
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 [³²P] 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

Nucleic Acids Research

K E L I H V N H L I G H N L K D R E T I R S D T P F I Y S K R I F K D G A I L S
 CAAGGAATTAATCATCTCAATCATTGATGGCCATAATTTGAAGGACCGTGAACCATCAGGTCAGACACATTTCTCATATACAGCAAACGAATCTTCAAGATGGAGCAATCTCTAG
 2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520
 Q V L K N S S K L V M V S G D L S E N T V M S C A N I A S T V A R L C E N G L P
 TCAAGTCTCAAAAATCATCTAAATAGTATGGTGCAGGTGATCTCAGTGAANAACCGTAATGTCTGTGCCAACATTTGCCTACTGTAGCACCGCTATGGAGAACCGGGCTTC
 2530 2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640
 K D F C Y Y L N Y I M S C V Q T Y P D S E F S Y N N N S H P D L N Q S W I E D I
 CAAGGACTCTGTACTATTTAACTATATATAGTGTGGTGCAGACATCTTGAAGTCTCTCTCAACAACAATTCGACCCCGATCTTAACCAAGTCGTGGATGGAGGACAT
 2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750 2760
 S F V H S Y V L T P A O L G G L S N L O Y S R L Y T R N I G D P G T T A F A E I
 CTCTTTGGCACTCATATGTTCTGACTCTGCAATAGGGGACTTAGTAACTTCAATCTCAAGGCTCTACACTAGAAATATCGGTGACCGGGGACTACTGCTTTTGAGAGAT
 2770 2780 2790 2800 2810 2820 2830 2840 2850 2860 2870 2880
 K R L E A V G L S P N I M T N I L T R P P G N G D M A S L C N D P Y S F N P E
 CAAGGCTAGAGAAGTGGGATTAAGTCTTAACCTATAGTAACTTCTAACTAGGCGCGCTGGGAATGGAGATGGGCGAGTCTTCAACCGCCATCATCTTCAATTTTGA
 2890 2900 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
 T V A S P N I V L K K H T Q R V L P E T C S N P L L S G V H T E D N E A E E K A
 GACTGTGCAAGCCCAACATGTTCTTAAGAAACATACGCAAGAGTCTCTAATTTGAAACTTGTCAAAATCCCTTATGTCTGGAGTGCACACAGAGGATAATGGCGGAGAGGAC
 3010 3020 3030 3040 3050 3060 3070 3080 3090 3100 3110 3120
 L A E F L L N Q E V I H P R V A H A I N E A S S V G R R K Q I O G G L V D T T N T
 ATTGGCTGAAATCTTGTCTAATCAAGAGGATTCATCCCGGCTGGCATGCTATCATGGAGCAAGCTCTGTAGGTAGGAAAGCAAATTCAGGGGACTTGTGACACAACAGAGCC
 3130 3140 3150 3160 3170 3180 3190 3200 3210 3220 3230 3240
 V I K I A L T R R P L G I K R L M R I V M Y S S H H A M L F R D D V F S S N R S
 CGTAATAGATGCACTTACTAGGAGCCACTAGGCACTCAAGAGCTGATGGGATGCTCAATTTCTAGCATGGATGCAATGCTTTCAGAGCAATGTTTTTCTTCAATTTTGA
 3250 3260 3270 3280 3290 3300 3310 3320 3330 3340 3350 3360
 N H P L V S S N H C S L T L A D Y A R N R S H S P L T G C R K I L G V S N P D T
