Dynamics of rabies virus quasispecies during serial passages in heterologous hosts

B. Kissi,1 H. Badrane,2 L. Audry,1 A. Lavenu,1 N. Tordo,2 M. Brahimi3 and H. Bourhy1

1, 2Rabies Unit1 and Lyssavirus Laboratory2, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France
3Institut Pasteur d’Algérie, Alger, Algeria

To understand the mutations and genetic rearrangements that allow rabies virus infections of new hosts and adaptation in nature, the quasispecies structure of the nucleoprotein and glycoprotein genes as well as two noncoding sequences of a rabies virus genome were determined. Gene sequences were obtained from the brain and from the salivary glands of the original host, a naturally infected European fox, and after serial passages in mice, dogs, cats and cell culture. A relative genetic stasis of the consensus sequences confirmed previous results about the stability of rabies virus. At the quasispecies level, the mutation frequency varies, in the following order: glycoprotein region (21.9 × 10⁻⁴ mutations per bp), noncoding sequence nucleoprotein–phosphoprotein region (7.2–7.9 × 10⁻⁴ mutations per bp) and nucleoprotein gene region (2.9–3.7 × 10⁻⁴ mutations per bp). These frequencies varied according to the number, type of heterologous passages and the genomic region considered. The shape of the quasispecies structure was dramatically modified by passages in mice, in which the mutation frequencies increased by 12–31 × 10⁻⁴ mutations per bp, depending on the region considered. Non-synonymous mutations were preponderant particularly in the glycoprotein gene, stressing the importance of positive selection in the maintenance and fixation of substitutions. Two mechanisms of genomic evolution of the rabies virus quasispecies, while adapting to environmental changes, have been identified: a limited accumulation of mutations with no replacement of the original master sequence and a less frequent but rapid selective overgrowth of favoured variants.

Introduction

Lyssaviruses participate in many independent enzootic cycles, the reservoir of which can be bats, dogs or some wildlife terrestrial mammals such as the red fox (Wandeler et al., 1994). Their RNA genome is of negative polarity (12 kb in size) and consists of five genes: the nucleoprotein (N), the phosphoprotein (P), the matrix protein, the glycoprotein (G) and the polymerase (L) (Tordo et al., 1988). Sequencing has been used previously to characterize mutations in rabies virus and for studies of phylogenetic and epidemiologic relationships among isolates (Sacramento et al., 1991). In lyssaviruses of genotype 1 (rabies virus) (Bourhy et al., 1993), several co-circulating phylogenetic lineages were distinguished, from which geographical localization was determined (Kissi et al., 1995; Nadin-Davis et al., 1993, 1994; Sacramento et al., 1992; Smith et al., 1992). Some lineages more specifically adapted to one animal species were also characterized (Kissi et al., 1995).

Several factors may be involved in generating sequence heterogeneity in rabies virus, including duration of infection, route of transmission, virus load, host immune response, virus and host protein cooperation. Because of the limited replication fidelity caused by an absence of proofreading/repair and post-replicative error correction, it is often reported that RNA virus populations (Holland et al., 1992; Drake, 1993; Domingo & Holland, 1994, 1997), and rabies virus in particular (Bennamsour et al., 1992), exhibit a heterogeneous population structure within single individuals, often referred to as ‘quasispecies’. This population is characterized by one or several master sequences and a large spectrum of related variants. This quasispecies model of mixed RNA virus populations implies a significant adaptive potential, because it allows the rapid selection of the mutant(s) with the highest
fitness in any new environmental condition (Morimoto et al., 1998). However, in rabies virus little is known about the relationship between this potential diversity available for selection and evolution observed in nature, especially the colonization of new ecological niches, i.e. new animal vector species. To understand how selection operates, sequence variation and the quasispecies structure of the N and G genes, as well as those of two noncoding sequences, one located between the N- and P-coding sequences (N–P region) and the other located between the G- and L-coding sequences (G–L region), were analysed after several passages of a fox rabies virus isolate in mice, dogs, cats and cell culture.

Methods

■ Origin of the virus. Submaxillary salivary glands (FOXSG) of a rabies virus-infected fox from France (9147FRA) (Kissi et al., 1995) were ground and clarified by centrifugation at 2000 g for 30 min at 4 °C. The hippocampus (FOXBR) of the infected fox was also sampled. The supernatant was filtered on a 0.45 µm membrane, aliquoted and stored at −80 °C. Nine other viruses (8661FRA, 8663FRA, 86111YOU, 8903FRA, 9004FRA, 9213ALL, 9223FRA, 9445FRA and 9244FRA) were used for sequencing the noncoding N–P region. All these isolates were shown to belong to the western European lineage of fox rabies virus (H. Bourhy, unpublished results).

