Volume 11 Number 18 1983 Nucleic Acids Research

Nucleotide sequence of reovirus genome segment S3, encoding non-structural protein sigma NS

Michael A.Richardson and Yasuhiro Furuichi

Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ 07110, USA

Received 16 June 1983; Revised and Accepted 22 August 1983

ABSTRACT

This report describes the complete nucleotide sequence of human reovirus (Dearing strain) genome segment S3. Previous studies indicated that this RNA encodes the major non-structural viral polypeptide sigma NS, a protein that binds ssRNAs (Huisman & Joklik, Virology 70, 411-424, 1976) and has a poly(C)-dependent poly(G) polymerase activity (Gomatos et al., J. Virol. 39, 115-124, 1981). The genome segment consists of 1,198 nucleotides and possesses an open reading frame that extends 366 codons from the first AUG triplet (residues 28-30). There is no significant sequence homology between the plus strand of genome segment S3 and that of genome segment S2 determined previously (Cashdollar et al., PNAS 79, 7644-7648, 1982). However, S3 RNA has significant dyad symmetry and regions that can potentially hybridize (ΔG = -26 KCal/mole) with S2 RNA. From the predicted amino acid sequence a possible secondary structure for sigma NS protein was determined. Structural features of reovirus RNA and sigma NS are discussed in relation to their role(s) in viral genome assembly.

INTRODUCTION

The reovirus genome consists of 10 segments of double-stranded (ds)RNA. The segments are transcribed by virus-associated RNA polymerase to form capped mRNAs (1) which also function as templates for a putative replicase in virus-infected cells (2,3).

In infected cells, reovirus morphogenesis commences 4 hr post infection (2). In the early stages of morphogenesis, a fraction of the mRNAs is enclosed in immature virus particles in an RNase-sensitive form (2,3). Within these particles the plus strands are transcribed into minus strands, forming progeny dsRNA molecules. The nature of the mechanism which ensures that each progeny virus particle receives one copy of each genome RNA segment is not known. This is one of the principle themes in reovirus research, and perhaps of research on segmented genome systems in general. Each viral transcript probably contains recognition sites for primary genome transcription, mRNA translation and dsRNA replication as well as specific signals to ensure the correct assembly of virus

particles.

Previously, Huisman and Joklik (4) found that a viral protein, absent from mature reovirus particles, formed complexes with all reovirus mRNAs found in infected cells. This non-structural protein, sigma NS, bound single-stranded (ss)RNAs but not dsRNAs (4,5). Recently, Gomatos et al. (6,7) found that in infected cells sigma NS formed small spherical or triangular reovirus-specific particles which sedimented at 13-19S. These particles were comprised solely of sigma NS and, in addition to binding reovirus mRNAs, exhibited a unique poly(C)-dependent poly(G) polymerase activity (7). These observations suggest that sigma NS may act during infection as a condensing factor, bringing together the 10 reovirus ssRNA templates in preparation for dsRNA synthesis. Consistent with this possibility, a temperature sensitive mutant (Group E) known to be defective in sigma NS protein was unable to synthesize dsRNA at the restrictive temperature (8,9).

We have recently developed a general method for preparing cDNAs from a mixture of dsRNAs (10). Reovirus cDNAs were cloned in pBR322, and recombinant plasmids which contained sequences representing the 10 genome segments were identified. In this report, we present the complete sequence of the S3 gene which codes for sigma NS protein. Possible mechanism(s) involved in the assembly of segmented genomes are discussed in relation to the structural features of sigma NS protein and reovirus RNAs.

MATERIALS AND METHODS

Cloning of reovirus gene segment S3 has been described (10). Purification of the recombinant plasmid DNA was carried out as described by Birnboim et al. (11), and the nucleotide sequence of cloned S3 DNA was determined by the Maxam-Gilbert sequencing method (12). All ^{32}P radioactive nucleoside triphosphates used for end-labeling of DNA fragments were purchased from Amersham. Restriction endonucleases were from either Bethesda Research Laboratories or New England Biolaboratories. Polynucleotide kinase, calf intestine alkaline phosphatase and DNA polymerase (Klenow fragment) were from Boehringer Mannheim. Chemicals used in the Maxam-Gilbert sequence reactions were from Aldrich (dimethylsulfoxide and hydrazine) and Fisher (piperidine and formic acid). Elu-Tip columns used for purifying DNA fragments were obtained from Schreier and Schull.

