Volume 15 Number 10 1987 Nucleic Acids Research

Nucleotide sequence analysis of the L gene of Newcastle disease virus: homologies with Sendai and vesicular stomatitis viruses

Khatijah Yusoff, Neil S.Millar, Philip Chambers and Peter T.Emmerson

Department of Biochemistry, The University, Newcastle upon Tyne NE1 7RU, UK

Received March 11, 1987; Revised and Accepted April 23, 1987

ABSTRACT

The nucleotide sequence of the L gene of the Beaudette C strain of Newcastle disease virus (NDV) has been determined. The L gene is 6704 nucleotides long and encodes a protein of 2204 amino acids with a calculated molecular weight of 248822. Mung bean nuclease mapping of the 5' terminus of the L gene mRNA indicates that the transcription of the L gene is initiated 11 nucleotides upstream of the translational start site. Comparison with the amino acid sequences of the L genes of Sendai virus and vesicular stomatitis virus (VSV) suggests that there are several regions of homology between the sequences. These data provide further evidence for an evolutionary relationship between the Paramyxoviridae and the Rhabdoviridae. A non-coding sequence of 46 nucleotides downstream of the presumed polyadenylation site of the L gene may be part of a negative strand leader RNA.

INTRODUCTION

Paramyxoviruses such as NDV have genomes which are nonsegmented negative-sense strands of RNA, approximately 15 kb long (1). The gene organization, control of mRNA and protein synthesis, RNA replication and viral assembly of paramyxoviruses are broadly similar to the corresponding functions of the rhabdovirus vesicular stomatitis virus (VSV), which has been more intensively studied (2-6). Both families of viruses have large (L) genes that occupy most of the 5' half of the genome. The L gene encodes the L protein which has a molecular weight greater than 200 K (5). The L protein is the least abundant of virion proteins of VSV (7,8) and the paramyxovirus Sendai virus (9) which suggests that it has an enzymatic rather than a structural role. Although the precise functions of the L protein are not known, it is assumed to be the viral RNA-dependent RNA polymerase. The L and P proteins of NDV, which are analogous to L and NS proteins of VSV respectively, reconstitute an active transcriptive complex when added to viral nucleocapsids which have been stripped of these proteins (10-11). There is evidence that the analogous proteins of VSV are responsible for

synthesis of viral mRNA and capping, methylation, and polyadenylation of newly synthesized viral mRNAs (12-15). Since the L protein is 5-10 times larger than the P or NS proteins it may well perform most of these processes.

The complete nucleotide sequences of the L genes of VSV and Sendai virus, another paramyxovirus, have been determined recently (16-18). Comparison of those sequences to that of NDV may help us to understand the functions of the L protein. Here, we report the nucleotide sequence of the L gene of NDV.

