nuclease S1 signal corresponded to nucleotides 1098–1104. These sites of nuclease S1 cleavage occurred within and just downstream of a long oligo(U) stretch, and we have not attempted to confirm these as RNA start sites by primer extension. Nevertheless, these data raise the possibility that the gene for the 147-kDa RNA polymerase subunit is transcribed at late as well as early times after infection.

Precise mapping of the 3′ end of the mRNA for the 147-kDa subunit proved to be difficult because of the lack of useful restriction sites. Nuclease S1 experiments using probes labeled at distant upstream sites indicated that the 3′ end of this message is located at nucleotide 5660 ± 20, in good agreement with a previous determination (11). Since the 3′ noncoding region for the 147-kDa-subunit mRNA is about 420 nucleotides long, the total length of this message is about 4350 nucleotides. This agrees well with a recent length determination of 4500 nucleotides by RNA blot analysis (P. Earl, unpublished results).

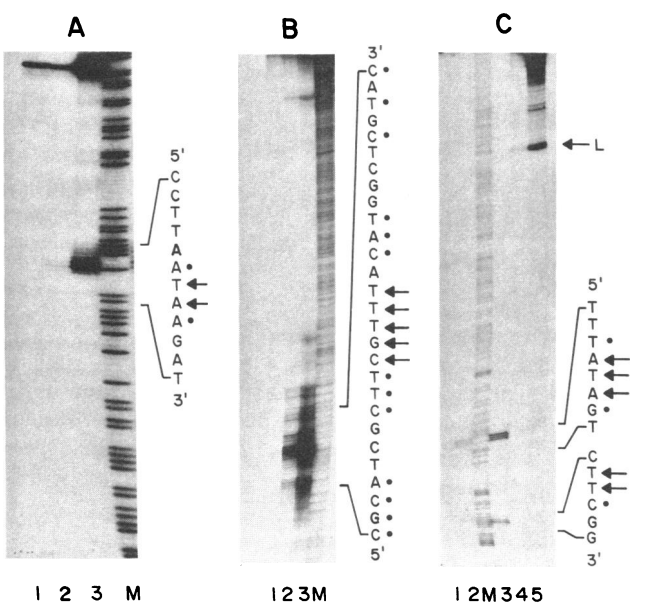

FIG. 2. Nuclease S1 analysis of the 5′ and 3′ ends of mRNA for the 22-kDa subunit (*A* and *B*, respectively) and of the 5′ end of the mRNA for the 147-kDa subunit (*C*). DNA probes labeled with ^{32}P at an appropriate restriction site (*A*, *Ava* II; *B*, *Hin*fI; *C*, Asp-718) were hybridized to 20 μg of yeast tRNA (lanes 1), 10 μg (lanes 2) or 20 μg (lanes 3) of vaccinia early RNA, or 10 μg (lane 4) or 20 μg (lane 5) of vaccinia late RNA and then were treated with nuclease S1. Digestion products were electrophoresed on high-resolution gels alongside G+A sequence "ladders" (lanes M) for reference. Locations of primary protection products are indicated in the nucleotide sequence by arrows. Minor bands are indicated by dots. In *C*, the major protection product produced in the presence of late RNA is indicated by L.

Evidence for Homology Between Vaccinia Virus and Cellular RNA Polymerase Subunits. The derived amino acid sequences of both vaccinia virus RNA polymerase subunits were compared with the National Biomedical Research Foundation Protein Library by using the FASTP computer program of Lipman and Pearson (18). The search against the 147-kDa sequence found significant homology with the β′-chain (22) of the *E. coli* RNA polymerase. This match was by far the highest-scoring one and was the only match considered to be of possible biological significance. Comparison of the 147-kDa sequence with the sequences of the large subunits of the *Saccharomyces cerevisiae* RNA polymerases II and III (23) and a partial sequence of the *Drosophila melanogaster* RNA polymerase II large subunit (24), however, revealed even greater homology. Optimal alignment of homologous regions was achieved by division of the vaccinia virus 147-kDa sequence into blocks of 200–300 amino acid residues prior to comparison using the FASTP program. This procedure revealed five major regions of homology common to the viral, bacterial, and yeast large subunits (Fig. 3). A sixth region of homology, near the amino termini of the vaccinia virus and the eukaryotic RNA polymerases, was undetectable in the *E. coli* β′ subunit. The limited data available for the *Drosophila* RNA polymerase II from the partial sequence follow the same pattern. The blocks of homology with the vaccinia 147-kDa subunit range from one end of the molecule to the other, in a pattern that is similar among all the RNA polymerases (Fig. 4). An exception is the repeating heptapeptide only present at the carboxyl terminus of the yeast RNA polymerase II large subunit (23).

