Arch Virol (1997) 142: 979–992

Archives of Virology

Polymorphism of rabies viruses within the phosphoprotein and matrix protein genes

S. A. Nadin-Davis, W. Huang*, and **A. I. Wandeler**

Rabies Unit, Animal Diseases Research Institute, Agriculture and Agri-Food Canada, Nepean, Ontario, Canada

Accepted September 11, 1996

Summary. Although the rabies virus P and M genes, encoding the viral phosphoprotein and matrix protein respectively, have been characterised for a few laboratory-adapted strains, there is essentially no information on the variability of these two genes in wild-type rabies viruses. In this study rabies viruses, responsible for epizootics in different wildlife species in three geographically distinct areas of North America, have been characterised at the P and M gene loci. These data reveal that the M gene and its encoded product are much more conserved than the P gene and its encoded phosphoprotein. The latter product harbors two variable domains which contribute to different hydropathy profiles for this protein for each of the rabies virus strains studied. Phylogenetic analysis of the nucleotide sequences generated in this study, together with data generated previously on the N and G genes of these rabies virus strains, indicated that similar evolutionary relationships are predicted regardless of the portion of the viral genome targeted.

Introduction

Historically studies on the lyssavirus genus, of which rabies virus is the representative type, have focused on two antigens, the nucleoprotein encoded by the N gene and the glycoprotein encoded by the G gene. The high immunogenicity of these two antigens has facilitated the production of monospecific antibodies directed against them. Antigenic variants of both of these proteins have been identified with these reagents thus forming the basis of rabies virus strain distinction [27]. The nucleoprotein is the most highly conserved of all rabies antigens [27] and is an important component of the nucleocapsid material targeted by diagnostic methods such as the fluorescent antibody test

* Present address: Genentech Inc., San Francisco, CA, U.S.A.

[34]. The glycoprotein, the sole surface antigen of the viral particle, is the only viral component to elicit the production of neutralising antibodies [10] and is therefore an important component of all rabies vaccines.

On the basis of antigenic variation, lyssaviruses were divided into four serotypes and two unclassified groupings of European bat lyssaviruses EBLsl and 2 [2, 28]. Molecular characterization of the rabies virus genome (reviewed in [36]) has facilitated the development of genetic typing methods. Phylogenetic analysis of representative lyssaviruses using nucleoprotein sequence data identified six distinct genotypes [3]. Within serotype 1, which includes all rabies viruses, variation at the N gene locus has been investigated using isolates representative of most major rabies epizootics around the world and the phylogenetic relationships of these strains have been examined [16, 29]. Some studies have reported on the sequence of the G gene for certain laboratory adapted and wild-type strains [1, 37] and phylogenetic relationships have been proposed based on sequence data at this locus [31] and at the G-L intergenic region [26].

In contrast, the phosphoprotein and matrix proteins, encoded respectively by the P (formerly NS) and M genes located within the viral genome between the N and G genes, have until recently received limited attention. Studies of these proteins have been hampered by the limited availability of monoclonal antibodies directed against them. Indeed, the description of a panel of 21 monoclonal antibodies directed against the matrix protein of the Nishigahara strain of rabies virus is the most extensive yet to be described [14]. The phosphoprotein is believed to constitute a part of the viral replicase activity by interacting with the L gene product [36] and specific binding of the P gene product to the nucleoprotein has been documented [6, 11]. The matrix protein is located inside the envelope of the viral particle where it likely interacts with the cytoplasmic domain of the glycoprotein [36]. Evidence indicating an inhibitory effect of the matrix protein on virion-associated transcription was recently reported [15].

The variability of the P and M genes of wild-type rabies viruses has not been examined; sequence information for these genes is reported for only a small number of laboratory-adapted strains [13, 17], and a single non-rabies lyssavirus, Mokola virus, which is representative of serotype 3 [3]. To further knowledge of variation within the P and M genes and their respective products, we report on the characterisation of these two genes for isolates of three distinct rabies virus strains. These include specimens of the so-called arctic strain of rabies virus, which circulates in red and polar foxes in the North American arctic and in red foxes in southern Ontario [30], skunk rabies from western Canada which is closely related to the skunk rabies strain prevalent in the northern central states of the U. S. [4] and isolates of the raccoon strain which has spread throughout the raccoon population of the entire eastern U. S. seaboard (reviewed in [35]). Features of these genes and their encoded products are described. In addition the combination of these data with sequence information already generated in previous studies allow a comprehensive

examination of the utility of several alternative portions of the viral genome for phylogenetic analyses.

Materials and methods

Rabies virus isolates

Most isolates have been described in previous studies and are summarised in Table 1. The six representative specimens of the arctic fox strain which originated from Ontario and arctic Canada were characterised previously at the N and G gene loci (20, 21, manuscript in preparation). The two western Canada skunk rabies virus isolates were obtained from the rabies diagnostic unit of ADRI, Lethbridge, Alberta, Canada. The two raccoon strain specimens were representative of isolates from New York and Florida [22]. RNA was extracted directly from the brain tissue of infected animals as previously described [21].

