Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 4152–4155, June 1989
Genetics

Retroviral protease-like gene in the vaccinia virus genome

(retrovirus/sequence comparisons/gene transfer)

M. B. SLABAUGH* AND N. A. ROSEMAN

Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331-6503

Communicated by Russell F. Doolittle, February 13, 1989

ABSTRACT **The retroviral protease-encoding region, *PR*, situated between the *gag* and *pol* genes, underwent gene duplication in the lineage now represented by simian retrovirus type 1; the sequence of the duplicated segment has diverged considerably from the present *PR* sequence [Power, M. D., Marx, P. A., Bryant, M. L., Gardner, M. B., Barr, P. J. & Luciw, P. A. (1986) *Science* 231, 1567–1572]. The *PR*-like duplicated gene segment was at some point translocated to a new site within the *pol* gene of a lentivirus (subsequent to the divergence of human immunodeficiency virus type 1), where it has been maintained. We have identified in the vaccinia virus genome a sequence that is homologous to the *PR*-like duplicated gene segment of both types of retrovirus in an open reading frame whose product is predicted to be a 16.2-kDa protein. The vaccinia *PR*-like gene is located in the *Hin*dIII F fragment, and its product displays 31–34% amino acid identity to the two types of retroviral duplicated protease sequences over a region encompassing 125 amino acid residues. Sequences flanking the vaccinia gene showed no significant homology at either the DNA or amino acid level to the retroviruses. Nuclease S1 and primer extension analyses determined that the vaccinia gene is transcribed early in infection.**

Transfer of genetic material between unrelated organisms is a poorly understood phenomenon by which genes are disseminated into new environments. The presence in certain bacteria of coding sequences related to plant genes has been reported (1, 2), and the high degree of homology between some vaccinia virus and cellular genes (3–5) suggests that acquisition of host coding sequences may be a route by which these viruses expand their repertoire of functions. In this paper, we report that a protein-encoding sequence that has already been shown to have translocated horizontally between distantly related retroviruses (6) is also present in the genome of vaccinia virus. Vaccinia, a member of the Poxviridae, is a large DNA virus with no previously detected homology to the retroviruses.

The sequence in question was first noted in the simian retrovirus type 1 (SRV-1) genome, in which the protease-encoding region (*PR*) between the *gag* and *pol* genes is preceded by a related but distinct sequence in-frame with *PR* (7). Although extensive sequence divergence precludes statistically significant alignment of the two regions, the presence of conserved remnants led to the conclusion that a gene duplication event had given rise to an independently evolving segment whose present function is unknown (6, 7).

A sequence that is homologous to the duplicated *PR* segment is also present in certain lentiviruses but is inserted into the 3′ end of the *pol* region between the ribonuclease H- and endonuclease/integrase-encoding regions rather than adjacent to authentic *PR*. The similarity between these *PR*-like regions has been noted by McClure *et al.* (6, 8), who designated the duplicated *PR* segment of SRV-1 and its relatives as "X1" and the corresponding segment of the lentiviruses as "X2." Pairwise comparisons of several *X1*, *X2*, and *PR* sequences revealed that the *PR*-like segments are more closely related among themselves than they are to authentic *PR* sequences. These authors concluded that duplication of the *PR* gene occurred in an ancestral retroviral lineage whose descendants include mouse mammary tumor virus (MMTV), hamster intercisternal A-type particle, and human endogenous retrovirus, as well as SRV-1, and that the duplicated sequence was subsequently translocated to a new site in a progenitor of the lentiviruses at a point in time subsequent to the divergence of human immunodeficiency virus (HIV). A mechanism for the horizontal relocation involving integrase-mediated excision and insertion during a dual infection by the donor and recipient retroviruses was proposed (6).

Comparison of a sequenced vaccinia virus open reading frame (ORF)† (see Fig. 1) and its deduced amino acid sequence to available protein sequences revealed a high degree of similarity to both the X1 and X2 retroviral sequences. The vaccinia ORF putatively encodes a 16.2-kDa protein, and the ORF 5′ and 3′ ends correspond closely to the ends of the retroviral *PR*-like segments.

