Available online at www.sciencedirect.com

SCIENCE DIRECT®

Virology 319 (2004) 201–211

VIROLOGY

www.elsevier.com/locate/yviro

Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution

Laurence Blanchard,[a,*] Nicolas Tarbouriech,[b] Martin Blackledge,[a] Peter Timmins,[c] Wilhelm P. Burmeister,[b,d] Rob W.H. Ruigrok,[b,d] and Dominique Marion[a]

[a] *Institut de Biologie Structurale 'Jean-Pierre Ebel' (UMR 5075, CEA-CNRS-UJF), 38027 Grenoble cedex 1, France*

[b] *EMBL, Grenoble Outstation, 38042 Grenoble cedex 9, France*

[c] *Institut Laue Langevin, 38042 Grenoble cedex 09, France*

[d] *Laboratoire de Virologie Moléculaire et Structurale, Faculté de Médecine de Grenoble, Université Joseph Fourier, 38700 La Tronche, France*

Received 11 August 2003; returned to author for revision 10 October 2003; accepted 17 October 2003

Abstract

The RNA-dependent RNA polymerase of the Sendai virus (SeV) consists of the large protein (L) and the phosphoprotein (P). P plays a crucial role in the enzyme by positioning L (which carries the polymerase activity) onto the matrix for transcription and replication formed by the RNA and the nucleoprotein, the N-RNA. P has a modular structure with distinct functional domains: an N-terminal domain involved in binding to N° (N that is not yet bound to RNA) and a C-terminal domain that carries the oligomerisation domain, the N-RNA binding domain and the L binding domain and that, combined with L, is active in transcription. Structural data have previously been obtained on the N-terminal domain and on the oligomerisation domain of P, but not yet on its N-RNA binding domain (also-called the X protein). Here we present an NMR and a small angle neutron scattering study of the SeV X protein. We show that this molecule presents two subdomains linked by an 11-residue linker, with the N-subdomain lacking a well-defined conformation. The 3D structure of the C-subdomain consists of three α-helices revealing an asymmetric charge distribution that may be important for binding to RNA-bound nucleoprotein. The structure of the entire C-terminal domain of P is modelled from its constituent parts in combination with small angle scattering data on this domain.

Keywords: Sendai virus; *Paramyxoviridae*; Phosphoprotein; RNA-dependent RNA polymerase; NMR structure; Partially unfolded protein

Introduction

Sendai virus (SeV) is the prototype virus of the *Paramyxoviridae*. It infects the respiratory track of mice and causes pneumonia. This family contains not only human respiratory viruses, such as parainfluenza viruses, that cause croup, bronchiolitis and pneumonia, but also measles and mumps viruses. The negative-strand RNA genome of SeV is encapsidated by the nucleoprotein (N) with a stoichiometry of one N monomer for six nucleotides (Calain and Roux, 1993; Egelman et al., 1989; Kolakofsky et al., 1998). N binds to the sugar-phosphate backbone of the RNA, exposes the nucleotide bases and modulates the signals for transcription regulation that are encoded in the RNA (Iseni et al., 2002). This N-RNA complex serves as the template for the RNA-dependent RNA polymerase of the virus, which is constituted by two proteins, the large (L) protein and the phosphoprotein (P). The L protein contains the polymerase activity as well as the capping and polyadenylation activities (Abraham et al., 1975; Hunt et al., 1984; Testa and Banerjee, 1977). The P protein binds to the C-terminal domain of N and positions the polymerase (L) onto the template (Horikami and Moyer, 1995). Like other paramyxovirus phosphoproteins, the SeV P protein (568 amino acid residues) has a modular structure with two distinct functional domains (Fig. 1). The phosphoprotein N-terminal domain (PNT) (1–319), which is poorly conserved in sequence, acts as a chaperone for newly synthesised N, so-called N°, and prevents it from binding to nonviral RNA in the infected

* Corresponding author. Institut de Biologie Structurale 'Jean-Pierre Ebel' (UMR 5075, CEA-CNRS-UJF), 41 rue Jules Horowitz, 38027 Grenoble cedex 1, France. Fax: +33-4-38-78-54-94.

E-mail address: Laurence.Blanchard@ibs.fr (L. Blanchard).

