Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA

Anne Kouznetzoff,¹ Malcolm Buckle² and Noël Tordo¹

Laboratoire des Lyssavirus¹, and Unité de Physicochimie des Macromolécules Biologiques, URA 1149 du CNRS², Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

In rabies virus, the ribonucleoprotein complex (RNP), the RNA genome (G) and the antigenome (M) are specifically coated by the viral nucleoprotein (N protein), forming the template for transcription and replication by the viral RNA polymerase. This specific encapsidation starts at the 5′ ends of the RNAs. To investigate domains of the N protein that govern binding specificity, we tested in vitro the ability of both full-length and truncated forms of the N protein to interact with a synthetic RNA probe corresponding to the 5′ end of the antigenome. UV-LASER cross-linking, which covalently links RNA and proteins in intimate contact, showed that the entire N protein (450 aa) and the NH₂-terminal 376 aa (t42) contained all of the determinants for specific interaction. It was demonstrated by affinity chromatography that a peptide near the COOH terminus of t42 (position 298–352), which is located in the most conserved region of Rhabdoviridae N proteins, bound directly to the viral RNA. However, no significant sequence similarity was detected between this peptide and known RNA binding proteins in the databases. This suggests both that N proteins may possess a new type of RNA binding motif and that protein folding contributes to the architecture of the RNA binding site.

Introduction

The rabies virus (genus Lyssavirus, family Rhabdoviridae) possesses a non-segmented, negative-stranded RNA genome and is grouped within the order Mononegavirales (Murphy et al., 1995). Such negative-stranded RNA viruses contain a ribonucleoprotein complex (RNP) which participates in transcription and replication. The RNP comprises the RNA genome tightly coated with the nucleoprotein (N protein) and associated with the RNA-dependent RNA polymerase (L) and its co-factor phosphoprotein (M1 or P). Encapsidation of both the genome and antigenome by interaction with the N protein is necessary to generate transcription/replication-competent progeny virions. The entry site of the N protein has been determined on the genome or antigenome for a few members of the Mononegavirales (Blumberg et al., 1983; Moyer et al., 1991), but little is known about the RNA binding site on the N proteins themselves. Comparisons between Mononegavirales N protein sequences revealed a short central region of identity between those of the filoviruses and paramyxoviruses and, with less significance, those of rhabdoviruses (aa 247–386 in Sendai virus and 223–362 in rabies virus strain PV) (Sanchez et al., 1992), based particularly on the conservation of hydrophobic stretches of amino acids (Barr et al., 1991; Kondo et al., 1990); this region was predicted to contain the RNA binding site. In fact, the central region of the vesicular stomatitis virus (VSV) N protein was shown to be in contact with the genome RNA by electron microscopy (Thomas et al., 1985). It has been demonstrated that the NH₂-terminal 90% of Sendai and measles virus N proteins contains the RNA binding site (Bankamp et al., 1996; Buchholz et al., 1993; Curran et al., 1993; Heggeness et al., 1981). Recently, two regions in the Sendai virus N protein were shown to be involved in binding to the viral RNA: one occupied the extreme NH₂ terminus (residues 1–82) whereas the other was located towards the COOH end (residues 258–399) (Buchholz et al., 1993; Myers et al., 1997).

Preliminary results from UV-LASER cross-linking studies demonstrated that the N protein from rabies virus RNP was able to bind very specifically to the 5′ end of the viral antigenome and particularly to the first 12 nucleotides. In this paper, we focus on the N protein. Firstly, its structure was studied by digestion with various proteases and an NH₂ portion was shown to be very resistant to protease digestion. Further, this trypsin-resistant domain (376 aa) was demonstrated to be sufficient for specific binding to a synthetic RNA probe corresponding to the 45 nt at the 5′ end of the rabies virus antigenome (Leader probe). Secondly, we identified a
55 aa peptide which is directly involved in binding to the RNA.