MATERIALS AND METHODS

DNA Sequencing. A clone of the *Hin*dIII F fragment of vaccinia virus strain WR was obtained from B. Moss (National Institute of Allergy and Infectious Diseases). DNA subfragments were inserted into M13 or phagemid vectors and sequenced by the dideoxynucleotide chain-termination method (9, 10).

Computer Analyses. The computer used in this study was a Sun 3/60 computer running the UNIX (Berkely 4.2) operating system. Comparison of the deduced amino acid sequence of ORF *F2L* with the National Biomedical Research Foundation and GenBank data bases was performed with the FASTA program (11) and Intelligenetics IFIND.

RNA Analyses. Vaccinia virus (strain WR) was grown on BSC40 monkey kidney cells, and total RNA was prepared as described (12). S1 nuclease mapping using 5′-end-labeled DNA fragments and primer extension analysis using an end-labeled oligonucleotide primer were performed as described (4, 13).

RESULTS AND DISCUSSION

Homology Between ORF *F2L* and Retroviral Duplicated *PR*-Like Sequences. The deduced amino acid sequence of *F2L*, an ORF in the left end of the *Hin*dIII F fragment of vaccinia

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: ORF, open reading frame; *PR*, retroviral protease gene; SRV-1, simian retrovirus type 1; MMTV, mouse mammary tumor virus; EIAV, equine infectious anemia virus; VLV, visna lentivirus.

*To whom reprint requests should be addressed.

†The sequence reported in this paper is being deposited in the EMBL/GenBank data base (Intelligenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. M23167).

virus genome (Fig. 1), was compared to the National Biomedical Research Foundation Protein Sequence Database‡ and the GenBank Genetic Sequence Database§. The search yielded five highly significant alignment scores, all to retroviral sequences. Three alignments matched *F2L* with the *PR*-like region (*X1*) at the 5′ end of the *pol* gene of SRV-1 (7), the closely-related Mason–Pfizer monkey virus (14), and human endogenous retrovirus (15). The other two alignments corresponded to an internal part of *pol* (*X2*) between the ribonuclease H- and endonuclease/integrase-encoding regions of two lentiviruses, equine infectious anemia virus (EIAV; ref. 16) and visna lentivirus (VLV; ref. 17). Fig. 2 shows the genomic locations of *PR*, *X1*, *X2*, and *F2L* sequences in SRV-1, EIAV, and vaccinia virus.

In Fig. 3 the deduced amino acid sequence of ORF F2L has been aligned with X1 and X2 amino acid sequences from four infectious retroviruses: two duplicated sequences from SRV-1 and MMTV (18), and two duplicated/translocated sequences from EIAV and VLV. (The endogenous proviral sequences, hamster intracisternal A-type particle, and human endogenous retrovirus have been omitted from this analysis.) A single gap in the vaccinia, EIAV, and VLV protein sequences and two gaps in the SRV-1 and MMTV protein sequences were required to obtain an optimal alignment. The termini of the retroviral protease-like sequences are as in ref. 6. Twenty residues, indicated by asterisks, are identical in all five viral proteins.

To determine whether ORF *F2L* was more closely related to the duplicated or duplicated/translocated *PR*-like sequences, we carried out pairwise comparisons of the corresponding five amino acid sequences shown in Fig. 3. We arbitrarily began the comparison at the codon in *F2L* that specifies Arg-21, since this is the point where interviral homology is first evident at the amino acid level, and we terminated the analysis with the last codon before the TAA which completes *F2L*. The scoring matrices, shown in Table 1, display the number of identities between pairs. As expected, the SRV-1/MMTV and EIAV/VLV pairs were most similar. Assessed by either amino acid or nucleotide (not shown) identities, the vaccinia sequence displayed a marginally higher degree of similarity to the lentivirus sequences than to SRV-1 and MMTV. Additionally, F2L was colinear with both EIAV and VLV X2, whereas gaps were required in the alignments with either SRV-1 or MMTV X1. These results suggest that *F2L* was derived from the same sequence from which the *X2* sequences of EIAV and VLV were derived. However, the number of amino acid identities between F2L and either SRV-1 or MMTV protease-like sequences, 39, was not significantly different from the number of amino acids shared by either of the lentiviruses and SRV-1 or MMTV, 34–38, leading us to tentatively conclude that acquisition of this segment by vaccinia virus occurred at about the same time as did the transfer of the *PR*-like segment to the lentiviruses. This assumes, of course, that rates of sequence divergence in retroviruses and poxviruses are similar.

