Vol. 136, No. 1, 1986 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 14, 1986 Pages 183-192

BIOSYNTHESIS OF REOVIRUS-SPECIFIED POLYPEPTIDES

Molecular cDNA Cloning and Nucleotide Sequence of the Reovirus Serotype 1 Lang Strain s4 mRNA which Encodes the Major Capsid Surface Polypeptide σ3

Jonathan A. Atwater[1], Susan M. Munemitsu[1] and Charles E. Samuel

Section of Biochemistry and Molecular Biology
Department of Biological Sciences
University of California, Santa Barbara, CA 93106

Received March 4, 1986

SUMMARY: Serotype 1 Lang strain s4 mRNA, which encodes the major capsid surface polypeptide σ3 of reovirions, was cloned as a cDNA:mRNA heteroduplex in Escherichia coli using phage M13. A complete consensus nucleotide sequence for s4 mRNA has been determined from cDNA clones. The Lang strain s4 mRNA is 1196 nucleotides in length and possesses an open reading frame with a coding capacity of 365 amino acids, sufficient to account for a σ3 polypeptide of 41,212 daltons. Comparison of the serotype 1 (Lang) s4 sequence with the serotype 3 (Dearing) S4 sequence reveals 94% homology at the nucleotide level; the predicted σ3 polypeptides of the Lang and Dearing strains display 96% homology at the amino acid level. Two third base C codons (leu:CUC and ser:AGC) are used about one-tenth as frequently in the reovirus s4 mRNAs as compared to mammalian cellular mRNAs.

The genome of reovirus consists of ten segments of double-stranded RNA and is surrounded by a double capsid shell of polypeptides (16). The outer capsid shell of reovirions is composed of three polypeptides, σ1, σ3 and μ1C (31), which are encoded by the S1, S4 and M2 genome segments, respectively (18,21,24). σ3 is one of the major surface polypeptides of the outer capsid shell (31).

In addition to the structural role that the S4-encoded σ3 polypeptide plays as a major component of the virion outer capsid shell, three functional roles have also been described for the product of the S4 gene. Biochemical studies have established that the σ3 polypeptide binds to dsRNA (13). Genetic studies have revealed that the S4 gene plays an important role in the establishment of persistent reovirus infections in mouse L-cells (1) and, in lytic infections of L-cells, is responsible for the inhibition of cellular protein synthesis (30).

[1] The order of the first two authors is arbitrary.

0006-291X/86 $1.50

The capacity of reovirus to inhibit cellular protein synthesis differs significantly between the serotypes of virus (23,30). Serotype 1 Lang strain does not inhibit cellular protein synthesis in mouse L cells even at late times after infection whereas cellular protein synthesis is significantly inhibited by serotype 3 Dearing strain at late times after infection (23,30,34).

As part of our study of the role of the S4 gene in the control of protein synthesis in reovirus-infected cells, we have utilized a simple, rapid method involving cDNA:mRNA heteroduplex insertion into bacteriophage M13 to obtain molecular clones of the reovirus Lang strain s4 mRNA. A complete consensus nucleotide sequence was determined for the Lang s4 mRNA and, when compared with the Dearing S4 dsRNA-derived sequence of Giantini et al. (11), revealed that the S4 genes of the serotype 1 (Lang) and 3 (Dearing) strains of reovirus are surprisingly similar to each other.

MATERIALS AND METHODS

Virus and mRNA. The Lang strain of reovirus serotype 1 was obtained from the American Type Culture Collection (ATCC VR-230). The virus was thrice-plaque purified before amplification in mouse L cells in suspension culture (23). Reovirus mRNA was transcribed *in vitro* from chymotrypsin-activated reovirions, purified and fractionated by centrifugation on a 15-30% sucrose density gradient (18).

Cloning Strategy. The s-class mRNA was polyadenylated with *E. coli* poly A polymerase (Bethesda Research Labs); an average of 20 to 30 residues was added (8). Single-stranded cDNA was synthesized essentially as described by Buell et al. (3) using avian myeloblastosis virus reverse transcriptase (Life Sciences, Inc.) with oligo dT_{12-18} (Collaborative Res., Inc.) as the primer. The reovirus mRNA:cDNA heteroduplex was then treated with mung bean nuclease (P.L. Biochemicals), and blunt end ligated using T4 DNA ligase (BRL)into the *Sma* 1 site of the replicative form of bacteriophage M13 mp9. Transformation was in *E. coli* JM103 (22).