■ Cell culture adaptation. Virus was propagated and titrated according to Bourhy & Sereau (1991). After six passages on N2a cells (ATCC CCL 131), virus was adapted to BSR cells, a clone of BHK-21 cells (Wiktor et al., 1977), and subsequently passaged up to 20 times (BSR/P20) in 25 cm² flasks with an m.o.i. of 0.1. Titration of the cell culture-adapted virus was obtained by infection of BSR cells with serial dilutions of the virus in a 96-well cell culture plate. After 48 h incubation, the number of fluorescent foci/ml was determined with an anti-rabies virus nucleocapsid antibody conjugated to fluorescein (Diagnostics Pasteur) (Bourhy & Bourhy, unpublished results).

■ Passages in animals. The titre of the filtered virus suspension (FOXSG) used to initiate passages in different hosts was 10¹⁸³ LD₅₀/0.2 ml in 4-week-old OF1 mice inoculated intracerebrally (Bourhy & Sereau, 1991). Litters of 48 h-old suckling mice or groups of 4-week-old OF1 mice (IFFA CREDO) were inoculated intracerebrally with 15 µl and 30 µl, respectively. After 6–10 days of incubation, the brains of sick mice were harvested, checked for positive immunofluorescence (Bourhy et al., 1989), ground and used for another passage. The virus suspension was serially passaged 16 times in suckling mice (SMP16). Adult mice were inoculated through the intramuscular route with 0.2 ml suspension and serially passaged up to three (AMP3) or up to six times (AMP6). Titres of the different virus suspensions used for each passage were between 10³ and 5 × 10⁴ LD₅₀/0.03 ml.

Dogs and cats between 1 and 2 years old were kept in captivity in the facilities of the Institut Pasteur of Alger. These animals were checked for the absence of rabies virus antibodies in their sera by RFFIT (Smith et al., 1973) before inoculation. The first passage was obtained by inoculation of 0.4 ml of the filtered virus suspension via the intracerebral route, inoculation of 0.1 ml by the intracocular route and/or inoculation of 0.2 ml by the intramuscular route. Animals developing rabies symptoms were euthanased. Rabies diagnosis was confirmed by direct immunofluorescence on impressions of Ammon’s horn and by cell culture inoculation (Bourhy et al., 1989). The submaxillary salivary glands of each positive animal were ground, clarified as described above and used for subsequent passages using the same route of inoculation. The virus suspension used for the subsequent passages was diluted to correspond to 10³–³ LD₅₀/0.02 ml. The second passage intramuscular in dogs (DG2) and the third passage intramuscular in cats (CTP3) were used for further studies.

■ RNA isolation, cDNA synthesis and PCR amplification. Four hundred μl of the different virus suspensions was subjected to RNA extraction as previously described (Sacramento et al., 1991). Total RNA was reverse-transcribed using 200 units of SuperScript reverse transcriptase (Gibco-BRL) and amplified by PCR using Tag DNA polymerase (Gibco-BRL) (5 units/μl). PCR was performed in a final volume of 50 μl (2.5 mM MgCl₂, 50 mM KCl, 20 mM Tris–HCl pH 8, 200 μM each dNTP, 0.1 mM EDTA, 25 mM TAPS pH 9.3, 1 mM DTT, 25 ng salmon sperm activated DNA). Thermal cycling was performed for 30 cycles. cDNA synthesis of the genomic RNA and PCR amplification were performed with primer sets N7 (nt 55–73) and N8 (nt 1584–1568), G3 (nt 3891–3908) and GH4 (nt 4621–4602), and G (nt 4665–4687) and LH (nt 5520–5539) for the N, G and G–L genomic regions, respectively (positions are given relative to the PV rabies virus genome; Tordo et al., 1986).

