Conservation of Potential Phosphorylation Sites in the NS Proteins of the New Jersey and Indiana Serotypes of Vesicular Stomatitis Virus

By BRENDAN P. RAE AND RICHARD M. ELLIOTT*
MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, U.K.

(Accepted 18 March 1986)

SUMMARY

A full length cDNA copy of the NS mRNA of the Missouri strain (Hazelhurst subtype, New Jersey serotype) of vesicular stomatitis virus (VSV) has been cloned and sequenced. The mRNA is 856 nucleotides long (excluding polyadenylic acid) and encodes a protein of 274 amino acids (mol. wt. 31000). Comparison with the NS gene of the Ogden strain (Concan subtype, New Jersey serotype) showed 15% difference at the nucleotide level and 10% difference at the amino acid level; the majority of the changes were located in the 3' half of the mRNA. Comparison with the NS genes of two strains representing the Indiana serotype showed about 50% nucleotide and 33% amino acid sequence homology between the serotypes. In a four-way comparison of the proteins, two regions of higher homology were noted which may be of functional importance. Eighteen potential phosphorylation sites (Ser or Thr) were conserved between the four proteins; five of these sites correspond to the residues which have been suggested to be constitutively phosphorylated and may be essential for NS activity.

INTRODUCTION

The NS (so-called non-structural) protein of vesicular stomatitis virus (VSV) is multifunctional and has been shown to play a role in mRNA transcription, genome replication and virion maturation (for review, see Pringle, 1986). In the virus particle NS is complexed with the L protein and is closely associated with the negative strand RNA–nucleocapsid (N) protein complex. NS is a phosphoprotein with a molecular weight of about 30000 (Gill & Banerjee, 1985) although in SDS–polyacrylamide gels its migration is anomalously slow probably due to its highly acidic nature (Marnell & Summers, 1984). It is documented that the degree of phosphorylation of NS plays a regulatory role in transcription in vitro (Kingsford & Emerson, 1980; Hsu et al., 1982) and the ability to bind nucleocapsids (Clinton et al., 1978), and recently Sánchez et al. (1985) reported that the viral L protein specifically phosphorylates NS in vitro.

Detergent treatment of VSV particles yields a transcribing ribonucleoprotein complex (TNP), comprising the virion RNA and L, N and NS proteins, which retains essentially all the transcriptase activity of the virion (Bishop & Roy, 1972; Emerson & Wagner, 1972; Szilágyi & Uryvayev, 1973). Fractionation and reconstitution experiments with TNP have been used in an attempt to define the roles of the different proteins in transcription but have generated conflicting results. Ongrádi et al. (1985a, b) have suggested that L alone is the transcriptase while NS exerts a controlling effect, whereas the laboratories of both Emerson and Banerjee have reported that a combination of L and NS is required to synthesize RNA in vitro (Emerson & Yu, 1975; Mellon & Emerson, 1978; Kingsford & Emerson, 1980; De & Banerjee, 1984, 1985).

As an alternative approach to studying the role of NS we undertook to clone a cDNA copy of the NS mRNA with the long term view to express the cDNA in a eukaryotic system. For these studies we chose the Missouri strain of the New Jersey serotype of VSV because temperature-sensitive mutants of this strain with lesions in the NS gene had previously been isolated and characterized in our Institute (Pringle et al., 1971; Evans et al., 1979). Here we report the cloning...
of a full length cDNA of the Missouri strain NS mRNA, its complete nucleotide sequence and a comparison with the NS genes of three other VSV strains which indicates conservation of certain potential phosphate-accepting amino acids.

**METHODS**

**Virus and cells.** The Missouri strain of VSV New Jersey serotype was obtained from J. F. Szilágyi. Virus stocks were prepared and titrated in BHK-21 cells as described by Pringle et al. (1971).

**Production of infected cell RNA.** BHK cells were infected with 10 p.f.u./cell of VSV in the presence of 10 μg/ml actinomycin D. After adsorption for 1 h at 31 °C, the monolayers were washed with medium and incubated for 6 h at 31 °C in the presence of actinomycin D. The cells were then harvested by scraping and washed with phosphate-buffered saline. Total cellular RNA was prepared by the guanidinium isothiocyanate/caesium chloride method of Chirgwin et al. (1979).