Identification of Clones. Recombinant M13 phage derived from clear plaques indicating insertional inactivation of β-galactosidase enzymic activity were screened for insert size by agarose gel electrophoresis. Recombinant phage replicative form (RF) DNA containing "large" inserts was then further screened by Northern blot hybridization to Lang strain genome dsRNA segments fractionated on a 1.4% agarose gel, blotted onto 0.45 μ BA 85 nitrocellulose paper, and probed with nick-translated (27) M13 RF DNAs. Hybridization conditions were essentially as previously described (20). Candidate S4 clones identified from Northern blot hybridization signals were further screened for their ability to hybrid-select reovirus mRNA which encoded polypeptide σ_3 when translated *in vitro* as previously described (15).

DNA Sequence Analysis. The complete nucleotide sequence of four large s4 cDNA clones was determined from a family of overlapping deletion subclones prepared as described by Dale et al. (7) and sequenced by the dideoxy chain

termination method (29). Of the four s4 clones analyzed in detail, all lacked the 5' end of the S4 gene as deduced from the absence of the known conserved 5'-terminal tetranucleotide GCUA (9). Therefore, to determine a consensus sequence for the 5'-terminal region of the Lang strain s4 mRNA, a s4-specific antisense 15 base oligomer complementary to nucleotide positions 93 to 107 was synthesized using an Applied Biosystems model 380A DNA synthesizer, hybridized to reovirus Lang strain mRNA, and extended with AMV reverse transcriptase in the presence of dideoxynucleotides (6).

RESULTS

Nucleotide Sequence of the Lang S4 Gene. The strategy used to determine the complete nucleotide sequence of each of four independent s4 molecular clones is shown in Figure 1. The complete consensus nucleotide cDNA sequence of the Lang strain s4 mRNA is shown in Figure 2. The serotype 1 Lang strain s4 mRNA is 1196 nucleotides in length. The first AUG codon was at position 33 to 35; it initiated an open reading frame that ended with a UAA termination codon at positions 1128 to 1130. Analysis of the deduced Lang s4 mRNA sequence identified only one long open reading frame; the predicted encoded polypeptide was 365 amino acids long (Fig. 2). No additional open reading frames capable of yielding a polypeptide greater than 40 amino acids were revealed upon examination of all other possible reading frames in both the plus and the minus sense.

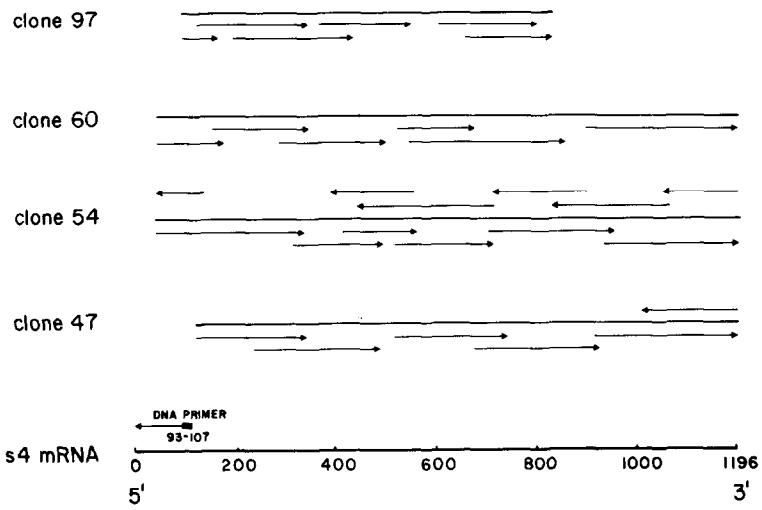

Figure 1. Strategy used to determine the sequence of the cDNA clones of the reovirus serotype 1 Lang S4 gene. A family of overlapping deletion subclones was generated from clones 47, 54, 60 and 97 of the Lang strain S4 gene and sequenced as described under "Materials and Methods". The arrows indicate the orientation and extent of the sequences determined with the subclones.