 CAACCCTTACTGTCTCTTAACTATGTTCTCTGACACTGGGAGACTATGCAAGGATGAAGCTGGTCACTTTCAGCGGGAGCGAGAAATCTGGGTGATCTAATCTGATCTAG
 3370 3380 3390 3400 3410 3420 3430 3440 3450 3460 3470 3480
 I E L V E G E I L S V S G C T R C D S G D E O P T M F H L P S N I E L T D D T
 GATAGACTGTAGAGGCTAGATTTCTAGTGAAGCGAGGGTGCACAAGATGGCCAGCGGAGATGAACAGCTTTACTGTTTCACTTTCAGCAAGATATAGAATTTGACCGATGACAC
 3490 3500 3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
 S K N P P H R V P Y L G S E T O E R R A A S L A R I A H S P H V K A A L R A S
 CAGCAAGATCTCGAGAGGAGTCAATCTCGGGTCAAGACACAGAGGAGGAGCTGCTCACTCGAAAATATGCTCATATCTCCCACTGTGAAGGCTCCCTTAGGGCTATC
 3610 3620 3630 3640 3650 3660 3670 3680 3690 3700 3710 3720
 S V L I W A Y G D N E V N M T A A L T I A K S R C H N I N L E Y L R L L S P L P T
 ATCCGTGTGATCTGGGCTATGGGATATGAAGTAATTTGGACTGCTCTCTAGCATTTGCAAAATCTCGATTAATAAACTTAGAGATCTCTGGCTTATGTCCTTCCCTTACCCAC
 3730 3740 3750 3760 3770 3780 3790 3800 3810 3820 3830 3840
 A G N L O H R L D D G I T O M T F T P A S L Y R C H L T F T Y P H I L K G Y S L
 CGCTGGGAATCTCAACATAGACTAGTATGGTAACTCAAGACATGACATTCACCCCTGCATCTCTACAGGCTGACCTTACATTTCAATATCAATGATTCTCAAGGCTATTCACT
 3850 3860 3870 3880 3890 3900 3910 3920 3930 3940 3950 3960
 K K P S R R G W F I H R V M L L G L S L I S I P M T T T R T Y D E I T L H
 GAAGAAAGTCAAAGAGGAAATGTGTTTATCAACAGAGTCACTGCTGGGTTTATCTCTAATCGAATGCTATTTCCAAATGACGCAACAGGATATGATGATGATGCAATGCA
 3970 3980 3990 4000 4010 4020 4030 4040 4050 4060 4070 4080
 L H S K F S C C I R E A P V A V P F E L L G V A P E L R T V T S N K F H Y D P S
 TCTACTAGTAAATTTAGTGTCTGCTATCAGGAAAGCCTGTGTGGGTTCTCTTCCAGCTACTTGGGGTGGCACCGGACTAAGGACAGCTACCTCAAAAGATTTATGATGATCTAG
 4090 4100 4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
 P V S E G D F A R L D L A I F K S Y E L N L E S Y P T I E L M N H I L S I S S G K
 CCCTGTATCGGAGGAGACTTTGGAGACTTGAATGATCTCTCAAGAGTATGATGCTAATCTGCAAGTATATCCACGATAGAGCTAATGACAACTTTTCAATATCCAGCGGAA
 4210 4220 4230 4240 4250 4260 4270 4280 4290 4300 4310 4320
 L I G O S V S Y D E D T S I E N D A I I V Y D N T R N W I S E A Q N S D G V R
 GTTGATGGCCAGTCTGGTTTCTAGTATGAAGTATCTCAATAAAGATGACGGCATAATAGTCTATGACAATATCCGAAATGGACAGTCAAGATTCAGAAATCCAGATGCGGTCCG