0042-6822/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2003.10.029

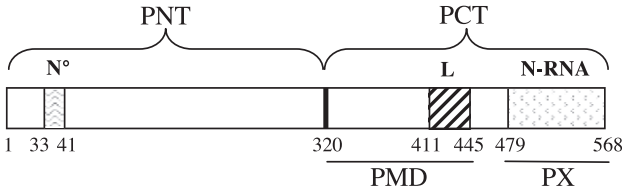


Fig. 1. Summary of the SeV phosphoprotein domains. The schematic representation of P shows its N- and C-terminal domains (PNT and PCT), the oligomerisation domain (residues 320–446) (also-called PMD) and the X protein (PX) (474–568). The location of discrete regions mapped for different protein–protein interactions is also indicated. These regions, from the N-terminus to the C-terminus are the N°-binding domain (residues 33–41), the L-binding domain (residues 411–445) and the N-RNA-binding domain (residues 479–568) (after Curran et al., 1995a).

cell (Curran et al., 1995b). Structural investigations on the measles virus PNT have shown that it is not structured in solution. This result was extended to several *Morbillivirus* PNTs and to SeV PNT (Karlin et al., 2002). The phosphoprotein C-terminal domain (PCT) is only functional as an oligomer and is sufficient for in vitro transcription (Curran et al., 1995a). Discrete regions in PCT have been mapped for various protein–protein interactions required for viral transcription. These regions are the L-binding domain (residues 411–445), which is part of the oligomerisation domain (also-called PMD) (residues 320–446), and the N-RNA-binding domain (residues 479–568) (Curran et al., 1995a). The high-resolution X-ray structure of a slightly shortened oligomerisation domain (residues 320–433) has been solved revealing a homotetrameric coiled coil structure (Tarbouriech et al., 2000b). So far, the N-RNA-binding part of the SeV phosphoprotein (also-called the X protein or PX) is the less characterised part from a structural point of view. Using site-directed mutagenesis, several residues of the SeV X protein were shown to be involved in RNA synthesis and responsible for the binding to the nucleoprotein (Tuckis et al., 2002). The X protein is produced in large amounts in infected cells from P mRNA by an internal ribosomal entry mechanism (Curran and Kolakofsky, 1988) and seems to constitute a domain on its own.

To better understand the mechanisms and contacts among the different proteins involved in viral transcription, a structural investigation of the SeV X protein was initiated. This protein was found to be extremely soluble and monomeric and to fold as an independent domain (Tarbouriech et al., 2000a). Because no crystals were obtained after extensive crystallisation trials, we chose NMR for solving its structure. Here we present an NMR and a small angle neutron scattering study of the SeV X protein. We show that this small protein presents two subdomains linked by an 11-residue linker, with the N-subdomain lacking a well-defined conformation. The solution structure of the C-subdomain of the X protein consists of three α-helices showing an asymmetric charge distribution that may be important for binding to RNA-bound nucleoprotein. The entire PCT could be modelled by combining the X-ray structure of the oligomerisation domain, the NMR structure of the PX C-subdomain and small angle scattering data on PX and intact PCT.

Results

Resonance assignment and characterisation of secondary structure elements of PX

The ^{1}H, ^{15}N and ^{13}C resonances of the X protein were assigned and reported elsewhere (Marion et al., 2001; http://www.bmrb.wisc.edu/ accession number 4999). After completion of these assignments, several weaker peaks were identified, corresponding to two short segments of amino acids surrounding two proline residues: P502 (residues E497 to A504) and P512 (residues P512 to K514) (Table 1). The identity of these peaks as a minor conformation caused by proline *cis–trans* isomerisation was confirmed by the characteristic Cβ and Cγ shifts of the proline side chains (Richarz and Wüthrich, 1978). Notably, the major conformation of all prolines was found to be trans.

Chemical shift and short-range nOe measurements indicate that PX consists of three α-helices (M519-E527, R533-K545, D549-S565) located at the C-terminal part of the molecule (Fig. 2). An additional α-helix (M519-E527) is identified compared to the predicted secondary structure (Tarbouriech et al., 2000a). Nearly half of the molecule (residue 474–518 as well as the N-terminal His tag) appears to have no persistent canonical secondary structure elements from chemical shift dispersion.