Reverse transcription – polymerase chain reaction (RT-PCR)

Amplification primers for the P and M genes (see Table 2) were designed as far as possible using sequence information already obtained for the N and G genes of the viral strains under study. The complete contiguous target sequence was defined by primers RabPfor (+ sense) and RabMrev (− sense) the sequence of which reside in the N and G genes respectively. Two internal PCR primers, RabPrev (− sense) and RabMfor (+ sense), were designed based on information available for laboratory-adapted strains by selecting sequences encoding relatively conserved protein sequences and/or amino acid residues having limited codon degeneracy. In all cases cDNA synthesis was primed with appropriate positive sense PCR primer using 2 μg RNA. Conditions for the reverse transcriptase reaction and subsequent PCR were essentially as detailed previously [20, 21].

Table 1. Summary of rabies virus isolates employed in these studies

Isolate	Strain/variant	Reference
ONT1	Arctic fox T1	[20], manuscript in prep.
ONT2	Arctic fox T2	
ONT3	Arctic fox T3	
ONT4	Arctic fox T4	
ONT5	Arctic fox T5	[21]
ARC5	Arctic fox T5	
WCS1	Western Canada skunk	this study
WCS2		
NYRAC	Mid-Atlantic raccoon	[22]
FLRAC	Southern raccoon	
PV	–	[33]
SAD_{B19}	–	[7]
CVS	–	[18] (N)
		[17] (P)
		[13] (M)
		[25] (G)
MOKOLA	–	3

Table 2. Oligonucleotides used for RT-PCR and sequencing of the P and M genes

Primer name	Position/sense	Sequence (5′–3′)
RabPfor	1249–1272 +	CTACTTCTCCGGGGAAACCAGAAG
RabPrev	2554–2531 −	GAGG(GA)TTTTTGAGTGTCCTCGTC
RabMfor	2381–2403 +	GACTTGAATCGCTATACATCTTG
RabMrev	3339–3318 −	ACAAAAGAGCCTGAGGAACCAT
Pseq1	1450–1472 +	TCCATAGATTGTGTATATCCTTC
Pseq2	1637–1659 +	GAACCTATAGAGGTGGACAATCT
Pseq3	1828–1850 +	CAGACAAGTGAGGTCAGGGGAGA
Pseq4	1999–2021 +	CCAGAGAGACAGTCAATCCTCGA
Pseq5	2197–2219 +	GCAGTTAAAAATGAACCTTGATG
Pseq6	2392–2369 −	GCGATTTAGATCGTCTTGCATGAT
Pseq7	2219–2196 −	CATCAAGGTTCATTTTTAACTGCT
Pseq8	1993–1971 −	AACGGATGTTGTCTCCTTCTTGA
Pseq9	1785–1763 −	AGGCTAGGATCCTCCCCTTCATC
Pseq10	1914–1893 −	GGGAAATTGATCATGACATAGG
Mseq1	2573–2595 +	AGATGACGATGACCTATCGCTCC
Mseq2	2831–2854 +	GGTATACAAGTTGAGGAGAACCCT
Mseq3	2840–2818 −	CTTGTATACCCAGTTCATGCCCT
Mseq4	3107–3085 −	TGATCATTCTAGAAGCAGAGAAG
Mseq5	2754–2774 +	CAAAGAATGATTGGGTTAGTC

Oligonucleotides designated Rab... were used for RT-PCR and sequencing; other primers were used for sequencing only. Positions are indicated with respect to the PV strain (33)

Amplification products were visualised by ethidium bromide staining and UV irradiation after electrophoresis of PCR aliquots through 1% agarose.

DNA sequencing

Prior to consensus sequencing, PCR products were recovered, either directly from the amplification reaction or occasionally subsequent to purification using low melting point agarose, using a Wizard PCR purification kit (Promega). Cycle sequencing was performed with a fmol sequencing kit (Promega) using either amplification or internal primers (see Table 2) which were end-labelled using gamma [^{32}P]-ATP and polynucleotide kinase. IBI Pustell software was used to translate nucleotide sequences to protein and to analyse the hydrophilicity profiles of these gene products. Nucleotide and amino acid sequences were aligned using the CLUSTALW package [12] and phylogenetic analysis was performed on 1000 bootstrap replicates of aligned sequences using the PHYLIP3.5 package [9]. Distance values were determined from aligned sequences using DNADIST. Neighbor joining (NJ) was performed using the DNADIST and NEIGHBOR programs; analysis by the maximum parsimony (MP) approach was performed using DNAPARS. In each case the consensus tree was identified using CONSENSE.

Results

Characterisation of P and M genes of several rabies virus strains

The RabPfor/rev and RabMfor/rev primers (see Table 2), designed as indicated, were initially evaluated on the arctic fox strain of rabies and found to generate

PCR products of the expected sizes from all five genetic variants of this strain. Subsequently, PCR products were obtained from rabies virus isolates of both the U.S. raccoon strain and the western Canada skunk specimens.