■ Analysis of the sequences. PCR products were cloned into pUC18 using the SureClone ligation kit (Pharmacia). Recombinant pUC18 vectors from each sample and PCR products were sequenced with appropriate primers on both strands either using the T7 sequencing kit (Pharmacia) or the ABI PRISM Ready Reaction kit (Perkin Elmer) and the ABI PRISM 377 DNA sequencer (Perkin Elmer). Sequences obtained directly from the PCR products were called ‘consensus sequences’ and individual sequences of recombinant pUC18 vectors were called ‘quasispecies sequences’. Multiple sequence alignments of the nucleotide sequences were generated with the CLUSTALW 1.60 program (Thompson et al., 1994). The nucleotide composition of each region was determined using the MEGA package version 1.01 (Kumar et al., 1995). The mutation frequency of different mutant positions is defined as the proportion of positions with mutations relative to the consensus nucleotide sequence. The relative frequency of mutation (RF) for 1000 bases among four nucleotides was calculated by normalizing the number of mutations for each genomic region according to its length and composition in nucleotides. The sporadic mutation frequency is calculated in the same way as RF but considering only substitutions found in one of the multiple clones sequenced.

■ Determination of the frequency of mutations introduced by the methodology. PCR products obtained with primer sets N7–N829 (positions 829–811 relative to the PV rabies virus genome; Tordo et al., 1986) were cloned into pCR 2.1 vector (Invitrogen) using the SureClone ligation kit (Pharmacia). The transcription of the PCR products was obtained with 1 μg of recombinant plasmid, 50 units of T7 RNA polymerase (BioLabs) and 30 units of RNA guard (Pharmacia) incubated for 1 h at 37 °C in 50 μl buffer (40 mM Tris–HCl pH 7.9, 6 mM MgCl₂, 2 mM spermidine, 10 mM DTT) supplemented with 0.5 mM each dATP, dUTP, dGTP, dCTP. The products of transcription (RNA) were then incubated for 2 h at 37 °C with 200 units of PolⅢ (Boehringer) (three PolⅢ sites are located in the N7–N12 DNA insert), then for 30 min at 37 °C with 10 units of RNase-free DNase (Pharmacia). The RNA was finally purified by a phenol–chloroform extraction, precipitated in ethanol and resuspended in 20 μl of distilled water containing 1 μg of total RNA extracted from a rabies virus-negative mouse brain. Each of these
reconstituted suspensions was reverse-transcribed as described above. One µl of the cDNA previously diluted 1/100 in distilled water was then used for PCR amplification. The most diluted samples yielding clonable PCR product were used for cloning and sequencing as described for RNA originating from biological specimens. A PCR control performed directly on RNA without previous reverse transcription remained negative, attesting that the RNA preparation was not contaminated with entire residual N7–N829 DNA products. The mutation frequency introduced by our methodology was calculated by comparing the original N7–N829 plasmid DNA sequence to the sequences cloned after reverse transcription and PCR of the T7 transcription products.

Results

Genetic stasis of the consensus sequences

The consensus sequences of the fox isolate 9147FRA were obtained from FOXBR, FOXSG, SMP16, AMP3, DGP2, CTP3 and BSRP20. The sequence nt 71–1421 (positions given according to the PV rabies virus genome; Tordo et al., 1986) encoding the entire N protein, nt 1422–1514 corresponding to the entire N–P noncoding region, nt 3909–4388 encoding 36% of the G protein ectodomain, and nt 5058–5417 corresponding to 69% of the G–L non-protein coding region were determined (GenBank accession nos U22474, AF047509, AF047516 and AF047517). The sequence was always unambiguous, reflecting a greater intensity of signal from one of the four DNA bases. The consensus sequence was then clearly discernible. The N and G–L sequence of the fox isolate 9147FRA is related to the European phylogenetic lineage of lyssaviruses of genotype 1 (Sacramento et al., 1992; Kissi et al., 1995). The nucleotide sequence of the G region and its deduced amino acid sequence (166 AA) show 6% and 3% divergence, respectively, from their homologous region in the PV rabies virus (Tordo et al., 1986). No nucleotide substitutions were noticed in the consensus sequences whatever the organ (brain or salivary glands) of the original animal, the type of heterologous host (cells, suckling and adult mice, dogs and cats) or the number of passages studied. The only exception was one transversion (T → G), observed in position 4117 in AMP3. It changed the amino acid in position 247 of the G protein from aspartate to glutamic acid. This virus population was subsequently serially passaged three more times in adult mice by the intramuscular route. The consensus sequence of the sixth passage (AMP6) remained the same as AMP3.