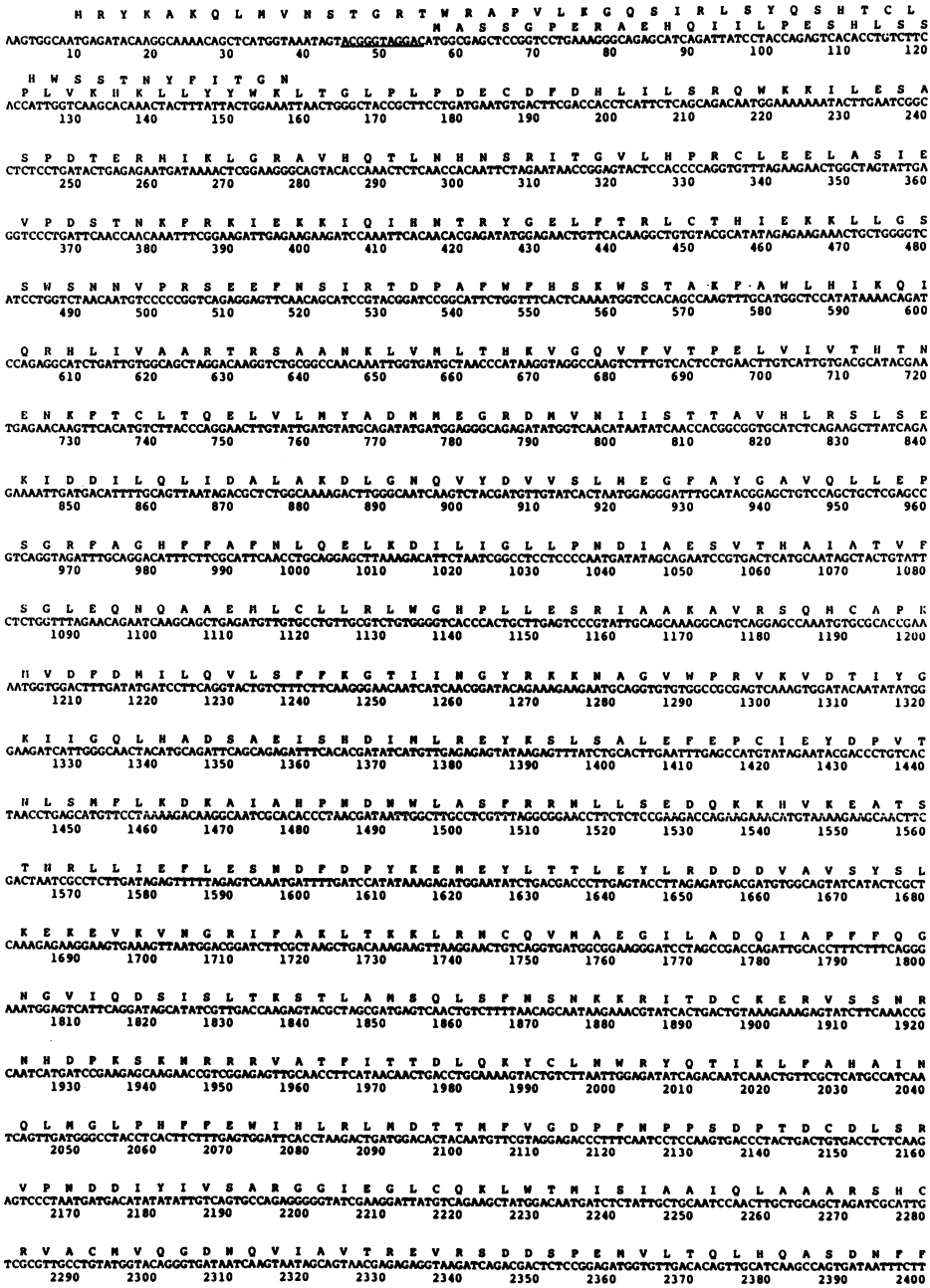
MATERIALS AND METHODS

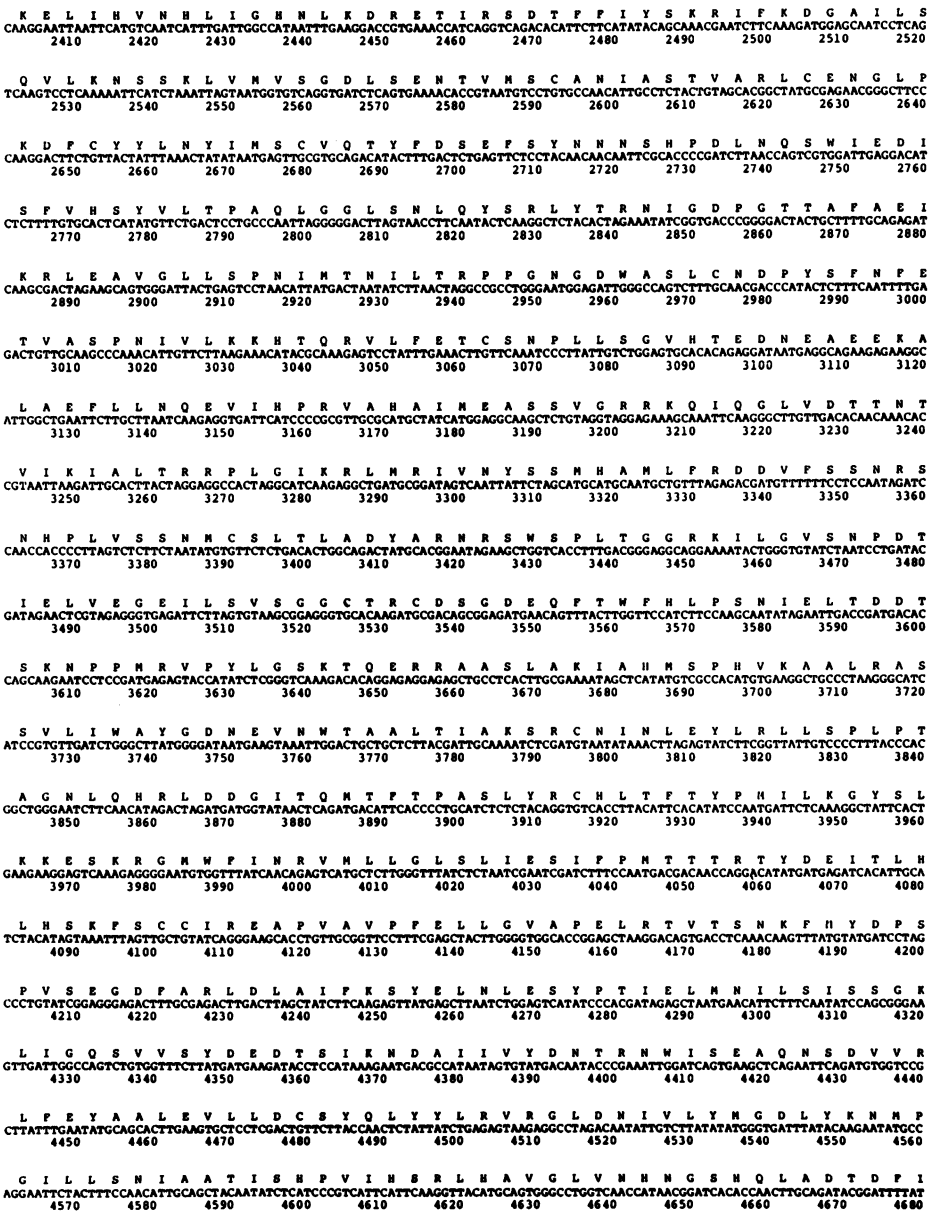
DNA sequencing

Details of the construction and characterization of cDNA clones to the L gene of NDV strain Beaudette C have been reported (19). In addition to the nine clones previously assigned to the L gene, a clone (designated 4.14) was selected from the colony bank by hybridization with an 868 bp PstI-HindIII fragment of clone 3.23 which extended the sequence past the 3' end of the L gene. The dideoxy chain termination method was used for DNA sequencing (20), but in several instances it was necessary to substitute deoxyguanosine triphosphate with deoxy-7-deazaguanosine triphosphate (Boehringer Mannheim) in order to resolve ambiguous regions of the sequence (21). Oligonucleotide sequencing primers made on an Applied Biosystems model 381A DNA synthesiser were used in addition to a 17-base universal primer (Pharmacia). The sequence of both strands of DNA was determined by using M13 mp8, mp9, mp18 and mp19 phage vectors. The sequence data were assembled and analysed using the computer programs of Staden (22) and Queen and Korn (23).