To supplement the structural characterisation of PX, we measured heteronuclear relaxation at 600 MHz. Of the 95

Table 1
Resonance assignments of PX minor conformation

Residue	*N*	Cα	Cβ	Other
E497		55.4 (4.25)	30.1	Cγ 33.8 (2.31)
R498	123.0 (8.44)	55.7	31.2	Cδ 43.51 (3.17)
D499	119.6 (8.10)	54.4	41.2 (2.77, 2.67)	
T500	113.4 (8.04)	61.6 (4.22)	70.3 (4.27)	$C\gamma_2$21.8
E501	124.3 (8.29)			
P502		62.4 (4.85)	**34.3** (2.40, 2.14)	Cγ **25.0** (1.98, 1.91); Cδ 50.2
R503	123.1 (8.57)	56.6 (4.32)	30.8 (1.89, 1.82)	Cγ 27.5 (1.69); Cδ 43.5 (3.23)
A504	125.2 (8.51)			
P512		62.7 (4.60)	**34.7** (2.44, 2.14)	Cγ **24.9** (1.96); Cδ 50.4
S513	116.4 (8.62)	58.3	64.1 (3.98, 3.89)	
K514	123.5 (8.50)			

Proton chemical shifts are given in parentheses.

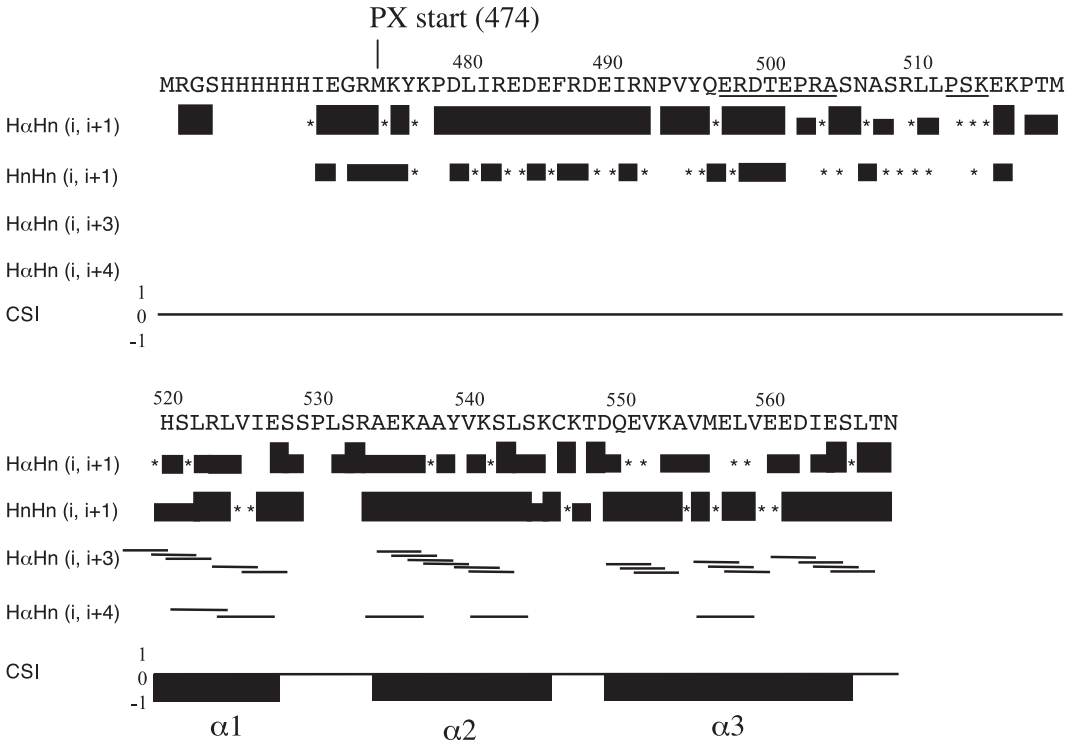


Fig. 2. Characterisation of secondary structure elements of the X protein. From top to bottom: PX sequence (including the N-terminal tag: $MRGSH_6IEGR$), underlined residues with resonance splitting, sequential ($H\alpha_i - Hn_{i+1}$) and ($Hn_i - Hn_{i+1}$) nOe correlations (reflecting the intensity of the corresponding cross-peak in the 3D ^{15}N-edited NOESY spectrum, missing assignments or peak overlaps are denoted with a star), ($H\alpha_i - Hn_{i+3}$) and ($H\alpha z - Hn_{i+4}$) nOe correlations. Consensus chemical shift index (CSI). Secondary structure elements.