Heterogeneity of the virus population

To further analyse the evolution of the N, N–P and G sequences, in particular the quasispecies differing in these genes, we decided to clone the PCR products and investigate the number of substitutions observed in the different clones. Therefore for each gene and each type of passage, the sequences of 9–38 clones were determined and compared (Fig. 1). However, to determine the maximum number of reverse transcriptase and Taq polymerase errors introduced by our methodology, target RNA molecules (RNA)t synthesized with T7 polymerase from a single plasmid clone were subjected to reverse transcription, PCR and then sequencing. PCR controls performed directly on the RNA tum molecules indicated that the preparations were free of any residual complete target DNA. The RT–PCR products obtained from the most diluted RNA tum molecules mixed in total RNA suspensions (30 amplification cycles) were then cloned in pCR2.1 vector and 19 clones were sequenced. A total of 12930 bases were analysed and 12 mutations were detected among six different clones. This gives a maximum error rate of 9.3 × 10⁻⁴ per bp (0.31 × 10⁻⁴ per bp per cycle). These mutations are all transitions and consisted of eight A → G, two C → U and two U → C. The reproducibility of our type of analysis was also investigated. Two different cDNAs of the N gene were obtained from the same RNA extracted from the salivary glands of the naturally infected fox (FOXSG). These cDNAs were amplified separately by PCR and cloned. Three variants were obtained out of 20 clones from one cDNA and three out of 18 for the second cDNA (Fig. 1). Therefore, in our conditions, different reverse transcription and PCR assays performed on the same RNA do not significantly alter the number of variants (as measured by a χ² test).

Whatever the type of sample analysed (i.e. FOXSG, FOXBR, BSRP20, SMP16 and DGP2), the N and N–P regions were characterized by a relatively low proportion of mutants (0–27% and 0–40%, respectively), in contrast to the G region which contained 50–60% of the mutants (Fig. 1). The most frequent mutants were found in 30% of clones in the N–P region (corresponding to T30 and C31), 20% in the G region (A → T181) and 11% in the N region (T → C107). In this case, none of the mutants gradually displaced the parental particles and became dominant, so the population remained polymorphic. However, an important change in the proportion of mutants was seen after three to six passages in adult mice. Here, the mutants reached 61%, 43% and 100% in the N, N–P and G regions, respectively. The substitution T → G207 (position 4117 on the genome) in the G region was found in 100% of the clones. No significant variation (assessed by Fisher’s test) in heterogeneity was observed if we analysed the mutants according to the different types of passages and original tissues, except passages intramuscularly in adult mice. Therefore, in the following part of the analysis, the results obtained after AMP3 and AMP6 will be compared to the five other samples grouped together (FOXBR, FOXSG, BSRP20, SMP16 and DGP2), which we can consider as representative of a state of lower genetic diversity. In this case, χ² tests (corrected for small samples in the cases of N–P and G, χ²p) were significant at the 0.1% (χ² = 27·16), 5% (χ² = 4·06) and 5% (χ² = 3·94) levels for the N, N–P and G regions, respectively. This rejection of the null hypothesis of homogeneity of sequence variation among samples indicated a strong association between the number of mutations and the type of sample analysed across all genes: the five types of sample grouped together versus AMP6 (or AMP3).
B. Kissi and others

Fig. 1. Evolution of the quasispecies sequences during propagation of the virus in vivo and in vitro. Molecular clones of fox isolate 9147FRA were obtained from the hippocampus (FOXBR) and the salivary glands (FOXSG) of a naturally infected fox, the 16th passage by the intracerebral route in suckling mice (SMP16), the third passage in intramuscularly infected adult mice (AMP3*), the sixth passage in intramuscularly infected adult mice (AMP6), the second passage in intramasseterly infected dogs (DGP2), the third passage in intramasseterly infected cats (CTP3) and the 20th passage on BSR cells (BSRP20). All the mutations found in one clone are indicated on the same line. Numbers in brackets indicate the position of the mutation in the sequence analysed. N, Nucleoprotein; P, phosphoprotein; G, glycoprotein; b, base; Nb, number; A>G(147), mutation A to G in position 147; Del., deletion; T(30), insertion of a T in position 30. The signs 2 < and 3 < indicate that the mutation is found in two or three different clones. The GenBank accession numbers of the N, N–P and G sequences are U22474, AF047509 and AF047516, respectively.

It was important to evaluate the amount of RNA used to determine the error rate and the amounts of viral RNA present in the different biological specimens. If we consider that the efficiency of all the methodological steps from the RNA production to PCR was 100%, the equivalent in DNA of a maximum of 6.5 x 10^5 molecules of RNA were incorporated into the PCR reaction. In the different virus suspensions, the only way to quantify virus-specific RNA is to consider the number of infectious particles. However, this obviously underestimates the number of virus-specific RNAs as during the grinding of tissues a large proportion of the viral RNA is present in non-infectious virus particles but is still available for reverse transcription. According to the titres of the different virus suspensions, a minimum of 1.3 x 10^4 viral RNA molecules were used in each case for cDNA synthesis and then for PCR. The only exception was the virus suspension obtained from the salivary glands of the fox, from which at least 2 x 10^4 viral RNA molecules were used. Therefore large amounts of specific viral RNA molecules were used to determine the error rate introduced by the methodology and the frequency of mutations in the biological specimens.