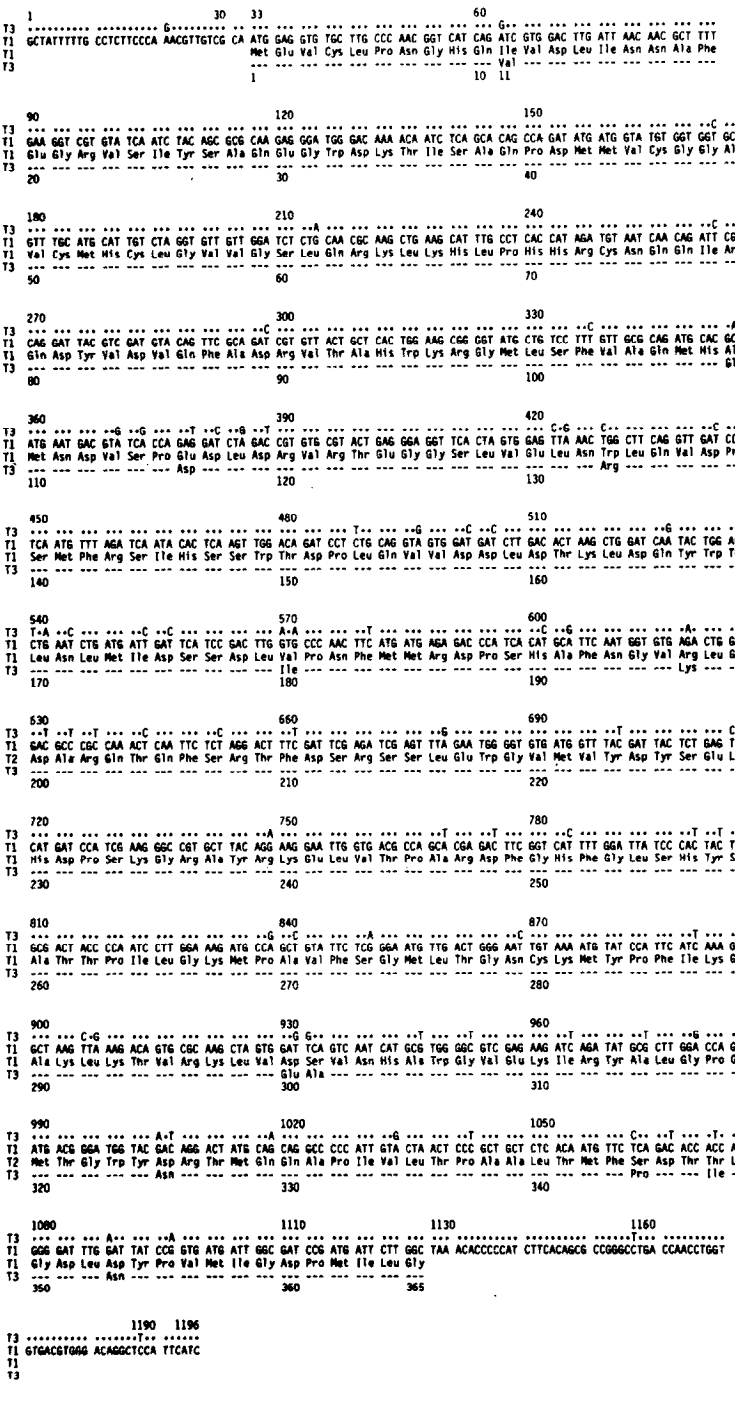

Figure 2. Consensus nucleotide sequence of the S4 gene of reovirus serotype 1 Lang strain (T1) and deduced amino acid sequence of the σ3 polypeptide. The nucleotide sequence is presented in the mRNA sense. Numbers above each line refer to nucleotide position and numbers beneath each line to amino acid position. For purposes of comparison with the reovirus serotype 1 Lang strain (T1) sequence, the serotype 3 Dearing strain (T3) S4 sequence of Giantini et al. (11) is also included. Identical nucleotides are indicated by dots, and identical amino acids by dashes. Both S4 genes are 1,196 nucleotides long and both predict a 365 amino acid σ3 polypeptide product.