RESULTS

Nucleotide Sequence of Reovirus S3 DNA

The recombinant DNA clone (S3-1485) of RNA genome segment S3 was digested

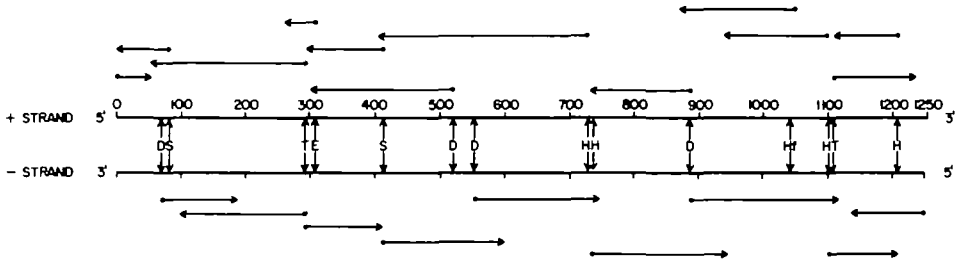

Fig. 1. Restriction endonuclease sites in cloned S3 DNA and a summary of the strategy used in determining its sequence. D, S, T, E, H, and Hf represent positions of Dde1, Sau3AI, Taq1, EcoRV, Hha1, and Hinf1 cleavage sites, respectively. The arrows indicate the extent and direction of sequencing.

with PstI restriction endonuclease. The S3-1485 DNA was separated from pBR322 DNA by agarose gel electrophoresis, eluted from the gel and further purified by an Elu-Tip column. Dot-blot hybridization and polyacrylamide gel electrophoresis indicated that the cloned cDNA was specific for the S3 gene segment and probably represented a full-length copy (results not shown).

The nucleotide sequence of the cloned S3 DNA was determined by the Maxam-Gilbert method (12). Figure 1 shows a restriction endonuclease map of S3 DNA with a summary of the strategy used. The complete sequence is shown in Fig. 2. Sequence determination was performed on both strands, except for the 3'-terminal 40 bases of the minus strand. The 5' and 3' terminal sequences of the cloned S3 DNA were in total agreement with those reported for S3 genome RNA by Antczak et al. (13). These results confirmed that the cloned S3 DNA contained a full length copy of the S3 gene.

Genome segment S3 consists of 1,198 base pairs. It possesses one long open reading frame that starts with the first AUG (at residues 28-30) and extends 366 codons. The other two reading frames contain multiple termination codons, with the longest reading frame extending only 43 codons (residues 668-796). This 5'-proximal AUG appears to be an efficient initiator since it is surrounded by a purine at position -3 and a guanine at position +4 (14). In fact, Kozak and Shatkin have previously shown in ribonuclease protection studies that this region forms part of a protein synthesis initiation complex (15). The second and third AUGs (positions 106 and 193) are relatively weak initiators according to the Kozak rule (14), since both contain C in the +4 position. A single termination codon occurs at position 1,128 in contrast to four in-phase termination codons present in the reovirus S2 gene (16) which encodes the