**Analysis of mutations**

Mutation frequencies in biological specimens were deduced for the three different genomic regions considered (Table 1). The results obtained from FOXBR, FOXSG, BSRP20, SMP16 and DGP2 allowed us to compare the respective frequencies of the three genomic regions. There were more mutations in the G gene (2.1 x 10^-3 mutations per bp) than in the N–P region (7.2 x 10^-4 mutations per bp) and the N region (2.9 x 10^-4 mutations per bp). After six or three passages in mice (AMP6 and AMP3), the mutation frequencies in the different regions varied in the same order. The number of mutations per bp was
The mutation frequencies of fox isolate 9147FRA were deduced from analysis of the isolate obtained from the hippocampus (FOXBR), the salivary glands (FOXSG), the sixth passage by the intramuscular route in suckling mice (SMP16), the sixth intramuscular passage in adult mice (AMP6), the second intramuscular passage in dogs (DGP2), the third intramuscular passage in cats (CQP3) and the 20th passage on BSR cells (BRSP20).

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>N–P</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of clones</td>
<td>Sporadic mutation (× 10⁻⁴)</td>
<td>Mutation frequency (× 10⁻⁴)</td>
</tr>
<tr>
<td>FOXBR, FOXSG, BSRP20, SMP16 and DGP2</td>
<td>101</td>
<td>2.3</td>
<td>2.9</td>
</tr>
<tr>
<td>AMP6</td>
<td>23</td>
<td>—</td>
<td>15.2</td>
</tr>
<tr>
<td>FOXBR, FOXSG, BSRP20, SMP16 and DGP2</td>
<td>41‡</td>
<td>—</td>
<td>3.7</td>
</tr>
<tr>
<td>AMP6</td>
<td>13‡</td>
<td>—</td>
<td>13.5</td>
</tr>
</tbody>
</table>

* Mutation frequencies with sporadic mutations included. Under the same conditions, the maximum frequency of artefactual nucleotide misincorporation due to the technique was evaluated to be 9.3 × 10⁻⁴ mutations per bp (this analysis was performed on 19 clones covering 12930 nt).
† Corresponds to the third passage in mice AMP3.
‡ Comparison of the mutation frequencies of the N and N–P regions obtained from the analysis of 54 clones from which the sequences of the two regions were simultaneously analysed.

44 × 10⁻⁴, 38.4 × 10⁻⁴ and 15.2 × 10⁻⁴ in the G, N–P and N regions, respectively. Another comparison of the mutation frequencies of the N and N–P region obtained from the analysis of 54 clones from which the sequences of the two regions were simultaneously analysed gave similar results (Table 1). Considering the maximum mutation frequency possibly introduced by Taq errors (9.3 × 10⁻⁴ per bp), it is possible that most of the mutations and particularly the sporadic mutations (2.3 × 10⁻⁴ per bp) observed in the N region in FOXSG, FOXBR, BSRP20, SMP16 and DGP2 are artefacts introduced by our methodology. However, this is not true for the N–P region where sporadic mutations are very rare, and for the G region where the frequency is high compared to the frequency of artefactual nucleotide misincorporation.

Alignments of 31 N nucleotide sequences previously described (Kissi et al., 1995) (GenBank accession nos AF033884–AF033886, AF033904, AF033905, U22839, U42605–U42607, U42700, U42702, U42706, U42708–U42713, U42717, U42718, U42992–U42996, U43006, U43008, U43009 and U43433–U43435) and of nine N–P sequence isolates (GenBank accession nos AF047506–AF047508 and AF047510–AF047515) obtained from fox rabies virus were performed and compared to the corresponding sequences of isolate 9147FRA (not shown). The distribution of the mutations observed in the variant clones was compared to the substitutions fixed in nature (Fig. 2). The substitutions are randomly distributed in the N region considered, but not in the N–P region, where they are preferentially located at the 5′- and 3′-ends leaving intact the transcription signals of lyssaviruses (Bourhy et al., 1993). Conversely, the mutations observed in the variant clones are not randomly distributed. In particular, they are absent in the 3′-end of the N region and from both the 3′- and 5′-ends of the N–P region. This indicates different distributions of the mutations observed in the variant clones compared to the substitutions fixed during natural transmission of the virus in foxes.

