



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Structure and Function of the C-terminal Domain of the Polymerase Cofactor of Rabies Virus

Manos Mavrakis^{1†}, Andrew A. McCarthy^{1†}, Stéphane Roche²
Danielle Blondel² and Rob W. H. Ruigrok^{1,3*}

¹EMBL Grenoble Outstation
BP181, 38042 Grenoble Cedex 9
France

²Unité Mixte de Virologie
Moléculaire et Structurale
UMR 2472; CNRS, 91198 Gif
sur Yvette Cedex, France

³Laboratoire de Virologie
Moléculaire et Structurale; EA
2939; Université Joseph Fourier
Grenoble, France

The phosphoprotein (P) of rabies virus binds the viral polymerase to the nucleoprotein (N)-RNA template for transcription and replication. By limited protease digestion we defined a monomeric C-terminal domain of P that can bind to N-RNA. The atomic structure of this domain was determined and previously described mutations that interfere with binding of P to N-RNA could now be interpreted. There appears to be two features involved in this activity situated at opposite surfaces of the molecule: a positively charged patch and a hydrophobic pocket with an exposed tryptophan side-chain. Other previously published work suggests a conformational change in P when it binds to N-RNA, which may imply the repositioning of two helices that would expose a hydrophobic groove for interaction with N. This domain of rabies virus P is structurally unrelated to the N-RNA binding domains of the phosphoproteins of Sendai and measles virus that are members of the same order of viruses, the non-segmented negative strand RNA viruses. The implications of this finding for the evolution of this virus group are discussed.

© 2004 Elsevier Ltd. All rights reserved.

*Corresponding author

Keywords: rabies virus; replication; transcription; polymerase; evolution

Introduction

Non-segmented negative strand RNA viruses are enveloped viruses with viral RNA (vRNA) in the opposite sense to that of mRNA. This order of viruses (*Mononegavirales*) contains the *Paramyxoviridae* (measles and Sendai virus), *Bornaviridae*, *Rhabdoviridae* (rabies and vesicular stomatitis virus, VSV) and *Filoviridae* (Ebola virus). Negative strand RNA viruses cause important human and animal disease. In 2003, measles virus killed 760,000 children, mainly in Africa and Asia (WHO report, 2003). Rabies virus kills about 50,000 people per year in

these continents but is also an increasing health concern in the USA.¹ The vRNA of these viruses is bound to the viral nucleoprotein (N) with a virus-family-specific nucleotide-N stoichiometry.^{2–5} The viral RNA-dependent RNA polymerase consists of two subunits, the large protein (L) that contains the polymerase activity and a polymerase cofactor, the phosphoprotein (P) that binds L to the N-RNA.^{6–11} The polymerase complex cannot transcribe or replicate the naked vRNA; vRNA has to be bound to N in order to be a functional template.¹²

P has a second role in the viral life-cycle as a chaperone for N that is not yet bound to vRNA (N⁰). When recombinant N is expressed alone in eukaryotic cells it binds non-specifically to cellular RNA to form N-RNA complexes that have the same structure as the viral N-RNA structures.^{5,13–15} However, when N and P are co-expressed, a soluble complex between N⁰ and P is formed that does not contain RNA.^{16–19} This N⁰-P complex provides the soluble N required for binding to newly replicated vRNA.

Sequence alignment of the phosphoproteins of the *Lyssavirus* genus (rhabdovirus genus to which the rabies virus belongs) indicates conservation and prediction of α -helical structure at the very amino terminus of the molecule (Figure 1; sequence

† M.M. & A.M. contributed equally to this work.

Present addresses: M. Mavrakis, Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA; R.W.H. Ruigrok, EMBL Grenoble Outstation, 6 rue Jules Horowitz, 38000 Grenoble, France.

Abbreviations used: P, phosphoprotein; N, nucleoprotein; L, viral polymerase; vRNA, viral RNA; PML, promyelocytic leukaemia; SAD, single anomalous dispersion; mAb, monoclonal antibody; EGS, ethylene glycolbis(succinimidyl succinate).

E-mail address of the corresponding author: ruigrok@embl-grenoble.fr

conservation for residues 1–51 and predicted α -helix for residues 10–30). There is some sequence conservation and prediction of an α -helical region between residues 93 and 130, and extensive conservation from about residue 200 until the C-terminal end (Figure 1). Most published work on the structure and function of rabies virus P suggests the existence of at least two domains, an N-terminal domain (residues 1–177/184) and a C-terminal domain (residues 178/185–297).^{20–22} The L-binding domain is present in the first 19 amino acid residues of P.²³ The domain that binds to N⁰ is located in the N-terminal 177 residues,^{20–22} whereas the C-terminal domain binds to N–RNA,^{21,22} (see Figure 2). However, deletion of the C-terminal 24 residues did not abolish binding of P to N–RNA.²¹

P also has an oligomerisation domain. When P is expressed in bacteria or in insect cells, the molecule forms oligomers but it is not clear whether these are trimers or tetramers.^{19,24} When Gigant *et al.*²⁴ expressed an N-terminal deletion mutant of P missing the first 52 amino acid residues, the molecule was still oligomeric.

The gene for the rabies virus phosphoprotein codes for a total of five polypeptides; full-length P plus N-terminal truncated forms that arise through

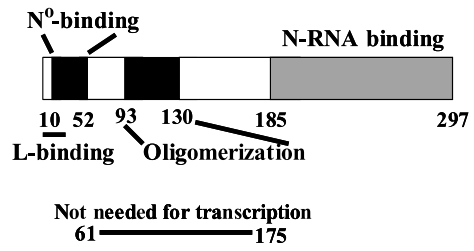


Figure 2. Proposed domain structure of rabies virus phosphoprotein. The domain structure of the protein is given in a bar code with the start and end of each domain indicated. The function of the domains, derived from previously published literature and from the experiments described here, are given above or below the bar.

a leaky scanning activity of the ribosome on P mRNA.²⁵ The shorter forms start at amino acid residues 20 (P2), 53 (P3), 69 (P4) or 83 (P5). Although the replication of rabies virus takes place in the cytoplasm, P3–5 are mainly found in the nucleus whereas P1 and P2 are cytoplasmic. The nuclear products of the P gene interact with PML (promyelocytic leukaemia) nuclear bodies that have been