Methods

- **Proteins.** RNPs were purified from infected cells extracts as described previously (Perrin, 1996). Protein samples were diluted in 62.5 mM Tris–HCl (pH 8.8), 2% SDS, 5% β-mercaptoethanol, 10% glycerol and 0.002% bromophenol blue, boiled for 5 min and then analysed by 15% SDS–PAGE at 20 mA.

Proteins were transferred onto a nitrocellulose membrane by semi-dry transfer for 1 h at 0.8 mA/cm² in 48 mM Tris base, 39 mM glycine and 20% methanol (v/v). The membrane was incubated with a specific antibody (rabbit polyclonal or mouse monoclonal) for 1 h at 37 °C in PBS–Tween buffer (PBS with 0.05% Tween 20) containing 1–5% BSA. After washing in PBS–Tween buffer, the membrane was incubated for 1 h at 37 °C with goat anti-rabbit or goat anti-mouse IgG antibody coupled to alkaline phosphatase (Promega) and then developed with nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate (Boehringer).

- **Protein quantification.** Total protein concentration was determined with the Pierce (BCA) protein assay (Smith et al., 1985). N protein concentration was measured by ELISA (Perrin et al., 1986) or by comparison with standards on Coomassie blue-stained SDS–PAGE gels.

- **SDS PhastGel high density gel electrophoresis.** Samples were dialysed for 12 h at 4 °C against 10 mM Tris–HCl (pH 7.5), lyophilized and resuspended in 4 µl 62.5 mM Tris–HCl (pH 8.8), 2% SDS, 5% β-mercaptoethanol, 10% glycerol and 0.002% bromophenol blue. Samples were then boiled for 5 min and electrophoresed (10 mA, 15 °C) on SDS PhastGel high density gel (Pharmacia) in 200 mM Tricine, 200 mM β-mercaptoethanol, 10% glycerol and 0.002% bromophenol blue. Samples were then dialysed and electrophoresed (10 mA, 15 °C) on SDS PhastGel high density gel (Pharmacia) in 200 mM Tricine, 200 mM Tris, 0.5% (v/v) SDS, pH 8.1.

- **Protease digestion of RNPs.** Digestion of purified RNPs (115 ng/µl) under native conditions was carried out for 12 h in 5 mM Tris–HCl (pH 7.5) containing 400 mM NaCl using modified sequencing-grade proteases (Boehringer): trypsin (23 ng/µl, 37 °C); lysine-specific endoprotease (lys-C; 5.8 ng/µl, 37 °C); aspartate-specific endoprotease (asp-N; 5.8 ng/µl, 37 °C); and glutamate-specific endoprotease (glu-C or V8; 5.8 ng/µl, 25 °C). Lys-C digestion was also performed under denaturing conditions in incubation buffer containing 0% SDS. Protease digestion was stopped by addition of 1.35 µM (1 µl) T7CK.HCI (Boehringer) to trypsin and lys-C reactions, 5 mM (1 µl) EDTA (pH 8) to asp-N reactions, and 0.1 mM (1 µl) 3,4-dichloroisocoumarin (Boehringer) to V8 reactions. The resulting peptides from native trypsin digestion were separated by SDS–PAGE, and then transferred for 15 h (18 V, room temperature) onto PVDF membranes (ProBlott; Applied Biosystems) in 50 mM Tris base, 50 mM boric acid. Peptides visualized with Coomassie blue staining were excised from the membrane and subjected to NH₂-terminal sequence analysis by automated Edman degradation on an Applied Biosystems 473A sequencer.