The sequence obtained from the G region covered 17 of the 30 amino acids shown to be involved in its antigenic sites (Flamand et al., 1993). These comprised the whole of sites Ia, I, VI and a portion of antigenic sites II and III. These sites were not modified by mutations. Deletions and insertions were only found in the noncoding N–P region. Premature stop codons were not noted in the coding regions. Transitions and transversions occurred with a similar rate (50–60%) in the N–P and G genes. More transitions (96.3%) than transversions were found in the N gene. Analysis of the type of nucleotide changes revealed that there are more nonsynonymous mutations (NSY) than synonymous mutations (SYN) in the N (NSY/SYN = 2) and the G (NSY/SYN = 5.8) genes. However, these rates varied also according to the type of passage.
Fig. 2. Distribution of nucleotide mutations in the N and the noncoding N–P regions. (A) Substitutions fixed in nature. The sequences of the N and N–P regions of isolate 9147FRA (GenBank accession nos U22474 and AF047509, respectively) were compared to 31 sequences of the N gene (GenBank accession nos AF033884–AF033886, AF033904, AF033905, U22839, U42605–U42607, U42700, U42702, U42706, U42708–U42713, U42717, U42718, U42992–U42996, U43006, U43008, U43009 and U43433–U43435) and to nine sequences of the N–P region (GenBank accession nos AF047506–AF047508 and AF047510–AF047515). Multiple sequence alignments of the nucleotide sequences were generated with CLUSTALW 1.60 program (Thompson et al., 1994). (B) Mutations observed in the quasispecies.

Table 2. Relative frequency of substitutions for 1000 bases

The proportions of substitutions relative to the consensus nucleotide sequences were normalized for the N, the noncoding N–P and the G regions according to their length, nucleotide composition and synonymous (SYN) and nonsynonymous (NSY) sites. The size on which this frequency was calculated is indicated in brackets next to the name of each genomic region. No mutations G → C or C → G mutations were found.

<table>
<thead>
<tr>
<th>Mutations</th>
<th>Biological specimens</th>
<th>N (49.6 kb)</th>
<th>N–P (5.5 kb)</th>
<th>G (13.9 kb)</th>
<th>Maximum frequency of artefactual substitutions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total SYN NSY</td>
<td>Total SYN NSY</td>
<td>Total SYN NSY</td>
<td></td>
</tr>
<tr>
<td>A → G</td>
<td></td>
<td>0.77 0.67 1.05</td>
<td>0.53</td>
<td>1.46 0.72 1.80</td>
<td>1.94</td>
</tr>
<tr>
<td>G → A</td>
<td></td>
<td>0.24 0.29 0.22</td>
<td>2.12</td>
<td>1.07 0 1.52</td>
<td>0</td>
</tr>
<tr>
<td>U → C</td>
<td></td>
<td>0.71 0.67 0.73</td>
<td>1.47</td>
<td>2.70 1.53 3.45</td>
<td>0.56</td>
</tr>
<tr>
<td>C → U</td>
<td></td>
<td>0.29 0.20 0.37</td>
<td>0</td>
<td>0.37 0.69 0</td>
<td>0.71</td>
</tr>
<tr>
<td>Transitions</td>
<td></td>
<td>0.52 0.46 0.56</td>
<td>0.91</td>
<td>1.44 0.76 1.85</td>
<td>0.84</td>
</tr>
<tr>
<td>U → G</td>
<td></td>
<td>0.09 0 0.15</td>
<td>0</td>
<td>3.00 0 5.42</td>
<td>0</td>
</tr>
<tr>
<td>G → U</td>
<td></td>
<td>0 0 0</td>
<td>2.12</td>
<td>0.53 0 0.76</td>
<td>0</td>
</tr>
<tr>
<td>A → C</td>
<td></td>
<td>0 0 0</td>
<td>0.53</td>
<td>0 0 0</td>
<td>0</td>
</tr>
<tr>
<td>C → A</td>
<td></td>
<td>0 0 0</td>
<td>0</td>
<td>0 0 0</td>
<td>0</td>
</tr>
<tr>
<td>U → A</td>
<td></td>
<td>0 0 0</td>
<td>0</td>
<td>1.20 1.53 0.99</td>
<td>0</td>
</tr>
<tr>
<td>A → U</td>
<td></td>
<td>0 0 0</td>
<td>0</td>
<td>0.97 0 1.47</td>
<td>0</td>
</tr>
<tr>
<td>Transversions</td>
<td></td>
<td>0.02 0 0.03 0.55</td>
<td>1.44 0.38 2.19</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>0.54 0.46 0.59</td>
<td>1.46</td>
<td>2.87 1.14 4.04</td>
<td>0.89</td>
</tr>
</tbody>
</table>
Overall, the results obtained from FOXBR, FOXSG, BSRP20, SMP16 and DGP2 indicated a low NSY/SYN for the N gene (NSY/SYN = 1:2) and the G gene (NSY/SYN = 3:2). The same rate calculated after three or six passages in mice is higher in the N gene (NSY/SYN = 6) and more so in the G gene (NSY/SYN = 18). A predominance of the transitions U → C and A → G was noted in the N and G genes (76-9% and 75%, respectively). This was not observed to the same extent in the noncoding N–P region (transitions = 60%). In this region, insertions were observed in positions 30 and 31 and one deletion was reported in position 63.