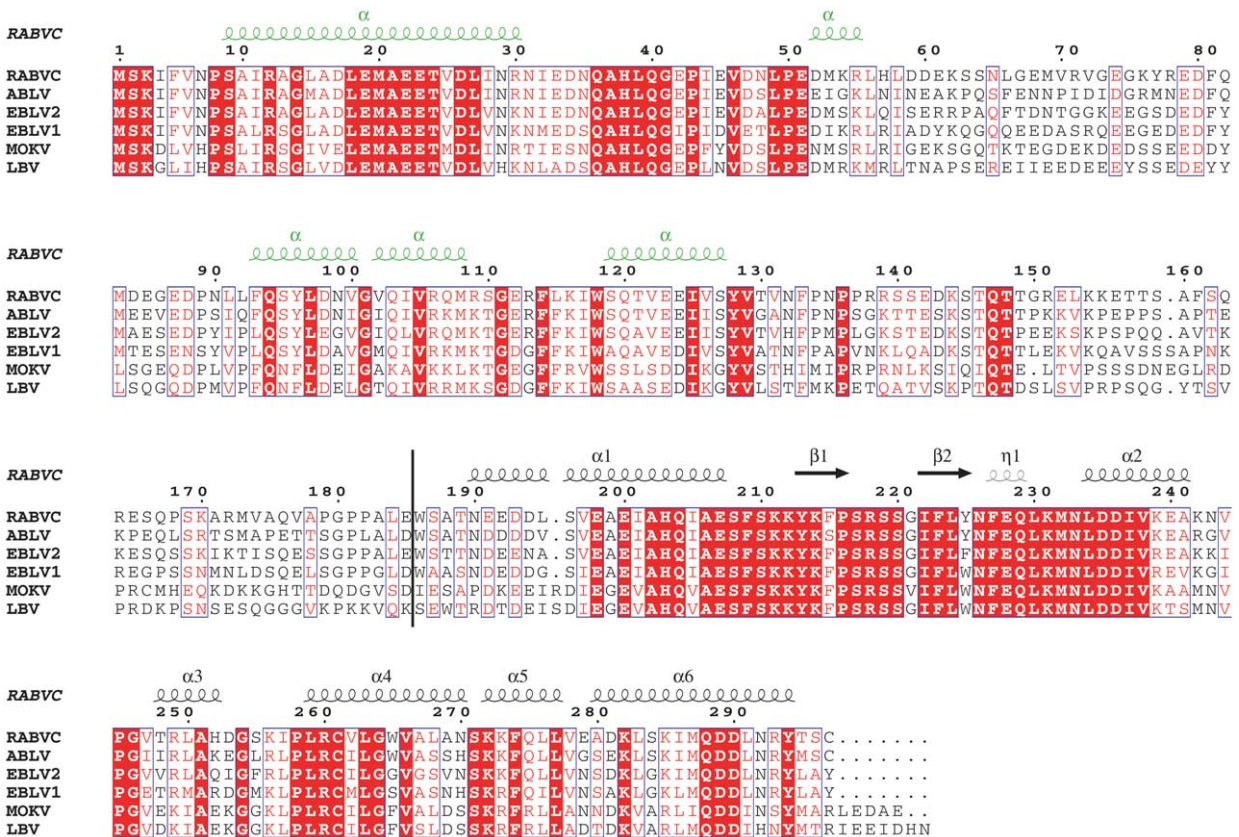


Figure 1. Multiple sequence alignment and secondary structure (predicted or observed) for the Lyssavirus phosphoproteins. The alignment was made with ClustalW for the phosphoproteins of RABVC (rabies virus CVS strain; SWISSPROT acc. P22363), ABLV (Australian bat lyssavirus; SPTREMBL acc. Q9QSP3), EBLV1/2 (European bat lyssavirus 1 and 2; SPTREMBL acc. O56776 and SPTREMBL acc. O56780, respectively), MOKV (Mokola virus; SPTREMBL acc O56775) and LBV (Lagos bat virus; SPTREMBL acc. O56773). The secondary structures that are indicated above the alignment are predicted (indicated in green) for residues 1–185 using PHD and ESPript⁷⁵ and observed (indicated in black) for the crystallised domain of rabies virus P presented here, residues 186–297.

associated with the cellular defence mechanism against viral infections.²⁶ The interaction of the P products with PML bodies may overcome the antiviral response of the cell. The Sendai virus P gene also codes for a number of different polypeptides that are generated by a stuttering mechanism of the polymerase that adds one or several G residues at a specific site in the mRNA for P.^{27,28} Some of the viral polypeptides that result from this editing mechanism are also active in fighting the innate antiviral activity of the infected cell.^{29,30}

Here, we present limited protease digestion experiments on rabies virus P that show that only the C-terminal part of the protein is protease resistant. This C-terminal domain is monomeric and binds to N-RNA. In crystallisation trials with intact P, spontaneous proteolysis occurred in the crystallisation drops and only the C-terminal domain crystallised, for which the structure was determined. The previously described mutations leading to loss of binding of this domain to N²² can now be interpreted in the light of this structure. This rabies virus P domain has a structure that is

different from that of the functionally analogous domains of Sendai and measles virus P^{31,32} and implications for the evolution of negative strand RNA viruses are discussed.

Results

Full-length phosphoprotein produced in *Escherichia coli* is oligomeric²⁴ and binds to rabies virus N-RNA¹⁵ (see also Figure 4, below). The domain structure of this recombinant protein was probed by limited protease digestion using trypsin, endoprotease Lys-C and thermolysin. These proteases digest P to protease-resistant fragments that migrate on SDS-PAGE with apparent molecular masses of 15–16 kDa (Figure 3A–C). The fragments react with monoclonal antibody 25C2 that recognises residues 225–246 of P³³ (Figure 3D). N-terminal sequencing of the fragment obtained after trypsin digestion showed that cleavage had occurred after R172. The apparent mass of the fragment from SDS-PAGE suggests that it contains

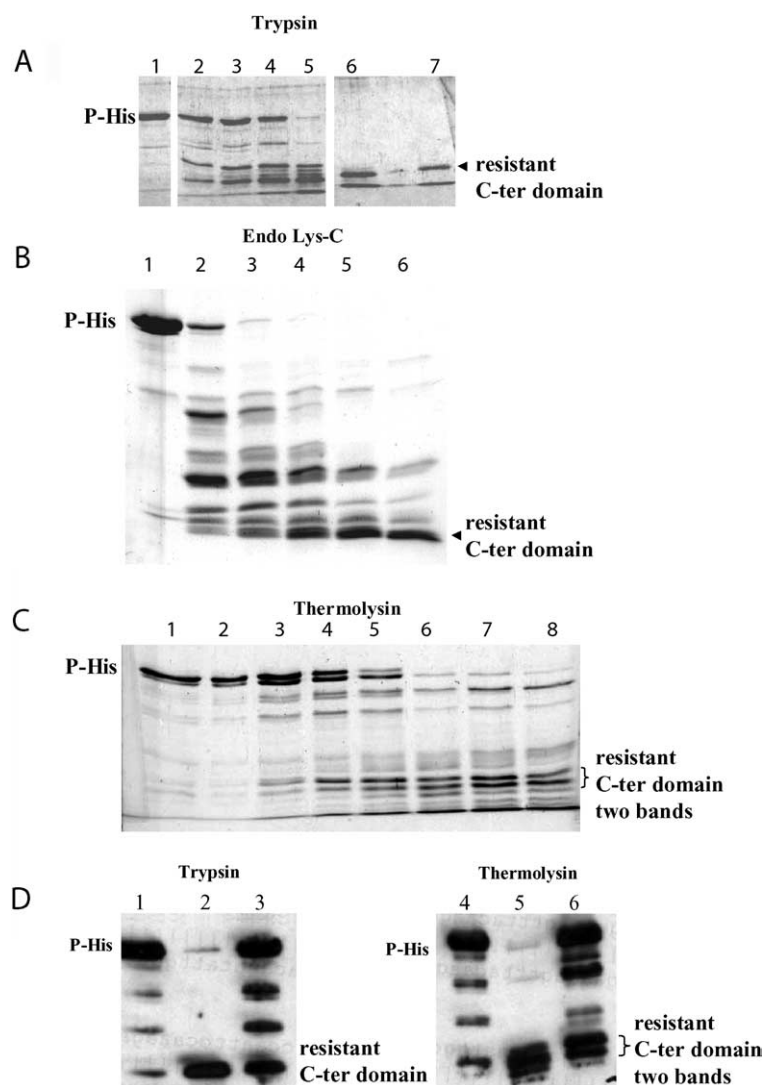


Figure 3. Limited proteolysis experiments on His-tagged rabies P. His-tagged P was purified from *E. coli* and digested for one hour at 37 °C with different SDS proteases and analysed by SDS-16% PAGE. A, Digestion with trypsin. Lane 1, no trypsin. Lanes 2–7, trypsin/P-His ratios (w/w) of 1:8000, 1:4000, 1:2000, 1:1000, 1:500 and 1:250. The reaction was quenched by adding PMSF (10 mM final concentration). B, Digestion with endoprotease Lys-C. Lane 1, no protease. Lanes 2–6, quantities of protease used per 1 µg of P-His in units (one unit hydrolyses 1 µmol of *N*-*p*-tosyl-Gly-Pro-Lys *p*-nitroanilide per minute at pH 7.7 at 25 °C). Lane 2, 3.75 × 10⁻⁵ units; lane 3, 7.5 × 10⁻⁵ units; lane 4, 1.5 × 10⁻⁴ units; lane 5, 3 × 10⁻⁴ units; lane 6, 6 × 10⁻⁴ units. The reaction was quenched by the addition of EDTA (5 mM final concentration). C, Digestion with thermolysin. Lanes 1–8, thermolysin:P-His ratios (w/w) of: lane 1, 1:64; lane 2, 1:32; lane 3, 1:16; lane 4, 1:8; lane 5, 1:4; lane 6, 1:2; lane 7, 1:1; lane 8, 2:1. The reaction was quenched by adding EDTA (5 mM final concentration). D, Western blot of P-His after digestion with trypsin (left) and thermolysin (right) using MAb 25C2. Lane 1, no trypsin. Lanes 2 and 3, trypsin/P ratios of 1:2000 and 1:8000, respectively. Lane 4, no thermolysin. Lanes 5 and 6, thermolysin/P ratios of 1:1 and 1:4, respectively.