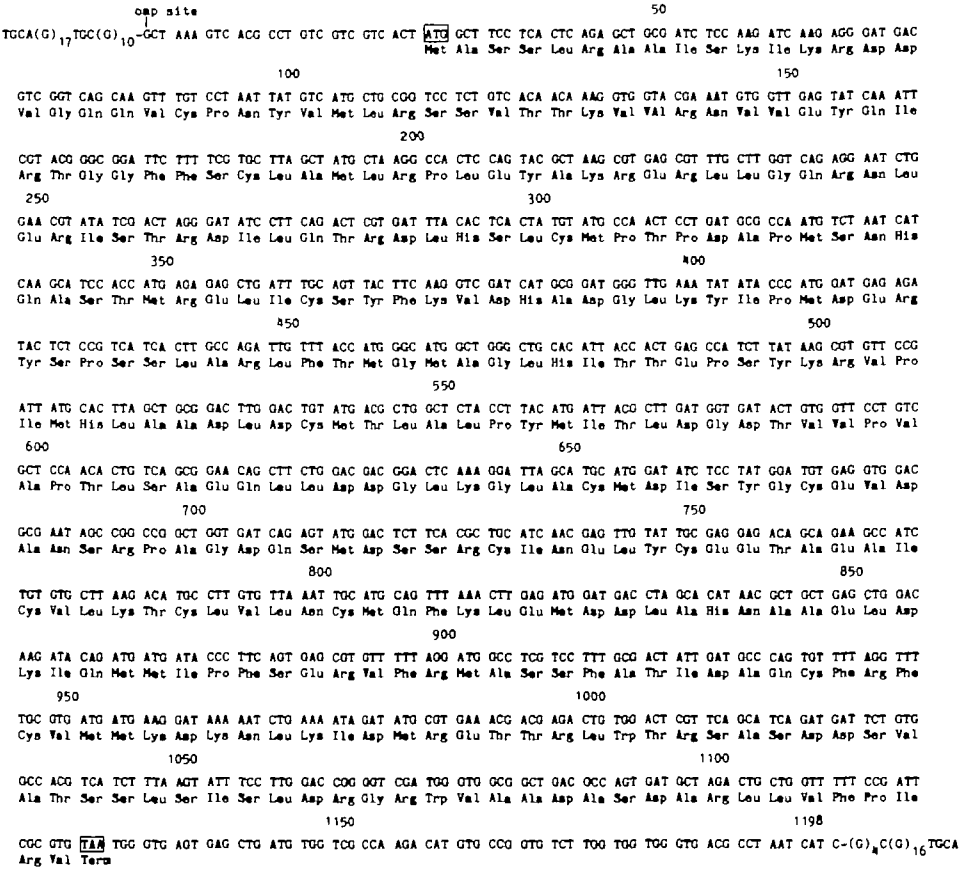

Fig. 2. Sequence of cloned DNA of reovirus genome segment S3. The positive sense strand (corresponding to mRNA) is shown. Also indicated are cap site, initiation and termination sites of protein synthesis, and the predicted amino acid sequence of the S3 gene product, sigma NS. The $TGCA(G)_{17}TGC(G)_{10}$ and $(G)_4C(G)_{10}TGCA$ nucleotides at the 5' and 3' termini respectively were added during the cloning process (10).

structure protein sigma 2 (5). Assuming the first AUG is the initiator of protein synthesis, the S3 gene product was estimated to have a molecular weight of 41,061, consistent with a reported molecular weight for sigma NS of approximately 40,000 (2,5).

Structural Aspects of S3 mRNA

Based on the sequence data a possible secondary structure was deduced for sigma NS mRNA. The positive strand of genome S3 RNA can form several stable

Table 1.

Intra-molecular base-pairing of S3 mRNA

```
       509                                  541
(1)      CGATTATGCACTTAGC-TGC-GGACTTGGACTGTA
         || |||||| |||| ||| ||| ||| || ||       ΔG = -15
         GCGAGTGCG--AATCGCATGACCT-CACCGGA-AT
       227                                  197

       644                                    677
(2)      GA-TTAGCATGCATGGAT-ATCTC-CT-ATGGATGTGA
         ||  |||| |||| ||| | ||| || |||| || ||       ΔG = -14
         CTTCATTG-ACGT--TTAGTCGAGAGAGTACC-AC-CT
       375                                    343

       926                                        964
(3)      ATGCCCAGTGTTTTAGG--TTTTGCGTGA-TGATGAAGGA---TA
         ||||||| ||   |||| |||| |||| |||||| |||  ||       ΔG = -22
         TACGGGT-AC--CATTTGTTAGAC-CGTTCACTACTGCCTCTCAT
       464                                        424

      1021                                                    1075
(4)      GATGA-TTCTG-TGGCC-ACGTCATCTTTAAGTATTTCCTTG-GACCG-GGGTC-GA-TG
         ||||| || || ||||| |||||||||||    |||| | | || || || || || || ΔG = -14
         CTGTTTGA-ACGACTGGCTGCAGTAGGGAGAACTAGA--ACCTCTAGCGTCGAGACTCAC
        94                                                        37

      1120                                      1155
(5)      CGCGTGTAA--TGGGTGAGTGAGCTGATGTG--GTC--GCCA
         ||| |||||  ||| | ||||  |||||||  |||  ||||       ΔG = -22
         GCGAATATTCTACCGAGTCAC--CATTACACGTCGGGTCGGT
       503                                      464
```