Extent of Retroviral Sequences Present in Vaccinia Virus. As is typical of vaccinia virus genome organization, the ORFs in the left end of *Hin*dIII F fragment are tightly packed and tandemly oriented. Thirty-five and 14 nucleotides of noncoding sequence separate the flanking ORFs from *F2L* (manuscript in preparation). To determine whether these genes might be part of the translocated DNA, we compared their deduced amino acid sequences to the National Biomedical Research Foundation‡ and GenBank§ data bases, but no evidence for homology to retroviral proteins was found. We then compared the vaccinia virus DNA sequence to each of the four retrovirus sequences upstream and downstream of the homologous regions. Whereas vaccinia *F2L* and retroviral *X1* or *X2* regions were 44–49% identical, sequence identity fell to 25–30% immediately upstream of the codon encoding Arg-21 and downstream of the codon encoding Gly-142. Although differential selection pressures might favor conservation of some acquired sequences and divergence of others, we conclude that the translocated region consisted of approximately 360–370 base pairs (bp), although we cannot exclude the possibility that an additional 5′ sequence was involved, since the 5′ ends of the retroviral *PR*-like sequences exhibit considerable divergence.

```
1     GTA CAA TCA TCA CAC TTA TTC ATG GAA TAT ATG GGA TGG TAA
43    ATA ATT TTG AAA TAA AAT ATT AGT TTT ATG TTC AAC ATG AAT
                                              MET Asn
85    ATT AAC TCA CCA GTT AGA TTT GTT AAG GAA ACT AAC AGA GCT
      Ile Asn Ser Pro Val Arg Phe Val Lys Glu Thr Asn Arg Ala
127   AAA TCT CCT ACT AGG CAA TCA CCT TAC GCC GCC GGA TAT GAT
      Lys Ser Pro Thr Arg Gln Ser Pro Tyr Ala Ala Gly Tyr Asp
169   TTA TAT AGC GCT TAC GAT TAT ACT ATC CCT CCA GGA GAA CGA
      Leu Tyr Ser Ala Tyr Asp Tyr Thr Ile Pro Pro Gly Glu Arg
211   CAG TTA ATT AAG ACA GAT ATT AGT ATG TCC ATG CCT AAG TTC
      Gln Leu Ile Lys Thr Asp Ile Ser MET Ser MET Pro Lys Phe
253   TGC TAT GGT AGA ATA GCT CCT AGG TCT GGT CTG TCC CTA AAA
      Cys Tyr Gly Arg Ile Ala Pro Arg Ser Gly Leu Ser Leu Lys
295   GGC ATT GAT ATA GGA GGC GGT GTA ATA GAC GAA GAT TAT AGG
      Gly Ile Asp Ile Gly Gly Gly Val Ile Asp Glu Asp Tyr Arg
337   GGA AAC ATA GGA GTC ATT CTT ATT AAT AAT GGA AAA TGT ACG
      Gly Asn Ile Gly Val Ile Leu Ile Asn Asn Gly Lys Cys Thr
379   TTT AAT GTA AAT ACT GGA GAT AGA ATA GCT CAG CTA ATC TAT
      Phe Asn Val Asn Thr Gly Asp Arg Ile Ala Gln Leu Ile Tyr
421   CAA CGT ATA TAT TAT CCA GAA CTG GAA GAA GTA CAA TCT CTA
      Gln Arg Ile Tyr Tyr Pro Glu Leu Glu Glu Val Gln Ser Leu
463   GAT AGT ACA AAT AGA GGA GAT CAA GGG TTT GGA TCA ACA GGA
      Asp Ser Thr Asn Arg Gly Asp Gln Gly Phe Gly Ser Thr Gly
505   CTT AGA TAA TAA ACA ATA GTA TGT TGT CGA TGT TTA TGT GTA
      Leu Arg TER
547   ATA ATA TCG TAG ATT ATG TAG ATG ATA TAG ATA ATG GTA TAG
```

FIG. 1. Nucleotide sequence of vaccinia virus ORF *F2L* and the deduced amino acid sequence. The RNA start sites as determined by S1 nuclease mapping and primer extension are marked by arrowheads. The position of the oligonucleotide used in the primer extension reaction is overlined.