Mapping of the mRNA 5'terminus

Chick embryo fibroblast (CEF) cells were infected with NDV strain Beaudette C at a multiplicity of about 50 p.f.u./cell (24). Total cellular RNA was isolated from cells at nine hours post-infection by phenol extraction at 70°C. A restriction fragment of plasmid 3.73, which spans the HN/L gene junction, was isolated from an 8% polyacrylamide gel by electroelution (19). The restriction fragment extends from an EcoRI site [86-81 bp upstream from the start of the HN/L intergenic region (25)] to an XbaI site (position 305 to 310, Fig. 1). The 5' termini were labelled with [^{32}P] ATP using polynucleotide kinase. The labelled end at the EcoRI site was removed by cleavage with HincII [67 bp upstream from the start of the HN/L intergenic region (25)] to give a 376 bp fragment used for nuclease mapping. Total RNA from infected or uninfected cells was added to labelled restriction fragment





```
  L  Q  L  I  S  R  L  C  C  L  Y  T  V  L  F  A  T  T  R  E  I  P  K  I  R  G  L  S  A  E  E  K  C  S  V  L  T  E  Y  L
GCTTCAGCTGATATCCCGGTTATGCTGTCTGTACACGGTACTCTTTGCTACAACAAGAGAAATCCCGAAAATAAGAGGCTTATCTGCAGAAGAGAAATGTTCAGTACTTACTGAGTATCT
     4810      4820      4830      4840      4850      4860      4870      4880      4890      4900      4910      4920

  L  S  D  A  V  K  P  L  L  S  P  D  Q  V  S  S  I  N  S  P  N  I  I  T  F  P  A  N  L  Y  Y  H  S  R  K  S  L  N  L  I
ACTGTCGGATGCTGTGAAACCATTACTTAGCCCTGATCAGGTGAGCTCTATCATGTCTCCTAACATAATTACATTCCCAGCTAATCTGTACTACATGTCTCGGAAGAGCCTCAATTTGAT
     4930      4940      4950      4960      4970      4980      4990      5000      5010      5020      5030      5040

  R  E  R  E  D  K  D  S  I  L  A  L  L  F  P  Q  E  P  L  L  E  F  P  S  V  Q  D  I  G  A  R  V  K  D  P  F  T  R  Q  P
CAGGGAAAGGGAGGACAAGGATTCTATCCTGGCGTTGTTGTTCCCCCAAGAGCCATTATTAGAGTTCCCTTCTGTGCAAGATATTGGTGCTCGAGTGAAAGATCCATTCACCCGACAACC
     5050      5060      5070      5080      5090      5100      5110      5120      5130      5140      5150      5160

  A  A  F  L  Q  E  L  D  L  S  A  P  A  R  Y  D  A  F  T  L  S  Q  I  H  P  E  L  T  S  P  N  P  E  E  D  Y  L  V  R  Y
TGCGGCATTTTTGCAAGAGTTAGATTTGAGTGCTCCAGCAAGGTATGACGCATTCACACTTAGTCAGATTCATCCTGAGCTCACATCACCAAATCCGGAGGAAGACTACTTAGTACGATA
     5170      5180      5190      5200      5210      5220      5230      5240      5250      5260      5270      5280

  L  F  R  G  I  G  T  A  S  S  S  W  Y  K  A  S  H  L  L  S  V  P  E  V  R  C  A  R  H  G  N  S  L  Y  L  A  E  G  S  G
CTTGTTCAGAGGAATAGGGACTGCATCCTCCTCTTGGTATAAGGCATCCCATCTCCTTTCTGTACCCGAGGTAAGATGTGCAAGACACGGGAACTCCTTATACTTAGCTGAAGGAAGCGG
     5290      5300      5310      5320      5330      5340      5350      5360      5370      5380      5390      5400

  A  I  M  S  L  L  E  L  H  V  P  H  E  T  I  Y  Y  N  T  L  F  S  N  E  M  N  P  P  Q  R  H  F  G  P  T  P  T  Q  F  L
AGCCATCATGAGTCTTCTCGAACTGCATGTACCACATGAAACTATCTATTACAATACGCTCTTTTCAAATGAGATGAACCCCCCGCAGCGACATTTCGGGCCGACCCCAACCCAGTTTTT
     5410      5420      5430      5440      5450      5460      5470      5480      5490      5500      5510      5520

  N  S  V  V  Y  R  N  L  Q  A  E  V  T  C  K  D  G  F  V  Q  E  F  R  P  L  W  R  E  N  T  E  E  S  D  L  T  S  D  K  A
GAATTCGGTTGTTTATAGGAACCTACAGGCGGAGGTAACATGCAAGGATGGATTTGTCCAAGAGTTCCGTCCACTATGGAGAGAAAATACAGAGGAAAGCGACCTGACCTCAGATAAAGC
     5530      5540      5550      5560      5570      5580      5590      5600      5610      5620      5630      5640

  V  G  Y  I  T  S  A  V  P  Y  R  S  V  S  L  L  H  C  D  I  E  I  P  P  G  S  N  Q  S  L  L  D  Q  L  A  I  N  L  S  L
AGTGGGGTATATTACATCTGCAGTGCCCTACAGATCTGTATCATTGCTGCATTGTGACATTGAAATCCCTCCAGGGTCCAATCAAAGCTTACTAGATCAATTAGCTATCAATTTATCTCT
     5650      5660      5670      5680      5690      5700      5710      5720      5730      5740      5750      5760

  I  A  M  H  S  V  R  E  G  G  V  V  I  I  K  V  L  Y  A  M  G  Y  Y  F  H  L  L  M  N  L  F  A  P  C  S  T  K  G  Y  I
GATTGCCATGCATTCCGTAAGGGAGGGCGGGGTAGTGATCATCAAAGTGTTGTATGCAATGGGATACTACTTTCATCTACTCATGAACTTGTTCGCTCCGTGTTCCACAAAAGGATACAT
     5770      5780      5790      5800      5810      5820      5830      5840      5850      5860      5870      5880

  L  S  N  G  Y  A  C  R  G  D  M  E  C  Y  L  V  F  V  M  G  Y  L  G  G  P  T  F  V  H  E  V  V  R  H  A  K  T  L  V  Q
TCTCTCTAATGGTTATGCATGTAGAGGGGATATGGAGTGTTACCTGGTATTTGTCATGGGTTACCTGGGCGGGCCTACATTTGTACACGAGGTGGTGAGGATGGCAAAAACTCTGGTGCA
     5890      5900      5910      5920      5930      5940      5950      5960      5970      5980      5990      6000

  R  H  G  T  L  L  S  K  S  D  E  I  T  L  T  R  L  F  T  S  Q  R  Q  R  V  T  D  I  L  S  S  P  L  P  R  L  I  K  Y  L
GCGGCACGGTACGCTTTTGTCCAAATCAGATGAGATCACACTGACCAGGTTATTCACCTCACAGCGGCAGCGTGTGACAGACATCCTATCCAGTCCTTTACCAAGATTAATAAAGTACTT
     6010      6020      6030      6040      6050      6060      6070      6080      6090      6100      6110      6120

  R  K  N  I  D  T  A  L  I  E  A  G  G  Q  P  V  R  P  F  C  A  E  S  L  V  S  T  L  A  D  I  T  Q  I  T  Q  I  I  A  S
GAGAAAGAATATTGACACTGCGCTGATTGAAGCTGGGGGACAGCCCGTCCGTCCATTCTGTGCAGAGAGTTTGGTGAGCACGCTGGCGGACATAACTCAGATAACCCAGATCATTGCTAG
     6130      6140      6150      6160      6170      6180      6190      6200      6210      6220      6230      6240

  H  I  D  T  V  I  R  S  V  I  Y  H  E  A  E  G  D  L  A  D  T  V  F  L  F  T  P  Y  N  L  S  T  D  G  K  K  R  T  S  L
TCACATTGACACAGTCATCCGGTCTGTGATATATATGGAAGCTGAGGGTGATCTCGCTGACACAGTATTTCTATTTACCCCTTACAATCTCTCTACTGACGGGAAAAAGAGAACATCACT
     6250      6260      6270      6280      6290      6300      6310      6320      6330      6340      6350      6360

  K  Q  C  T  R  Q  I  L  E  V  T  I  L  G  L  R  V  E  D  L  N  K  I  G  D  V  I  S  L  V  L  K  G  H  I  S  H  E  D  L
TAAACAGTGCACGAGACAGATCCTAGAGGTTACAATATTGGGTCTTAGAGTCGAAGATCTCAATAAAATAGGCGATGTAATCAGCCTAGTGCTTAAAGGCATGATCTCCATGGAGGACCT
     6370      6380      6390      6400      6410      6420      6430      6440      6450      6460      6470      6480

  I  P  L  R  T  Y  L  K  H  S  T  C  P  K  Y  L  K  A  V  L  G  I  T  K  L  K  E  M  F  T  D  T  S  V  L  Y  L  T  K  A
TATCCCACTAAGGACATACTTGAAGCATAGTACCTGCCCTAAATATTTGAAGGCTGTCCTAGGTATTACCAAACTCAAAGAAATGTTTACAGACACCTCTGTATTGTACTTGACTCGTGC
     6490      6500      6510      6520      6530      6540      6550      6560      6570      6580      6590      6600

  Q  Q  K  F  Y  H  K  T  I  G  N  A  V  K  G  Y  Y  S  N  C  D  S
TCAACAAAAATTCTACATGAAAACTATAGGCAATGCAGTCAAAGGATATTACAGTAACTGTGACTCTTAACGAAAATCACATATTAATAGGCTCCTTTTCTGGCCAATTGTATCCTTGGT
     6610      6620      6630      6640      6650      6660      6670      6680      6690      6700      6710      6720

GATTTAATTATACTATGTTAGAAAAAAATTGAACTCCGACTCCTTAGAGCTCGAATTCGAACTCAAATAAATGTCTT
     6730      6740      6750      6760      6770      6780      6790
```