Inter-molecular base-pairing between S3 and S2 mRNA

```
       904                 (S3 RNA)                 945
(6)      TGGCCTCGTCCTTTGC-GACTATTGATGCCCAGT-GTTTTAGGT
         ||||  |||  ||||| |||||||||| |||||||| |||||  |||      ΔG = -26
         ACCG-TGTA--AAACGTCTGGTGGTT-TGGGTTGTCAGAACTTA
        78                 (S2 RNA)                 39

      1034                 (S3 RNA)                1069
(7)      GGCCACGTCATCTT--TAAGTATTT--CCTTGGACCG-GGG
         ||| || |||| ||  || |||||   |||||  |||  ||      ΔG = -26
         CCGTTG-AGTA-AACTAT-CATAGAACTGGGTC-GGCAGCC
       229                 (S2 RNA)                 193
```

For convenience, sequences are shown in DNA form rather than RNA form. (1-5): Dyad symmetry within the S3 positive strand; (6-7): Potential regions of hybridization between S3 and S2 RNAs. The sequences are numbered from the 5'-terminus. ΔG represents free energy KCal/mol (1 Cal = 4.184 J).

intra-molecular base-paired regions as shown in Table 1 (lines 1-5). The 5'-terminal cap and the putative AUG initiator (position 28) are located in single stranded regions. These structural features of S3 mRNA also occur in reovirus S2 mRNA (data not shown). By contrast, in human rotavirus genome segment 11 the 5' and 3' termini of the plus strand are part of a stable pan-handle structure (10).

When the S3 and S2 genes were compared, no homologous region of statistical significance was found, apart from the 5'm^7GpppG^{m}pCpUpA and 3'UpCpApUpC terminal sequences common to the plus strand of all genome segments (17). In contrast, S2 and S3 mRNAs were found to contain two regions capable of stable (ΔG = -26 KCal/mole) inter-molecular base pairing (Table 1, lines 6-7). It remains to be investigated if these possible inter-molecular and intra-molecular secondary structures are involved in viral morphogenesis and/or regulation of protein synthesis.

Structural Features of Sigma NS Protein

Sigma NS possesses 366 amino acids including 45 acidic, 48 basic, 162 nonpolar and 111 neutral polar amino acids. The amino acid composition predicts that sigma NS is a neutral protein (charge 0) at pH 7.0, assuming glutamic and aspartic acid are each -1, arginine and lysine each +1, and histidine +0.5 at this pH (18). Codon usage by sigma NS does not show any bias against the CG dinucleotide, unlike influenza virus and simian rotavirus gene-8 (18,19). This lack of bias against CG utilization also occurs with reovirus S2 gene (16), although we found that human rotavirus gene-11 does display such bias. There is no apparent leader signal peptide at the N-terminus of sigma NS protein.

Sigma NS can form aggregates and bind ssRNA (7). To gain some insight into the conformation/function relationships important to these properties of sigma NS, we have examined the secondary structure of this protein and the distribution of (1) hydrophobic and hydrophilic amino acids, and (2) positively and negatively charged amino acids. The predicted secondary structure and distribution of charged amino acids are shown in Fig. 3. Sigma NS protein comprised 36% α-helix, 31% β-sheet and 33% coiled regions. The distribution of charged amino acids showed distinct clusters of basic and acidic amino acids. These clusters are present in helical regions; helical regions 1-14 and 285-311 are rich in basic amino acid residues, whereas helical regions 242-247 and 265-278 are dominated by acidic residues. A hydropathicity plot (21) showed that there were no marked hydrophobic or hydrophilic regions (data not shown). By contrast, reovirus sigma 2 protein which we similarly analyzed was found to be