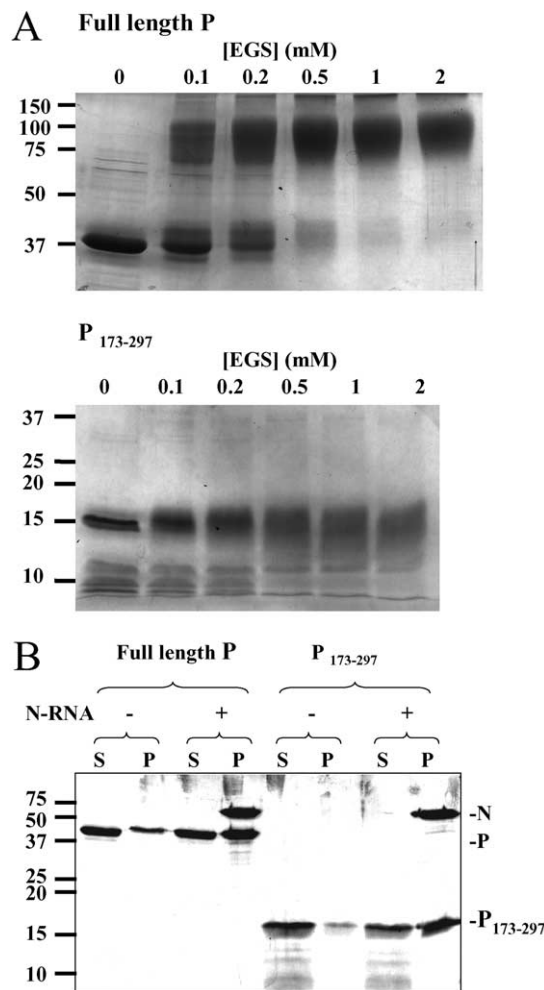


Figure 4. Oligomerisation states and N-RNA binding activities of full-length P and P₁₇₃₋₂₉₇. **A**, Full-length P and P₁₇₃₋₂₉₇ were incubated with different concentrations of the cross-linking agent EGS as indicated and analysed by 12% Tris-glycine SDS-PAGE for intact P and SDS-16% PAGE for P₁₇₃₋₂₉₇. **B**, N-RNA templates isolated from infected BSR cells were incubated with full-length P and P₁₇₃₋₂₉₇. After centrifugation of the N-RNA-P mixture through sucrose cushions, proteins present in the supernatants (S) and pellets (P) were analysed by SDS-16% PAGE.

residues 173–297. The N-terminal sequence of the endoprotease Lys-C digested fragment indicates digestion after K170. These experiments show that there exists a C-terminal domain of rabies virus P that is resistant to proteolysis.

Cross-linking experiments showed that, under conditions that totally cross-link full-length P, the C-terminal domain derived after trypsin digestion (P₁₇₃₋₂₉₇) remained monomeric (Figure 4A), suggesting that this domain is a monomer in solution. Binding of P₁₇₃₋₂₉₇ to N-RNA was compared to that of full-length P. The two proteins were incubated with N-RNA isolated from virus-infected cells. The N-RNA was then sedimented through a sucrose cushion taking with it any bound

Table 1. Data collection, phasing and refinement statistics

Crystal	Native	Se-Met
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	44.3, 44.3, 106.2	44.05, 44.05, 105.9
Wavelength (Å)	0.9393	0.9795
Resolution (Å)	30–1.5 (1.6–1.5)	30–1.9 (2.1–1.9)
Completeness (%)	98.5 (93.9)	92.7 (74.5)
R_{merge}	6.3 (20.7)	4.9 (18.3)
$R_{\text{merge}}^{\text{ano}}$		4.3 (16.8)
$\langle I/\sigma(I) \rangle$	16.8 (7.1)	36.1 (12.9)
Redundancy	6.4 (6.1)	11.9 (4.0)
Unique reflections	19, 783	16, 801
Total reflections	120, 720	216, 310
Figure of merit		0.31
R_{cryst} (%)	17.5	
R_{free} (%)	19.8	

Values in parentheses are for the outermost resolution shell. $R_{\text{merge}} = \sum_h \sum_i |I_{h,i} - \langle I \rangle_h| / \sum_h |I_h|$ calculated for the whole data set. $R_{\text{merge}}^{\text{ano}} = R_{\text{merge}}$ with Friedel-related reflections excluded. Figure of merit = $|\sum P(\alpha)e_{i\alpha} / \sum P(\alpha)|$, where $P(\alpha)$ is the phase probability distribution and α is the phase. $R_{\text{cryst}} = \sum |F_o| - |F_c| / \sum |F_o|$. R_{free} was calculated as for R_{cryst} with 5% of the data omitted from the structural refinement.

P proteins. Both full-length P and P₁₇₃₋₂₉₇ were found to co-sediment with the N-RNA (Figure 4B), indicating that the C-terminal domain of Rabies virus P is the N-RNA-binding domain, as is the case for the C-terminal domain of Sendai virus P.

Non-tagged, full-length P was produced in insect cells and put in crystallisation drops. The molecule that crystallised was not the intact protein but a C-terminal domain corresponding to residues 186–297. Degradation of intact P to this domain by contaminating proteases was reproducible and was observed for native and for selenomethionine-substituted protein. The structure was solved using the single anomalous dispersion (SAD) method as indicated in Materials and Methods and the crystallographic details are given in Table 1. All but the last two C-terminal residues could be modelled.

The C-terminal domain folds as a single compact domain, with six α -helices ($\alpha 1$ – $\alpha 6$), one 3_{10} helix ($\eta 1$), and a small, two-stranded antiparallel β -sheet (Figure 5). The overall structure resembles a pear cut along its length with a flat face and a round face (Figure 6A and B). The top part of the half-pear is comprised of the first three helices ($\alpha 1$, $\alpha 2$, $\alpha 3$), with $\alpha 2$ and $\alpha 3$ packing perpendicularly to $\alpha 1$. The abrupt change in direction between $\alpha 2$ and $\alpha 3$, provided by P245, is necessary for their packing against $\alpha 1$. The lower part is comprised of the small β -sheet, with two helices ($\alpha 4$ and $\alpha 5$) packing against one side and $\alpha 6$ capping one edge of the β -sheet. The two parts are packed together by hydrophobic interactions between $\alpha 2$ and $\alpha 4$, with some additional hydrophobic residues contributed by the C terminus of $\alpha 1$. The flat face is composed mainly from α -helices, whereas the round face is made from the C terminus of $\alpha 1$, the short β -sheet, the 3_{10} helix and the loops connecting these secondary structures. A positive surface potential

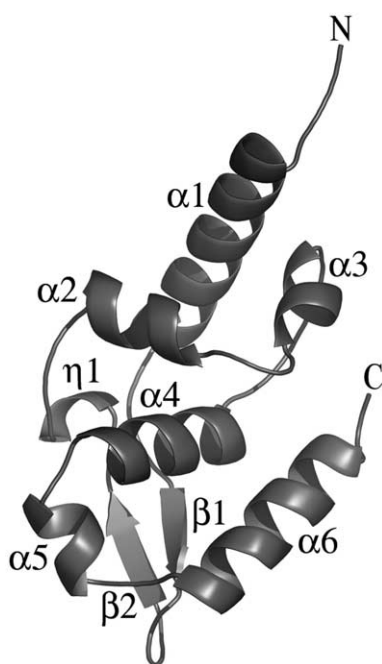


Figure 5. Structure of the carboxy-terminal domain of rabies virus phosphoprotein. Ribbon diagram of the secondary structure of residues 186–297 of rabies virus P, CVS strain. The amino and carboxy-terminal ends are indicated as are the secondary structure elements. The Figure was generated with PYMOL (<http://www.pymol.org>).

area exists on this round face, as calculated in GRASP³⁴ (Figure 6A–C). On the other hand, the flat face has a more symmetric charge distribution, with the main surface feature being a shallow uncharged groove, where one side is comprised of the loop between $\alpha 2$ and $\alpha 3$ and the C terminus of $\alpha 4$, and the other side is comprised of $\alpha 6$. This groove also contains a prominent hydrophobic pocket, the W-hole, composed of L244, P245, C261, W265, M287 and L291 (Figure 7A).

The C-terminal domain crystallised in trigonal space group $P3_121$ and the asymmetric unit contains one molecule. The flat face mediates practically all of the intermolecular contacts, where two major regions of the molecule are involved in contacts with neighbours. The largest region of symmetry interaction involves residues of $\alpha 1$ and $\alpha 2$ from one molecule, which interact with residues from the last three helices ($\alpha 4$, $\alpha 5$ and $\alpha 6$) of a symmetry-related molecule. The smaller region is composed of the four N-terminal residues, which interact with residues in or around the W-hole from a symmetry-related neighbour. Here, W186, at the extreme N terminus of the crystallised fragment is lodged into the W-hole of its symmetry-related molecule and is therefore immobilised and well defined in the electron density map (Figure 7B).