Figure 1. Nucleotide sequence of the L gene of NDV strain Beaudette C, and deduced amino acid sequence of the L protein. The cDNA sequence is shown in the positive (mRNA) sense. The semi-conserved 11-nucleotide start and 10-nucleotide polyadenlyation signals are underlined. The deduced amino acid sequence of the major open reading frame as well as a second, smaller potential open reading frame, are shown above the nucleotide sequence.

and dried under vacuum on a 'UNIVAP' rotary evaporator and dissolved in 30 ul hybridization buffer (0.04 M PIPES, pH 6.4, 0.4 M NaCl, 0.001 M EDTA, 80% formamide). Samples were incubated at 85°C for 15 minutes to denature the

DNA and then at 50°C for 3 hours to allow hybrid formation, essentially as described by Maniatis et al. (26). The hybridized products were then diluted ten-fold into ice-cold nuclease buffer (30 mM Na acetate, 50 mM NaCl, 1 mM $ZnCl_2$, 5% glycerol, 0.001% Triton X-100) and incubated for 30 minutes at 37°C with 50 units of mung bean nuclease. After phenol extraction and isopropanol precipitation, the DNA fragments protected from nuclease action were analysed by electrophoresis on 6% denaturing polyacrylamide gels (27). Fragment sizes were determined by comparison with the migration of polynucleotides of known length generated by a dideoxy sequencing reaction run in parallel.

RESULTS

The nucleotide sequence and predicted amino acid sequence of the NDV L gene is shown in Fig. 1. The L gene is 6704 nucleotides long and is presumed to extend from a typical NDV mRNA transcriptional start site (position 45 to 55) to a typical polyadenylation site (position 6738 to 6748). The sequence of this mRNA start site 5'-ACGGGTAGGAC-3' matches the consensus sequence 5'-ACGGGTAGAAG-3' of the mRNA start sites for the NP, P, M, F and HN genes of NDV at nine of the eleven positions (25,28-31).

The L gene nucleotide sequence contains a very large open reading frame starting at an ATG codon at position 57 to 59 and continuing to a termination codon TAA at position 6668 to 6670 (Fig. 1). The sequence around this ATG codon conforms well to the consensus (5'-purine-N-N-A-U-G-G-3') for functional initiation codons in eukaryotes (32,33). This open reading frame encodes a protein of 2204 amino acids with a predicted molecular weight of 248822 which is in good agreement with the estimated molecular weight of the L protein of NDV (220 K), derived from its mobility on SDS gels (5). A small open reading frame in a different phase starts upstream of the proposed start site of the L message at the ATG codon position 9 to 11, Fig. 1. Translation of this small open reading frame would generate a rather hydrophobic polypeptide of 50 amino acids with a predicted molecular weight of 5.9 K which does not correspond to any known viral polypeptide.

In an earlier report (25), it was suggested on the basis of a provisional S1 mapping experiment, that the mRNA start site for the NDV L gene is at position 1 in the sequence shown (Fig. 1). Using mung bean nuclease (Fig. 2) we have located the transcriptional start site at position 45 (Fig. 1). A fragment of 264 bp was protected against mung bean nuclease by RNA from NDV-infected CEF (Fig. 2, lane 2) while no protected bands other than a trace of the starting material were seen in a control using uninfected