Sequences derived from clones 47, 60 and 97 were identical and the consensus sequence is derived from them; clone 54 possessed a total of 7 nucleotide changes from the consensus sequence. Clones 54, 60 and 97 were independent isolates from the same cDNA reaction and transformation. Clone 47 was obtained from a separate cloning experiment in which the cDNA was synthesized using a different preparation of mRNA.

Relatedness of the Lang and Dearing S4 Genes. The overall sequence homology between the serotype 1 Lang strain s4 cDNA derived from mRNA and serotype 3 Dearing strain S4 cDNA derived from dsRNA is 94% at the nucleotide level. The S4 genes of the two serotypes are the same length (1196 nucleotides), and the positions of the presumed initiation AUG codon (nucleotides 33-35) and termination UAA codon (nucleotides 1128-1130) are conserved. The 5'- and 3'-untranslated sequences are slightly more homologous than the translated sequences, showing 97% homology at the nucleotide level; however, the predicted σ3 polypeptides display 96% homology at the amino acid level as compared to 93% homology at the nucleotide level in the translated region.

Codon Usage. The predicted codon usage in the synthesis of polypeptide σ3 from the s4 mRNA is summarized in Table I for the serotype 1 Lang strain. For comparative purposes, the predicted codon usage for the serotype 3 Dearing strain σ3 polypeptide was also analyzed. The codon use frequencies in s4 mRNA do not differ significantly between the two reovirus strains with the exception of two codons with UA.

Statistical analyses have revealed that the use of degenerate codons in the expression of mRNAs in both prokaryotes and eukaryotes is not random (12). Two points are apparent when the codon frequencies of the human reovirus Lang and Dearing s4 mRNAs are compared to those of mammalian genes. First, codons with CG (Ser:UCG, Thr:ACG, Ala:GCG, and Pro:CCG) are used about twice as frequently as a group in reovirus s4 mRNAs than in mammalian mRNAs. Second, 14 of the possible 16 codons with C as the third base are used less frequently in the reovirus s4 mRNAs than in mammalian mRNAs. Two of the third base C codons (Leu:CUC and Ser:AGC) are used very rarely in the reovirus s4 mRNAs, once out of

Table I

Codon Usage and Amino Acid Composition of Polypeptide Sigma 3 of Reovirus Serotype 1 (Lang strain)[a] and Serotype 3 (Dearing strain)[b] Predicted from S4 Gene Sequences

	Codon	Lang	Dearing		Codon	Lang	Dearing
Gly	GGA	10	9	Arg	CGA	1	1
	GGC	5	5		CGC	3	2
	GGG	2	3		CGG	1	2
	GGU	11	11		CGU	6	7
					AGA	6	6
Ala	GCA	4	3		AGG	4	3
	GCC	4	4				
	GCG	6	4	Lys	AAA	3	4
	GCU	7	10		AAG	12	12
Val	GUA	7	4	His	CAC	5	6
	GUC	4	5		CAU	9	8
	GUG	12	14				
	GUU	7	7	Ser	UCA	11	9
					UCC	3	4
Leu	CUA	5	5		UCG	4	4
	CUC	1	1		UCU	4	3
	CUG	8	9		AGC	1	1
	CUU	5	5		AGU	2	2
	UUA	5	2				
	UUG	7	9	Thr	ACA	4	5
					ACC	3	3
Ile	AUA	1	2		ACG	4	3
	AUC	6	6		ACU	9	8
	AUU	6	6				
				Asp	GAC	11	12
Cys	UGC	2	2		GAU	18	15
	UGU	4	4				
				Glu	GAA	3	3
Met	AUG	21	21		GAG	9	10
Pro	CCA	10	11	Asn	AAC	5	7
	CCC	4	3		AAU	7	7
	CCG	2	2				
	CCU	2	3	Gln	CAA	6	6
					CAG	10	10
Phe	UUC	10	8				
	UUU	4	6	Tyr	UAC	8	6
					UAU	3	5
Trp	UGG	8	7				

[a] The consensus sequence of the Lang strain S4 gene (Fig. 1) predicts a 365 amino acid sigma 3 polypeptide, molecular weight 41,212.