- **In vitro synthesis of rabies virus Leader RNA probe.** Oligodeoxynucleotides comprising either the T7 promoter (−17 to +2) upstream of the region complementary to the rabies virus Leader RNA (bottom strand) or the complement of the −15 to +1 region of the T7 promoter (top strand) were synthesized (Milligan et al., 1987). The partially double-stranded templates were prepared by boiling equimolar amounts of top and bottom strands for 5 min and then annealing by slow cooling at room temperature. The [32P]UTP-Leader RNA probe was synthesized from 0.52 pmol partially double-stranded DNA template incubated in 30 µl 40 mM Tris–HCl (pH 7.9), 6 mM MgCl₂, 2 mM spermidine, 10 mM NaCl, 1 mM DTT, 1 U/µl RNasin (Promega), 400 µM ATP, 400 µM GTP, 400 µM CTP and 21 µM UTP (Pharmacia). Amounts of [α-32P]UTP (800 Ci/mmol; Amersham) were adjusted to obtain probes with specific activities of 6.34 × 10⁶ to 2.8 × 10⁸ c.p.m./mol (Cerenkov). Transcription was carried out by incubating twice for 15 min at 37 °C with 30 U T7 RNA polymerase (Pharmacia). The non-radioactive Leader RNA probe was synthesized from 47 pmol partially double-stranded DNA template in 200 µl of a similar reaction mixture to that used above except that it contained 400 µM UTP. Transcription was carried out by incubating twice for 30 min at 37 °C with 150 U T7 RNA polymerase. After elimination of the DNA templates with RQ1 DNase I (Promega), RNA probes were purified by 8% urea–PAGE and eluted in TE buffer (10 mM Tris–HCl (pH 7.5), 1 mM EDTA (pH 8)). Probe concentrations were estimated in scintillation counts for labelled probes or by ethidium bromide staining for non-radioactive probes. The Leader RNA probe starts with one or two non-viral G residues corresponding to positions +1 and +2 of the T7 promoter.

- **RNA–protein interactions and analysis of RNA–protein complexes by UV-LASER cross-linking.** The [32P]-labelled probe (0.079 pmol) was incubated with 1 or 1.5 µg RNP preparation in the absence or presence of E. coli rRNA at room temperature for 20 min in 10 µl final reaction volumes containing 450 mM NaCl, 5 mM Tris–HCl (pH 7.5). Samples from the RNA–protein interaction experiments (10 µl in 0.5 ml Eppendorf tubes) were irradiated at 260 nm using a Nd YAG LASER with a single high intensity pulse (5 ns duration, 0.5 × 10¹¹ W/m² intensity; approximate dose 100 J/m²). Covalent RNA–protein complexes were analysed by 15% SDS–PAGE after digestion with 200 µg/ml RNase A (Boehringer) and 1000 U/ml RNase T1 (Sigma) for 1 h at 37 °C. Radioactive signals were quantified by scanning gels and autoradiograms with a PhosphorImager (ABI) and a Masterscan (Scanalyser, respectively.

- **Affinity chromatography.** Leader RNA–N protein complexes resulting from UV-LASER irradiation of the Leader probe (2.5 × 10¹⁰ mol) and RNPs (720 µg) were extensively digested with lys-C endoprotease under denaturing conditions (0.07% SDS). An equimolar quantity (2.5 × 10¹⁰ mol) of the cDNA complementary to the probe, biotinylated at its 5' end, was added and allowed to hybridize for 30 min at 37 °C. Leader probe–cDNA hybrids were collected on streptavidin magnetic beads (Dyna) for 30 min at 25 °C with gentle stirring, according to the manufacturer’s recommendations. The beads were harvested with a magnet and washed at room temperature (three times in 5 mM Tris–HCl (pH 7.5), 1 mM EDTA with decreasing NaCl concentrations (800, 400 and 150 mM) and then in distilled water). Sequential alkaline hydrolysis of the Leader probe–cDNA hybrids first in 200 µl and then in 100 µl 100 mM NaOH, each for 20 min at room temperature, released the N protein peptides covalently linked to the Leader probe from the beads. The two supernatants were pooled, neutralized with 100 µl 1 M Tris–HCl (pH 7.5) and subjected to NH₂-terminal sequencing.