To determine whether there was a bias for any type of mutation, we categorized the mutations and normalized them according to the different nucleotide compositions of the sequences (Table 2). The most common mutations introduced by our methodology are A → G (RF = 1:94) and C → U (RF = 0:71). However, RF of substitutions A → G and C → U observed in the biological specimens confirmed that mutations are biased in the G gene. In the G region, the RF of substitution U → C reaches 3:45 at NSY sites. In the N and N–P regions, substitutions A → G and C → U, when normalized, are less important than in the G region. In the G region, transversions mostly involved NSY mutations.

Discussion

Understanding the evolution of rabies virus is important for determining the basis of its genetic and phenotypic flexibility. The optimization of virus quasispecies in new environments, and the following modification of fitness, have been widely and empirically used for the production of modified live vaccines (for a review see Vodopija & Clark, 1991; Lafay et al., 1994). In wild isolates, previous results indicated a relative genetic stasis and a limited process of genetic ‘radiation’ around a prototype sequence (Amengual et al., 1997; Kissi et al., 1995) together with an adaptive process to different animal species (H. Bourhy, unpublished results). To further analyse the evolution of lyssaviruses, we undertook comparative analyses of sequences from rabies viruses isolated from the brain and the salivary glands of a naturally infected fox, which was also passaged serially in various laboratory and natural hosts. In total, the N, N–P, G and G–L regions analysed in this study represent 19-1% of the genome of rabies virus (Tordo et al., 1988). The regions analysed in the G ectodomain and in the G–L noncoding region are the most variable parts of the genome (Sacramento et al., 1992; Tordo et al., 1993b).

No tissue- or host-specific sequence variation was demonstrated at the level of the consensus sequence of the N, N–P and G–L regions. Nevertheless, one substitution was observed in the G region (Asp247 → Glu) after only three passages by the intramuscular route in adult mice. This conservative substitution (Li, 1997) has never been described previously in mice (Seif et al., 1985). At the same position, a mutation Asp → Asn has already been reported between glycoproteins of different laboratory viruses and wild isolates (Tordo et al., 1993a). Unfortunately, the importance of this position in the structure of the G protein is unknown. The extent of genetic stability is known to depend on the passage conditions of the viruses (Spindler et al., 1982; Steinhauer & Holland, 1987; Steinhauer et al., 1989). The virus population size used for each passage was controlled and maintained at a sufficiently high level to avoid the transmission of small numbers of infectious particles (bottlenecks) that may modify the heterogeneity of the virus population within a particular sample. The lowest population size used was $8 \times 10^6$ LD₅₀. Conversely, some rapid changes observed in an earlier study with rabies virus isolates from Algeria (Benmansour et al., 1992) could be due to passages with a lower population size or population bottlenecks, allowing rapid changes in the extent and nature of genetic polymorphism (Domingo & Holland, 1997).