amino acid residues (474–568), 68 $\{^1H\}$-^{15}N nOes (Fig. 3), 68 R_1 and 68 R_2 (data not shown) could be measured. The resulting data confirm that residues 474–515 indeed present differential dynamic behaviour compared to residues 516–568. In particular, the $\{^1H\}$-^{15}N nOe reveals a rigid domain (516–568) linked to a flexible domain (474–504) by an even more flexible linker (505–515). Assuming that the internal motion is rapid compared to the overall rotational diffusion, the ratio of transverse and longitudinal ^{15}N spin relaxation rates R_2/R_1 provides an estimate of characteristic rotational correlation time τ_C of molecular domains. On this basis, the region 474–504 exhibits a τ_C of 4.55 ± 0.03 ns,

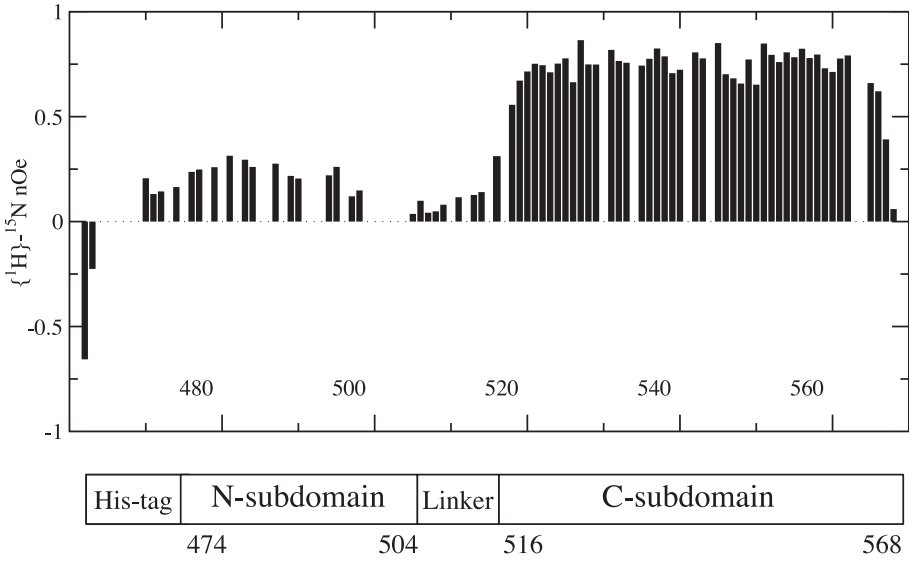


Fig. 3. Backbone $\{^1H\}$-^{15}N heteronuclear nOe values for PX and its domains. Residues without data correspond either to prolines or to amino-acids for which the HSQC peak overlaps with other signals.

while for the region 516–565, a value of 7.74 ± 0.04 ns is determined. The average $\{^1H\}$-^{15}N nOe values for these two subdomains are 0.24 ± 0.06 and 0.76 ± 0.05, respectively, indicating that residues 474–504 are also dynamic on the 100 ps to 1 ns time scale while residues 516–568 present motional restriction of the backbone as expected for a well-folded protein.

Taken together, these data suggest that PX presents two subdomains linked by a flexible 11-residue region where the N-subdomain (474–504) is more dynamic than the C-subdomain (516–568). In view of the absence of medium and long range nOes in the N-subdomain and its apparent dynamic behaviour, it is impossible, and inappropriate, to determine a single structure for this domain using the available NMR data. We have therefore determined the structure of the C-subdomain alone.

NMR structure of PX C-subdomain

A careful assignment of the inter-side chain nOes among the α-helices is a prerequisite to confidently identify intra-domain contacts. As the resonances from the PX N-subdomain (and from the 11-residue linker) overlap with those from the C-subdomain, assigning unambiguously long-range nOes was a difficult task. The structure of the PX C-subdomain is nevertheless unambiguously defined by these distances (Table 2) and is formed by three α-helices folded around a hydrophobic core (Figs. 4A, B and C). This core is formed by L522, and I526 on α-helix 1, L531 in a turn, Y539, V540 and L543 on α-helix 2 and V552, V555 and V559 on α-helix 3. Conservation of these hydrophobic residues among the paramyxovirus phosphoproteins (Tarbouriech et al., 2000a, Fig. 5) indicates that there is a similarity in the 3D structures of their N-RNA binding C-subdomains.