From the nucleotide sequence data generated from these PCR products, distance values between all wild-type strains and laboratory isolates under consideration were determined for both the P and M gene coding regions. The similarity values calculated for the rabies P gene varied from 99.6% for the two western Canada skunk specimens down to 70.3% between the New York raccoon and the SAD strains. Between the Ontario and arctic variants of the Canadian fox strain similarity was about 97%, When the Mokola sequence was included in this comparison similarity values dropped to as low as 21.3% (with the NY raccoon strain specimen). Similarity values for the M gene locus were generally higher, ranging from 100% for the western Canada skunk isolates down to 73.6% between the PV and Florida raccoon strains; between the Ontario and arctic variants of the Canadian fox strain similarity was about 98.3%. Between Mokola and all rabies viruses similarity at this locus was around 58-60% except for the raccoon strain for which it was higher (65%). The raccoon strain was clearly the most distant of all the rabies viruses examined. Between these coding regions, the sequences identified as the transcriptional start and stop signals [3] were highly conserved but the remaining non-coding sequences were quite divergent. In all cases the ORFs of both genes were preserved and translations of the P and M genes predicted the primary structure of their respective products for all isolates. These protein sequences are compared in Figs. 1 and 2 respectively to those sequences determined previously for laboratory-adapted rabies and Mokola virus isolates. Similarity values for the phosphoprotein ranged from $> 98\%$ for sequences of the same strain down to around 80% for sequences of different rabies virus strains but dropped to approximately 25% upon inclusion of the Mokola virus sequence. Most notably for this protein, amino acid differences tended to cluster to certain regions. For example, the N-terminal 50 amino acids of the phosphoprotein were well conserved in all rabies viruses and moderately so even when the Mokola sequence was considered. However, the next 35 residues were far more variable. Following a small conserved domain a second variable region was located between residues 130–180. The C-terminal region of the phosphoprotein contained only a few amino acid differences; however for both raccoon strain isolates this protein was longer by four residues (301 amino acids) compared to other rabies viruses (297 amino acids). In contrast the matrix protein was quite conserved between all viruses with a similarity value of 70% or greater between any sequence pair and intra-strain similarity approaching 100%. Amino acid residue differences were scattered throughout the sequence of this protein.

To examine the potential effect of these variations in protein primary structure on protein folding patterns, the hydrophilicity profiles of both the P and M gene products for several viral strains were determined. Surface probability profiles for the phosphoprotein of PV, arctic fox (T1), raccoon,

```
                    *         *         *         *         *        60
P.PV        MSKIFVNPSAIRAGLADLEMAEETVDLINRNIEDNQAHLQGEPIEVDNLPEDMGRLHLDD
P.SAD       ............................................................
P.CVS       ...........................................................K......
P.ONT1      .........................K.....................KQ.Q...
P.ONT2      .........................K.....................KQ.Q...
P.ONT3      .........................K.....................KQ.Q...
P.ONT4      .........................K.....................KQ.Q...
P.ONT5      .........................K.....................K..Q...
P.ARC5      .........................K.....................K..Q...
P.WCS1      ...........................................................R......
P.WCS2      ...........................................................R......
P.NYRAC     .........................A.................S..D..R.....N
P.FLRAC     .........................A.................S..D..R......
P.MOK       ...DL.H..L....IVE.......T.....T..S.........LY..S.....S..RIE.

                    *         *         *         *         *        120
P.PV        GKSPNPGEMAKVGEGKYREDFQMDEGEDPSLLFQSYLDNVGVQIVRQIRSGERFLKIWSQ
P.SAD       .....H..I.....................F......E.........M............
P.CVS       E..S.L...VR..................N.................M............
P.ONT1      R.PSGL....RA....C....L.......................M............
P.ONT2      K.PSGL....RA....C....L.......................V............
P.ONT3      K.PSGL....RA....C....L.......................V............
P.ONT4      K.PSGL....RA....C....L.......................V............
P.ONT5      K.PSGL....RA....C....L.......................M............
P.ARC5      K.PSGL....RA....C....L.......................M............
P.WCS1      ...S.L.........R.................L.............M......F.....
P.WCS2      ...S.L.........R......I.........L.............M......F.....
P.NYRAC     E.PSGFDKVTEE..S.CH..........................M............
P.FLRAC     E.PSGFDKVT.E..S.CH..........................M............
P.MOK       KSRRTKT.EEERD..SSE..NYLS..Q..LIP..NF..EI.ARA.KRLKT..G.FRV..A

                    *         *         *         *                   *
P.PV        TVEEIISYVAVNFPNPPGKSSEDKSTQTTGRELKKETTPTPSQRESQ----SSKARMAAQ
P.SAD       ............................................................
P.CVS       .....V...T.......RR.................SAF..........P.....V..
P.ONT1      .........T.........L.................SIS...D..............
P.ONT2      .........T.........................SVS...D............V..
P.ONT3      .........T.........L...............SVS...D............V..
P.ONT4      .........T.........................SVS...D............V..
P.ONT5      .........T.........L...............SIS...D............V..
P.ARC5      .........T.........L...............SIS...D............V..
P.WCS1      .........M.....S..R............VP.....S...................
P.WCS2      .........M.....S..R............VP.....S.....D.............
P.NYRAC     .........MI...GSL.RP....A...AN..PR.GVASVS..L.G..........AS..
P.FLRAC     .........MI...GSL.RP....A....N..P..GVASVS..L.G..........V...
P.MOK       LSDD.KG..ST.IMTSGER--DT..I.IQTEPTASVSSGNE.RHD.ESMHDPNDKKDHTP

             180        *         *         *         *         *
P.PV        TASGPPALEWSATNEEDDLSVEAEIAHQIAESFSKKYKFPSRSSGILLYNFEQLKMNLDD
P.SAD       I...........................................................
P.CVS       V.P.........................................F.............
P.ONT1      A.........................V.................F.............
P.ONT2      A.........................V.................F.............
P.ONT3      A...........N.............V.................F.............
P.ONT4      A.........................V.................F.............
P.ONT5      A.........................V.................F.............
P.ARC5      A.........................V.................F.............
P.WCS1      A...........................................F.............
P.WCS2      A...........................................F.............
P.NYRAC     ...........................................F.............
P.FLRAC     ...........................................F.............
P.MOK       DHDVV.DI.S.TDKG.-IRDI.G.V...V................F.W.........

             240        *         *         *         *         *
P.PV        IVKEAKNVPGVTRLARDGSKLPLRCVLGWVALANSKKFQLLVESNKLSKIMQDDLNRYTS
P.SAD       ..............H.............................D.............
P.CVS       ..............H....I.......................AD.............
P.ONT1      ..............H.............G............PD..N..........A.
P.ONT2      ..............H.............G............PD..N..........A.
P.ONT3      ..............HE............G............PD..N..........A.
P.ONT4      ..............H.............G............PD..N..........A.
P.ONT5      ..............H.............G............PD..N...........
P.ARC5      ..............H.............G............PD..N...........
P.WCS1      ..............H.........................AD..N..........A.
P.WCS2      ..............H.........................AD..N..........A.
P.NYRAC     ...GS.........HE..................R.......PD..N..........L.
P.FLRAC     ...GS.........HE..................R.......PD..N..........L.
P.MOK       ...A.M.....E.I.EK.G......I..F...DS..R.R..ADND.VARLI.E.I.S.MA
```

```
          297
P.PV        C-----
P.SAD       ......
P.CVS       ......
P.ONT1      ......
P.ONT2      S.....
P.ONT3      S.....
P.ONT4      N.....
P.ONT5      S.....
P.ARC5      S.....
P.WCS1      ......
P.WCS2      ......
P.NYRAC     RQLNL.
P.FLRAC     RQLNL.
P.MOK       RLEEAE
```