A search of the Protein Data Bank with the full C-terminal domain using the program DALI³⁵ did not reveal any significant homology to other

proteins. Three hits with a Z-score >2.5 were found but close scrutiny showed only one with any significant structural similarity. This was from the C-terminal domain of the α -subunit from F_1 -ATPase with a Z-score of 2.8 and an rms value of 3.9 Å over 66 α -carbon positions, where the structural conservation is between the first four helices, ($\alpha 1$, $\alpha 2$, $\alpha 3$ and $\alpha 4$) from the C-terminal domain of P and the final four helices from the C-terminal α -helical bundle from the α -subunit of the ATPase. So the fold of the C-terminal domain of P could be described as an embellished α -helical bundle, where the small β -sheet and the 3_{10} helix have been inserted between the first and second helix. However, the structure described here could be considered as a new fold given the importance of the β -sheet and the final two helices to the hydrophobic core packing. There is no structural homology with the N-RNA binding domains of the phosphoproteins of Sendai and measles virus.^{31,32}

Jacob *et al.*²² have produced random mutations in Mokola virus P and tested if those mutants could still bind to Mokola virus nucleoprotein in a yeast two-hybrid assay. Mokola virus is closely related to rabies virus and the sequences of the P proteins of the two viruses are very similar (Figure 2). The mutated residues in the Mokola virus P protein were compared to the corresponding residues in the rabies virus P protein. We have analysed the structural consequences of the mutations identified as having the largest effect on N-RNA binding (Table 2). Eleven of these mutants are expected to have a destabilising effect on the structure. Seven mutations form two major clusters on the surface. The first cluster comprises the highly conserved lysine residues (K211, K212 and K214) and L224 (Figure 6C and D). Two of the lysine residues contribute to the positive potential on the round face described above. The second cluster is comprised of three residues and is centred on the less conserved W-hole and includes C261, W265 and M287 (Figure 7). The remaining four mutations that replace lysine residues by uncharged or negatively charged amino acids are randomly spread with no discernible order (Figures 6D and 7A). From this mutational analysis it would seem that both the positively charged patch on the round face of the domain and the W-hole on the flat face are important for binding to N-RNA.

Rabies virus phosphoprotein contains four potential phosphorylation sites.³⁶ There is evidence that phosphorylation of the first two sites, S63 and S64, is mediated by a unique cellular protein kinase of 71 kDa that is packaged inside rabies virus particles, and is necessary during the infection process.³⁶ The other two possible phosphorylation sites are S210 and S271 that can be phosphorylated *in vitro* by specific isomers of protein kinase C although it is not known whether this actually happens in the infected cell.³⁶ S210 lies on a surface exposed loop between $\alpha 1$ and $\beta 1$ and a phosphorylated serine residue can easily be created without any conformational change. Interestingly, this

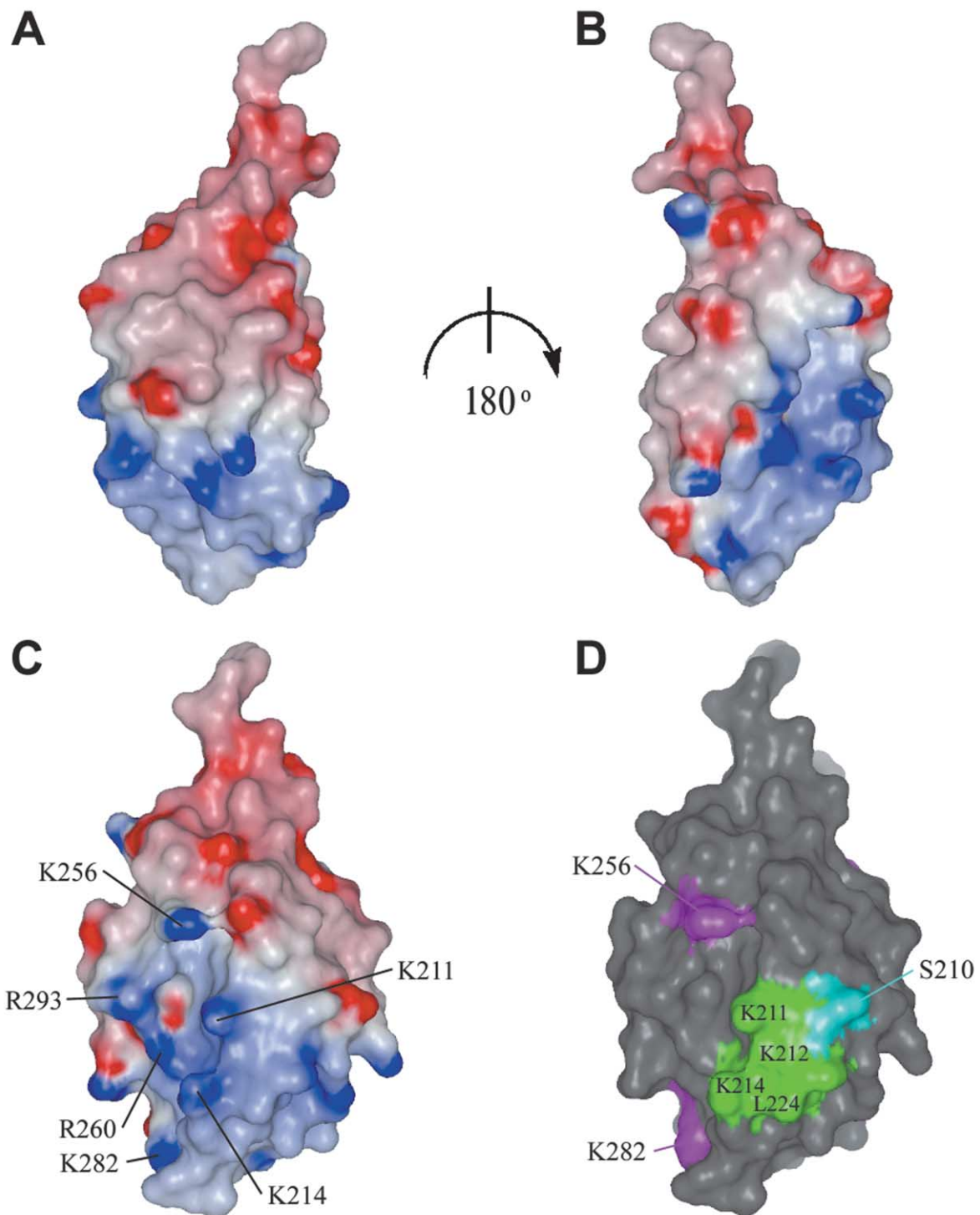


Figure 6. Surface representation of the C-terminal domain of rabies virus phosphoprotein. A, Side-view of the electrostatic potential map of the C-terminal domain showing its “half-pear” shape and the distribution of charges on its surface; blue indicates positively charged residues (12 kT) and red negatively charged residues (−12 kT). The amino acid residue at the top of the molecule is the well-defined W186 that is wedged in the W-hole of a neighbouring molecule in the crystal (see Figure 7). B, As A rotated by 180°. C, Front view of the round face (rotated by 90° compared to the view in A) indicating the residues that are involved in binding to N-RNA (see Table 2). D, Same view as in C indicating the residues in cluster 1 (see the text) in green. Residues that are involved in binding to N-RNA but that are not in cluster 1 are shown in magenta. S210, one of the phosphorylation sites on this domain, is in blue. The Figure was generated with Grasp³⁴ and GLR (L. Esser, personal communication).