**Synthesis of complementary DNA.** Double-stranded cDNA was synthesized according to the method of Gubler & Hoffman (1983), using 50 μg total cellular RNA as template, 1 μg of oligo(dT)$_{12-18}$ as primer and 40 units of reverse transcriptase in a 40 μl first-strand reaction. Double-stranded cDNA was tailed with dCMP residues, annealed to dGMP-tailed PstI-cut pBR322 DNA (Bethesda Research Laboratories), and recombinant plasmids used to transform competent Escherichia coli MC1061 (Pringle et al., 1984).

**Identification of NS gene-specific recombinant plasmids.** Tetracycline-resistant, ampicillin-sensitive transformed E. coli were initially screened with a 32p-labelled, partially alkali-digested vesicular stomatitis virion RNA probe by colony hybridization (Pringle et al., 1984) to identify VSV cDNA-containing clones. NS gene-specific plasmids were identified by a combination of Northern blot hybridization and message selection/in vitro translation of infected cell RNA using methods detailed by Pringle et al. (1984) and Elliott (1985).

**Nucleotide sequence determination and analysis.** The NS gene cDNA insert was excised from the pBR322 vector by PstI digestion, and purified by agarose gel electrophoresis. The insert was further digested with HpaII, HaeIII or HinfI and ligated into bacteriophage M13mp18 DNA. Nucleotide sequence determination followed the dideoxy chain termination method of Sanger et al. (1977, 1980). Sequence data were manipulated and analysed with a DEC PDP11 computer as described previously (Lees et al., 1986).

**RESULTS AND DISCUSSION**

**Cloning and identification of NS gene cDNA**

Total cellular RNA extracted from actinomycin D-treated, VSV-infected BHK cells was used as the template for cDNA synthesis. With oligo(dT)$_{12-18}$ as primer and the reaction conditions of Gubler & Hoffman (1983) double-stranded DNA was produced and, following dCMP tailing, the cDNA was annealed to dG-tailed, PstI-cut pBR322 plasmid DNA. Colony hybridization with 32P-labelled, partially alkali-digested vesicular stomatitis virion RNA probe by colony hybridization (Pringle et al., 1984) to identify VSV cDNA-containing clones. NS gene-specific plasmids were identified by a combination of Northern blot hybridization and message selection/in vitro translation of infected cell RNA using methods detailed by Pringle et al. (1984) and Elliott (1985).

**Nucleotide sequence determination and analysis.** The NS gene cDNA insert was excised from the pBR322 vector by PstI digestion, and purified by agarose gel electrophoresis. The insert was further digested with HpaII, HaeIII or HinfI and ligated into bacteriophage M13mp18 DNA. Nucleotide sequence determination followed the dideoxy chain termination method of Sanger et al. (1977, 1980). Sequence data were manipulated and analysed with a DEC PDP11 computer as described previously (Lees et al., 1986).
Fig. 1. Characterization of a full length cDNA clone of the NS mRNA of the Missouri strain of VSV. (a) Northern blot analysis. Total RNA extracted from VSV-infected cells was fractionated by formaldehyde-agarose gel electrophoresis and transferred to a nitrocellulose membrane (Pringle et al., 1984). The filter was hybridized with $^{32}$P-labelled plasmid DNA: (1) pNJN1, reacting with the N mRNA; (2) pNJNS1, reacting with either the NS or M mRNA; (3) a marker track of $^{32}$P-labelled VSV mRNAs. (b) Message selection/in vitro translation. Plasmid DNA was bound to nitrocellulose and hybridized to total infected cell RNAs. The hybridized RNAs were eluted and used to programme a rabbit reticulocyte in vitro system. The translation products were analysed by polyacrylamide gel electrophoresis. (1) Selected with pBR322 DNA; (2) selected with pNJNS1 DNA; (3) selected with pNJN1 DNA; (4) translation in vitro of infected cell RNA; (5) marker track of $[^{35}S]$methionine-labelled VSV proteins. VSV proteins are indicated at the right. pG is the in vitro equivalent of the VSV G protein. The solid arrowheads indicate the translation products of the selected mRNAs. (c) Agarose gel electrophoresis of a partial PstI digest of pNJNS1. (1) HindIII-digested λ DNA markers; (2) partial PstI digest of pNJNS1 showing the intact insert (in1; approx. 900 bp) and the complete PstI digestion products in2 and in3. pBR is the pBR322 vector DNA.
Fig. 2. Nucleotide sequence and deduced amino acid sequence of the NS mRNA of the Missouri strain of VSV. The complete sequence of the NS mRNA was obtained by subcloning fragments of the insert of pNJNS1 into M13mp18 DNA and using the methodology of Sanger et al. (1977, 1980). The sequence is compared with that of the Ogden strain (Gill & Banerjee, 1985) and differences are indicated below the Missouri strain sequence.