 4330 4340 4350 4360 4370 4380 4390 4400 4410 4420 4430 4440
 L P F Y A A L E V L L D C S Y O L Y L R V R G L D M I V L Y N G D L Y K N H P
 CTTATTTGAATGCAAGTGAAGTCTCTGAGCTTCTTACCACTTATATCTGAGAGTAAAGGCGCTAGACAATATTTCTTATATATGATGGGATTTATGCAAGAAATATGCC
 4450 4460 4470 4480 4490 4500 4510 4520 4530 4540 4550 4560
 G I L L S N I A A T I S H P V I H R L H A V G L V M H N G S N Q L A D T D F I
 AGGAATCTACTTCCAACTGACATATATCTCATCCGCTCATTCAGGTTACATGCAAGTGGCGCTGGTCAACCATAGCGGATCACACCACTGCGAGTACCGGATTTAT
 4570 4580 4590 4600 4610 4620 4630 4640 4650 4660 4670 4680
 E M S A K L L V S C T R R V I S G L Y S G N K Y D L L P P S V L D D N L N E K M
 CGAAATGCTCAAAATCTTAGTATCTTGCACCTGAGGCTGATCTCGGCTTATATTCAGGGAATAGTATGCTGTTCCCATCTGCTTGTAGATGATAAATCTCAATGAGAAT
 4690 4700 4710 4720 4730 4740 4750 4760 4770 4780 4790 4800

```

L Q L I S R L C C L Y T V L P 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
GCTTCAGCTGATATCCCGGTTATGCTGCTGTACACGGTACTCTTTGGCTACAACAAGAGAAATCCCGAAATAAGAGGCTTATCTCGCAGAGAGAAATGTCAGTACTACTGAGTATCT
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 P P A H L Y Y H S R K S L N L I
ACTGTCGGATGCTGTGAACACTACTACTCCGCTGATCAGGTGAGCTCTATCATGCTCTCCTAACTAATTAACATCCAGCTAATCTGTACTACATGCTCGGAGAGGCTCAATTTGAT
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 P P Q E P L L E F P S V Q D I G A R V K D P P T R O P
CAGGGAAGGGAGGACAAAGGATTCTATCTCTGGCGTGTGTGTCCTCCCAAGGCCATTATTAGAGTCCCTCTGTGCAAGATATTGGTCTCGATCGATGAAGATCATTACCCGACAAAC
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
TGGCCGATTTTGAAGAGTTAGATTTGAGTCTCCAGCAAGGTATGAGCATTACACTTAGTCAGATTATCTGAGCTCACAATCAACAAATCCGGAGAGGACTACTAGTACGATA
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
CTTGTTCAGAGAAATAGGACTGCATCTCTCTGGTATAGGCATCCCATCTCTTCTGTACCGAGGTAAAGATGTGCAAGACACGGGAATCTCTATACTAGTGAAGGAGCGG
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 P S N E M N P P Q R H F G P T P T Q F L
AGCCATCATGAGTCTTGAAGTGCATGTACACATGAACTATCTATACAATAGGCTCTTTTCAAATGAGATGAACCCCGGAGCGACATTTGGGGCGACCCCAACCCAGTTTTT
5410 5420 5430 5440 5450 5460 5470 5480 5490 5500 5510 5520

H 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
GAATTCGGTGTATTAAGAACCTACAGGGGAGGTAAACGTCAGAGTGGATTTGTCCAAGAGTCCGTCACATGGAGAGAAATACAGAGAGGAGGACCTGACTCAGATTAAGC
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