Expression of *F2L* During Vaccinia Virus Infection. In contrast to its retroviral progenitor, which would be expressed as part of a polyprotein whether in the duplicated or duplicated/translocated position, the vaccinia *F2L* gene encodes translational initiation and termination signals. To determine whether *F2L* is transcribed during a viral infection, a primer extension analysis was done. A 5′-end-labeled oligonucleotide was hybridized to immediate early RNA and extended with reverse transcriptase. The primer extension product was coelectrophoresed with a dideoxy sequencing reaction primed from the same oligonucleotide (Fig. 4*A*). The results mapped a transcriptional start site within a methionine codon that is in-frame of ORF *F2L* and nine nucleotides upstream of a second ATG, thus establishing the second methionine codon as the putative translational start site (Fig. 1). In contrast to the ATG to which the transcriptional start site mapped, the second ATG is situated in a favorable environment for translation initiation, with purines at the −3 and +4 positions (19).

To define the temporal gene class of *F2L*, an S1 nuclease protection study was carried out with RNA isolated at various times after infection. A 5′ single end-labeled *Xba* I–*Nde* I probe (Fig. 4*C*) was hybridized to viral RNA isolated from cells infected with vaccinia virus in the presence of

‡Protein Identification Resource (1988) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 16.0.
§EMBL/GenBank Genetic Sequence Database (1988) (Intelligenetics, Mountain View, CA), Tape Release 55.

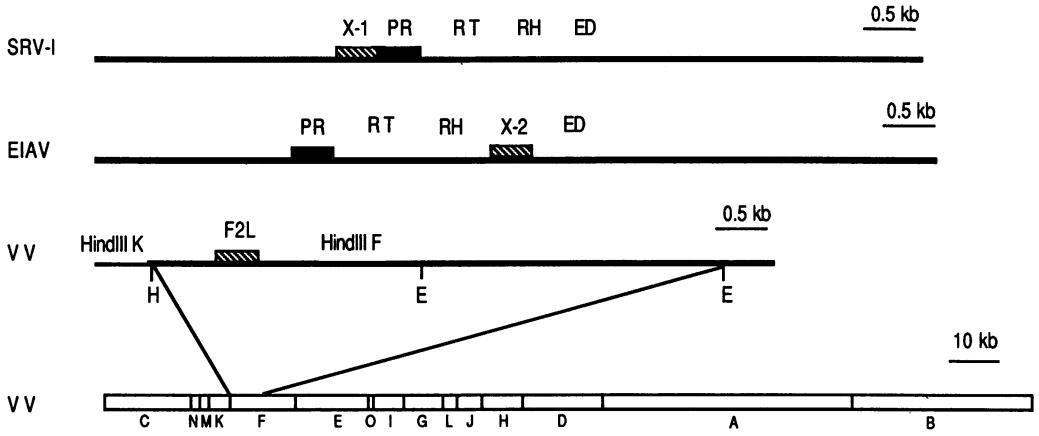

FIG. 2. Position of the authentic retroviral protease gene (*PR*) and the *PR*-like sequence in SRV-1 (*X1*), EIAV (*X2*), and vaccinia virus (VV) (*F2L*). The approximate positions of genes encoding reverse transcriptase (RT), ribonuclease H (RH), and endonuclease/integrate (ED) in the *pol* region are indicated for SRV-1 and EIAV.

cycloheximide or at the indicated times after infection in the absence of drugs. After S1 nuclease digestion, electrophoretic analysis (Fig. 4*B*) revealed a protected fragment that migrated with the 396-nucleotide marker. We consider this result to be consistent with the primer extension analysis, which predicted fragments of 392–393 nucleotides protected from S1 nuclease cleavage. *F2L* was transcribed in the presence of cycloheximide (lane C), and the time course showed that the transcript was most abundant 1 and 3 hr after infection, thus defining *F2L* as a member of the early gene class, which is transcribed prior to DNA replication.