The C-subdomain of PX contains nine positive (R, K, H) and nine negative (D, E) residues and has a calculated isoelectric point of 5.74. These residues create an asymmetric charge distribution in the molecule with negative residues close to each other on one side and positive residues on the other side of the protein (Figs. 4D and F). Only three of the charged residues are highly conserved (Fig. 5), whereas most of the other conserved residues concern the hydrophobic core.

Structural characterisation of PX N-subdomain

The N-subdomain is not completely without any structure because the spin relaxation characteristics do not correspond to the expected behaviour of a 40-residue tail sampling random coil conformations (Lopez Garcia et al., 2000; Nagadoi et al., 1999; Sayers et al., 2000). In fact, this subdomain exhibits some of the NMR spectral properties characteristic of previously observed, partially folded states (Brutscher et al., 1997), suggesting rapidly fluctuating local structure sampling a range of local conformations, and the presence of a persistent overall fold. While the overall fold of this subdomain is difficult to determine from the current NMR data, small angle neutron scattering can provide complementary information about the average shape of the overall molecular envelope.

According to the UV circular dichroism (CD) spectrum of PX (Tarbouriech et al., 2000a), which has been reinterpreted using the CDnn program (Bohm et al., 1992), the secondary structure content is 34% α-helix (in agreement with the 34% helix observed in PX), 18% β-sheet and 17% β-turn. The β-sheet and β-turn contents do not correspond to structures in the C-subdomain and must thus be present in the N-subdomain. This suggests a structure containing β-bridges and tight turns which could be stabilised by four proline residues present in the sequence.

The radius of gyration (R_g) for the PX C-subdomain can be calculated from the atomic coordinates determined from its solution structure and was found to be 11.3 Å. The radius of gyration of intact PX (including the 14 amino-terminal residues corresponding to the His-tag) was 25 ± 1 Å, as determined by small angle neutron scattering,

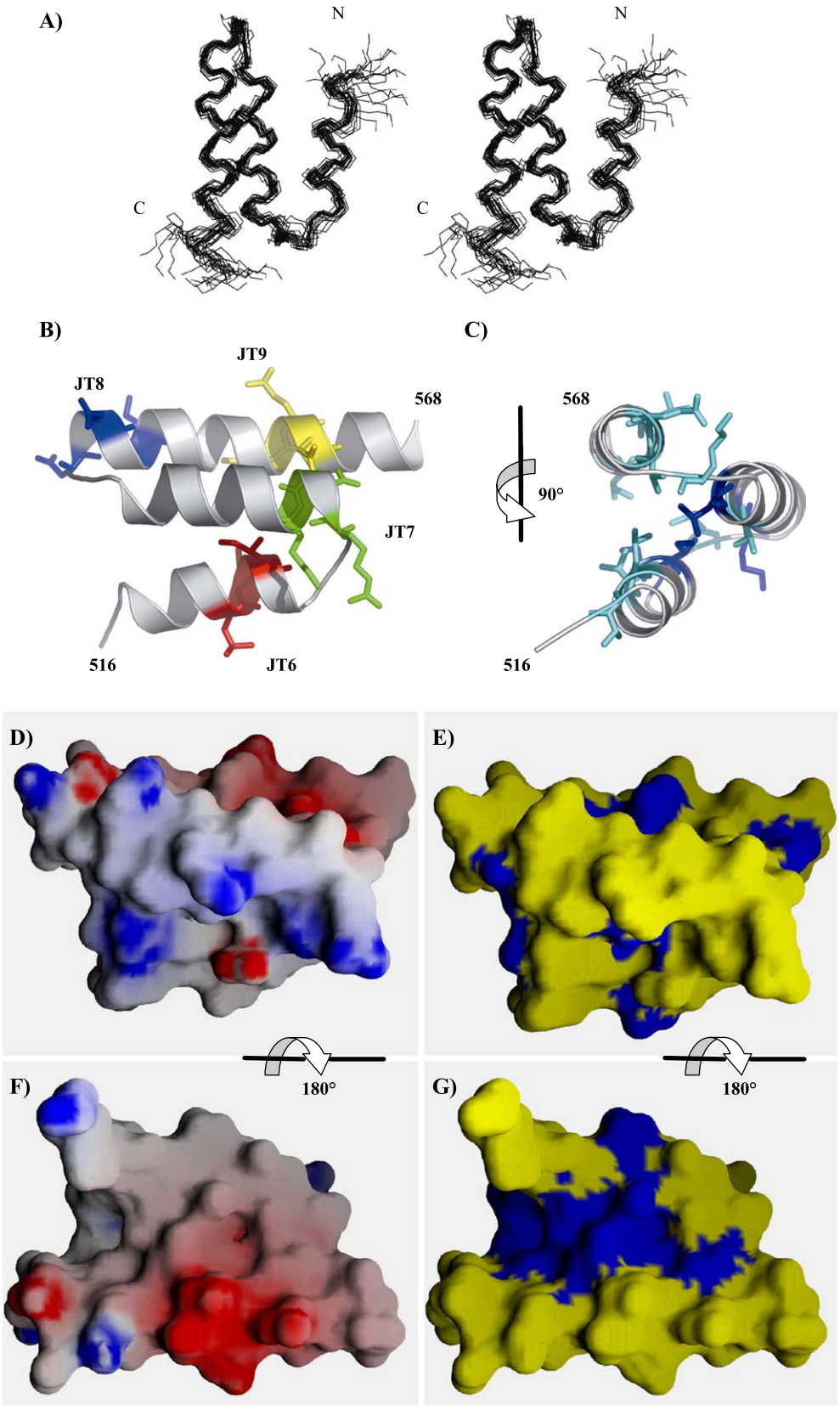
Table 2
Structural statistics for the 20 lowest energy structures