- **Sequence analysis of proteins.** Sequence analyses were performed using packages available at the Pasteur Institute Scientific Computer Service, Paris, France. Multiple alignments were made using the Clustal W program (GCG package) with a 10 gap opening penalty and the Henikoff matrix (Henikoff & Henikoff, 1993). Similarity profiles were designated from multiple alignments by the plotsimilarity program (GCG) (window = 50).

Results

**Structural analysis of the rabies virus N protein with proteases**

Preliminary work demonstrated that high ionic strength...
(450 mM NaCl) destabilized the macromolecular structure of the RNPs, which were the source of N protein capable of specifically interacting with $^{32}$P-labelled RNA probes. In order to outline the domains of the N protein responsible for this interaction, its molecular structure was evaluated by digestion with various proteases in 400 mM NaCl. Although the proteases had numerous potential cutting sites throughout the protein (Fig. 1A), SDS–PAGE analysis (Fig. 1B) revealed resistant polypeptides of 42, 45, 43 and 45 kDa after digestion with trypsin (lane a), asp-N (lane b), V8 (lane c) and lys-C (lane d) endoproteases, respectively. With V8 and lys-C endoproteases, only a small proportion of the N protein was digested, demonstrating its strong resistance to proteolysis. The resistant protein was confirmed as truncated N protein by Western blotting of the trypsin digestion products with anti-RNP polyclonal (lane f) and 5D53 and 8D2 anti-N protein monoclonal antibodies (lanes h and j, respectively). Using this sensitive method, traces of undigested N protein were also observed after trypsin digestion (lanes f and h). Extensive modification of the digestion conditions (substrate/ enzyme ratio, temperature, time or medium composition) failed to increase the susceptibility of the protein to the proteases. In fact, the most complete digestion of the N protein was obtained with lys-C under heavily denaturing conditions (0.07% SDS). This protocol generated peptides of 4–29 kDa (Fig. 3A, lane b). Only chemical attack with cyanogen bromide achieved almost complete hydrolysis, generating fragments with sizes of ≤ 4 kDa, which is as expected from the sequence (not shown). These data illustrate the compact structure of the N protein from RNPs.

Trypsin digestion, which generated the smallest fragment (t42) of 42 kDa, also produced an additional 17 kDa fragment (t17) (Fig. 1B, lane a), which was not detected by Western blot with either anti-RNP polyclonal or anti-N protein monoclonal antibodies (lanes f, h and j). The sum of the two fragments was close to the molecular mass of the N protein (55 kDa), suggesting a structure with two compact domains separated by a protease-sensitive link. To accurately delineate each domain, the undigested N protein and the t42 and t17 fragments were submitted to NH$_2$-terminal sequencing. The N protein and t42 were refractory to sequencing, suggesting that they shared the same blocked NH$_2$ terminus. In contrast, t17 gave the sequence T-D-V-A-L, which corresponds to position 377–381 in the N protein sequence, indicating that the main cleavage site for trypsin is at position K-376. Thus, trypsin digestion cleaved between two resistant regions: a large NH$_2$-terminal side mapping between K-364 (insensitive to trypsin) and K-376 (sensitive) and the smaller 73 residue domain of the COOH terminus (Fig. 1A).