**Nucleotide sequence of the NS mRNA of VSV (Missouri strain)**

The two insert fragments of pNJNS1 were digested with *HpaII*, *HaeIII* or *HinfI*, subcloned into bacteriophage M13mp18 and their sequence was determined by the dideoxy chain termination method (Sanger et al., 1977, 1980). The sequence spanning the internal *PstI* site was...
obtained by digesting intact pNJNS1 with HpaII, isolating the appropriate fragment from a polyacrylamide gel and ligating the purified fragment into M13mp18 DNA.

The complete sequence of the insert from pNJNS1 is shown in Fig. 2, and reveals that pNJNS1 contains a full length cDNA copy of the NS mRNA because (i) at the 5' end of the insert, following a string of approximately 25 G residues (not shown on figure) is the sequence AACAGAGATC, which corresponds to the consensus sequence 5'-AACAGNNAUC-3' reported to be at the 5' end of all VSV mRNAs (McGeoch, 1979; Rose, 1980; Franze-Fernandez & Banerjee, 1978), and (ii) at the 3' end there is the sequence TATG followed by a polyadenylate stretch of 12 residues [corresponding to the oligo(dT)12-18 primer] before a 'tail' of approximately 25 C residues (not shown in the figure). Rose (1980) and Rhodes & Banerjee (1980) reported that the sequence... UAUGAAAA... occurs at the mRNA–poly(A) junction in each mRNA. Furthermore, the sequence of approximately 200 bases at the 5' end of the NS mRNA is identical to the sequence obtained independently by D. J. McGeoch and A. Dolan by primer extension on the viral genomic RNA from within the N gene, across the intergenic region and into the NS gene (D. J. McGeoch, personal communication).

The NS mRNA is 856 nucleotides long excluding the 3' poly(A) tract, and encodes a protein of 274 amino acids (mol. wt. 31000) beginning at AUG (bases 11 to 13) and terminating at UAA (bases 833 to 835). The next largest AUG-initiated open reading frame in either the (+) [mRNA]- or (−) [genomic RNA]-sense RNA would encode a protein of 65 amino acids, beginning at bases 51 to 53 in the mRNA. The amino acid composition of the NS protein indicates that the protein is relatively rich in aspartate and methionine residues, and relatively deficient in alanine and glycine residues compared to an 'average' protein (Klapper, 1977), i.e. Asp 9.2% compared to 5.5%, Met 4.2% compared to 1.7%, Ala 3.5% compared to 9.0%, and Gly 4.2% compared to 7.5%. There are a total of 46 potential phosphorylated residues in the NS protein (33 Ser and 13 Thr). The NS protein has an overall charge of −18.5 at pH 7.0, assuming +1 charge for Lys and Arg, +0.5 for His and −1 charge for Asp and Glu at this pH.

Comparison of VSV NS genes and proteins

Nucleotide sequences of the NS genes of four strains of VSV, representative of the New Jersey (NJ) and Indiana (IND) serotypes, are now available, and a summary of the comparisons of these sequences is given in Table 1.

The mRNAs of the New Jersey serotype strains are both 856 nucleotides in length (excluding polyadenylic acid) compared to 814 nucleotides for the Indiana serotype strains. There is about 50% nucleotide homology between the NS genes of the two serotypes and the distribution of homology is displayed as a computer-generated dot matrix in Fig. 3 (a). This shows that there are only short regions of significant homology scattered throughout the gene sequences. For comparison the nucleotide sequences of the G and N genes of the NJ and IND serotypes show 54% and 68% homology respectively (Gallione & Rose, 1983; Banerjee et al., 1984).