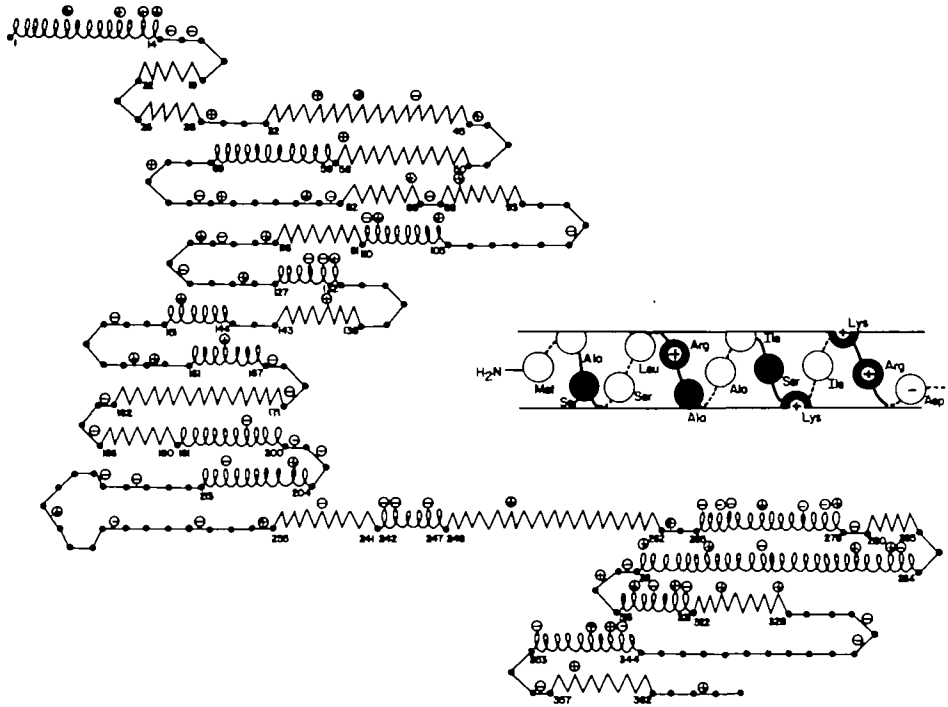

Fig. 3. Schematic representation of the secondary structure of sigma NS protein. The plus (+) and minus (-) indicate positions of basic and acidic amino acids, respectively. This structure was constructed according to the rules of Chou and Fasman (20). Inset: Three dimensional representation of α-helix region at the N-terminus.

rich in β-sheet structures (β-sheet 49%, α-helix 21% and coiled region 30%) and hydrophobic.

DISCUSSION

How segmented viral genomes are assembled and packaged into virus particles is an unresolved mystery. A random selection and subsequent packaging of ssRNA genes has been postulated for influenza virus (22), but the molecular mechanism for such a process has not been thus far elucidated. For viruses of the Reoviridae family, a complete set of dsRNA segments is apparently necessary for formation of an infectious virus particle (23). The selection of ssRNA and packaging of the individual RNA segments appears to be extremely efficient since the ratio of the number of virus particles to the number of infectious particles

can be as small as 1 (24). In addition, each infectious particle contains only one copy of each viral gene (2). A mechanism based on random assortment of viral genes, therefore, appears unlikely. Rather, a specific mechanism(s) that directs the selection of virus specific RNAs, followed by the organization and packaging of a unique set of 10 RNA species, should be expected.

To gain some insight into this problem, we have examined the structural features of reovirus RNAs and sigma NS protein since several biochemical and genetic data suggested that sigma NS might be involved in genome assembly during viral morphogenesis (4,7). Consistent with previous observations by Warrington et al. (25), both S2 and S3 mRNAs were found to contain regions capable of forming stable intra-molecular base-paired domains, with ΔG values of -14 to -22 KCal/mole (Table 1). Apart from the 5' and 3' terminal oligonucleotides common to all reovirus genome segments (17), there is no significant sequence homology between S2 and S3 RNAs. While common termini may allow selection of reovirus specific mRNAs from the pool of ssRNAs in infected cells, such common sequences per se do not provide sufficient information to enable the organization of a set of 10 viral RNA species into a precursor virion particle. Possibly, the greater structural complexity resulting from intra-molecular base-pairing (Table 1) may provide specific "identification sites" for recognition of individual mRNAs.

Reovirus S3 and S2 mRNAs can potentially hybridize to each other (Table 1, lines 6-7). It is intriguing to speculate that such intermolecular interactions may result in the formation of large RNA networks, each containing a complete set of reovirus genes. Indeed, similar analyses by Antczak et al. (13) on the terminal sequences of all 10 reovirus RNA segments indicated that complex interactions between termini can occur. More extensive sequence data on reovirus genes is required, however, to determine the extent of these types of inter-molecular interactions between reovirus mRNA species.