**The NH$_2$-terminal domain of the N protein specifically binds to Leader probe**

We compared the ability of the $^{32}$P-labelled Leader probe to bind to both the full-length (Fig. 2A, lane b) and trypsin-digested N protein (lanes c and d) by UV-LASER cross-linking. Tryptic digestion after UV-LASER irradiation showed that t42 (t42-2) was the only labelled fragment identified by SDS–PAGE autoradiography (lane c). Thus, cross-linking to the probe did not modify the accessibility of the N protein to the protease. Similarly, the t42 protein generated by trypsin digestion before UV-LASER irradiation (t42-1) bound the $^{32}$P-labelled Leader probe efficiently (lane d). Thus, elimination of the COOH-tail did not modify the capacity of the N protein to bind RNA. These results indicated that t42 contained the RNA
binding domain. Next, we compared the RNA binding specificity of t42 with that of the entire N protein (Fig. 2B). A constant amount of $^{32}$P-labelled Leader probe was incubated with saturating quantity (> 100-fold molar excess) of protein (either N or t42) in the presence of increasing amounts of E. coli rRNA. Trypsin digestion of RNPs was done either before (t42-1) or after (t42-2) UV-LASER cross-linking to the $^{32}$P-labelled Leader probe. Reaction mixtures were incubated at 37 °C either before (N1) or after (N2) UV-LASER cross-linking to the $^{32}$P-labelled Leader probe or not incubated at 37 °C (N). Covalent complexes were digested with RNases A + T1 and analysed. Signals were quantified with a PhosphorImager. Dashed lines indicate the (E. coli rRNA)$_{50}$ value (see text).

Fig. 2. Identification of the domain of the N protein responsible for specific binding to the Leader probe. ($^{32}$P]UTP-Leader probe (0-079 pmol = 1-2 ng) was incubated with RNPs or trypsin-digested RNPs in the presence of various amounts of E. coli rRNA. After UV-LASER cross-linking, samples were digested with RNases A + T1 and analysed by 15% SDS–PAGE. Gels were quantified by scanning with a PhosphorImager. (A) Identification of the trypsin-resistant RNA binding domain of the N protein. Covalent complexes were made with 1.5 µg protein extract in the presence of E. coli rRNA (1 µg). Lane: RNP, RNPs; a, trypsin digest of RNPs (t42); b, RNPs cross-linked to the $^{32}$P-labelled Leader probe; c, trypsin digest of RNPs after cross-linking to the $^{32}$P-labelled Leader probe; d, trypsin digest of RNPs before cross-linking to the $^{32}$P-labelled Leader probe. The asterisk indicates the position of the protease. (B) Inhibition of Leader probe complex formation with N protein or t42 by E. coli rRNA. Complexes were formed with protein extract (1 µg) and with increasing amounts of E. coli rRNA. Trypsin digestion of RNPs was done either before (t42-1) or after (t42-2) UV-LASER cross-linking to the $^{32}$P-labelled Leader probe. Reaction mixtures were incubated at 37 °C either before (N1) or after (N2) UV-LASER cross-linking to the $^{32}$P-labelled Leader probe or not incubated at 37 °C (N). Covalent complexes were digested with RNases A + T1 and analysed. Signals were quantified with a PhosphorImager. Dashed lines indicate the (E. coli rRNA)$_{50}$ value (see text).