The two NJ strains are 85% homologous at the nucleotide level, and as shown in Fig. 2 the majority of the nucleotide differences occur in the 3' half of the mRNA. A limited comparison of 151 bases at the 5' end and 208 bases at the 3' end of the N protein mRNAs of the Missouri and Ogden strains showed a similar degree of homology (D. J. McGeoch, personal communication; Pringle, 1986).

Comparison of the amino acid sequences of the NS proteins showed about 33% homology between the two serotypes, although the dot matrix comparison (Fig. 3b) indicated a clear similarity between the two proteins. The conservation of the NS proteins is thus considerably lower than that of the G proteins (51%) or the N proteins (69%) of the two serotypes (Gallione & Rose, 1983; Banerjee et al., 1984). There are 27 amino acid differences between the Missouri and Ogden strains (Fig. 2), notably a stretch of six contiguous mismatched residues at positions 190 to 195. The 10% difference in the amino acid sequences between the NS proteins of the Missouri and Ogden strains [which represent the Hazelhurst and Concan subtypes of the NJ serotype (Reichmann et al., 1978)] manifests itself as a difference in the electrophoretic mobility of the proteins in SDS–polyacrylamide gels (Schnitzlein & Reichmann, 1985). The only other available comparison of VSV gene sequences is from Gopalakrishna & Lenard (1985) who compared the
Table 1. **Comparison of the NS genes and proteins of four strains of VSV**

<table>
<thead>
<tr>
<th>Serotype</th>
<th>Subtype</th>
<th>Strain</th>
<th>Length of mRNA (no. of nucleotides)</th>
<th>Size of protein (no. of amino acids)</th>
<th>Homology (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Jersey</td>
<td>Hazelhurst</td>
<td>Missouri (MIS)</td>
<td>856</td>
<td>274</td>
<td>MIS 100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Concan</td>
<td>856</td>
<td>274</td>
<td>100</td>
</tr>
<tr>
<td>Indiana</td>
<td>Hazelhurst</td>
<td>Ogden (OG)</td>
<td>814</td>
<td>265</td>
<td>100</td>
</tr>
<tr>
<td>Cocal</td>
<td>Mudd–Summers (M-S)</td>
<td>San Juan (SJ)</td>
<td>814</td>
<td>265</td>
<td>100</td>
</tr>
</tbody>
</table>

Reference: This paper, Gill & Banerjee (1985), Gillione et al. (1981), with corrections reported in Gill & Banerjee, 1985, Hudson et al. (1986).
VSV NS genes

Fig. 3. Computer-generated dot matrix comparisons of (a) the nucleotide sequences and (b) the amino acid sequences of the NS genes of the Missouri strain (New Jersey serotype) and the San Juan strain (Indiana serotype) of VSV. In (a) a window of 12 nucleotides was used and in (b) a window of 7 amino acid residues was used. Regions of homology are denoted by a diagonal line, and the level of homology by the thickness of the line (60% homology is the minimum plotted).

M genes of three strains of VSV IND (namely San Juan, Glasgow and Orsay), and reported 1.8% nucleotide and 2.6% amino acid difference between the three strains.

The classification of VSV isolates into the IND and NJ serotypes was based on the finding of little or no reciprocal cross-neutralization of infectivity (Cartwright & Brown, 1972). The IND serotype was further subdivided into four subtypes (Indiana, Argentina, Brazil and Cocal) and the NJ serotype into the Hazelhurst and Concan subtypes on the basis of neutralization, RNA hybridization, oligonucleotide fingerprinting and interference by defective interfering particles (Cartwright & Brown, 1972; Clewley et al., 1977; Crick & Brown, 1973; Federer et al., 1967; Reichmann et al., 1978; Repik et al., 1974). It is apparent from the literature that the various isolates of VSV have been passaged many times in different laboratories (see Clewley et al., 1977; Reichmann et al., 1978) and hence conclusions drawn from sequence comparisons of these laboratory-passaged strains may not be strictly extrapolated to the original isolates. However, the information given above suggests that the IND serotype contains more closely related strains than the NJ serotype. It will be interesting to see whether this trend is maintained when sequence data from other strains become available. However, it is worth noting that Schnitzlein & Reichmann (1985) regard the Hazelhurst subtype as 'an uncommon manifestation of the New Jersey serotype' because of its apparently more limited host range, and this may account for the greater sequence divergence than in the IND serotype.