The degree of homology of the prokaryotic and eukaryotic large subunits with that of the vaccinia virus RNA polymerase varies in the different regions. The identity ranges from a low of 11% for domain VI of the *E. coli* β′ subunit to 38% for domain II of the yeast RNA polymerase III (Table 1). The level of homology rises 2- to 3-fold when conservative amino acid changes are taken into account. The longest contiguous amino acid match among all four RNA polymerases is the heptapeptide Asn-Ala-Asp-Phe-Asp-Gly-Asp located at residues 413–419 of the vaccinia virus 147-kDa sequence. A FASTP computer search comparing this short sequence against the protein library identified no other perfect matches.

DISCUSSION

The amino acid sequences of the vaccinia virus 22-kDa and 147-kDa RNA polymerase subunits were inferred from the DNA structure. The only remarkable feature of either polypeptide is the high (40%) β-pleated-sheet content of the 22-kDa subunit. Both the 5′ and 3′ noncoding sequences of the mRNAs for the two subunits are longer than those previously described (30–32) for vaccinia virus early mRNAs. However, the 5′ end of the 22-kDa-subunit message and the 3′ end of the 147-kDa-subunit message map within adjacent ORFs, indicating dual utilization of the same nucleotides for protein-coding sequences and transcriptional control signals.

The vaccinia virus RNA polymerase bears a remarkable resemblance to cellular RNA polymerases. All possess two large subunits (>100 kDa) and several polypeptides of lesser size. It is now evident that the similarity between vaccinia and cellular RNA polymerases extends to the molecular structure of at least one subunit. The amino acid sequence of