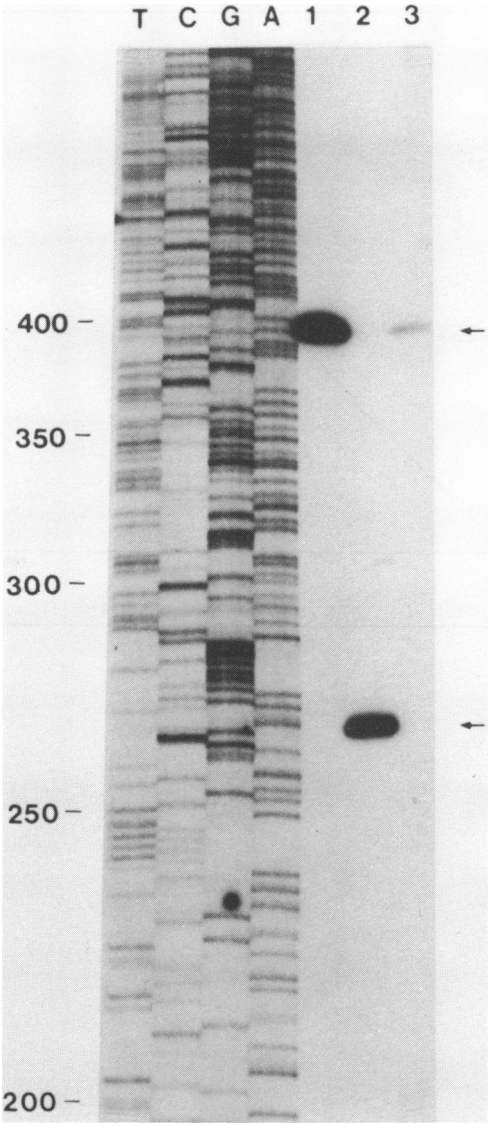


Figure 2. Mapping of the 5' terminius of the L mRNA. Lane 1: the original ^{32}P 5' end-labelled, 376 bp HincII-XbaI restriction fragment (see Methods), indicated by upper arrow. Lane 2: restriction fragment annealed to the RNA from NDV-infected CEF cells, mung bean nuclease protected fragment of 264 bp, indicated by lower arrow. Lane 3: control using RNA from uninfected cells. A dideoxy sequencing reaction (lanes T,C,G and A) was run in parallel to allow determination of fragment sizes.

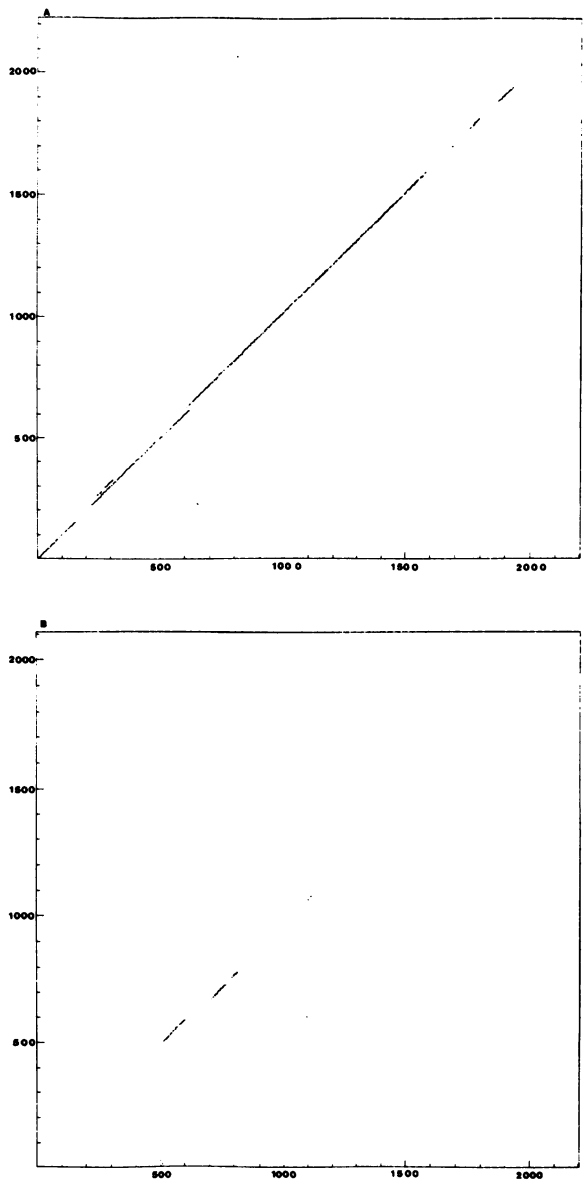


Figure 3. Dot matrix homology plots of the L amino acid sequences. Comparisons of NDV with Sendai virus and VSV are shown in (a) and (b), respectively. In both cases a window of 99 and a proportional score of 1025 were used with the DAIGON computer program (Staden, 1984).

CEF cellular RNA (Fig.2, lane 3). The sequence resembling the mRNA start site at base 1 is discussed later.

The polyadenylation site, 5'-TTAGAAAAAAA-3', matches a consensus sequence reported previously for NDV (19). There is a non-coding region of 70 nucleotides between the end of the long open reading frame and the proposed polyadenylation site (Fig. 1). Downstream of the proposed polyadenylation site is another non-coding region of 49 nucleotides which is discussed later.

Comparison of the predicted NDV L amino acid sequence to the L amino acid sequences of Sendai virus (16) and VSV (18) using the dot matrix homology computer program DIAGON (22) are shown in Fig. 3. The diagonal lines on the plots suggest extensive homology between the L amino acid sequences of NDV and Sendai virus (Fig. 3a) and to a lesser extent with that of VSV (Fig. 3b).

An alignment of the L amino acid sequence of NDV to those of Sendai virus and VSV is shown in Fig. 4. Gaps were inserted where necessary to maximise homology. If the extra gaps positioned to align the VSV sequence to those of NDV and Sendai virus are disregarded, the overall level of amino acid identities between NDV and Sendai virus is 27% in this alignment but the proposed sequence alignment can be roughly sub-divided into four regions with different levels of homology. The first 660 positions (matches plus gaps) show 24% identity, the second 650 positions show 39% identity, the third 600 positions show 21% identity and the remaining 350 positions show 17% identity.

Additional gaps must be inserted to align the VSV L amino acid sequence (residues 226 to 1243) into the sequence alignment (Fig. 4). Two regions of homology between VSV and the paramyxovirus sequences have been identified corresponding to amino acids 226 to 581 and 598 to 1243 of the VSV L sequence. There are 127 positions in which the amino acid residues are identical in the three proteins. These 'three-way matches' occur in the two blocks of sequence where the L protein of VSV has been aligned to those of NDV and Sendai virus. The number of glycine residues conserved in three-way matches is striking: 19/127 of the three-way matches are glycines (15% of the total) compared to the overall abundance of glycine in these regions of the three viral sequences (about 5% in all cases). Lysine and arginine residues are also present in three-way matches at a somewhat higher level than might be expected on the basis of their relative abundances.