Fig. 1. Comparison of the predicted amino acid sequence of the phosphoprotein for thirteen rabies virus isolates. Isolates are indicated thus: PV, PV strain (33), SAD, $SADB_{19}$ strain (7), CVS, CVS-11 strain (17), ONT1-5, arctic fox strain isolates recovered in Ontario and representing variants 1–5 respectively, ARC5, arctic fox strain isolate recovered from the arctic, WCS1 and WCS2, two western Canada skunk isolates, NYRAC and FLRAC, raccoon strain isolates from New York and Florida respectively, Mok, Mokola virus (3). Sequences were aligned using CLUSTALW; ... indicate that residue is conserved with the PV sequence; - - - indicate gaps in the alignment

western Canada skunk and Mokola strains are shown in Fig. 3. These profiles are similar over much of the protein's length with the exception of the two most variable domains described above. Within the first variable domain all strains exhibited distinct profiles. The raccoon rabies and Mokola strains generated distinct profiles in the second variable domain, whereas the three other rabies strains (ie. PV, western Canada skunk and arctic fox) generated profiles that were quite similar in this region. When surface probability profiles were determined for the matrix protein sequences of these same five viruses, all profiles were virtually identical (data not shown).

Phylogenetic analysis using different genomic regions

The nucleotide sequences generated in this study facilitated a comprehensive analysis of the utility of different portions of the genome sequence for assessing phylogenetic relationships of rabies virus strains. The coding sequences of four genes ie. N, P. M and G (less sequence encoding the signal peptide) for all isolates under study were analysed by both NJ and MP methods. Since sequence information on the N and G gene loci of the western Canada skunk strain was not previously available, these genes were also amplified and sequenced as detailed previously [20, 21]. Nucleotide sequence homologies to the PV strain were approximately 93% and 91% for the N and G genes respectively and the encoded proteins differed from those of the PV strain by various single but unremarkable substitutions.