Function of *F2L*. Whether the *F2L* sequence presently encodes a proteolytic activity remains to be determined. In cellular acid proteases and authentic retroviral proteases, an Asp-(Thr or Ser)-Gly motif is absolutely conserved (20), and it has been shown that these residues are present in the active site of pepsin (21). Site-specific mutagenesis of the invariant aspartic residue in the avian sarcoma–leukosis virus protease and human immunodeficiency virus type 1 protease abolishes cleavage of precursor polyproteins (22–24) and infectivity (25). The Asp-Thr-Gly motif has not been completely retained in products of *F2L*, *X1*, or *X2*, although its remnant, amino acids 26–28 of Fig. 3, can be identified (6). It seems unlikely, therefore, that these sequences presently encode a protease. However, seven glycine residues are conserved in the homologous protein sequences, and these residues correspond in relative position to conserved glycine residues in authentic retroviral and cellular proteases (6, 20), suggesting that proteins encoded by *F2L*, *X1* and *X2* might retain the overall three-dimensional structure of the acid proteases.

Mechanism of Translocation. How the *PR*-like gene segment was transferred among viruses is open to speculation. Examination of the DNA sequences flanking both the *X1* and *X2* gene segments in several retroviruses led to the suggestion that the *PR* gene duplication event fortuitously created a pair of excision signals, which promoted illegitimate transposition to a coinfecting lentivirus (6). The presence of sequences homologous to *X1* and *X2* in vaccinia virus lends support to the idea that the duplicated *PR*-like segment may be susceptible to some sort of transposition mechanism. It has been demonstrated recently that retroviral and *Ty1* element cDNA intermediates contained in cytoplasmic extracts are correctly inserted *in vitro* into linear phage λ (26, 27) or φX174 replicative form DNA (28). Therefore, a cytoplasmic double-stranded DNA viral genome (e.g., vaccinia virus) might conceivably be a target for an integrase-mediated reaction during the course of a dual infection.

We have examined the vaccinia virus DNA sequence in the vicinity of *F2L* for evidence of a retrotransposon-type integration event—specifically, the presence of direct repeats flanking short inverted repeats. Although no significant inverted repeats were found by a computer search that examined the DNA sequence 200 bp upstream and downstream of

```
          1        10        20        30        40        50        60
Vaccinia  MNINSPVRFVKETNRAKSPTRQSPYAAGYDLYSAYDYTIFPGERQL IKTDISMSMPKGCYGRIAPRS
VLV             SEIFLAKEGRGILQKRAEDAGYDLICPQEISIPAGQVKR IAIDLKINLKKDQWAMIGTKS
EIAV            EEIMLAYQGTQIKEKRDEDAGFDLCVPYDIMIPVSDTKI IPTDVKIQVPPNSFGWVTGKS
SRV-1     SKWGGQLCSSQQKQPISKLTRATPGSAGLDLSSTSHTVLTPEMGPQALSTGIYGPLPPNTFGLILGRS
MMTV      GVKGSGLNPEAPPFTIHDLPRGTPGSAGLDLSSQKDLILSLEDGVSLVPTLVKGTLPEGTTGLIIGRS
                                    ** **                                  *

 70        80        90        100       110       120       130       140
GLSLKGIDIGGGVIDEDYRGNIGVILINNGKCTFNVNTGDRIAQLIYQRIYYPELEEVQSLDSTNRGDQGFGSTGLR
SFANKGVFVQGGIIDSGYQGTIQVVIYNSNNKEVVIPQGRKFAQLILMPLIHEELEPWGETRKTERGEQGFGSTGMY
SMAKQGLLINGGIIDEGYTGEIQVICTNIGKSNIKLIEGQKFAQLIILQHHSNSRQPWDENKISQRGDKGFGSTGVF
SITIKGLQVYPGVIDNDYTGEIKI MAKAVNNIVTVPQGNRIAQLILLPLIETDNKVQQP    YRGQGSFGSSDIY
SNYKKGLEVLPGVIDSDFQGEIKV MVKAAKNAVIIHKGERIAQLLLLPYLKLPNPVIKE    ERGSEGFGSTSHV
     *     * **    * *               *   ***                    **   ***
```