AGTGGGTATATACATCTGCAAGTGCATACAGATCTGATCATGCTGATGTGACATTTGAATCCCTCCAGGGTCCAATCAAGCTTACTAGATCAATAGCTATCAATTTATCTCT
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
GATTCGCATGATCCGTAAGGAGGGGGGGTATGATCAAAAGTGTGTATGGAATGGGACTACTTCTCACTCATGAGACTTGTTCGCTCCGTTCCCAAAAGGATACAT
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
TCTCTCAATGGTATGATGCTAGAGGGATATGAGTGTACTGATTTGCTCATGGTTACCTGGGGCGGCTACATTTGTACAGAGGTTGAGGATGGCAAAATCTCGGTGCA
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
CGGGACGGTACGCTTTTGCCAAATCAGATGAGATCAACATGACGAGTATTACCTCAACAGGCGAGCTGTGACAGACATCTATCAGTCTTTACCAAGATTAATAAGTACT
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 O I T O I A S
GAGAAGAATATGACTCGCTGATTAAGCTGGGGACACCCCTCCCTCAATCTGTGACAGAGTTGGTGACAGCCTGGGGACATAACTCAGATACCCAGATCATCTGTATG
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 C G K R T S L
TCACATGACACAGTACTCGGCTGTGTATATATGAACTGAGGGTGTACTCGTGACACAGTATTCTATTACCGCTTACATCTCTCTACTGACGGGAAAGAGAACATCAGT
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 C H I S H E D L
TAAACAGTCCAGACAGATCTAGAGGTTACAATATGGGCTTTAGAGTGAAGATCTCAATAAATAGGGCGATGAATCAGCTTACTGCTTAAAGCATGATCTCTGAGGAGCT
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 R L R E H F T D T S V L Y L T R A
TATCCCAATGACATACTTAAGCATAGTACTCGCTCAATTAATGAGGCTGTCTAGGTATTACCAAACTCAAGAANTGTTTACAGACACTCTGTATTGACTTGAATCTGCTC
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
TCAACAAAATTTACATGAAAATATAGGCAATGCAAGGATATACAGTAATCTGACTTTAACGAAAATCACAATATAATAGGCTCTTTCTGGCCAATGATCTCTGGT
6610 6620 6630 6640 6650 6660 6670 6680 6690 6700 6710 6720

GATTTAATATACTATCTTAGAAAAAATGAAGCTCCGACTCCTTAGAGCTCGAATTCGAAGTCAAAATAATGCTCT
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 polyadenylation 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 ZnCl₂, 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

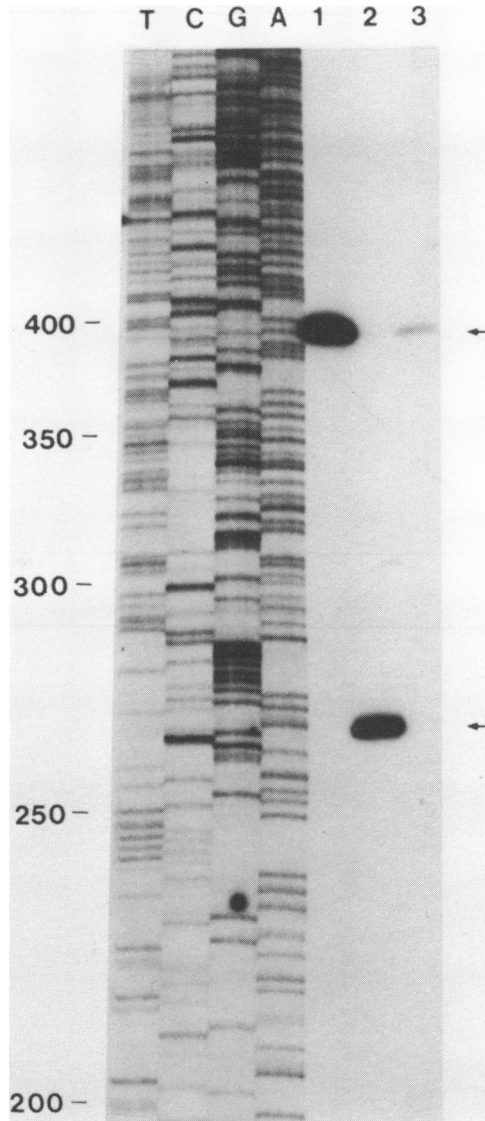


Figure 2. Mapping of the 5' terminus of the L mRNA. Lane 1: the original ^{32}P 5' end-labelled, 376 bp *HincII-XbaI* restriction fragment (see Methods), indicated by upper arrow. Lane 2: restriction fragment annealed to the RNA from NDV-infected CEF cells, mung bean nuclease protected fragment of 264 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.