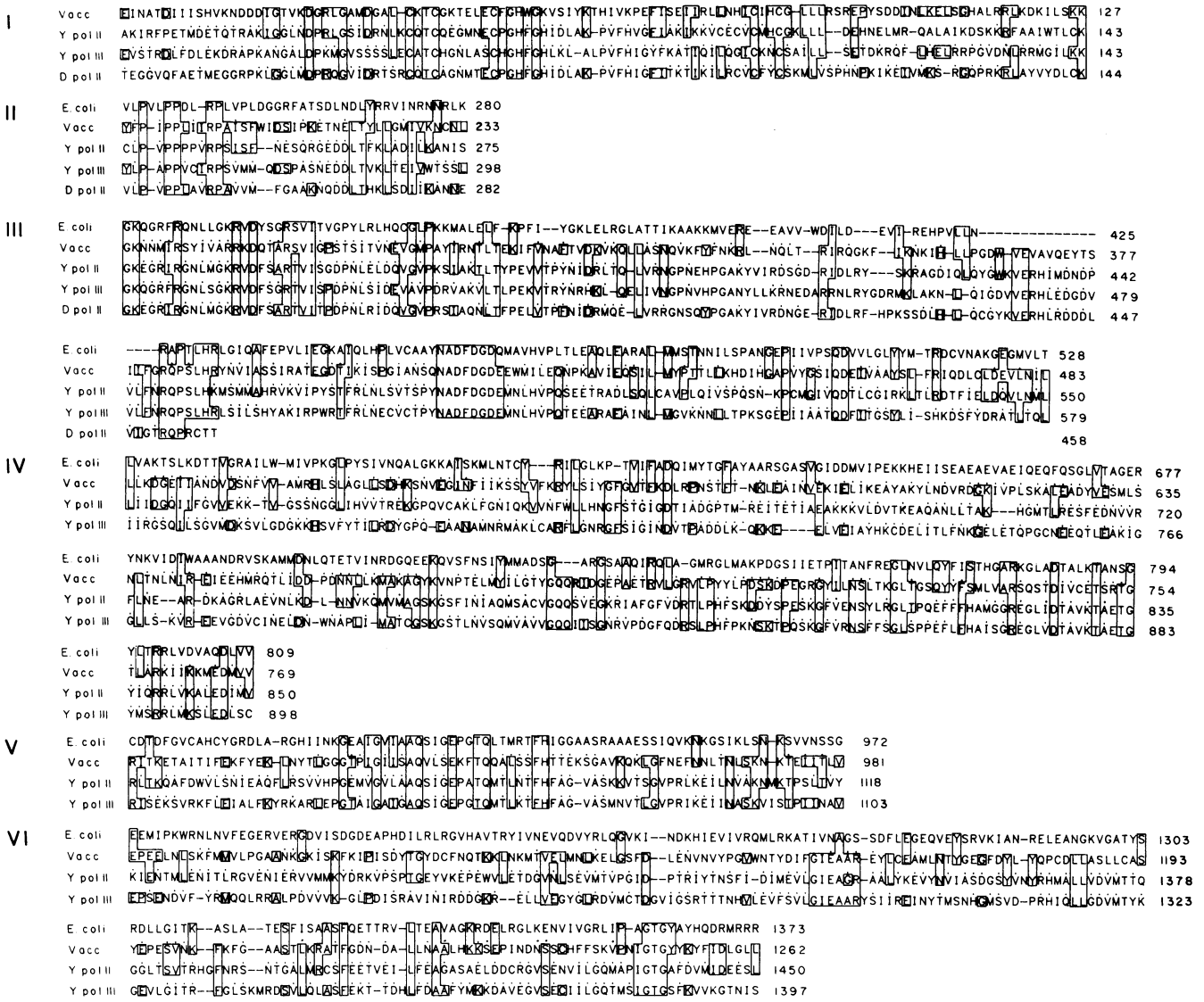

FIG. 3. Homology between vaccinia virus RNA polymerase 147-kDa subunit and *E. coli* RNA polymerase β′ chain, the large subunits of yeast RNA polymerases II and III, and the large subunit of *Drosophila* RNA polymerase II. Six different regions of homology (I–VI) are shown, using the single-letter amino acid abbreviations. Amino acid residue numbers are at right. Matches between the vaccinia virus 147-kDa-subunit sequences and the others are boxed. Conservative amino acid changes are shown by dots.

the largest subunit of the vaccinia virus RNA polymerase, the 147-kDa subunit, shares extensive homology with the largest subunits of the prokaryotic RNA polymerase and eukaryotic RNA polymerases II and III.

Because of their structural similarities, it seems likely that the large subunits of all four RNA polymerases have similar functions. The *E. coli* β′ chain has been implicated as a DNA-binding component of RNA polymerase (24–28) and

FIG. 4. Locations of domains of homology between the vaccinia virus RNA polymerase 147-kDa subunit, the *E. coli* RNA polymerase β′ chain, the large subunits of yeast RNA polymerases II and III, and the large subunit of *Drosophila* RNA polymerase II. The six regions of homology between the RNA polymerases (Fig. 3) are indicated between the primary sequences of the polypeptides, aligned at their amino termini.

Table 1. Homology of RNA polymerase large subunits to the vaccinia virus RNA polymerase 147-kDa polypeptide

	% homology*					
	I	II	III	IV	V	VI
E. coli β′	NS†	18 (33)	22 (58)	15 (66)	18 (43)	11 (28)
Yeast pol II	25 (54)	33 (70)	28 (70)	21 (66)	27 (70)	19 (78)
Yeast pol III	21 (68)	38 (78)	25 (62)	22 (69)	30 (57)	22 (64)
Drosophila pol II	24 (65)	35 (66)	27 (70)‡			

The homologous domains I–VI are shown in Fig. 3.
*Numbers in parentheses show percent homology when conservative amino acid changes are considered.
†No significant homology detected.
‡Only 119 of the 216 amino acids in this domain are available for comparison.

may participate in DNA unwinding (27). In addition, a mutation in the largest subunit of *Drosophila* RNA polymerase II confers a reduced rate of RNA chain elongation *in vitro* (29).

The degree of relatedness among the four types of RNA polymerases has implications for the evolution of the vaccinia virus enzyme. In all regions of the polypeptide, the vaccinia 147-kDa subunit more closely resembles the eukaryotic large subunits than the prokaryotic large subunit. Some regions of the vaccinia 147-kDa subunit are more closely related to the yeast RNA polymerase II, while others more closely resemble RNA polymerase III. It is difficult to ascertain, therefore, whether the vaccinia virus RNA polymerase evolved from RNA polymerase II, RNA polymerase III, or their ancestor.