FIG. 3. Alignment of deduced amino acid sequence (single-letter code) of F2L and those of four retroviral protease-like sequences. Amino and carboxyl termini of the retroviral sequences are as in ref. 6. Asterisks identify amino acids conserved in all sequences.

Table 1. Numbers of amino acid identities (% identities in parentheses) between pairs of protease-like sequences

	VLV	EIAV	SRV-1	MMTV
Vaccinia	41 (33)	42 (34)	39 (31)	39 (31)
VLV		52 (42)	36 (29)	37 (30)
EIAV			34 (27)	38 (31)
SRV-1				62 (52)

Pairs of sequences were compared between Arg-21 and Arg-144 (vaccinia virus *F2L* numbering), with gaps introduced as in Fig. 3.

ORF *F2L*, a nearly perfect 9-bp direct repeat at the 5′ and 3′ extremities of *F2L* was identified:

TCAACATGA.....TCAACAGGATTAGATAA.

(Boldface nucleotides represent the putative initiation and termination codons of *F2L*.) However, the significance of this finding is uncertain, since the 3′ repeat encodes highly conserved amino acids at the carboxyl terminus of the sequence and, therefore, is likely to have been part of the translocated DNA segment.

CONCLUSION

We have sequenced and analyzed a gene within the vaccinia virus genome, ORF *F2L*, which is related to a retroviral *PR*-like gene segment. This sequence is notable because it appears to have relocated into at least two distinct sites: a distantly related retrovirus and a poxvirus. An RNA species containing transcribed *F2L* sequences was detected early in a vaccinia infection, indicating that the gene is expressed from the poxvirus genome.

These findings raise several intriguing questions. (*i*) Do the *F2L*, *X1*, or *X2* sequences encode functional protease activities? (*ii*) If so, are the substrates cellular or viral proteins? (*iii*) If not, what are the present functions of these proteins and how do they relate to the viral life cycles? (*iv*) By what mechanisms have these sequences relocated into new genetic environments?