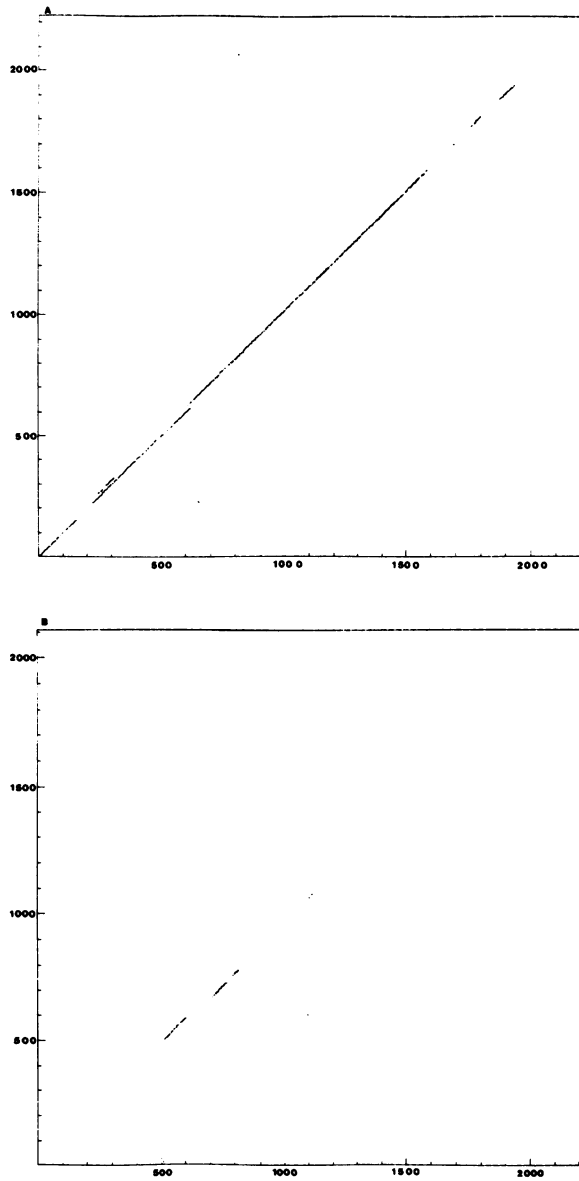


Figure 3. Dot matrix homology plots of the L amino acid sequences. Comparisons of NDV with Sendai virus and VSV are shown in (a) and (b), respectively. In both cases a window of 99 and a proportional score of 1025 were used with the DAIGON computer program (Staden, 1984).

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.

Nucleic Acids Research

1	1	M D G Q E S S S Q N P S D T L L Y P E C H L N S P I V R G K I A Q L H V L L D V N Q P Y R L X D P H I L I N I T K H K I R N G	Sen
181		H A S S G P E R A E H Q T L L P E S H L S P C L L Y Y W K L T G L P L P D E C D F H H I L I S R Q W K K I L E	NDV
61		G L S P R Q I X I R S L Q K A L Q R I I K D L D R Y T F E P Y P T Y S Q E L L R L D I P E I C D K I R S V F A V S D R L	Sen
61		S A S P D T E R M I K L G R A V H S R I T L N H N S R I T C V L H P R C L E E L A S I E V E P D S T N K F R K I E K K I Q I H	NDV
121		T R E L S S Q F Q D U L W L N T F K R Q G N I E G H E G Y D P L Q D I G T I P E I T D K Y S R N R I Y R P P L W F S I R	Sen
121		N T R Y C E L C T R D L W L N T F K R Q G N I E G H E G Y D P L Q D I G T I P E I T D K Y S R N R I Y R P P L W F S I R	NDV
181		Y D M R W M Q K T R P G G P L D T S M S H N L T E C R S Y T L V T Y G D L V Y M I L N K L L T L G Y I L T P E L V L M Y C	Sen
181		Q I Q R H L I V A A R T S R A A M K L V M L T H N K Y G Q V F T P E L V Y V T H T N E N K F C L T O E L V L M E A	NDV
226		H D R N F L L H V K	VSV
241		D V Y E C R W N M S A A G H L D K K S I C I T S K G E E L W E L V D S C F S S L C E E I Y H V I A T L E P L S L A	Sen
239		D H M E G R D M V N I I S T T A V H L R S L S E K I I D D T L Q L I D A L A K D L G N Q V Y D V V S L H E F I A Y G	NDV
236		D T I T G C N H Q T V L S H V C R I D W L F S E Q D I F S L L N L I Y A T G D K I V E R G C H F S Y D T T K H V E F I C N L	VSV
298		L I G L N D P V I P L R G A F M H R H V L T E L Q T V T I S R D Y Y T D A E A D T I V E S L L A F H G T S I D E K A E	Sen
296		A V G L L E S G R E A G H F F A F N L Q E L K D I L I G L P N D I A E S V T H A I A T V F S G L E Q N Q A A E	NDV
296		K L N K L A R E S R P V P Q P P H F E N H I K T S V D E G A K L T R G R F L H D Q H S S V K T V D L T L V	VSV
357		I F S F E R T F G H F S L E A V A T A D K V R A H M Y A Q A I L K L T L Y E C H A V F C T I I N G Y R R H G C Q V	Sen