Fig. 3. Identification of the region of the N protein directly bound to the viral RNA by lys-C digestion of Leader probe–N covalent complexes under denaturing conditions. The Leader probe was incubated with RNPs. After LASER irradiation, the complexes were digested with lys-C endoprotease in 0.07% SDS. (A) Characterization of the lys-C peptides by both Coomassie blue staining (lanes a and b) and scanning with a PhosphorImager (lanes c and d) of 20% SDS PhastGel high density gel. Covalent complexes were made with ($^{32}$P]UTP-Leader probe (0.4 pmol = 6 ng), RNPs (7.5 µg) and E. coli rRNA (5 µg). They were digested with RNases A + T1 before gel analysis. Lanes: a and c, RNPs cross-linked to the Leader probe; b and d, lys-C RNPs cross-linked to the Leader probe. (B) Identification of the regions of the N protein directly bound to the RNA by affinity chromatography. The Leader probe (250 pmol = 3.9 µg) was incubated with RNP (720 µg). Covalent RNA–lys-C-generated peptide complexes were purified by RNA–cDNA affinity chromatography. The NH$_2$ termini of unpurified and purified products were sequenced (six cycles). For each putative tryptic peptide in the N protein (Table 1), the probability of its presence in the medium was calculated before (Pre-chromatography) and after (Post-chromatography) the purification procedure. This diagram shows the presence probabilities of each of the putative peptides (top) plotted against the protein sequence (bottom).
specificity was evaluated by the concentration of E. coli rRNA which inhibited the formation of 50% of the Leader–protein complexes ([E. coli rRNA]_{50} value). Since trypsin digestion requires a 37 °C incubation step, two additional controls were included for the entire N protein: 37 °C incubation before cross-linking (N1) and 37 °C incubation after cross-linking (N2) (compared to no incubation step at 37 °C (N)). Results of competition experiments showed that t42 and the N protein displayed very similar binding specificities for the Leader probe: ([E. coli rRNA]_{50} value for t42-1 and N1 was 250–275 mg/l and that for t42-2, N2 and N was 600 mg/l. This indicated that all the elements required for the viral RNA binding specificity are located in t42, whereas the COOH-tail t17 seems to play no significant role. The inhibitory effect of pre-incubation at 37 °C on both t42 and N binding specificity was significant and resulted in a 60% reduction in the ([E. coli rRNA]_{50} value for t42-1 and N1 compared with t42-2, N2 or N. This observation suggested a strong denaturing effect of temperature on N protein (and t42) structure. Although the induced conformational change is slight, since lengthy incubation at 37 °C did not modify the sensitivity of the protein to proteases, it is sufficient to have a dramatic effect on the specific RNA binding activity of the N protein. These data suggest the high sensitivity of rabies virus N protein to disorganization.

Identification of the region of the N protein directly bound to viral RNA

In order to map with greater precision the region(s) of the N protein involved in intimate contact with the viral RNA, covalently linked 32P-labelled Leader probe–N protein complexes were extensively digested with lys-C protease under strong denaturing conditions (0.07% SDS). After digestion with RNases to remove excess probe, samples were analysed on an SDS PhastGel high density gel (Fig. 3 A). Coomassie blue staining showed that proteolysis generated a peptide population of 2.5–29 kDa (compare Fig. 3 A, lanes a and b). Analysis of the gel with a PhosphorImager showed that the radioactive
signal, which was exclusively associated with the N protein before proteolysis (Fig. 3A, lane c), correlated with a unique 6–9 kDa peptide after proteolysis (Fig. 3A, lane d). The length of the residual RNA probe bound to the lys-C peptide after RNase treatment was estimated as 1–3 nt by 20% urea–PAGE analysis (not shown). Assuming molecular masses of 330 and 110 kDa for the nucleotides and amino acids, respectively, the length of the 6–9 kDa radioactive peptide was estimated as 54–59 aa.

Affinity chromatography was used to purify this RNA binding peptide. The Leader probe–N protein covalent complexes were digested with lys-C under strong denaturing conditions as described above, except that no RNase treatment of the probe was performed. Then, the Leader probe covalently linked to the peptide was hybridized with the complementary oligodeoxynucleotide and biotinylated at its 5’ end. cDNA–RNA hybrids were retained on streptavidin magnetic beads, then hydrolysed with NaOH; the released peptides were identified by six successive cycles of NH$_2$-terminal sequencing. The only other peptide which was also significantly concentrated by six successive cycles of NH$_2$-terminal sequencing was the S-P-Y-S-S-N (position 298–303) sequence NH$_2$-terminal domain containing the RNA binding domain. This peptide was localized in the COOH-tail and remained a minor component of the elution products. From these data, the location of the N protein directly linked to the Leader probe can be identified by six successive cycles of NH$_2$-terminal sequencing (not shown). Assuming molecular masses of 330 and 110 kDa for the nucleotides and amino acids, respectively, the length of the 6–9 kDa radioactive peptide was estimated as 54–59 aa.