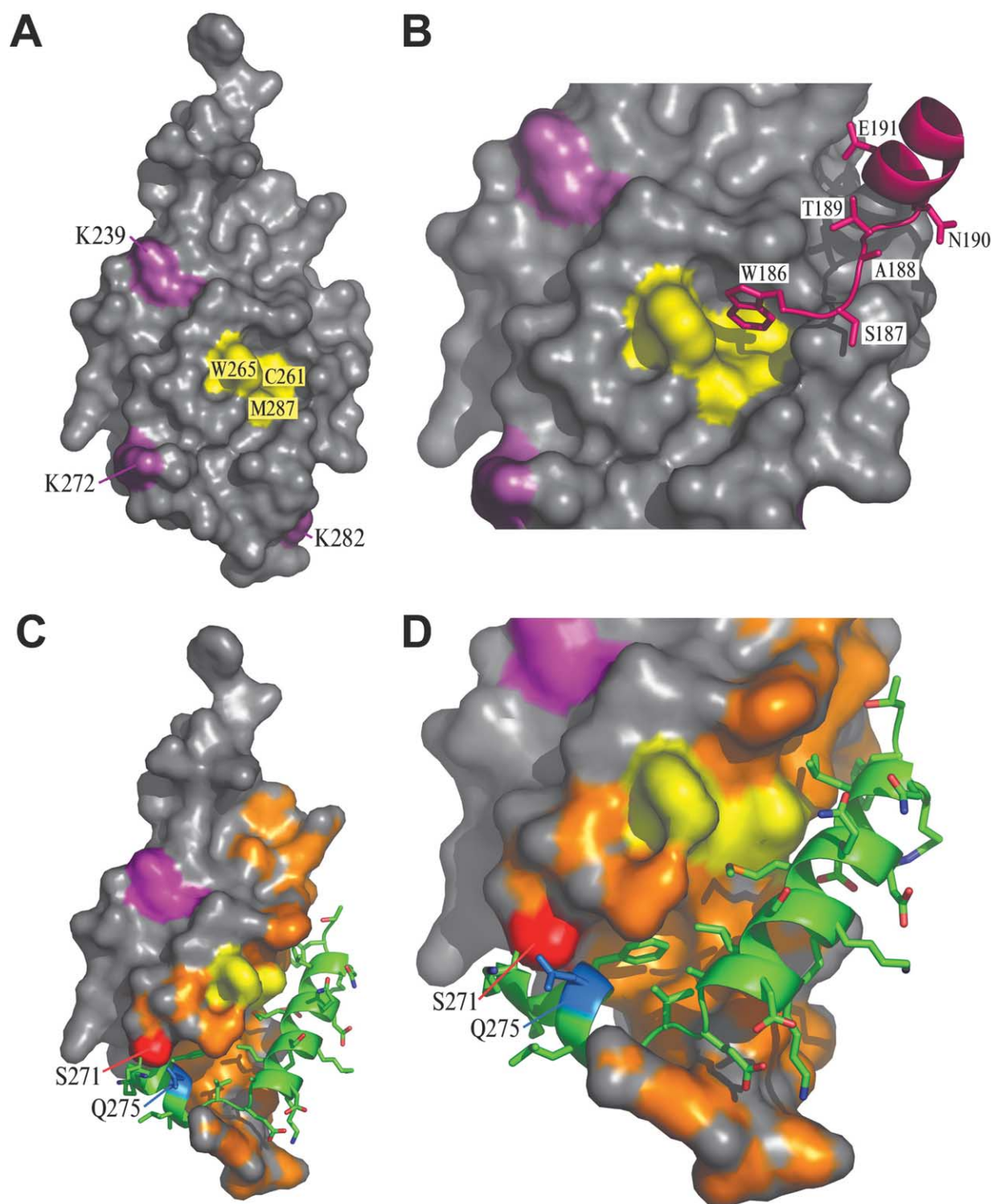


Figure 7. View of the flat face of the C-terminal domain of Rabies virus phosphoprotein. A, View of the flat face with residues in cluster 2 (W-hole, see the text) in yellow and other residues contributing to the binding to N-RNA in magenta. B, Detail of the W-hole and its interaction with the N-terminal W186 of a neighbouring subunit in the crystal. C, Same representation as in A with helices 5 and 6 outlined in green. The uncharged or hydrophobic surfaces on this face of the molecule and underneath helices 5 and 6 are indicated in orange. The potential phosphorylation site S271 and the centre of the epitope recognised by the mAb described by Toriumi *et al.*³⁷ are indicated in red and blue, respectively. D, Higher magnification detail of C. The Figure was generated with PYMOL (<http://www.pymol.org>).

Table 2. Mutations in Mokola virus phosphoprotein that prevent binding to Mokola virus N are compared with corresponding residues in rabies virus P and placed in the structure of rabies virus P₁₈₆₋₂₉₇

RAB ^a	MOK ^b	Position in structure	Effect on structure
<i>A. Mutations that likely lead to destabilisation of the structure</i>			
S196	D→G	α 1	Helix destabilisation
V197	I→L, V	α 1	Core packing destabilisation
F209	F→L	Loop	Core packing destabilisation
Y213	Y→C	β 1	Core packing destabilisation
F215	F→Y	β 1	Core packing destabilisation
M232	M→T	Loop	Core packing destabilisation
D235	D→G	α 2	Helix destabilisation
R260	R→G	α 4	Loss of H-bond between α 4/ α 6
L276	L→P	α 5	Helix destabilisation
N292	N→S	α 6	Helix destabilisation
Y294	Y→C	α	Core packing destabilisation
<i>B. Mutations that are located on the surface of P but not in clusters 1 and 2</i>			
K239	K→Q	α 2	
K256	K→E	loop	
K272	K→A	α 5	
K282	K→N, E	α 6	
<i>C. Mutations that are located on the surface of P and grouped in clusters 1 and 2</i>			
K211	K→E	Loop	Cluster 1
K212	K→E, R	Loop	Cluster 1
K214	K→E	β 1	Cluster 1
L224	L→S	β 2	Cluster 1
C261	C→R, S	α 4	Cluster 2 (W-hole)
W265	F→L	α 4	Cluster 2 (W-hole)
M287	I→T	α 6	Cluster 2 (W-hole)

^a Rabies virus phosphoprotein residue type and number.

^b Corresponding residue in Mokola virus phosphoprotein that, when mutated to the residue indicated, no longer binds to Mokola virus N in a yeast two-hybrid assay.²²

serine is also adjacent to the positively charged patch and the first cluster of mutations described above (Figure 6D). Thus, it is possible that phosphorylation of S210 would alter the N-RNA binding characteristics of P. S271 is located between α 4 and α 5 and is only partly solvent-exposed and quite constricted (Figure 7C and D). It is therefore probable that some conformational rearrangement would be required for phosphorylation to occur. S271 is located between the two clusters of mutations and close to K272 that, when mutated to A, also prevents binding of P to N (Table 2 and Figure 7A).

Discussion

Binding of rabies virus P to N-RNA

The phosphoprotein binds the viral polymerase (L) to the nucleocapsid. Here, we show that the domain of P that performs this function is a C-terminal domain from residue 186 until the C-terminal end. Residues that are implicated in the interaction with N can be found on the round and flat faces of the domain (Figures 6 and 7). Toriumi *et al.* have described a monoclonal antibody that can distinguish two conformations of rabies virus P, one when it is soluble and one when it is bound to N-RNA.³⁷ The presumed epitope for this mAb is located around Q275 (Figure 7C and D). This suggests that binding of P to N-RNA may be associated with a conformational change resulting

in the rearrangement of helices 5 and 6. These C-terminal 24 residues of P can be deleted without inhibiting the binding of P to N-RNA.²¹ It is, therefore, likely that α 5 and α 6 can occupy a conformation that is different from the one that is found in the crystal structure. Phosphorylation of S271 may have an influence on the structure of the protein in this area. Figure 7C and D show the molecule with the W-hole in yellow and helices 5 and 6 in green. These two helices lie in a hydrophobic groove (orange) and their displacement would expose this groove to the solvent. The conformation of the monomeric molecule described here is probably not stable when the two helices are displaced. Binding of P to N-RNA may involve complementary interactions from N, shielding the groove from the solvent. As mentioned above, a shallow uncharged groove is slightly exposed even with the helices in place, the W-hole making part of this surface structure.

Oligomerisation of P

Gigant *et al.*²⁴ have shown that the first 52 amino acid residues of rabies virus P are not involved in oligomerisation and here we have shown that P₁₇₃₋₂₉₇ is monomeric as well. Because oligomerisation of Sendai virus P takes place through the formation of an inter-monomer helical coiled-coil,³⁸ we suggest that the oligomerisation domain be situated in the predicted helical region between residues 93 and 130 (Figures 1 and 2). Jacob *et al.* have shown that a P mutant with a deletion from

residue 61 to 175, so most likely monomeric, was still active in supporting transcription activity by L (Figure 2).²² This deletion mutant still has the L-binding and the N-RNA binding domains. This is in contrast to the transcription and replication activities of the Sendai virus phosphoprotein that does require the presence of the oligomerisation domain.^{39,40} The difference between these two phosphoproteins could be due to a higher affinity of rabies virus phosphoprotein for its N-RNA than that of Sendai virus P for its N-RNA. The C-terminal domains of Sendai and measles virus P that bind to their N-RNAs⁴¹ are very small and flexible domains,^{31,32} much smaller than the corresponding domain of rabies virus P. Another important reason why the P protein of Sendai virus has to be oligomeric in order to be active is that the binding domain for Sendai L overlaps with the oligomerisation domain^{38,42} and that only oligomeric P can bind to L.