353		H L C L L R T Y G H P L E D S R T A A K A V R S Q M C A P K M Y F D M I L O V L S F E K G T I I N G Y R K K N A G V W	NDV
351		I Y G S F R H M G H P I D Y Y C G E K L H S Q V T H K K D I D V S Y A K A L A S L A R V L F Q Q N D H X V L	VSV
417		P C D F P P H V C L E L R N A Q G S N T A T S Y E C A V D M Y T S F I G F K F R K X I J E P Q L D E D L T Y H M K D K	Sen
413		F N W K V P T I Y G K I I G Q L M A D S A E L S H D I N T R E L K S L S A D E P F I S Y D V T N L S H F L K D K	NDV
410		F Y M K D L L P H D H P P G S H W K E N T W P T A C Q V Q D F G D K W H R L P L T C G P E I P D L D P S T I Y S D K	VSV
476		A L S P R K E A T D S Y P D S N Y Y K A P S E E F R L L E V F L N D E N F N P E I I N V Y E S G D W L	Sen
472		J I A H P N D N W L A S F R R N L S I S E D Q K K H V K E A T S I T R L L I E L E S N D D P Y K E M E Y L T L D E L	NDV
469		S H S H N R S E V L K H V R H M N P T P I P S K V L Q T H L D T K A T W K E L G I D E K C	VSV
532		K E E F N L S Y S L K E K E I R Q E G R L F A K M T Y K K L R N A V Q V L A B I L L A G K G I C E F R E N G G V Y K G E D	Sen
532		R D D D V A V I Y S S L K E K E V I W N G R I F A K L T Y K K L R N C Q V M A E G I L A D Q I F A F F G G N G V I Q D S S L	NDV
519		D D D L T G L K Q K E R E L K L A C H N F E S L S W K L R E Y F V I T E Y I T K H F A V L P L A D D L T A	VSV
592		L L V R L T L S V S G V P R T D S Y N M S K S S E K R N E C H E V K K S G G Y W D E K K R S R H E F K A T D S S T	Sen
592		L T K S L L A M S Q L S F N S N K K R I T D C K E R V S S N R H H D P K S K N R R R	NDV
579		V I K K M L D S S S G G L K S	VSV
651		D G Y E T L S C F L T T D L K Y C L N W R F E S T A L F P G Q R C H E I F D F K T F F N W H M P V L B R C T I Y V G D P	Sen
634		F N P S D P T D C D L S R V F N D D L Y V S A R G C E G C O K L W T L S I S A I H L A A A B S H C R V A C H V	NDV
595		E A T C I A N H T Y E E N N H Q K L S N G E P V R M G Q F L S L T R T R I H F P F K S T I Y T Y G R	VSV
711		Y C F V A D R M R H Q Q D H A D S G P F J H N P R G C T E L C O K L W T L S I S A I H L A A A B S H C R V A C H V	Sen
689		F N P S D P T D C D L S R V F N D D L Y V S A R G C E G C O K L W T L S I S A I H L A A A B S H C R V A C H V	NDV
653		P D H R Y H N N T L I N S T S O R V C V Q G E G C O K L W T L S I S A I H L A A A B S H C R V A C H V	VSV
771		Q G D N Q A I A V T S R V P V A Q T Y K Q K K R H V Y E E I T K Y F G A L R H V H F D V G R E K L N E T I S S	Sen