Other subunits of the vaccinia virus RNA polymerase also may prove to share homology with the RNA polymerases of prokaryotes and eukaryotes. For example, we predict that the second large subunit of the vaccinia RNA polymerase is related to the β chain of the *E. coli* RNA polymerase. The map locations of the vaccinia RNA polymerase subunits are known and are therefore amenable to sequence analysis. By identifying regions of homology with *E. coli* RNA polymerase subunits, highly conserved sequence probes might be used for identification of additional eukaryotic RNA polymerase genes.

We thank Drs. J. Weir, G. Rohrmann, A. Mahr, and B. Roberts for cloned vaccinia DNA; Drs. P. Earl, J. Rosel, M. Merchlinsky, and K. Brechling for helpful discussions; J. Baldick for synthesis of oligonucleotide primers, and M. Rust for typing the manuscript.

1. Moss, B. (1985) in *Virology*, eds. Fields, B. N., Knipe, D. M., Chanock, R. M., Melnick, J. L., Roizman, B. & Shope, R. E. (Raven, New York), pp. 685–703.
2. Nevins, J. R. & Joklik, W. (1977) *J. Biol. Chem.* **252**, 6930–6938.
3. Baroudy, B. M. & Moss, B. (1980) *J. Biol. Chem.* **255**, 4372–4380.
4. Spencer, I., Shuman, S. & Hurwitz, J. (1980) *J. Biol. Chem.* **255**, 5388–5395.
5. Studier, F. W. & Dunn, J. J. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 999–1007.
6. Puckett, C. & Moss, B. (1983) *Cell* **35**, 441–449.
7. Foglesong, D. J. (1985) *J. Virol.* **53**, 822–826.
8. Golini, F. & Kates, J. R. (1985) *J. Virol.* **53**, 205–213.
9. Rohrmann, G. & Moss, B. (1985) *J. Virol.* **56**, 349–355.
10. Belle Isle, H., Venkatesan, S. & Moss, B. (1981) *Virology* **112**, 306–317.
11. Mahr, A. & Roberts, B. E. (1984) *J. Virol.* **49**, 497–509.
12. Bajszar, G., Wittek, R., Weir, J. P. & Moss, B. (1983) *J. Virol.* **45**, 62–72.
13. Messing, J. & Viera, J. (1982) *Gene* **19**, 269–276.
14. Henikoff, S. (1984) *Gene* **28**, 351–359.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1984) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5469.
16. Stone, T. W. & Potter, K. N. (1984) *Nucleic Acids Res.* **12**, 367–378.
17. Wilbur, W. J. & Lipman, D. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 726–730.
18. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441.
19. Morrison, D. K., Carter, J. K. & Moyer, R. W. (1985) *J. Virol.* **55**, 670–680.
20. Mahr, A. & Roberts, B. E. (1984) *J. Virol.* **49**, 510–520.
21. Plucienniczak, A., Schroeder, E., Zettlmeissl, G. & Streeck, R. E. (1985) *Nucleic Acids Res.* **13**, 985–999.
22. Ovchinnikov, Y. A., Monastyrskaya, G. S., Gubanov, V. V., Guryev, S. O., Salomatina, I. S., Shuvaeva, T. M., Lipkin, V. M. & Sverdlov, E. D. (1982) *Nucleic Acids Res.* **10**, 4035–4044.
23. Allison, L. A., Moyle, M., Shales, M. & Ingles, C. J. (1985) *Cell* **42**, 599–610.
24. Biggs, J., Searles, L. L. & Greenleaf, A. L. (1985) *Cell* **42**, 611–621.
25. Fukuda, R. & Ishihama, A. (1974) *J. Mol. Biol.* **87**, 523–540.
26. Yarbrough, L. R. & Hurwitz, J. (1974) *J. Biol. Chem.* **249**, 5400–5404.
27. Panny, R., Heil, A., Mazus, B., Palm, P., Zillig, W., Midlin, S. Z., Ilyina, T. S. & Khesin, R. S. (1974) *FEBS Lett.* **48**, 241–245.
28. Gros, G., Fields, D. A. & Bautz, E. K. F. (1976) *Mol. Gen. Genet.* **147**, 337–341.
29. Coulter, D. E. & Greenleaf, A. L. (1985) *J. Biol. Chem.* **260**, 13190–13198.
30. Venkatesan, S., Gershowitz, A. & Moss, B. (1982) *J. Virol.* **44**, 637–646.
31. Venkatesan, S., Baroudy, B. M. & Moss, B. (1981) *Cell* **25**, 805–813.
32. Weir, J. P. & Moss, B. (1983) *J. Virol.* **46**, 530–537.