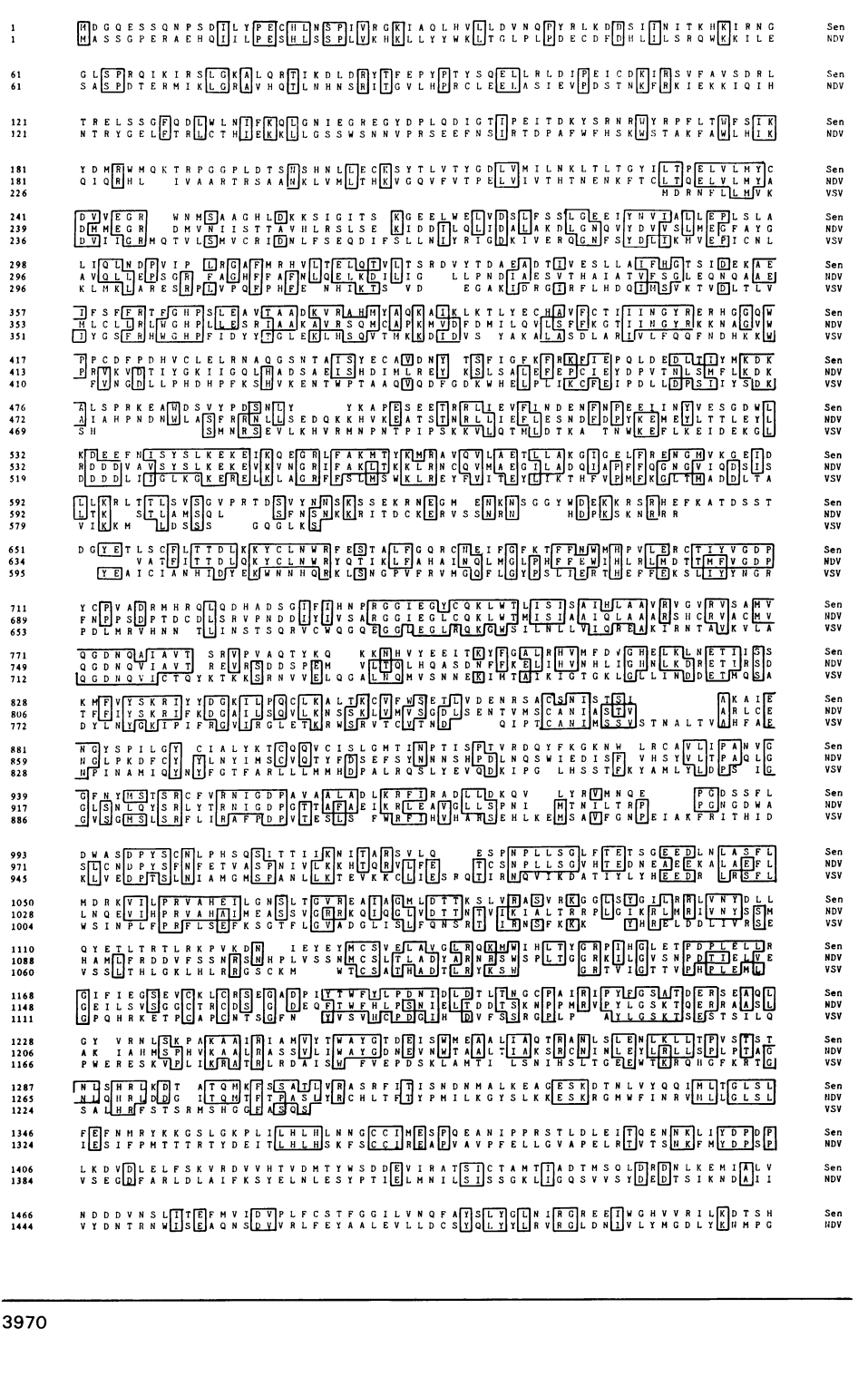


Figure 4. Comparison of the L amino acid sequences of NDV, Sendai virus (Sen) and VSV. The amino acid sequences are numbered from their N-termini, shown on the left hand side of the figure. Only the central region of VSV (residues 226-1243) which shows good homology to NDV and Sendai virus is shown. At positions where identical amino acids occur the residues are boxed.

The unmodified L protein has a net charge of +27 at pH 7.0 (assuming a charge of +0.5 on histidine residues) and an estimated pI of 7.28. A hydropathy plot of the NDV L amino acid sequence using the procedure of Kyte and Doolittle (34) is shown in Fig. 5. This plot is similar to the hydropathy plot of the L protein of Sendai virus and to a lesser extent to that of VSV (data not shown). There is a highly hydrophilic region in the NDV L protein sequence at position 602 to 633. The corresponding hydrophilic region is considerably more extensive in the L protein of Sendai virus but is absent from that of VSV. The amino acid alignment of Fig. 4 locates this hydrophilic region in an area where extensive insertions or deletions have occurred between the various proteins. The two regions where good homology can be detected between all three L protein sequences flank this highly variable area.

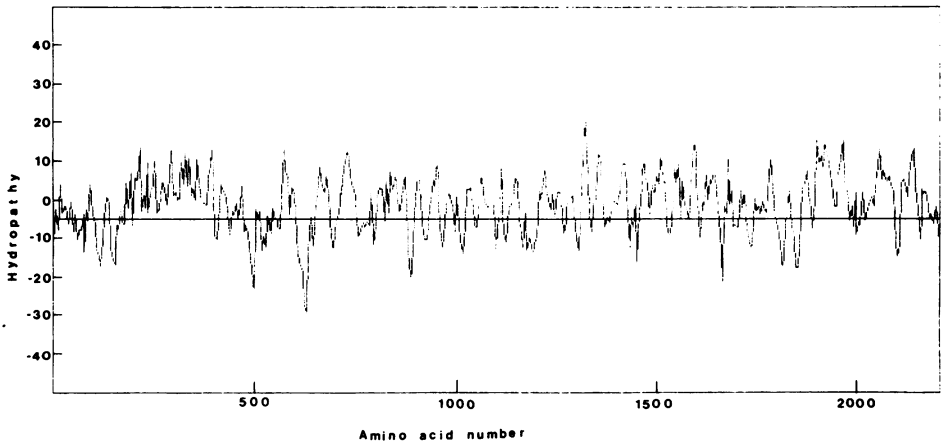


Figure 5. Hydropathy plots of the NDV L amino acid sequence. Hydrophobic regions are above and hydrophilic regions below the horizontal line, which is the average hydropathy of a large number of sequenced proteins.