Experimental nOe and dihedral angle violation energy (kcal mol^{-1})[a,b]	12.4 ± 0.3
Experimental RDC violation energy (kcal mol^{-1})[c]	85.1 ± 2.4
Total physical energy (kcal mol^{-1})[d]	−323.7 ± 5.5
Average RMSD to the mean structure	
Backbone (residues 516–568) (Å)	1.44 ± 0.30
Heavy atoms (residues 516–568) (Å)	1.95 ± 0.25
Backbone (519–565) (Å)	0.90 ± 0.22
Heavy atoms (519–565) (Å)	1.22 ± 0.26
Ramachandran plot[e]	
Residues in most favoured regions (%)	81.2
Residues in additionally allowed regions (%)	15.2
Residues in generously allowed regions (%)	2.7
Residues in disallowed regions (%)	0.9

[a] Using a force constant of 50 kcal mol^{-1} Å^{-2} and 100 kcal mol^{-1} deg^{-2}, respectively.
[b] No violation of the nOe-derived constraints larger than 0.8 Å is observed in any of the 20 structures.
[c] Force constants given in the text.
[d] Calculated using the AMBER force field.
[e] Calculated for the 20 structures using PROCHECK.

Fig. 4. Solution structure of the C-subdomain of PX. (A) Stereodiagram showing the superposed backbone of the 20 NMR-derived structures with the lowest overall energy (figures generated with InsightII, Accelrys). (B) Secondary structure of the C-subdomain and the residues mutated by Tuckis et al. (2002). (C) C-subdomain turned by 90° compared to B with the side chains of the conserved and strictly conserved residues (see also Fig. 5) shown in light blue and blue. Figures generated with PyMOL (DeLano, 2003). (D and F) Molecular surface of the C-subdomain coloured by its electrostatic potential calculated with GRASP (Nicholls et al., 1991). Colours ranging from − 10 (red) to 10 kT (blue). D is of same orientation as B. (E and G) Molecular surface. Hydrophobic residues (V, L, I, M, F, Y, W, C) are coloured in blue. E is of same orientation as B.

which excludes that PX is a globular protein (see Materials and methods section). The amount of information on its shape which can be extracted from R_g is very limited so that different geometrical approximations can be used. When the N-subdomain of PX (comprising His-tag and linker) is considered as a rod and the C-subdomain as a