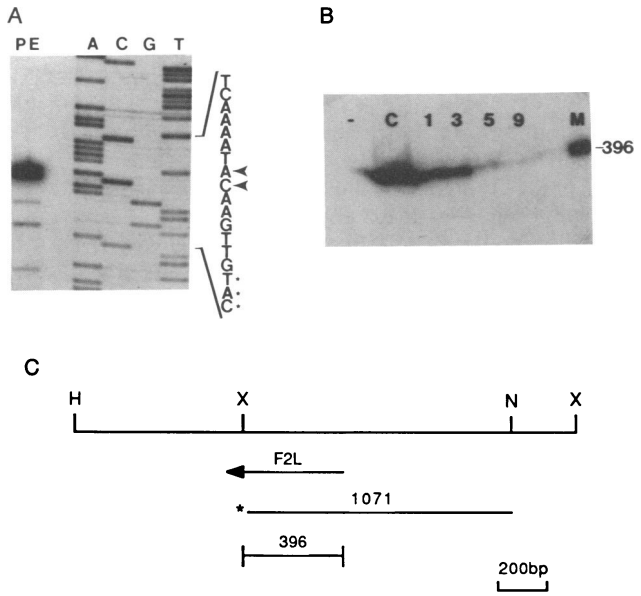

FIG. 4. Fine mapping and kinetics of transcription of *F2L*. (*A*) Primer extension mapping of the mRNA start site: Lanes: PE, primer extension product; A, C, G, and T, dideoxynucleotide sequencing reactions. The position of the oligonucleotide primer used is indicated in Fig. 1. Arrowheads indicate the mRNA start sites in context of the sequence. Dots indicate the putative translational start site. (*B*) A 5′ single end-labeled probe was hybridized to 20 μg of viral RNA isolated in the presence of cycloheximide (lane C) or at 1, 3, 5, and 9 hr after infection in the absence of drug. The first lane contains tRNA as a control, and end-labeled 1-kilobase markers (BRL) were run in lane M. (*C*) Restriction map of the left-hand end of *Hin*dIII F fragment, with the position and orientation of ORF *F2L* indicated. The *Xba* I–*Nde* I S1 nuclease probe is shown along with the fragment protected by the probe. X, *Xba* I; H, *Hin*dIII; N, *Nde* I.

We thank Christopher Mathews for his continuing interest in this project and for providing laboratory space in which this work was done, Jack Kramer for assistance with computer analyses, and R. F. Doolittle for helpful discussions. This work was supported by National Institutes of Health Grant AI24594.

1. Wakabayashi, S., Matsubara, H. & Webster, D. A. (1986) *Nature (London)* **322,** 481–483.
2. Carlson, T. A. & Chelm, B. K. (1986) *Nature (London)* **322,** 568–570.
3. Upton, C. & McFadden, G. (1986) *J. Virol.* **60,** 920–927.
4. Slabaugh, M., Roseman, N., Davis, R. & Matthews, C. (1988) *J. Virol.* **62,** 519–527.
5. Tengelsen, L. A., Slabaugh, M. B., Bibler, J. K. & Hruby, D. E. (1988) *Virology* **164,** 121–131.
6. McClure, M. A., Johnson, M. S. & Doolittle, R. F. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 2693–2697.
7. Power, M. D., Marx, P. A., Bryant, M. L., Gardner, M. B., Barr, P. J. & Luciw, P. A. (1986) *Science* **231,** 1567–1572.
8. McClure, M. A., Johnson, M. S., Feng, D.-F. & Doolittle, R. F. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2469–2473.
9. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
10. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4767–4771.
11. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.
12. Weinrich, S. L., Niles, E. G. & Hruby, D. E. (1985) *J. Virol.* **55,** 450–457.
13. Roseman, N. A. & Hruby, D. E. (1987) *J. Virol.* **61,** 1398–1406.
14. Sonigo, P., Barker, C., Hunter, E. & Wain-Hobson, S. (1986) *Cell* **45,** 375–385.
15. Ono, M., Yasunaga, T., Miyata, T. & Ushikubo, H. (1986) *J. Virol.* **60,** 589–598.
16. Kawakami, T., Sherman, L., Dahlberg, J., Gazit, A., Yaniv, A., Tronick, S. P. & Aaronson, S. A. (1987) *Virology* **158,** 300–312.
17. Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Haase, A. & Wain-Hobson, S. (1985) *Cell* **42,** 369–382.
18. Moore, R., Dixon, M., Smith, R., Peters, G. & Dickson, C. (1987) *J. Virol.* **61,** 480–490.
19. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
20. Pearl, L. H. & Taylor, W. R. (1987) *Nature (London)* **329,** 351–354.
21. James, M. N. G. (1977) *Nature (London)* **267,** 808–813.
22. Kotler, M. R., Katz, A. & Shalka, A. M. (1988) *J. Virol.* **62,** 2696–2700.
23. Mous, J., Heimer, E. P. & LeGrice, S. T. J. (1988) *J. Virol.* **62,** 1433–1436.
24. Nam, S. H., Kidokoro, M., Shida, H. & Hatanaka, M. (1988) *J. Virol.* **62,** 3718–3728.
25. Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A. F., Scolnick, E. M. & Sigal, I. S. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 4686–4690.
26. Brown, P. O., Bowerman, B., Varmus, H. E. & Bishop, J. M. (1987) *Cell* **49,** 347–356.
27. Eichinger, D. J. & Boeke, J. D. (1988) *Cell* **54,** 955–966.
28. Fujiwara, T. & Mizuuchi, K. (1988) *Cell* **54,** 497–504.