DISCUSSION

The nucleotide sequence of the NDV L gene is 6704 nucleotides long and encodes the L protein of approximately 249 K. Apart from the major open reading frame of 2204 amino acids, corresponding to the L protein, a second open reading frame of 50 amino acids is shown in Fig. 1. There is evidence from a previous S1 nuclease mapping experiment using a labelled fragment ending at the TaqI site (position 192, Fig. 1) which suggested that nucleotides 1 and 45 both represented the 5' ends of viral transcripts (25). If a genuine mRNA start occurred at position 1, corresponding to the sequence resembling an mRNA start sequence at bases 1 to 11, the resulting mRNA would encode the smaller open reading frame, which begins at the ATG codon at position 9 to 11. A sequence of seven A residues follows this open reading frame (position 218 to 227) and this resembles the consensus NDV polyadenylation sequence that contains six A residues (25). If such a transcript did terminate at this site it would not appear as a protected band in the mung bean nuclease mapping experiment described here (Fig. 2), using a fragment labelled at the XbaI site (position 309, Fig. 1). Work is now underway to investigate the possibility of further transcriptional start sites in the NDV genome. This small hydrophobic, 50 amino acid sequence is neither homologous to the small hydrophobic (SH) protein recently detected in SV5 (35) nor to the polypeptides which may be encoded upstream of the L gene

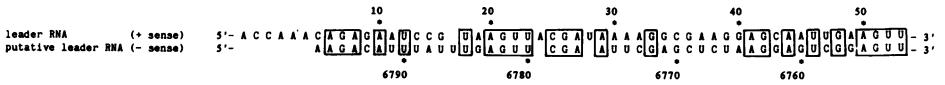


Figure 6. Comparison of positive strand leader RNA sequence with putative negative strand leader RNA.

of Sendai virus (17) and at present there is no evidence that this polypeptide is generated.

We have also mapped the 5' end of the HN mRNA using mung bean nuclease. The results indicate that there is a 31 bp intergenic region between the F and HN genes, rather than the dinucleotide intergenic region which was previously proposed on the basis of sequence homology (19). These results agree with the recently published sequence of HN from the B1 strain of NDV (31). At the junction of the F and HN gene there is, however, a similar situation to that described at the HN/L junction. A sequence resembling an mRNA start sequence in the non-coding region is followed by an open reading frame of 41 amino acids that overlaps HN, which in turn is followed by a sequence of six A residues (25). The significance of these overlapping reading frames at the F/HN and HN/L junctions is not yet clear.

The complementary sequence to the non-coding region at position 6749 to 6777, downstream of the polyadenylation site, is compared to the positive strand leader RNA sequence of NDV (36) in Fig. 6, and may be part of a negative strand leader RNA similar to that demonstrated for VSV (37). Approximately 50% of the bases are identical including nine of the thirteen most 3' proximal residues. These sequences could act as signals for termination of transcription in both positive strand and putative negative strand leader RNAs. The positive and negative strand leader RNAs of Sendai virus are identical in eleven of the twelve most 5' proximal bases (14,38). The six most 5' proximal bases found in the positive strand leader of NDV (36) do not show homology to our sequence (Fig. 6) which suggests that our clone (4.14) may not quite reach the 5' terminus of the NDV genome.

In our alignment of the three L amino acid sequences (Fig. 4) the central region (position 634 to 1283 of NDV) is 39% conserved between NDV and Sendai virus, compared to approximately 20% conservation of the viral NP, M, F and HN proteins (25,28-30) and is thus probably the most highly conserved viral protein sequence, given the known variability of the P proteins (39). Four peptides that show strong homology between the three viruses are located in the NDV sequence at positions 543 to 554 (7/12 three-way matches), 715 to

725 (7/10 three-way matches), 749 to 755 (6/7 three-way matches) and 1192 to 1199 (5/8 three-way matches). The region corresponding to positions 543 to 560 of the NDV L protein was identified by Morgan and Rakestraw (17) as being a well conserved region between Sendai virus and VSV.

The L protein of NDV is probably multifunctional in the processes of viral transcription and replication, perhaps including initiation, elongation, termination, polyadenylation, capping and methylation activities, as has been shown for VSV (12-15). It may be that the regions of high homology between NDV, Sendai and VSV are those that perform functions common to all the L proteins. These conserved regions could act as catalytic sites or binding sites for small host metabolites, while the variable regions of L may be those that interact with other viral proteins such as P or NS. The P and NS proteins are themselves known to be highly variable and appear to mutate at a high frequency (39,40). None of the well conserved regions that we have detected, match any of the presently identified consensus sequences for nucleotide-protein interacting sites, although the sequences Gly-Gly-Ile-Glu-Gly (NDV positions 715 to 719) and Gly-Ser-Lys-Thr (NDV positions 1194 to 1197) are reminiscent of the ATP binding sites with the consensus sequence Gly-X-Gly-X-X-Gly and Gly-Lys-Thr/Ser respectively (41-43).

Conserved regions which are present in all three viral sequences may have important enzymatic or conformational functions. Conserved glycines may be important in maintaining protein structure, as glycines occur at tight turns around alpha-helices or between the strands of beta-sheets (43-45). The abundance of conserved glycines in the regions of homology between NDV, Sendai virus and VSV L proteins suggests similar conformations in these regions. For example, the conserved region in the amino acid alignment corresponding to NDV positions 531 to 594 is predicted to be rich in alpha-helix (44). The conserved Gly-Arg-(hydrophobic) residues within this region (at position 551 to 553 in the NDV sequence) could form a turn in this largely alpha-helical part of the L proteins. This region could thus have a conformation of helix-turn-helix similar to that present at the DNA binding site of several bacterial repressor proteins (45), and could be important in the interaction of the L proteins with viral RNA.

The above considerations suggest that most of the RNA synthetic and modification activities are located in the N-terminal two-thirds of the L protein. The strongly hydrophilic region extending from position 602 to 633 in the NDV L protein, which is variable amongst the three viruses, is located between the two regions conserved in all three sequences and may form a

bridge linking these two conserved regions, which could thus be considered as separate domains. The domain nearer the C-terminus is more highly conserved than the domain nearer the N-terminus (39% and 26% identities between NDV and Sendai virus; 14% and 9% three-way matches, respectively). The most variable regions of the L proteins are located at the C-terminal third of the sequences. These regions may have virus-specific functions such as interactions with the respective NP and P or NS polypeptides.