Conserved characteristics of the *Mononegavirales*

Viruses in this order have in common that the vRNA is in the opposite sense to that of messenger RNA, that the vRNA is bound to nucleoprotein and that the viral RNA polymerase complex is only functional on this N-RNA matrix. The N-RNA complexes of all these viruses form helical structures with different helical parameters but with N-subunits that have the same overall shape.⁴³ It is likely that the nucleoproteins of all virus families in this order are homologous but proof for this will only come when several nucleoprotein structures have been determined. The viral polymerases (L) have also conserved sequence motifs and are, thus, homologous.^{44,45} Another conserved feature of this group of viruses is the order of the genes for the structural proteins on the vRNA.

No structural similarities amongst the matrix proteins of the *Mononegavirales* or between the phosphoproteins of rabies virus and Sendai/measles virus

The matrix proteins of this group of viruses play an important role in the formation of the virus particles.⁴⁶ They interact both with the membrane-embedded viral glycoproteins and with the nucleocapsids and their tendency to polymerise is thought to drive particle formation. These proteins also have so-called late domains that interact with host-proteins involved in cellular membrane modifications.⁴⁷ However, the sequences of the matrix proteins show very little identity (sometimes less than 10% even for closely related genera like *Lyssa*- and *Vesiculoviruses*, both *Rhabdoviridae*) and the three structures of negative strand RNA virus matrix proteins determined so far, that of influenza virus (a segmented negative strand RNA virus),^{48,49} Ebola virus⁵⁰ and that of VSV,⁵¹ are all unrelated and all three present a new fold.⁴⁷ The fact that the

structures are not related means that these proteins do not come from the same evolutionary lineage.

As for the matrix proteins, there is very little sequence identity between the phosphoprotein sequences from the different virus families. Even within families, like the *Rhabdoviridae*, there is very little sequence identity. Here, we have presented the first structure of part of a rabies virus protein, the C-terminal domain of P. There is no structural similarity between the C-terminal domains of Sendai/measles virus P and that of rabies virus P described here. Again, this suggests that the C-terminal domains of these P proteins do not come from the same evolutionary lineage.

Evolution of negative strand RNA viruses

The fact that the three known structures of negative strand RNA virus matrix protein are unrelated and that the two known structures of the N-RNA-binding domains of the phosphoproteins are also different, even though they fulfil the same functions, either implies that RNA virus evolution is so extensive that homologous proteins have evolved to independent folds (although from the known matrix protein structures it is not possible to see how one structure could evolve into the other) or it means that heterologous structures have been incorporated into these viruses through heterologous recombination. Heterologous recombination has been documented for *Nidovirales* (*Arteri-* and *Coronaviruses*), positive strand RNA viruses, where homologous RNA recombination by the viral polymerase is part of the mechanism for the production of mRNAs.^{52,53} Homologous recombination inside gene segments of segmented negative strand RNA viruses has also been shown or suggested; recombination between the polymerase genes of two Arenavirus strains has been documented⁵⁴ whereas recombination in the haemagglutinin gene of the 1918 Spanish influenza virus has been suggested from sequence comparisons.⁵⁵ For heterologous recombination to occur, the viral polymerase has to jump from the viral N-RNA template to a nucleoprotein-encapsidated host mRNA and then back again to the viral RNA template. Non-segmented negative strand virus nucleoprotein expressed alone in bacterial or eukaryotic cells binds in a non-specific fashion to cellular RNAs.^{5,13,56-58} Although during a virus infection most newly made N is first bound to the phosphoprotein in the N⁰-P complex for delivery onto newly made vRNA, it is very likely that some N will nevertheless bind to cellular RNA providing an N-cellular RNA template for the viral polymerase. It is also known that these polymerases can jump either inside the viral N-RNA template⁵⁹ or from the viral template to the newly copied, N-bound complementary strand in order to form copy-back interfering particles.⁶⁰ Heterologous recombination for the influenza viruses was suggested from the structure of the influenza C virus glycoprotein.⁶¹ From a comparison between

the influenza A and C haemagglutinin structures it was suggested that the influenza virus haemagglutinin consists of an ancestral fusion protein combined with an independent haemagglutinin domain, for influenza C virus further complemented by an esterase domain spliced into this haemagglutinin domain. The esterase domain is homologous to bovine brain and bacterial esterases.⁶¹ Finally, when Khatchikian *et al.*⁶² tried to select for a pathogenic turkey influenza virus to grow in chicken cells, they obtained a virulent mutant that had acquired 54 nt coming from the 28 S ribosomal RNA of the infected cells. This insertion led to a change in the presentation of the cleavage site on haemagglutinin to the cellular protease needed for proteolytic activation, allowing the virus to grow in the new host cells.

If the hypothesis on heterologous recombination will turn out to be correct, then the conserved parts of the ancestor common to all negative strand RNA viruses may be the nucleoprotein or N-RNA and the RNA-dependent RNA polymerase, i.e. the basic components of the replication machinery. The matrix proteins are not homologous nor are the phosphoproteins. The glycoproteins are probably also not conserved among all viruses. Although its structure is not yet known, the fusogenic glycoprotein of the rhabdoviruses has probably a very different structure from that of the fusion proteins of other negative strand RNA viruses because their fusion characteristics are so very different.⁶³ Apart from these major viral proteins, many viruses of this group code for additional proteins for which no structures are known. The presence of heterologous matrix proteins in these viruses may be the reason why the shapes of the virus particles in this group are so variable, from more or less spherical (influenza, paramyxoviruses) to bullet-shaped (rhabdoviruses) and filamentous particles (filoviruses).

Materials and Methods

Limited proteolysis experiments

C-terminal His-tagged rabies virus phosphoprotein was produced in *E. coli* BL21 cells as described.²⁴ Purified protein was taken up in 500 mM NaCl, 20 mM Tris-HCl (pH 8), 0.5 mM EDTA. Proteolysis experiments were performed in this buffer at 37 °C for one hour using the proteases and concentrations as indicated in the legend to Figure 3. The reactions were quenched as indicated and analysed on Tris-glycine SDS-16% (w/v) PAGE.

The Western blot shown in Figure 3D was performed with anti-P mAb 25C2 that was produced in mice immunised with recombinant P (PV strain) produced in *E. coli* as described.³³ The epitope of this mAb was mapped to amino acid residue positions (225–246) by a peptide array assay. Briefly, peptide arrays representing the rabies virus phosphoprotein (kindly provided by M. Schwemmler, Department of Virology, Freiburg) were chemically synthesised on cellulose sheets by the SPOT synthesis technique and treated with mAb 25C2 as described.⁶⁴

Functional characterisation of the C-terminal domain of P

Cross-linking analysis

About 10 µg of full-length recombinant P (see above) or P₁₇₃₋₂₉₇ was incubated with the cross-linker ethylene glycolbis(succinimidyl succinate) (EGS) at final concentrations as indicated for Figure 4 for 30 minutes at room temperature. Reactions were quenched by the addition of 200 mM glycine and incubated for an additional 30 minutes at room temperature. The samples were then analysed by SDS-PAGE.

Binding of P and P₁₇₃₋₂₉₇ to viral nucleocapsids

Viral nucleocapsids were isolated from virus-infected cells using a CsCl gradient as described by Schoehn *et al.*¹⁵ Full-length recombinant P protein or P₁₇₃₋₂₉₇ was incubated with nucleocapsids in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl for 1.5 hours at room temperature. The mixture was then deposited on a 20% (w/w) sucrose cushion which was spun for 2.5 hours at 25,000 rpm and 4 °C in an SW55 rotor (Beckman). The pellet and supernatant fractions were analysed by SDS-16% (w/v) PAGE.

Expression and purification of native and selenomethionine-substituted P for crystallisation

Native P

High Five cells were infected with AcNPVM1 encoding full-length P⁶⁵ at a multiplicity of infection of 5 p.f.u./cell. Cells were harvested 60 hours post infection, pelleted at 350g for ten minutes, resuspended in hypotonic buffer (50 mM NaCl, 20 mM Tris-HCl (pH 7.5), 1 mM EDTA in the presence of protease inhibitor cocktail "complete-EDTA free" from Roche Diagnostics), and lysed by three cycles of freezing and thawing. Cell debris was pelleted at 12,000g for 15 minutes at 4 °C and the cleared lysate was loaded onto a DEAE Sepharose Fast Flow (Pharmacia Biotech) column. The protein was eluted with a continuous 150 mM–300 mM NaCl gradient and fractions analysed by 10% Tris-glycine SDS-PAGE. Fractions containing P were pooled, concentrated by ultrafiltration (Centrifugal filter devices of Millipore) and loaded onto a Superdex 200 HR10/30 (Pharmacia Biotech) column. The protein was eluted in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM DTT. The eluate was analysed by Tris-Tricine SDS-15% PAGE and two populations of full-length P were found in overlapping peaks (the elution volumes were far from the void volume, indicating that both populations represented soluble forms and not aggregates). Fractions were pooled separately, so as to contain either of the two populations or a mixture of both, and crystallisation trials were set up using all three fractions. Only the second population, i.e. the P eluting at the larger volume from the S200 HR10/30, crystallised.