[b] The published sequence for the Dearing strain S4 gene (11) predicts a 365 amino acid polypeptide, molecular weight 41,165.

31 for leucine and once out of 25 or 23 for serine, equivalent to frequency of about 3 to 4 percent; by contrast, the CUC and AGC codons specify about 25 to 30 percent of the leucine and serine residues, respectively, in mammalian genes.

Structural Features of the Sigma 3 Polypeptide. The σ3 polypeptide predicted from the S4 gene sequence of the Lang strain of reovirus consists of

365 amino acid residues and has a molecular weight of 41,212 (Fig. 2, Table I). The Dearing strain σ3 polypeptide likewise consists of 365 amino acids and has a very similar molecular weight, 41,165 (11), very comparable to the value of 41,500 estimated by $NaDodSO_4$ polyacrylamide gel electrophoresis (28).

The Lang strain σ3 polypeptide contains six cysteine residues (Table I). The positions of the cysteine residues are conserved in the serotype 1 (Fig. 2) and the serotype 3 Dearing (11) σ3 polypeptides. The Lang strain σ3 polypeptide does not contain any potential N-glycosylation sites of the type Asn-X-Ser/Thr (Fig. 2). By contrast, the Dearing strain σ3 contains a single possible N-glycoslylation site, Asn-Arg-Thr, at residues 325-327 (11). Comparison of the hydropathicity plots of the σ3 polypeptides obtained by the method of Kyte and Doolittle (17) indicates that the Lang and the Dearing strain σ3 polypeptides

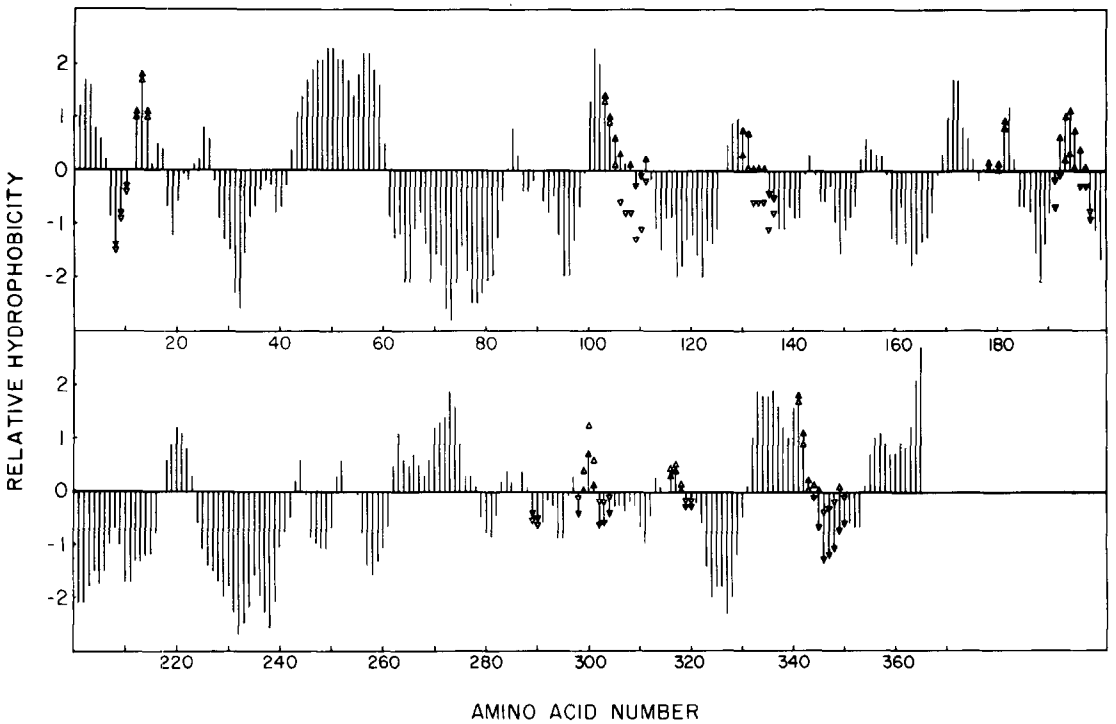

Figure 3. Hydropathicity profile of polypeptide σ3 of the serotype 1 Lang strain compared to σ3 of the serotype 3 Dearing strain. The program of Kyte and Doolittle (17) was used with a search length of six amino acids. The degree of hydrophobicity increases with distance above the the horizontal line, and hydrophilicity increases with the distance below the horizontal line. Lines without symbols indicate identity between the two serotypes; where differences between the two serotypes occur, solid triangles indicate serotype 1 and open triangles indicate serotype 3.

are very similar in terms of their predicted hydrophobic-hydrophilic domains (Fig. 3).