The amino acid sequence homologies detected between the L proteins of the two paramyxoviruses NDV and sendai virus and the rhabdovirus VSV, are evidence that an evolutionary relationship exists between the Paramyxoviridae and Rhabdoviridae, and strongly suggests that these groups of viruses have evolved from a common ancestor.

ACKNOWLEDGEMENTS

We thank June Ramsay and Kate Brown for technical assistance and the North of England Cancer Research Campaign for financial support. N.S.M. was supported by an SERC studentship.

REFERENCES

1. Kolakofsky, D., Boy de la Tour, E., and Delius, H. (1974) J. Virol. 13, 261-268.
2. Choppin, P. W. and Compans, R. W. (1975) In 'Comprehensive Virology' (Fraenkel-Conrat, H. and Wagner, R. R. eds.) pp.75-178, Plenum Press, New York.
3. Wagner, R. R. (1975) In 'Comprehensive Virology' (Fraenkel-Conrat, H. and Wagner, R. R. eds.) pp.1-80, Plenum Press, New York.
4. Ball, L. A. and White, C. N. (1976) Proc. Natl. Acad. Sci. USA 73,442-446.
5. Collins, P. L., Hightower, L. E. and Ball, L. A. (1980) J.Virol. 35, 682-693.
6. Abraham, G. and Banerjee, A. K. (1976) Proc. Natl. Acad. Sci. USA 73, 1504-1508.
7. Bishop, D. H. L. (1977) In 'Comprehensive Virology' (Fraenkel-Conrat, H. and Wagner, R. R. eds.) pp. 117-278, Plenum Press, New York.
8. Thomas, D., Newcomb, W. W., Brown, J. C., Wall, J. S., Hainfield, J. F., Trus, B. L. and Steven, A. C. (1985) J. Virol. 54, 598-607.
9. Lamb, R. A., Mahy, B. W. J. and Choppin, P. W. (1976) Virology 69, 116-131.
10. Hamaguchi, M., Yoshida, T., Nishikawa, K., Naruse, H. and Nagai, Y. (1983) Virology 128, 105-117.
11. Emerson, S. U. and Yu, Y.-H. (1975) J. Virol. 15, 1348-1356.
12. Horikami, S. M. and Moyer, S. A. (1982) Proc. Natl. Acad. Sci. USA 79, 7694-7698. 13. Schubert, M., Keene, J. D., Herman, R. C. and Lazzarani, R.A. (1980) J. Virol. 34, 550-559.
14. Schubert, M. and Lazzarani, R. A. (1981) J. Virol. 38, 256-262.
15. Hunt, D. M., Smith, E. F. and Buckley, D. W. (1984) J. Virol. 52, 515-521.

16. Shioda, T., Iwasaki, K. and Schibuta, H. (1986) Nucleic Acids Res. 14, 1545-1563.
17. Morgan, E. M. and Rakestraw, K. M. (1986) Virology 154, 31-40.
18. Schubert, M., Harmison, G. G. and Meier, E. (1984) J. Virol. 51, 505-514.
19. Chambers, P., Millar, N. S., Bingham, R. W. and Emmerson, P. T. (1986) J. Gen. Virol. 67, 475-486.
20. Sanger,F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
21. Mizusawa, S., Nishimura, S. and Seela, F. (1986) Nucleic Acids Res. 14, 1319-1324.
22. Staden, R. (1984) Nucleic Acids Res. 12, 521-538.
23. Queen, C. and Korn, L. J. (1984) Nucleic Acids Res. 12, 581-599.
24. Chambers, P. and Samson, A. C. R. (1982) J. Gen. Virol. 58, 1-12.
25. Millar, N. S., Chambers, P. and Emmerson, P. T. (1986) J. Gen. Virol. 67, 1917-1927.
26. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, New York.
27. Biggin, M. D., Gibson, T. J. and Hong, C. F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
28. Chambers, P., Millar, N. S., Platt, S. G. and Emmerson, P.T. (1986) Nucleic Acids Res. 14, 9051-9061.
29. Chambers, P., Millar, N. S. and Emmerson, P. T. (1986) J. Gen. Virol. 67, 2685-2694.
30. Ishida, N., Taira, H., Omata, T., Mizumoto, K., Hattori, S., Iwasaki, K. and Kawakita, M. (1986) Nucleic Acids Res. 14, 6551-6564.
31. Jorgensen, E. D., Collins, P. L. and Lomedico, P. T. (1987) Virology 156, 12-24.
32. Kozak, K. (1981) Nucleic Acids Res. 9,5233-5252.
33. Kozak, K. (1986) Cell 44, 283-292.
34. Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132.
35. Hiebert, S. W., Paterson, R. G. and Lamb, R. A. (1985) J. Virol. 55, 744-751.
36. Kurilla, M. G., Stone, H. O. and Keene, J. D. (1985) Virology 145, 203-212.
37. Lepert, M., Rittenhouse, L., Perrault, J., Summers, D. F. and Kolakofsky, D. (1979) Cell 18, 735-747.
38. Shioda, T., Hidaka, Y., Kanda, T., Shibuta, H., Nomoto, A. and Iwaski, K. (1983) Nucleic Acids Res. 11, 7317-7330.
39. Spriggs, M. K. and Collins, P. L. (1986) J. Gen. Virol. 51, 515-524.
40. Gill, D. S. and Banerjee, A. K. (1985) J. Virol. 55, 60-66.
41. Walker, J. E., Saraste, M., Runswick, M. J. and Gay, N. J. (1982) EMBO J. 1, 645-951.
42. Fry, D. C., Kuby, S. A. and Milvan, A. S. (1986) Proc. Natl. Acad. Sci. USA 83, 907-911.
43. Wierenga, R. K., Terpstra, P. and Hol, W. G. J. (1986) J. Mol. Biol. 187, 101-107.
44. Garnier, J., Osguthorpe, D. J. and Robson, B. (1978) J. Mol. Biol. 120, 97-120.
45. Pabo, C. O. and Sauer, R. T. (1984) Annu. Rev. Biochem. 53, 293-321.