All four coding regions generated phylogenetic trees of similar topologies with both methods of analysis so just one typical tree, generated from NJ analysis of the M gene sequences, is shown (Fig. 4). Clearly three main branches of rabies viruses were well supported. The first branch included all

```
                    *         *         *         *         *        60
M.PV        MNFLRKIVKNCRDEDTQKPSPVSAPLDDDDLWLPPPEYVPLKELTSKKNRRNFCINGGVK
M.SAD       ..L.......R.......S..A.....................G...M.......R..
M.CVS       ..V......K..............PY.....................M....V..E..
M.ONT1      ....................L....P.....................G...L.......E..
M.ONT2      ....................L....P.....................G...L.......E..
M.ONT3      ....................L....P.....................G...L.......EI.
M.ONT4      ....................L....P.....................G...L.......E..
M.ONT5      ....................L....P.....................G.N.L.......E..
M.ARC5      ....................L....P.....................G.N.L.......E..
M.WCS1      ....................T...P..............R.......M.......E..
M.WCS2      ....................T...P..............R.......M.......E..
M.NYRAC     ..............D....LA...P.....................G...M....V..E..
M.FLRAC     ..............D....LA...P.....................G...M....V..E..
M.MOK       ....K.MI.S.K..E...YPSA...P....I.M........TQVKG.ASV.....S.E..

                    *         *         *         *         *       120
M.PV        VCSPNGYSFGILRHILRSFDEIYSGNHRMVGLVKVVIGLALSGAPVPEGMNWVYKLRRTL
M.SAD       .........R......K............I.............S.....L.........F
M.CVS       A........R......G..N.........I......V.......................
M.ONT1      .........R.....................I.............................
M.ONT2      .........R.....................I.............................
M.ONT3      .........R.....................I.............................
M.ONT4      .........R.....................I......V......................
M.ONT5      .........R.....................I.............................
M.ARC5      .........R.....................I.............................
M.WCS1      .........R.....................I.............................
M.WCS2      .........R.....................I.............................
M.NYRAC     .........R......K.........Q..I......V...F....L..............
M.FLRAC     .........R......K.........Q..I......V...F....L..............
M.MOK       I........K......K...NV....R..I.........V...S................

                    *         *         *         *         *       180
M.PV        IFQWADSRGPLEGEELEYSQEITWDDNTEFVGLQIRVSAKQCHIRGRIWCINMNSRAGQL
M.SAD       ............................D..........I......Q..V......P..C..
M.CVS       ............................D..........G.R....Q.......S....C..
M.ONT1      ............................D............R....Q............C..
M.ONT2      ............................D............R....Q............C..
M.ONT3      ............................D............R....Q............C..
M.ONT4      ............................D............R....Q............C..
M.ONT5      ............................D............R....Q............C..
M.ARC5      ............................D............R....Q............C..
M.WCS1      .....N......................D............R....Q............C..
M.WCS2      .....N......................D............R....Q............C..
M.NYRAC     ............................D............R....Q............C..
M.FLRAC     ............................D............R....Q............C..
M.MOK       .....E.H....................EA...........R....Q..L.........C..

                    *        202
M.PV        WSDMSLQTQRSEEDKDSSLLLE
M.SAD       .......................
M.CVS       .......................
M.ONT1      .A.....................
M.ONT2      .......................
M.ONT3      .......................
M.ONT4      .......................
M.ONT5      .......................
M.ARC5      .......................
M.WCS1      .......................
M.WCS2      .......................
M.NYRAC     ........H.........V...
M.FLRAC     ........H.........V...
M.MOK       .A..I....Q.PD.ENT.....
```

Fig. 2. Comparison of the predicted amino acid sequence of the matrix protein for thirteen rabies virus isolates. Sequences were aligned with CLUSTALW; isolates and details are as indicated in Fig. 1

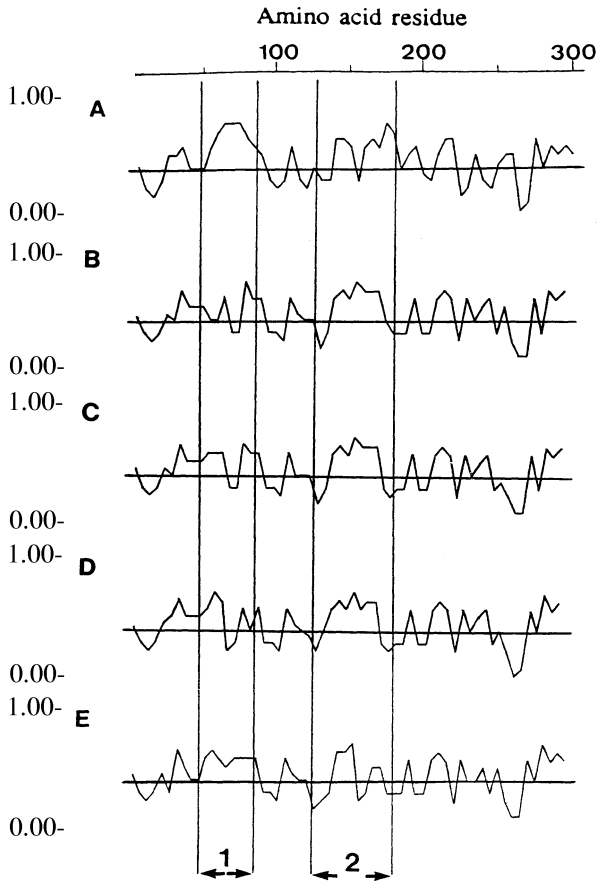

Fig. 3. Surface probability profiles predicted for the phosphoproteins of five lyssaviruses. **A** Mokola virus, **B** PV strain, **C** WCS1 isolate (western Canada skunk), **D** ONT1 (arctic fox strain), **E** NYRAC (mid-Atlantic raccoon strain). The primary sequence of the phosphoprotein of each isolate was analysed with IBI Pustell software to determine the hydrophilicity profile and hence predict surface probability plots. The closer the value of any point to 1.0, the higher the probability that the residues lie at the surface; conversely values approaching zero represent residues very unlikely to be located on the protein's surface. The two most variable domains are highlighted by vertical lines and marked 1 and 2 respectively

laboratory-adapted strains included in these studies together with the western Canada skunk isolates. The second branch corresponded to all isolates of the arctic fox strain with a consistent bifurcation between type 5 variants and other isolates. The third group comprised the raccoon strain and the Mokola virus was assigned as the outgroup. Confidence values for each branch approached 100% whilst within branches these values were generally >70%; the branch containing ONT variants 1–4 was the exception where, despite similar relationships being predicted for these isolates in virtually all trees, confidence values were sometimes lower. It was also noted that confidence in the position of the CVS strain was consistently lower than for other members of this branch.