DISCUSSION

The results presented here demonstrate three important points: first, the S4 genes and the encoded σ3 polypeptides of the Lang (serotype 1) and Dearing (serotype 3) strains of the human reoviruses are surprisingly more closely related at the structural level than anticipated from previous hybridization studies; second, the in vitro s4 mRNA transcript of the S4 genome segment is likely identical with the plus strand of the S4 genome dsRNA; and third, molecular cloning of heteroduplex cDNA:mRNA structures with the bacteriophage M13 vector provides a rapid and efficient method for cDNA cloning, but as in all cDNA cloning of animal virus RNAs, it is important to establish a consensus sequence by analysis of multiple independent clones.

Our results for the Lang strain together with those of Giantini et al. (11) for the Dearing strain strongly suggest that the S4 genes of serotypes 1 (Lang) and 3 (Dearing) have not significantly diverged during evolution. This is somewhat surprising in view of the fact that the differential ability of reovirus serotypes to inhibit cellular protein synthesis has been mapped to the S4 gene (30). The Lang s4 sequence and the Dearing S4 sequence possess 94% homology at the nucleotide level. The predicted σ3 polypeptide encoded by the two strains is 96% homologous at the amino acid level. Of the 74 nucleotide differences in the coding region, 60 of these did not cause a change in amino acid; 52 of the 60 silent differences were in the third base position of codons. There is very little sequence divergence in the 5'- and 3'-terminal noncoding regions of the S4 genes of the Lang and Dearing strains, with only 3 differences observed out of 101 untranslated nucleotides. By contrast to the sequence data, previous hybridization studies in which the relatedness of the Lang and Dearing S4 genes was assessed by RNase sensitivity of +RNA;-RNA hybrids containing the plus strand of one serotype and the minus strand of the other serotype suggested a relatedness of about 50 to 60% of the Lang and Dearing S4 genes (10).

All of the reovirus gene sequences published to date have been derived from cDNA synthesized using denatured genome double-stranded RNA segments as the template (2,4,5,11,25,26). In this paper, we present the first complete reovirus gene sequence derived from cDNA synthesized using virion-associated polymerase transcribed mRNA as the template. Comparison of the serotype 1 Lang s4 mRNA sequence presented here with the serotype 3 Dearing S4 dsRNA plus strand sequence (11) provide strong evidence for the complete conservation of genetic information during the transcription of mRNA from genome in vitro. The ability to translate denatured genome dsRNA in vitro to yield polypeptide products similar to those obtained in vivo, both for the apparently monocistronic genes (21) and the dicistronic S1 gene (14), together with the identical terminal sequence obtained for the S2 dsRNA and the s2 in vitro transcript (19), previously indicated that the sequence of the genome dsRNA plus strand and the corresponding mRNA transcript were likely identical.

Our cloning strategy demonstrates the utility of using the first strand cDNA:mRNA heteroduplex in molecular cloning with such vectors as bacteriophage M13. This strategy has the advantages of being rapid, and also of circumventing the problems frequently associated with second strand cDNA synthesis and subsequent dideoxy sequence analysis of molecular clones containing homopolymer sequences. However, the strategy does have the potential disadvantages of involving a blunt end ligation. To our knowledge, the only other reported examples of molecular cloning of heteroduplex structures with prokaryotic vectors involved the cloning of a poly A-tailed cDNA:mRNA heteroduplexes into either poly T-tailed $ColE_1$ (32) or pBR322 (33) vectors.

ACKNOWLEDGEMENT

This work was supported in part by Research Grants AI-12520 and AI-20611 from the National Institutes of Health.