749		Q G D N Q A I A V T R E V R S D D S P R M V L T Q L H Q A S D W F K Q L I H M L I G H M L K D G T R I R D	NDV
712		Q G D N Q V I C T Y K T K S R S D S P R M V L Q A L Y H M V S N N E M I H T I K I G T G K L G L I W D E T H S A	VSV
828		K M F V Y S K R T Y D K C A I L P Q C L K A L K C V M S E I T V D E N R S C S N S S I S L A K A I E	Sen
806		T F E I Y S K R I T R E V R S D D S P R M V L T Q L H Q A S D W F K Q L I H M L I G H M L K D G T R I R D	NDV
772		D Y L N Y G C K T P I F R E V R S D D S P R M V L T Q L H Q A S D W F K Q L I H M L I G H M L K D G T R I R D	VSV
882		N G Y S P I L G Y C I A L Y K T Q Q V C I S L G M T I N P T I S F I T V R D Q Y F G K G N W L R C A V I F A N Y G	Sen
859		H G L P K D F C Y T L N Y I M S Q Q T Y F P S E F S Y N N N S P D L N G S W H E D I S F V H S Y L T P A Q L G	NDV
828		H F I N A M I Q Y I F C T A R F A L L L M H D P A L R O S L Y E V Q D K I F G L I S D T E K Y A M L T L D P S I G	VSV
939		G F Y H S T S R L C F V R N I C D P A V A L A L D K R F I R A D L D K Q V L Y N M N Q E F G D S S F L	Sen
917		G L S N L Q V S R L Y T R N I C D P G T A F A E I K R L E A V G L L S P N I H T N I L T R F F C H G D N A	NDV
886		G V S G H S L S R F L I A F P D P V T E S S F W R P T H V H A N S E H L K E H S A V F G N E I A K F R I T H D	VSV
993		D W A S D F Y S C H L P H S Q S I T T I K N I A R S V L Q E S P N P L L S G L F E T S G E E D L N A S F L	Sen
971		S T C N U P S P N F E T V A S F N I V L K K H T Q R V T F E T C S N P L L S G V H T E D N E A E K A L A F L F L	NDV
945		K L Y E D P T S L N I A M G M S F A N L L K T E V K K C L I E S R Q T I R N G V I K D A T I Y L Y H E E D R L R S F L	VSV
1050		M D R K V I L P R V A H E T L G N S L T G V R E A T A G H L D T T K S L V R A S Y R K G C L S L G I L R R L V N Y D L L	Sen
1028		L N Q E V I H P R V A H T I M E A S S V C G R K Q I O G L V D T N T V T E C X L T R F L G C K R L M H V W Y S E M	NDV
1004		W S I N F L P R F L S E F K S O T F L G T A D G L I S L F Q W S R V T I M S P F K K T H R H L D D L T Y R E	VSV
1110		Q Y E L L T R L R K P Y K D N I E Y E Y M C S V E L A V G E L R Q K M H I H L T Y G R P T H G L E T P F L E L R	Sen
1088		H A M I T F R D D V F S S N R S M H P L V S S N M C S L L A D Y A R N B S W S P L T G G R K T I L G V S N P T T E L E E	NDV
1060		V S S L T H L G K L H L R H G S C H M W T C S A T R A D T L R Y K S W G R T V I Q T T V P R F L E M D	VSV