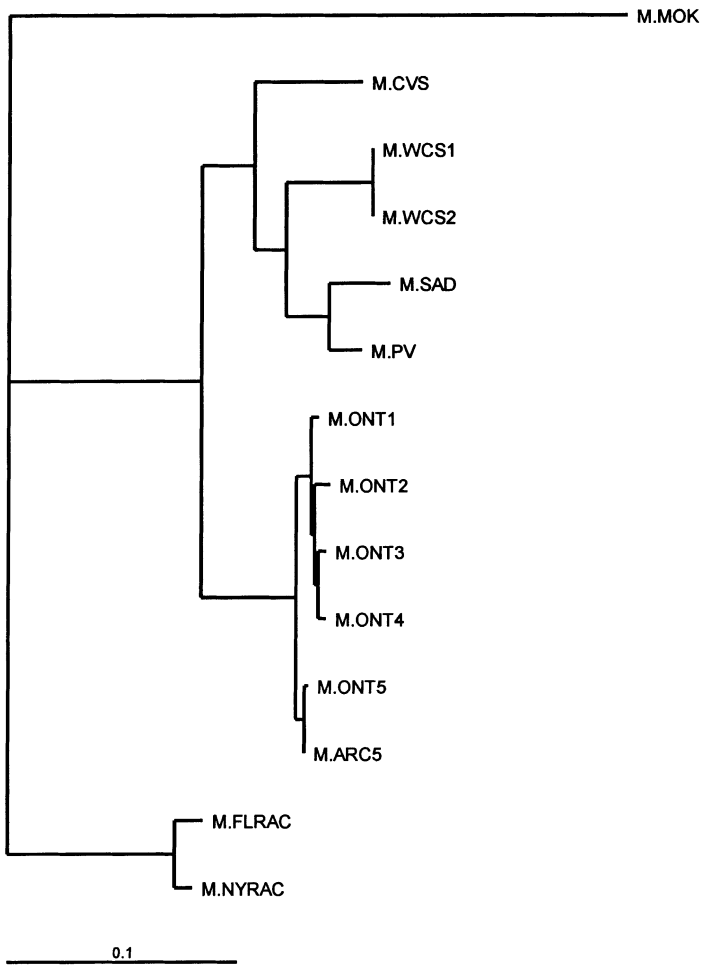

Fig. 4. Phylogram generated using M gene ORF sequences for 13 rabies viruses and Mokola virus. Strain designations are as for Fig. 1. Nucleotide sequences aligned using CLUSTALW were analysed by NJ using the PHYLIP3.5 package as described to generate a consensus tree. For accurate distance representation, distances were reapplied to this tree using the FITCH package and the original unbootstrapped distance data. The resulting tree was rerooted with the Mokola virus as outgroup and displayed with the help of the TREEVIEW package [24]

A similar analysis was performed for all isolates, except CVS for which the data were unavailable, on sequence from the non-coding P-M region. In order to generate trees with reasonably high confidence values additional flanking sequences either side of the non-coding region (240 bp total length) had to be employed; trees similar to that in Fig. 4 were then generated by both methods.

Discussion

The primary objective of these studies was to examine the degree of variation exhibited by distinct strains of wild-type rabies viruses at the P and M gene loci.

Although these genes had been characterised for a few laboratory-adapted rabies virus strains previously [13, 17], phylogenetic analyses performed on the G-L intergenic region of several of these isolates suggested many of them to be more closely related than historical records might suggest [26]. Thus the degree of diversity exhibited by these genetic loci in wild-type viruses was unclear. Isolates of three North American rabies virus strains differing in primary host and geographical origins were chosen for this study.

We have shown that the M gene and its encoded matrix protein are well conserved in several distinct rabies virus strains. The similar higher order structure predicted for all matrix proteins studied is in accord with, but extends, the observation that the matrix proteins of the CVS and Nishigahara strains exhibit similar hydropathy profiles [13]. The P gene of these same rabies virus strains exhibited much lower % similarity than the M gene. Indeed the identity of the P gene as the most poorly conserved of all lyssavirus genes was indicated previously [32, 19]. The present studies confirm these findings within serotype 1 viruses and identify specific protein domains, designated here as variable domains 1 and 2, within which phosphoprotein variability is concentrated. Whilst the surface probability profiles for these rabies virus phosphoproteins were similar over much of their length, thus suggesting a comparable overall structure for these proteins, the dissimilarities indicated for the variable domains may indicate localised variations in three-dimensional structure within these regions. Since even minor differences on amino acid residue placement within a protein can alter protein: protein interactions, the small variations identified here might influence phosphoprotein function in a strain specific manner. Clearly, however, further exploration of the functions of the phosphoprotein and in particular the variable domains described here will be required to address this issue.