Discussion

The tight structure of the rabies virus N protein was highlighted by its strong resistance to protease digestion. Proteolysis under native conditions with any of the four enzymes used consistently separated two domains, despite numerous potential cleavage sites: a large domain involving 80% of the NH$_2$ terminus and a COOH-tail. This architecture is archetypal in other Mononegavirales N proteins (Bankamp et al., 1996; Dietzschold et al., 1987; Heggeness et al., 1981; Lamb & Kolakofsky, 1996; Mountcastle et al., 1974). However, whereas the COOH-tail of the Sendai virus N protein is sensitive to trypsin (Heggeness et al., 1981), both domains are rather resistant in the rabies virus N protein. The presence of a unique cleavage site for trypsin was previously observed with the rabies virus strain CVS (Dietzschold et al., 1987); it was located about 100 residues from the COOH end, near residue 350. The size of the t17 fragment would have led us to the same conclusion, but NH$_2$-terminal sequencing of t17 mapped the cleavage site to K-376 and demonstrated that t17 was composed of only the 73 COOH-terminal residues. This discrepancy between the calculated (8 kDa) and apparent (17 kDa) molecular mass of the COOH-tail was also observed for the entire N protein (50 and 55 kDa, respectively). Such a discrepancy has been shown to be common for N proteins from Mononegavirales (Elango, 1989; Fang et al., 1994; Morgan et al., 1984; Sanchez et al., 1986, 1992; Wetzol et al., 1994; Zuidema et al., 1987). The discrepancy could result from post- translational modification(s) since S-389 was demonstrated to be phosphorylated in the viral N protein (Dietzschold et al., 1987). An alternative explanation could be the presence of a unique cluster of acidic amino acids at position 362–393 (particularly 378–393, at the NH$_2$ end of t17), as reported for the VSV and Sendai virus P proteins, or for the yeast GCN4 transcription factor, which affect electrophoretic mobility (Barik & Banerjee, 1991; Curran et al., 1991, 1994; Gallione et al., 1981; Hope & Struhl, 1986).

UV-LASER cross-linking of the $^{32}$P-labelled Leader probe to the N protein, either followed or preceded by trypsin cleavage, clearly demonstrated the importance of the NH$_2$-terminal domain [stopping between K-364 (insensitive) and K-376 (sensitive)] for specific RNA binding activity. The location of all the necessary elements for specific RNA binding in the NH$_2$-terminal domain of the rabies virus N protein is different from that in Sendai virus and VSV N proteins where the COOH terminus is also essential in vivo for the specificity of the viral RNA binding via an interaction with the P phosphoprotein (Buchholz et al., 1993; Curran et al., 1993; Masters & Banerjee, 1988; Takacs et al., 1993).

Affinity chromatography identified the rabies virus N protein peptide starting with the sequence NH$_2$ S-P-Y-S-S-N (55 aa, position 298–352) as being primarily involved in RNA binding with the Leader probe. This major RNA binding site is located at the COOH end of the NH$_2$-terminal domain as in the Sendai virus N protein (Buchholz et al., 1993; Myers et al., 1997). Also, the major RNA binding peptide encompasses the short region of identity between N proteins of viruses of the order Mononegavirales (positions 247–386 in Sendai virus and 223–362 in rabies virus strain PV) (Sanchez et al., 1992). Another peptide of the rabies virus N protein COOH-tail,
Fig. 4. Similarity between N proteins from Rhabdoviridae. Multiple alignments were made with the Clustal W program (opening gap penalty = 10) based on the multiple alignments: Lyssavirus (rabies virus strain PV and Mokola virus), Vesiculovirus (VSV ind, NJ, chan, piry), Ephemerovirus [Adelaide River virus (ARV) and bovine ephemeral fever virus (BEFV)] and insect virus (sigma virus). The horizontal dashed line indicates the medium score similarity; the numbered positions indicate the borders of the conserved regions; the grey region shows the hyperconserved motif; the black bar shows the lys-C-generated peptide (298–352) of the rabies virus strain PV N protein purified by affinity chromatography after binding to the Leader probe. (B) Hyperconserved motifs (underlined). The sequence of the rabies virus strain PV N protein is given as reference; conserved amino acids are given in capital letters, whereas divergent amino acids are in lower case. Numbers in parentheses for the genus Lyssavirus indicate the genotype of the virus. SYNV, Sonchus yellow net virus; RYSV, rice yellow stunt virus.