Sigma NS is known to bind reovirus ssRNAs but not dsRNAs (4,7). In an attempt to correlate structural features of sigma NS to its biological functions, we determined a possible secondary structure according to Chou and Fasman (20). As shown in Fig. 3, its overall structure appears to be a rather extended form, since it is not a hydrophobic protein, with α-helix, β-sheet and coil regions scattered throughout the entire molecule. The observation of Gomatos et al. (7) that sigma NS molecules form complexes of various size (13-19S) may be accounted for by this structural flexibility. RNA binding and poly(C)-dependent poly(G) polymerase activities associated with a fraction of sigma NS complex suggest that sigma NS monomer is a multi-functional "building

block". Complexes comprised of different numbers of sigma NS units may represent various catalytic, recognition and ssRNA binding sites. In this context, two structural aspects are noteworthy. Sigma NS protein is electrostatistically neutral at pH 7.0, but it contains three major and distinct, locally charged regions: a positive region (position 1-14) and two negative regions (positions 242-247 and 265-278). A stretch of α-helix at the N-terminus of sigma NS contains a cluster of four positively charged amino acids. A three-dimensional construction of this helix (Fig. 3, inset) revealed that all these basic amino acids were located on the same side. Such a configuration could be important for interaction with RNAs. Moreover, the same region also includes two serine residues which, upon phosphorylation, could neutralize the positive charges of basic amino acids. A necessary requirement for a reovirus assembly protein is the dissociation of protein and RNA before packaging. This could be accomplished via phosphorylation of these serine residues. Alternatively, conversion of ssRNA to dsRNA would result in the release of sigma NS from RNAs.

REFERENCES

1. Furuichi, Y., Morgan, M.A., Muthukrishnan, S. and Shatkin, A.J. (1975) Proc. Natl. Acad. Sci. USA 72, 362-366.
2. Joklik, W.K. (1981) Microbiol. Rev. 45, 483-501.
3. Silverstein, S.C., Christman, J.K. and Acs, G. (1976) Ann. Rev. Biochem. 45, 375-408.
4. Huisman, H. and Joklik, W.K. (1976) Virology 70, 411-424.
5. McCrae, M.A. and Joklik, W.K. (1978) Virology 89, 578-593.
6. Gomatos, P.J., Stamatos, N.M. and Sarker, N.H. (1980) J. Virol. 36, 556-565.
7. Gomatos, P.J., Prakash, O. and Stamatos, N.M. (1981) J. Virol. 39, 115-124.
8. Ito, Y. and Joklik, W.K. (1972) Virology 50, 189-201.
9. Schuerch, A.R. and Joklik, W.K. (1973) Virology 56, 218-229.
10. Imai, M., Richardson, M.A., Ikegami, N., Shatkin, A.J. and Furuichi, Y. (1983) Proc. Natl. Acad. Sci. USA 80, 373-377.
11. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
12. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-558.
13. Antczak, J.B., Chmelo, R., Pickup, D.J. and Joklik, W.K. (1982) Virology 121, 307-319.
14. Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252.
15. Kozak, M. and Shatkin, A.J. (1976) J. Biol. Chem. 251, 4259-4266.
16. Cashdollar, L.W., Esparza, J., Hudson, G.R., Chmelo, R., Lee, P.W.K. and Joklik, W.K. (1982) Proc. Natl. Acad. Sci. USA 79, 7644-7648.
17. Darzynkiewicz, E. and Shatkin, A.J. (1980) Nucleic Acids Res. 8, 337-350.
18. Huddleston, J.A. and Brownlee, G.G. (1982) Nucleic Acids Res. 10, 1029-1038.
19. Both, G.W., Bellamy, A.R., Street, J.E. and Siegman, L.J. (1982) Nucleic Acids Res. 10, 7075-7088.
20. Chou, P.Y. and Fasman, G.C. (1978) Advances Enzymol. 47, 45-148.

21. Hopp, T.P. and Wood, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824-3828.
22. Compans, R.W. and Choppin, P.W. (1971) in Comparative Virology, Maramorosch, K. and Kurska, F. Eds., pp. 407-432, Academic Press, New York.
23. Nonoyama, M. and Graham, A.F. (1970) J. Virol. 6, 693-694.
24. Spendlove, R.S., McClain, M.E. and Lennette, E.H. (1970) J. Gen. Virol. 8, 83-93.
25. Warrington, R.C., Hayward, C. and Kapuler, A.M. (1973) Biochim. Biophys. Acta 331, 231-242.