1168		G I F I E G S E V C K L C A S E G A P I Y T W F L P D N I D L D T L T I N C G P A I R I F Y L G S A T D E R S A A Q	Sen
1148		G E I L S V S G C G K R C D S G D E Q F T W F H L P S H I E L T D D T I S K N P P H A V P Y L L G S K T Q E R S A A Q	NDV
1111		G P Q H R K E T G A P N T S G I F M Y V S U C P D Q I H D V F S S R G P L A Y L G S K T E S T S I L Q	VSV
1228		G Y V R N L S L P A K A A I R A M V Y T W A Y G T D F I S M H E A L I A T Q I R A N L E L K L L V S T S T	Sen
1206		A K I A H M S P H V K A A L R A S S V L I W A Y G D N E V N W T A A L T I A K S R C N I L E Y L R L L S P L P T A G	NDV
1166		P W E R E S K V L I R A L R A D A I S L W F V E P D S K L A M T I L S N I H S T C T E G M T K R Q H G F K R T G	VSV
1287		N L S R R L K D T A T Q H M F S A S I V R A S R F I I S N D M H A L K E A C E S D I N L V Y Q Q H L T G L S L	Sen
1245		N L Q H R L D D G I T Q D T F T A S L Y B C H L T F I Y P H I L K G Y S L K K E S R G M V I N R V L L G L S L	NDV
1224		S A L H R F S T S R M S H G G F A G S	VSV
1346		F E F N H R Y K K G S L G K P L I L H L H L N N G C C I H E S P Q E A N I P P R S T L D L E I T T E N K L I Y D P D P	Sen
1324		I E S I F P H T T T Y D E I T L H L H S K F S C C I R E A D V A V P F L L G V A P E L R T V G S N X K I Y D P D P	NDV
1406		L K D D L E L F S K V R D V Y T D M T Y S D D I V I R A T S I C T A M T I A D T H S O L D R D N L K E I A L V	Sen
1384		V S E G D F A R L D L A I F V S Y E L N L E S Y P T I L M N I L S I S S C K L I Q Q S V S I T Q E R D S T I K N D A I V	NDV
1466		N D D D V N S L T I F E M V I D V Y L F C S T F G G I L V N Q F A T S L Y Q L N I R G R E I W G H V V R L K D T S H	Sen
1444		V Y D N T R N V I S E A C T S D V Y L F F E X A A L E V L D D C S Y Q L Y Q L N I R G R E I W G H V V R L K D T S H	NDV



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.

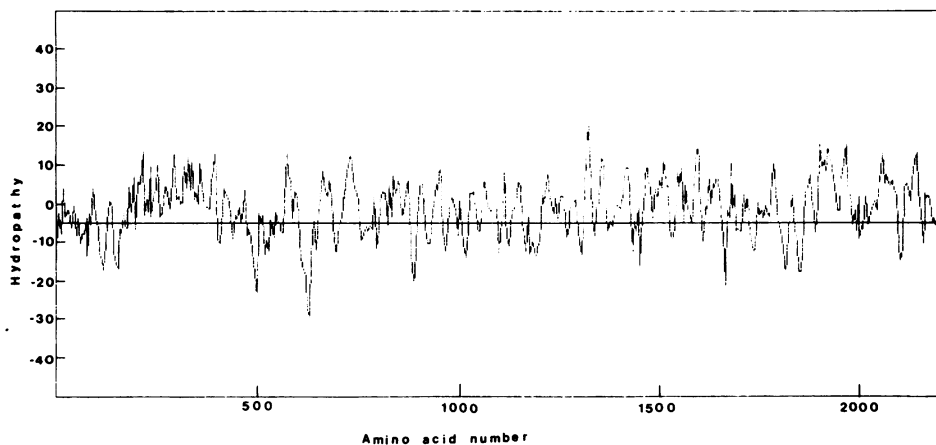


Figure 5. Hydropathy plots of the NDV L amino acid sequence. Hydrophobic regions are above and hydrophilic regions below the horizontal line, which is the average hydropathy of a large number of sequenced proteins.

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

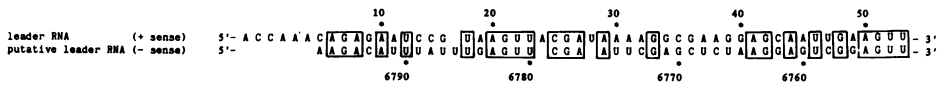


Figure 6. Comparison of positive strand leader RNA sequence with putative negative strand leader RNA.

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 Iwasaki, 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.