starting with the sequence NH$_2$ R-S-H-I-R-R (27 aa, position 418–444) was also concentrated by affinity chromatography, but remained a minor component of the eluted products. Considering that deletion of the COOH-tail does not affect RNA binding specificity, its contribution to specific binding to the Leader probe by UV-LASER cross-linking in 450 mM NaCl (i.e. under very stringent conditions favouring only interaction between residues which are in close physical contact) could be negligible with respect to that of the principal core and would be imperceptible in competition experiments presented in Fig. 2.

We compared N protein sequences of animal rhabdoviruses (lyssaviruses, vesiculoviruses, ephemeroviruses and insect rhabdoviruses). Fig. 4(A) displays a similarity profile obtained...
from the multiple alignments. Only two regions, residues 64–105 and 201–329 (with reference to the rabies virus strain PV) are reasonably conserved. The second region can be divided into two relatively more conserved domains: 210–242 and 271–317. The latter domain overlaps the NH₂-half of the major RNA binding peptide (298–352; black bar, Fig. 4A) starting with the sequence S-P-Y-S-S-N. This sequence is part of the most conserved motif in the domain (G-L-S-G-K-S-P-Y-S-S, underlined in Fig. 4B) found in mammalian rhabdoviruses and, to a lesser extent, in insect and plant rhabdoviruses. The conservation of this motif and its presence in the peptide cross-linked to the RNA probe are two independent clues which suggest that this region of the N protein plays a major role in genome binding. The hyperconserved rhabdoviral motif (Fig. 4B) did not have any equivalent among the known RNA binding motifs recorded in the databases. Thus, the Mononegavirales N proteins may possess a new type of RNA binding motif or protein folding may participate in the formation of the active binding site. This latter hypothesis is supported by several observations: (i) the inhibitory effect of temperature (37 °C) on the RNA binding specificity of N protein (Fig. 2B); (ii) the dramatic decrease of both RNA binding activity and antigenicity of N protein solubilized from purified RNP's or bacterial extracts by guanidinium.HCl treatment (not shown); (iii) the severe effect of the most limited structural changes in the NH₂-terminal core on its RNA binding activity (this work; Bankamp et al., 1996; Buchholz et al., 1993; Curran et al., 1993; Homann et al., 1991).

The identification by UV-LASER cross-linking of a peptide involved in direct contact with the viral RNA provides the first insight of the rhabdoviral N protein RNA binding site. Even if the peptide encompasses the most conserved motif in rhabdovirus N proteins, the primary structure alone seems insufficient to define the RNA binding site, and higher order organization is certainly implied in its formation. Further studies are essential for specification of the residues implicated in specific binding and to obtain pure protein with satisfactory RNA binding activity for crystallization of the RNA–protein complex.

We acknowledge Corinne Jallet for purifying RNP's of high and reproducible quality and Dr Danielle Blondel for providing the monoclonal antibodies. The Laboratory of Protein Microsequencing performed the N protein NH₂-terminal sequencing. We are deeply indebted to Drs Ben Blumberg and Anne-Marie Gilles for helpful discussions during the preparation of this work and to Dr Andrew Borman for critical reading of the manuscript. We thank Professor Henri Buc in whose laboratory this work was partially carried out. Anne Kouznetzoff was a recipient of a fellowship from the ‘Fondation Marcel Merieux’.

References


Received 19 September 1997; Accepted 26 January 1998