Attempts to map functional domains of the phosphoprotein have to date been restricted to localisation of the portions of the protein which bind to the viral nucleoprotein and thereby modulate replication/transcription activities. Two independent groups have reported mapping two distinct sites on the phosphoprotein responsible for such interaction. In both cases a C-terminal region was identified; Fu et al. [11] suggested this domain was located in the C-terminal 47 residues whilst Chenik et al. [6] concluded that the sequence responsible was contained in the C-terminal 30 residues. There was however disagreement on the probable location of the N-terminal binding domain: Fu et al. proposed it to be within the first 19 amino acids whilst Chenik and colleagues localised it to residues 69–177. The role played in this regard by the two variable domains described here thus remains inconclusive. The mapping of one antigenic site on the ERA strain phosphoprotein to residues 75–90 [8], which overlap the most 5′ variable region described here indicates that this domain of the phosphoprotein is immunogenic and structural analysis through epitope mapping may therefore be feasible. Chenik et al. [5] have reported the use, by the CVS strain when grown in cell culture, of a total of five inframe initiating codons within the P gene thus facilitating synthesis of five proteins of

various lengths. The first three of these methionine residues, but not the last two, are conserved in the phosphoprotein of all rabies viruses examined here. However it is presently unknown whether truncated versions of this protein are made by wild-type rabies viruses in vivo.

The sequence data compiled during these studies gave us the opportunity to examine methodically the use of different portions of the rabies virus genome for phylogenetic analysis. Five regions were examined corresponding to the coding regions for four genes and to a non-coding region (P-M intergenic region). Regardless of the locus examined or the method of analysis employed, provided a sufficiently long sequence window was employed in order to generate trees with high confidence values, very similar relationships were predicted both within and between all branches. Isolates that are epidemiologically more closely related exhibit the smallest degree of divergence as expected. The consistent dichotomy between ONT 1–4 isolates and the specimen from the arctic, ARC5, suggests that southern and northern populations of this strain have evolved independently for some time following the movement of rabies into Ontario from the arctic in the 1950s. We interpret the close association of ARC5 with the ONT5 specimen, found during a recent outbreak of rabies in northern Ontario, as probably reflecting recent movements of northern variants of the arctic strain into southern Ontario [21]. The more distant raccoon strain constitutes a distinct branch amongst the rabies strains examined. The demonstration that phylogeny prediction is independent of the region of the viral genome employed agrees with and extends the observations of Nel et al. [23] on the use of two distinct genomic regions for phylogenetic analysis of rabies viruses.

Acknowledgements

We thank C. Trimarchi and W. Bigler for gifts of the raccoon strain isolates used in these studies. Western Canada skunk isolates were provided by K. Loewen and J. Kush. We also acknowledge the excellent technical assistance of M. Prud'homme. These studies were supported in part by a contract from the Ontario Ministry of Natural Resources.

References

1. Benmansour A, Brahimi M, Tuffereau C, Coulon P, Lafay F, Flamand A (1992) Rapid sequence evolution of street rabies glycoprotein is related to the highly heterogeneous nature of the viral population. Virology 187: 33–45
2. Bourhy H, Kissi B, Lafon M, Sacramento D, Tordo N (1992) Antigenic and molecular characterization of bat rabies in Europe. J Clin Microbiol 30: 2 419–2 426
3. Bourhy H, Kissi B, Tordo N (1993) Molecular diversity of the lyssavirus genus. Virology 194: 70–81
4. Charlton KM, Webster WA, Casey GA, Rupprecht CE (1988) Skunk rabies. Rev Infect Dis 10: S626–S628
5. Chenik M, Chebli K, Blondel D (1995) Translation initiation at alternate in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a ribosomal leaky scanning mechanism. J Virol 69: 707–712