Fig. 4 is a four-way comparison of the amino acids of the NS proteins which have been aligned using the HOMOL program of Taylor (1984). Three regions of the sequences are of note. First, the regions designated A and C which have relatively high homology (62% and 68%) compared to the overall homology of 33%. The conservation of the amino acid sequences in these regions suggests that they may be of functional significance. Second, the region designated B where there is the most difference between the two NJ strains, and a deletion in the IND sequences was introduced in order to align the proteins. This suggests that this domain is a non-essential or dispensable portion of the protein. It would be of considerable interest to determine whether deletion of this region in the NJ protein affected its function.

Eighteen potential phosphorylation sites (Ser and/or Thr) are conserved between the four NS proteins. This represents 48% of the sites in the Mudd–Summers strain which contains the fewest Ser or Thr residues. None of the aligned potential phosphate acceptors is found in the
Fig. 4. Alignment of the NS proteins of four strains of VSV using the HOMOL program of Taylor (1984). The shaded regions denote amino acids conserved between the four strains and the filled circles denote conserved potential phosphorylation sites (Ser or Thr). The regions A, B and C are described in the text. OG, Ogden strain, NJ serotype; MIS, Missouri strain, NJ serotype; SJ, San Juan strain, IND serotype; M-S, Mudd-Summers strain, IND serotype.

conserved carboxy terminal domain (region C in Fig. 4). Phosphorylation of the NS protein has been studied in a number of laboratories by analysis of the products generated by enzymic or chemical cleavage of the protein. Hsu et al. (1982) reported that NS1 (the less phosphorylated class of NS molecules) was phosphorylated primarily at residues in two chymotryptic peptides, while the more highly phosphorylated molecules contained additional chymotryptic phosphopeptides. Studies by Marnell & Summers (1984) and Bell & Prevec (1985) indicated that the majority of phosphate residues resided in the amino terminal portion of the NS molecule, and Hsu & Kingsbury (1985) localized these residues to a peptide lying between amino acids 35 and 78. As shown in Fig. 4, our comparison reveals five potentially phosphorylated residues in this region which are conserved in the four VSV strains. Hsu & Kingsbury (1985) suggested that the same five sites were constitutively phosphorylated in NS and might be essential for a basal level of NS activity. The availability of a full length cDNA clone for expression in eukaryotic cells and the advent of oligonucleotide mediated site-directed mutagenesis (Zoller & Smith, 1982) will enable us to test this hypothesis experimentally.

We are most grateful to Duncan McGeoch for helpful discussions and access to his unpublished sequence data, and to Bob Lazzarini for supplying the Mudd-Summers sequence prior to publication. We thank Ulrich Desselberger and Duncan McGeoch for their comments, and Linda Shaw for typing the manuscript. B.P.R. received a postgraduate studentship from the Science and Engineering Research Council.

REFERENCES


---

**Text Input**

| NJ OG | 1 MDSVDRKTYLATYDNLDSALQDENESERREDKYLQDLFIEDQGDKPTPSYYQEEESSDSDTDYNAEHLTML |
| NJ MIS | 1 MDSVDRKTYLATYDNLDSALQDENESERREDKYLQDLFIEDQGDKPTPSYYQEEESSDSDTDYNAEHLTML |
| IND SJ | 1 MDNLTVKREVLYKSDQVEIDEIBAQRKSNENLFOEDGVEEHXTFSPQTTAEDDSITSEPEIDNQL |
| IND M-S | 1 MDNLTVKREVLYKSDQVEIDEIBAQRKSNENLFOEDGVEEHXTFSPQTTAEDDSITSEPEIDNQL |

**Text Content**

- **OG**, Ogden strain, NJ serotype; **MIS**, Missouri strain, NJ serotype; **SJ**, San Juan strain, IND serotype; **M-S**, Mudd-Summers strain, IND serotype.

---

**Fig. 4. Alignment of the NS proteins of four strains of VSV using the HOMOL program of Taylor (1984). The shaded regions denote amino acids conserved between the four strains and the filled circles denote conserved potential phosphorylation sites (Ser or Thr). The regions A, B and C are described in the text. OG, Ogden strain, NJ serotype; MIS, Missouri strain, NJ serotype; SJ, San Juan strain, IND serotype; M-S, Mudd-Summers strain, IND serotype.**


(Received 17 January 1986)