REFERENCES

1. Ahmed, R., and Fields, B.N. (1982). Cell 28, 605-612.
2. Bassel-Duby, R., Jayasuriya, A., Chatterjee, D., Sonenberg, N., Maizel, J.V., and Fields, B.N. (1985). Nature (London) 315, 421-423.
3. Buell, G.N., Wickens, M.P., Payvar, F.H., and Schimke, R.T. (1978). J. Biol. Chem. 253, 2471-2482.

4. Cashdollar, L. W., Esparza, J., Hudson, G.R., Chmelo, R., Lee, P.W.K., and Joklik, W.K. (1982). Proc. Natl. Acad. Sci. U.S.A. 79, 7644-7648.
5. Cashdollar, L.W., Chmelo, R. A., Wiener, J.R., and Joklik, W.K. (1985). Proc. Natl. Acad. Sci. U.S.A. 82, 24-28.
6. Collins, P.L., Huang, Y.T., and Wertz, G.W. (1984). Proc. Natl. Acad. Sci. U.S.A. 81, 7683-7687.
7. Dale, R.M.K., McClure, B.A., and Houchins, J.P. (1985). Plasmid 13, 31-40.
8. Devos, R., Gillis, E., and Fiers, W. (1976). Eur. J. Biochem. 62, 401-410.
9. Gaillard, R.K., Li, J.K.-K., Keene, J.D., and Joklik, W.K. (1982). Virology 121, 320-326.
10. Gaillard, R.K., and Joklik, W.K. (1982). Virology 123, 152-164.
11. Giantini, M., Seliger, L.S., Furuichi, Y., and Shatkin, A.J. (1984). J. Virol. 52, 984-987.
12. Grantham, R. Gautier, C., Gouy, M., Jacobzone, M., and Mercier, R. (1981). Nucleic Acids Res. 9, r43-r74.
13. Huismans, H., and Joklik, W.K. (1976). Virology 70, 411-424.
14. Jacobs, B.L., and Samuel, C.E. (1985). Virology 143, 63-74.
15. Jacobs, B.L., Atwater, J.A., Munemitsu, S.M., and Samuel, C.E. (1985). Virology 147, 9-18.
16. Joklik, W.K. (1984). "The Reoviridae." Plenum, New York.
17. Kyte, J., and Doolittle, R.F. (1982). J. Mol. Biol. 157, 105-132.
18. Levin, K.H., and Samuel, C.E. (1980). Virology 106,1-13.
19. Li, J.K.-K., Scheible, P.P., Keene, J.D., and Joklik, W.K. (1980). Virology 105, 282-286.
20. Masters, P.S., and Samuel, C.E. (1983). J. Biol. Chem. 258, 12026-12033.
21. McCrae, M.A., and Joklik, W.K. (1978). Virology 89, 578-593.
22. Messing, J., Crea, R., and Seeburg, P.H. (1981). Nucleic Acids Res. 9, 309-321.
23. Munemitsu, S.M., and Samuel, C.E. (1984). Virology 136, 133-143.
24. Mustoe, T.A., Ramig, R.F., Sharpe, A.H., and Fields, B.N. (1978). Virology 89, 594-604.
25. Nagata, L., Masri, S.A., Mah, D.C.W., and Lee, P.W.K. (1984). Nucleic Acids Res. 12, 8699-8710.
26. Richardson, M.A., and Furuichi, Y. (1983). Nucleic Acids Res. 11, 6399-6408.
27. Rigby, P.W., Dieckmann, M., Rhodes, C., and Berg, P. (1977). J. Mol. Biol. 113, 237-251.
28. Samuel, C.E. (1983). In R.H. Compans and D.H.L. Bishop (ed.), Double-stranded RNA viruses. Elsevier Sci. Publ., New York, pp. 219-230.
29. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
30. Sharpe, A.H., and Fields, B.N. (1982). Virology 122, 381-391.
31. Smith, R.E., Zweerink, H.J., and Joklik, W.K. (1969). Virology 39, 791-810.
32. Wood, K.O., and Lee, J.C. (1976). Nucleic Acids Res. 3, 1961-1971.
33. Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and Dunn, A.R. (1979). Cell 16, 851-861.
34. Zweerink, H.J., and Joklik, W.K. (1970). Virology 41, 501-518.