6. Chenik M, Chebli K, Gaudin Y, Blondel D (1994) In vivo interaction of rabies virus phosphoprotein (P) and nucleoprotein (N): existence of two N-binding sites on P protein. J Gen Virol 75: 2 889–2 896
7. Conzelmann K, Cox JH, Schneider LG, Thiel H (1990) Molecular cloning and complete nucleotide sequence of the attenuated rabies virus SAD B19. Virology 175: 485–499
8. Dietzschold B, Lafon M, Wang H, Otvos Jr L, Celis E, Wunner WH, Koprowski H (1987) Localization and immunological characterization of antigenic domains of the rabies virus internal N and NS proteins. Virus Res 8: 103–125
9. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap Evol 39: 783–791
10. Flamand A, Raux H, Gaudin Y, Ruigrok RWH (1993) Mechanisms of rabies virus neutralisation. Virology 194: 302–313
11. Fu ZF, Zheng Y, Wunner WH, Koprowski H, Dietzschold B (1994) Both the N- and the C-terminal domains of the nominal phosphoprotein of rabies virus are involved in binding to the nucleoprotein. Virology 200: 590–597
12. Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTALV: improved software for multiple sequence alignment. CABIOS 8: 189–191
13. Hiramatsu K, Mannen K, Mifune K, Nishizono A, Takita-Sonoda Y (1993) Comparative sequence analysis of the M gene among rabies virus strains and its expression by recombinant vaccinia virus. Virus Genes 7: 83–88
14. Hiramatsu K, Mifune K, Mannen K, Nishizono A, Kawano H, Ito Y, Kawai A (1992) Mapping of the antigenic determinants recognised by monoclonal antibodies against the M2 protein of rabies virus. Virology 187: 472–479
15. Ito Y, Nishizono A, Mannen K, Hiramatsu K, Mifune K (1996) Rabies virus M protein expressed in *Escherichia coli* and its regulatory role in virion-associated transcriptase activity. Arch Virol 141: 671–683
16. Kissi B, Tordo N, Bourhy H (1995) Genetic polymorphism in the rabies virus nucleoprotein gene. Virology 209: 526–537
17. Larson JK, Wunner WH (1990) Nucleotide and deduced amino acid sequences of the nominal nonstructural phosphoprotein of the ERA, PM and CVS-11 strains of rabies virus. Nucleic Acids Res 18: 7 172
18. Mannen K, Hiramatsu K, Mifune K, Sakamoto S (1991) Conserved nucleotide sequence of rabies virus cDNA encoding the nucleoprotein. Virus Genes 5: 69–73
19. Mebatsion T, Cox JH, Conzelmann KK (1993) Molecular analysis of rabies-related viruses from Ethiopia. Onderstepoort J Vet Res 60: 289–294
20. Nadin-Davis SA, Casey GA, Wandeler A (1993) Identification of regional variants of the rabies virus within the Canadian province of Ontario. J Gen Virol 74: 829–837
21. Nadin-Davis SA, Casey GA, Wandeler AI (1994) A molecular epidemiological study of rabies virus in central Ontario and western Quebec. J Gen Virol 75: 2 575–2 583
22. Nadin-Davis SA, Huang W, Wandeler AI (1996) The design of strain-specific polymerase chain reactions for discrimination of the raccoon rabies virus strain from indigenous rabies viruses of Ontario. J Virol Methods 57: 141–156
23. Nel LH, Thomson GR, Von Teichman BF (1993) Molecular epidemiology of rabies virus in South Africa. Onderstepoort J Vet Res 60: 301–306
24. Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. Comp Appl Biosc 12: 357–358
25. Prehaud P, Coulon P, Diallo A, Martinet-Edelist C, Flamand A (1989) Characterization of a new temperature-sensitive and avirulent mutant of the rabies virus. J Gen Virol 70: 133–143

26. Ravkov EV, Smith JS, Nichol ST (1995) Rabies virus glycoprotein gene contains a long 3′ noncoding region which lacks pseudogene properties. Virology 206: 718–723
27. Rupprecht CE, Dietzschold B, Wunner WH, Koprowski H (1991) Antigenic relationships of lyssaviruses. In: Baer GM (ed) The natural history of rabies, 2nd ed. CRC Press, Boca Raton, pp 69–100
28. Schneider LG, Dietzschold B, Dierks RE, Matthaaeus W, Enzmann PJ, Strohmaier K (1973) Rabies group-specific ribonucleoprotein antigen and a test system for grouping and typing of rhabdoviruses. J Virol 11: 748–755
29. Smith JS, Orciari LA, Yagar PA, Seidel HD, Warner CK (1992) Epidemiologic and historical relationships among 87 rabies virus isolates as determined by limited sequence analysis. J Infect Dis 166: 296–307
30. Tabel H, Corner AH, Webster WA, Casey GA (1974) History and epizootiology of rabies in Canada. Can Vet J 15: 271–281
31. Tordo N, Badrane H, Bourhy H, Sacramento D (1993) Molecular epidemiology of lyssaviruses: focus on the glycoprotein and pseudogenes. Onderstepoort J Vet Res 60: 315–323
32. Tordo N, Kouknetzoff A (1993) The rabies virus genome: an overview. Onderstepoort J Vet Res 60: 263–269
33. Tordo N, Poch O, Ermine A, Keith G, Rougeon F (1986) Walking along the rabies genome: is the large G-L intergenic region a remnant gene? Proc Natl Acad Sci USA 83: 3 914–3 918
34. Trimarchi CV, Debbie JG (1991) The fluorescent antibody in rabies. In: Baer GM (ed) The natural history of rabies, 2nd ed. CRC Press, Boca Raton, pp 219–233
35. Winkler WG, Jenkins SR (1991) Raccoon rabies. In: Baer GM (ed) The natural history of rabies, 2nd ed. CRC Press, Boca Raton, pp 325–340
36. Wunner WH (1991) The chemical composition and molecular structure of rabies viruses. In: Baer GM (ed) The natural history of rabies, 2nd ed. CRC Press, Boca Raton, pp 31–67
37. Xianhe B, Warner CK, Fekadu M (1993) Comparisons of nucleotide and deduced amino acid sequences of the glycoprotein genes of a Chinese street strain (CGX89-1) and a Chinese vaccine strain (3aG) of rabies virus. Virus Res 27: 101–112

Authors' address: Dr. S. A. Nadin-Davis, Rabies Unit, Animal Diseases Research Institute, 3851 Fallowfield Road, P. O. Box 11300, Station H, Nepean, Ontario K2H 8P9, Canada.

Received July 16, 1996