In the control experiment performed following guidelines described by Bracho et al. (1998), the maximum frequency of artefactual nucleotide misincorporation, due to the technique of investigation was evaluated to be $0.31 \times 10^{-4}$ substitutions per nucleotide per cycle ($9.3 \times 10^{-4}$ substitutions per nucleotide), which is in agreement with previous studies (Barnes, 1992; Plyusnin et al., 1995; reviewed in Smith et al., 1997; Bracho et al., 1998). The determination of the quasispecies structure of RNA viruses (Domingo et al., 1978, 1980, 1985; del Torre & Holland, 1990; Benmansour et al., 1992) indicates that multiple variants could coexist during propagation of the virus in vivo or in vitro. In rabies virus, our results demonstrated that a variety of mutant forms of viral RNA existed, including point mutants, deletion mutants and insertion mutants. No termination mutants were seen. Despite this, in the N and N–P regions, none of the mutants displaced the parental particles and became dominant. All were preserved at low levels (0–40%) in series of passages. In the G region, however, this proportion reached 60–100% so that the wild-type sequence is eventually replaced by a new master sequence (i.e. a substitution has occurred). The mutation frequencies of the N, N–P and G regions in the case of low diversity (FOXBR, FOXSG, BSRP20, SMP16 and DGP2) were $2.9 \times 10^{-3}$, $7.2 \times 10^{-4}$ (these are below the maximum frequency of artefactual nucleotide misincorporation) and $2.19 \times 10^{-4}$ mutations per bp, respectively. These frequencies can increase considerably when changes in the adaptive environment occur. For example, after passages in adult mice, there was an increase of $12.3 \times 10^{-4}$, $31.2 \times 10^{-4}$ and $22.1 \times 10^{-4}$ extra mutations per bp in the N, N–P and G regions, respectively. The frequencies of mutation in the rabies virus N and G regions are in the range of those found for other nucleocapsid and glycoprotein genes (Adami et al., 1995; Drake, 1993).

There was a predominance of U → C transitions in the G region. Some of them are in fact Tag substitutions as demonstrated here and by others (Martinez et al., 1994; Bracho et al., 1998). However, ‘biased hypermutations’ have also been previously described in association with some virus infections
The mechanism of dsRNA unwinding/modifying activity has been proposed to be responsible for these ‘biased hypermutations’ (Bass et al., 1989; Wagner et al., 1989). Yet, this would not explain the large differences found by the type of passage used and the RNA segments investigated, even if different selective constraints act on particular genomic RNA segments during their synthesis and replicative efficiency (Britten, 1993; Rima et al., 1995). Differences in the number of virus replications in the various systems of infection used could also play a role in the observed variation in frequency of mutations. Considering that the generation time is similar in all hosts, it is highly improbable that the number of replication cycles would be lower in dog or in cat compared to mouse. Furthermore, a high frequency of mutations would not explain why there is a predominance of NSY mutations (Table 2). One explanation would be that most of the mutations that arise are slightly deleterious and so will persist for a few generations before being removed by natural selection. However, the proportion of NSY mutations so greatly exceeded the 60% expected that this explanation is not realistic and, furthermore, a fixed NSY change has taken place in the G region. Another possible explanation would be that the G region may be driven by strong positive selection, such as that described for influenza viruses (for a review see Webster et al., 1992), thereby explaining the occurrence of mutations.

Genetic mechanisms underlying the change of the rabies virus genome appear to involve both the regular occurrence of background mutations with no replacement of the original wild-type virus (as seen in all three virus regions) and with the occasional rapid and selective overgrowth of variants from this mutational spectrum (as seen in the G region). Sudden exclusion of the previous dominant subpopulation was also determined in vitro with vesicular stomatitis virus (Clarke et al., 1983). The extent of the mutant repertoire, particularly in the NSY changes in the sequence of the G region, is not the only condition for the rapid fixation of a mutation. First, most of the mutations observed in the quasispecies are located in regions of the gene that are different from those where substitutions occurred mostly in wild isolates. Second, no fixation of mutation was reported in the N gene despite a large number of quasispecies in this gene after passages in mice, suggesting that some proteins (G) are under more positive selective pressure than others. Mutations in the G protein are known to modify the virulence of rabies virus in vitro and in vivo (Dietzschold et al., 1983; Seif et al., 1985). The trimeric G protein is the only viral protein exposed on the surface of the virion (Whitt et al., 1991). The consequences, during evolution and the adaptation to new hosts, of changes in the binding of G protein to cellular receptors (Tuffereau et al., 1998; Toulouse et al., 1998) are presently unknown. The G protein is also the major antigen of the rabies virus and an immunological response directed against this protein is able to protect from rabies virus (Wiktor et al., 1984). In our study, positive selective pressure was not particularly evident at the level of the different major B epitopes of the G region found in the mouse (Benmansour et al., 1991). Therefore, either nonspecific immunological mechanisms, or the rising specific cellular immunological response during infection, may be responsible for the positive selection pressure in the virus and hence the important predominance of the NSY changes in the sequence of the G region.

The authors thank E. C. Holmes for critical reading of the manuscript and helpful discussions.

References


B. Kissi and others


Received 17 December 1998; Accepted 10 May 1999