Selenomethionine-substituted P

Sf 9 insect cells adapted to serum-free medium were infected as for the native protein. After 24 hours the medium (Sf900II plus penicillin/streptomycin; Gibco-BRL, Life Technologies) was removed and replaced with methionine-free medium (Sf900II without L-methionine and without L-cysteine; Gibco-BRL, Life Technologies).

The cells were left in this medium for four hours to promote methionine depletion after which it was replaced with medium containing 50 mg/l of L-selenomethionine and 20 mg/l of L-cysteine and left for 67 hours before harvesting. Lysis and purification was as described for the native protein.

Crystallisation and data collection

Crystallization trials were all performed using full-length P in sitting drops at 20 °C, using a robot to set up drops of 2 µl (1 µl of protein solution plus 1 µl of reservoir solution). The protein concentration was 16 mg/ml in 20 mM Tris-HCl (pH 7.50), 150 mM NaCl, 5 mM DTT. Native, rod-shaped crystals grew between three and seven weeks under three reservoir conditions differing only in the salt used. Condition 1, 0.2 M MgCl₂, 0.1 M Bis-Tris (pH 5.5), 25% (w/v) polyethylene glycol (PEG)3350. Condition 2, 0.2 M NaCl, 0.1 M Bis-Tris (pH 5.5), 25% PEG3350. Condition 3, 0.2 M (NH₄)₂SO₄, 0.1 M Bis-Tris (pH 5.5), 25% PEG3350. Crystals of selenomethionine-substituted P grew between three and eight weeks after setting up drops in 0.2 M MgCl₂, 0.1 M Bis-Tris (pH 5.5), 25% PEG3350. The crystals were then transferred from their mother liquor to mother liquor plus 20% (v/v) glycerol prior to flash-freezing.

All X-ray data were collected at 100 K on beam line ID14-4 at the European Synchrotron Radiation Facility (ESRF). A high-resolution native data set to 1.5 Å was collected from a single crystal. A highly redundant data set from a single Se-Met crystal was collected at the peak of the Se-Met signal, as measured by X-ray fluorescence. The protein crystallised in space group P3₁21 with one molecule in the asymmetric unit. All the data were integrated and scaled using the XDS suite⁶⁶ and data statistics are given in Table 1.

Structure determination and refinement

Two Se-Met sites were located using SOLVE⁶⁷ via the single anomalous dispersion (SAD) method. The experimental phases were then improved by solvent flattening with RESOLVE.⁶⁸ This gave an interpretable map and an initial poly-alanine model. Scaling of the structure factor amplitudes between the native and Se-Met data gave an *R*-factor of 55.5%, indicating non-isomorphism. Molecular replacement was therefore used to reposition the molecule in the unit cell. The poly-alanine model from RESOLVE⁶⁸ was positioned in the cell using the AMoRe molecular replacement program,⁶⁹ as implemented in CCP4.⁷⁰ Input of the solution model into ARP/warp⁷¹ resulted in a nearly complete model.

Crystal structure refinement was performed using REFMAC⁷² and model building was performed with O.⁷³ A randomly chosen subset of 5% of reflections for the calculation of the free *R*-factor was used as a monitor of refinement.⁷⁴ Ordered water molecules were added at locations where there was $|F_o| - |F_c|$ density greater than 3σ above the mean and could form potential hydrogen-bonding contacts with the protein structure.

Protein Data Bank accession code

Co-ordinates have been deposited in the RCSB PDB as entry 1vvi.

Acknowledgements

We thank Drs Yves Gaudin (Gif-sur-Yvette), Dan Kolakofsky (Geneva) and Winfried Weissenhorn (EMBL-Grenoble) for numerous discussions. We gratefully acknowledge use of the crystallisation robot at the Institut de Biologie Structurale in Grenoble and thank the European Synchrotron Radiation Facility (ESRF) at Grenoble for beam-time at ID 14-4.

References

- Messenger, S. L., Smith, J. S., Orciari, L. A., Yager, P. A. & Rupprecht, C. E. (2003). Emerging pattern of rabies deaths and increased viral infectivity. *Emerg. Infect. Dis.* **9**, 151–154.
- Egelman, E. H., Wu, S. S., Amrein, M., Portner, A. & Murti, G. (1989). The Sendai virus nucleocapsid exists in at least four different helical states. *J. Virol.* **63**, 2233–2243.
- Thomas, D., Newcomb, W. W., Brown, J. C., Wall, J. S., Hainfeld, J. F., Trus, B. L. & Steven, A. C. (1985). Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. *J. Virol.* **54**, 598–607.
- Flamand, A., Raux, H., Gaudin, Y. & Ruigrok, R. W. H. (1993). Mechanisms of rabies virus neutralisation. *Virology*, **194**, 302–313.
- Mavrakis, M., Kolesnikova, L., Schoehn, G., Becker, S. & Ruigrok, R. W. H. (2002). Morphology of Marburg virus NP-RNA. *Virology*, **296**, 300–307.
- Mellon, M. G. & Emerson, S. U. (1978). Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus. *J. Virol.* **27**, 560–567.
- Hamaguchi, M., Yoshida, T., Nishikawa, K., Naruse, H. & Nagai, Y. (1983). Transcriptive complex of Newcastle disease virus. Both L and P proteins are required to constitute an active complex. *Virology*, **128**, 105–117.
- Horikami, S. M., Curran, J., Kolakofsky, D. & Moyer, S. A. (1992). Complexes of Sendai virus NP-P and P-L proteins are required for defective interfering particle genome replication *in vitro*. *J. Virol.* **66**, 4901–4940.
- Portner, A., Murti, K. G., Morgan, E. M. & Kingsbury, D. W. (1988). Antibodies against Sendai virus L protein: distribution of the protein in nucleocapsids revealed by immunoelectron microscopy. *Virology*, **163**, 236–239.
- Curran, J., Pelet, T. & Kolakofsky, D. (1994). An acidic activation-like domain of the Sendai virus P protein is required for RNA synthesis and encapsidation. *Virology*, **202**, 875–884.
- Smallwood, S., Ryan, K. W. & Moyer, S. A. (1994). Deletion analysis defines a carboxyl-proximal region of Sendai virus P protein that binds to the polymerase L protein. *Virology*, **202**, 154–163.
- Emerson, S. U. (1987). Transcription of vesicular stomatitis virus. In *The Rhabdoviruses* (Wagner, R. R., ed), pp. 245–269, Plenum, New York.
- Spehner, D., Kirn, A. & Drillien, R. (1991). Assembly of nucleocapsid like structures in animal cells infected with a vaccinia virus recombinant encoding the measles virus nucleoprotein. *J. Virol.* **65**, 6296–6300.
- Iseni, F., Barge, A., Baudin, F., Blondel, D. & Ruigrok,

- R. W. H. (1998). Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. *J. Gen. Virol.* **79**, 2909–2919.
15. Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D. & Ruigrok, R. W. H. (2001). Structure of recombinant rabies virus N-RNA and identification of the phosphoprotein binding site. *J. Virol.* **75**, 490–498.
 16. Masters, P. S. & Banerjee, A. K. (1988). Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to non-specific RNA. *J. Virol.* **62**, 2658–2664.
 17. Peluso, R. W. (1988). Kinetic, quantitative, and functional analysis of multiple forms of the vesicular stomatitis virus nucleocapsid protein in infected cells. *J. Virol.* **62**, 2799–2807.
 18. Peluso, R. W. & Moyer, S. A. (1988). Viral proteins required for the *in vitro* replication of vesicular stomatitis virus defective interfering particle genome RNA. *Virology*, **162**, 369–376.
 19. Mavrakis, M., Iseni, F., Mazza, C., Schoehn, G., Ebel, C., Gentzel, M. *et al.* (2003). Isolation and characterisation of the rabies virus N^o-P complex produced in insect cells. *Virology*, **305**, 406–414.
 20. 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**, 2889–2896.
 21. Fu, Z. F., Zheng, Y., Wunner, W. H., 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.
 22. Jacob, Y., Real, E. & Tordo, N. (2001). Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. *J. Virol.* **75**, 9613–9622.
 23. Chenik, M., Schnell, M., Conzelmann, K. K. & Blondel, D. (1998). Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. *J. Virol.* **72**, 1925–1930.
 24. Gigant, B., Iseni, F., Gaudin, Y., Knossow, M. & Blondel, D. (2000). Neither phosphorylation nor the amino-terminal part of rabies virus phosphoprotein is required for its oligomerization. *J. Gen. Virol.* **81**, 1757–1761.
 25. 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.
 26. Blondel, D., Regad, T., Poisson, N., Pavie, B., Harper, F., Pandolfi, P. P. *et al.* (2002). Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies. *Oncogene*, **21**, 7957–7970.
 27. Vidal, S., Curran, J. & Kolakofsky, D. (1990). A stuttering model for paramyxovirus P mRNA editing. *EMBO J.* **9**, 2017–2022.
 28. Hausmann, S., Garcin, D., Delenda, C. & Kolakofsky, D. (1999). The versatility of paramyxovirus RNA polymerase stuttering. *J. Virol.* **73**, 5568–5576.
 29. Garcin, D., Latorre, P. & Kolakofsky, D. (1999). Sendai virus C proteins counteract the interferon-mediated induction of an antiviral state. *J. Virol.* **73**, 6559–6565.
 30. Kato, A., Kiyotani, K., Sakai, Y., Yoshida, T. & Nagai, Y. (1997). The paramyxovirus, Sendai virus, V protein encodes a luxury function required for viral pathogenesis. *EMBO J.* **16**, 578–587.
 31. Johansson, K., Bourhis, J. M., Campanacci, V., Cambillau, C., Canard, B. & Longhi, S. (2003). Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J. Biol. Chem.* **278**, 44567–44573.
 32. Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W. P., Ruigrok, R. W. H. & Marion, D. (2004). Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology*, **319**, 201–211.
 33. Raux, H., Iseni, F., Lafay, F. & Blondel, D. (1997). Mapping of monoclonal antibody epitopes of the rabies virus P protein. *J. Gen. Virol.* **78**, 119–124.
 34. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296.
 35. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
 36. Gupta, A. K., Blondel, D., Ghoudhary, S. & Banerjee, A. K. (2000). The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C. *J. Virol.* **74**, 91–98.
 37. Toriumi, H., Honda, Y., Morimoto, K., Tochikura, T. S. & Kawai, A. (2002). Structural relationship between nucleocapsid-binding activity of the rabies virus phosphoprotein (P) and exposure of epitope 402-13 located at the C terminus. *J. Gen. Virol.* **83**, 3035–3043.
 38. Tarbouriech, N., Curran, J., Ruigrok, R. W. H. & Burmeister, W. P. (2000). Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nature Struct. Biol.* **7**, 777–781.
 39. Curran, J., Boeck, R., Lin-Marq, N., Lupas, A. & Kolakofsky, D. (1995). Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, *via* predicted coiled coils. *Virology*, **214**, 139–149.
 40. Kolakofsky, D., Le Mercier, P., Iseni, F. & Garcin, D. (2004). Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology*, **318**, 463–473.
 41. Curran, J., Homann, H., Buchholz, C., Rochat, S., Neubert, W. & Kolakofsky, D. (1993). The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation. *J. Virol.* **67**, 4358–4364.
 42. Bowman, M. C., Smallwood, S. & Moyer, S. A. (1999). Dissection of individual functions of the Sendai virus phosphoprotein in transcription. *J. Virol.* **73**, 6474–6483.
 43. Schoehn, G., Mavrakis, M., Albertini, A., Wade, R., Hoenger, A. & Ruigrok, R. W. H. (2004). 12 Å Structure of trypsin treated measles virus N-RNA. *J. Mol. Biol.* **339**, 301–312.
 44. Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8**, 3867–3874.
 45. Poch, O., Blumberg, B. M., Bougueleret, L. & Tordo, N. (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J. Gen. Virol.* **71**, 1153–1162.
 46. Schmitt, A. P. & Lamb, R. A. (2004). Escaping from the cell: assembly and budding of negative-strand RNA viruses. *Curr. Top. Microbiol. Immunol.* **283**, 145–196.
 47. Timmins, J., Ruigrok, R. W. H. & Weissenhorn, W. W.

- (2004). Structural studies on the Ebola virus matrix protein VP40 indicate that matrix proteins of enveloped RNA viruses are analogous but not homologous. *FEMS Microbiol. Letters*, **233**, 179–186.
48. Sha, B. & Luo, M. (1997). Structure of a bifunctional membrane-RNA binding protein, influenza virus matrix protein M1. *Nature Struct. Biol.* **4**, 239–244.
 49. Arzt, S., Baudin, F., Barge, A., Timmins, P., Burmeister, W. P. & Ruigrok, R. W. H. (2001). Combined results from solution studies on intact influenza virus M1 protein and from a new crystal form of its N-terminal domain show that M1 is an elongated monomer. *Virology*, **279**, 439–446.
 50. Dessen, A., Volchkov, V., Dolnik, O., Klenk, H. D. & Weissenhorn, W. (2000). Crystal structure of the matrix protein VP40 from Ebola virus. *EMBO J.* **19**, 4228–4236.
 51. Gaudier, M., Gaudin, Y. & Knossow, M. (2002). Crystal structure of vesicular stomatitis virus matrix protein. *EMBO J.* **21**, 2886–2892.
 52. Lai, M. M. (1992). RNA recombination in animal and plant viruses. *Microbiol. Rev.* **56**, 61–79.
 53. Lai, M. C. (1996). Recombination in large RNA viruses: Coronaviruses. *Semin. Virol.* **7**, 381–388.
 54. Charrel, R. N., Lemasson, J.-J., Garbutt, M., Khelifa, R., De Micc, P., Feldmann, H. & de Lamballerie, X. (2003). New insights into the evolutionary relationships between arenaviruses provided by comparative analysis of small and large segment sequences. *Virology*, **317**, 191–196.
 55. Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. (2001). Recombination in the haemagglutinin gene of the 1918 “Spanish flu”. *Science*, **293**, 1842–1845.
 56. Fooks, A. R., Schadeck, E., Liebert, U. G., Dowsett, A. B., Rima, B. K., Steward, M. *et al.* (1995). High-level expression of the measles virus nucleocapsid protein by using a replication-deficient adenovirus vector: induction of an MHC-1-restricted CTL response and protection in a murine model. *Virology*, **210**, 456–465.
 57. Bhella, D., Ralph, A., Murphy, L. B. & Yeo, R. P. (2002). Significant differences in nucleocapsid morphology within the Paramyxoviridae. *J. Gen. Virol.* **83**, 1831–1839.
 58. Murphy, L. B., Loney, C., Murray, J., Bhella, D., Ashton, P. & Yeo, R. P. (2003). Investigations into the amino-terminal domain of the respiratory syncytial virus nucleocapsid protein reveal elements important for nucleocapsid formation and interaction with the phosphoprotein. *Virology*, **307**, 143–153.
 59. Jennings, P. A., Finch, J. T., Winter, G. & Robertson, J. S. (1983). Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? *Cell*, **34**, 619–627.
 60. Perrault, J. (1981). Origin and replication of defective interfering particles. *Curr. Topics. Microbiol. Immunol.* **93**, 151–207.
 61. Rosenthal, P. B., Zhang, X., Formanowski, F., Fitz, W., Wong, C. H., Meier-Ewert, H. *et al.* (1998). Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. *Nature*, **396**, 92–96.
 62. Khatchikian, D., Orlich, M. & Rott, R. (1989). Increased viral pathogenicity after insertion of a 28S ribosomal RNA sequence into the haemagglutinin gene of an influenza virus. *Nature*, **340**, 156–157.
 63. Gaudin, Y., Tuffereau, C., Durrer, P., Brunner, J., Flamand, A. & Ruigrok, R. (1999). Rabies virus-induced membrane fusion. *Mol. Membr. Biol.* **16**, 21–31.
 64. Billich, C., Sauder, C., Frank, R., Herzog, S., Bechter, K., Takahashi, K. *et al.* (2002). High-avidity human serum antibodies recognizing linear epitopes of Borna disease virus proteins. *Biol. Psychiatry*, **51**, 979–987.
 65. Préhaud, C., Nel, K. & Bishop, D. H. L. (1992). Baculovirus-expressed rabies virus M1 protein is not phosphorylated: it forms multiple complexes with expressed rabies N protein. *Virology*, **189**, 766–770.
 66. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallog.* **26**, 795–800.
 67. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallog. sect. D*, **55**, 849–861.
 68. Terwilliger, T. C. (2000). Maximum-likelihood density modification. *Acta Crystallog. sect. D*, **56**, 965–972.
 69. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
 70. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
 71. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
 72. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
 73. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
 74. Brünger, A. T. (1992). The free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
 75. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.

Edited by A. Klug

(Received 28 May 2004; received in revised form 23 August 2004; accepted 24 August 2004)