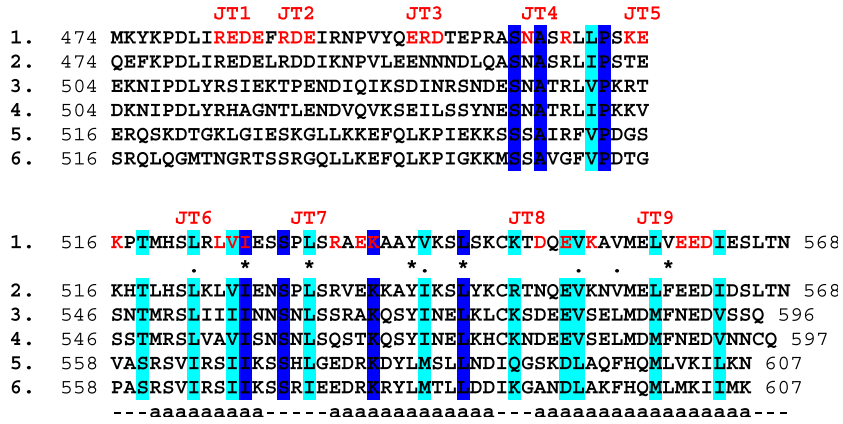


Fig. 5. Sequence alignment of the nucleocapsid-binding domain of six paramyxovirus phosphoproteins. (1) SeV X protein (P04859), (2) human parainfluenza 1 virus (P32530), (3) bovine parainfluenza 3 virus (P06163), (4) human parainfluenza virus 3 (O89234), (5) rinderpest virus (P35945), (6) measles virus (P03422). Strictly conserved residues in dark blue, conserved residues in light blue, residues contributing to the hydrophobic core of PX C-subdomain are marked with an asterisk (fully buried) or a dot (partly buried). The positions of the mutations studied by Tuckis et al. (2002) are shown in red. The bottom part of the figure corresponds to the alignment with the C-subdomain of which the secondary structure is given.

sphere with a diameter of 29 Å (R_g = 11.3 Å), or when PX is approximated by a thin rod, the length of PX (comprising the His-tag) can be calculated as 85 Å. The length of PX without His-tag can be estimated to be around 70 Å (Fig. 6C), which leaves about 31 Å for the length of the N-subdomain. As the NMR data exclude a random coil conformation for the N-subdomain, we assume that it forms a slightly elongated but compact structure. However, a main source of uncertainty is the structure of the 14 residues of the His tag, which is difficult to model, as well as the structure of the flexible linker between the N- and the C-subdomain.

Discussion

In this study, we have determined the solution state structure and dynamics of SeV PX using NMR spectroscopy. PX clearly presents two identifiable subdomains with differential dynamic characteristics linked by a flexible 11-residue region. The C-subdomain (516–568) is a relatively rigid, if comparatively open, three-helix bundle, stabilised by several hydrophobic interactions. The N-subdomain (474–504) on the other hand is much more flexible, undergoing large amplitude local motions on the 100 ps–1 ns time scale, and presenting some of the characteristics of a protein in a partially folded state. The flexibility of this domain may have been responsible for the inability to crystallise the full-length PX for X-ray studies. Using this structural and dynamic characterisation of PX, we can now go a step further in modelling the overall structure of the C-terminal domain of SeV phosphoprotein (PCT).

While the N-terminal domain of P (PNT) is rather poorly characterised, and possibly poorly structured (Karlin et al., 2002), the C-terminal domain (PCT) has already been studied in detail both on the atomic and the macromolecular scale. All structural information on the known parts of PCT is summarised in Fig. 6.

The most N-terminal part of PCT consists of the oligomerisation domain, PMD of which the crystal structure is known (Tarbouriech et al., 2000b; Fig. 6A). It has a length of 102 Å and from the structure we calculated its radius of gyration R_g (measure for its overall size) and its cross-sectional radius of gyration R_c (measure for the width of the domain) to be 33 and 13 Å, respectively. We also measured these values directly by small angle X-ray scattering and found the same value of 13 Å for R_c but 37 Å for R_g (Tarbouriech et al., 2000a). The difference between the calculated and the measured values for R_g can be explained by the fact that some residues of PMD were absent in the crystal structure (434–445) but were present in the protein we used for the scattering experiment (dotted lines on the right of the tetrameric molecule in Fig. 6A).

The most C-terminal part of PCT is the PX C-subdomain (Fig. 6B), of which the solution structure is presented here, and which has an average diameter of 29 Å and a calculated R_g of 11.3 Å (see Results). The size of PX was calculated from the NMR structure of its C-subdomain combined with the small angle neutron scattering data on the entire PX (Fig. 6C). The length of this domain is about 70 Å (excluding the His-tag; see Materials and methods).

Finally, as mentioned above, there are several residues of PMD that are not present in the crystal structure plus many residues connecting PMD with PX (446–473). We have no data concerning the local structure in this domain which precludes any modelling. However, because the total number of residues is about the same as that of PX minus its C-subdomain, we show it with the same shape and size as the N-subdomain (Fig. 6D).

All components are joined together in the complete PCT (Fig. 6E). The only data on the complete structure of PCT come from small angle scattering that gave a value for R_g of

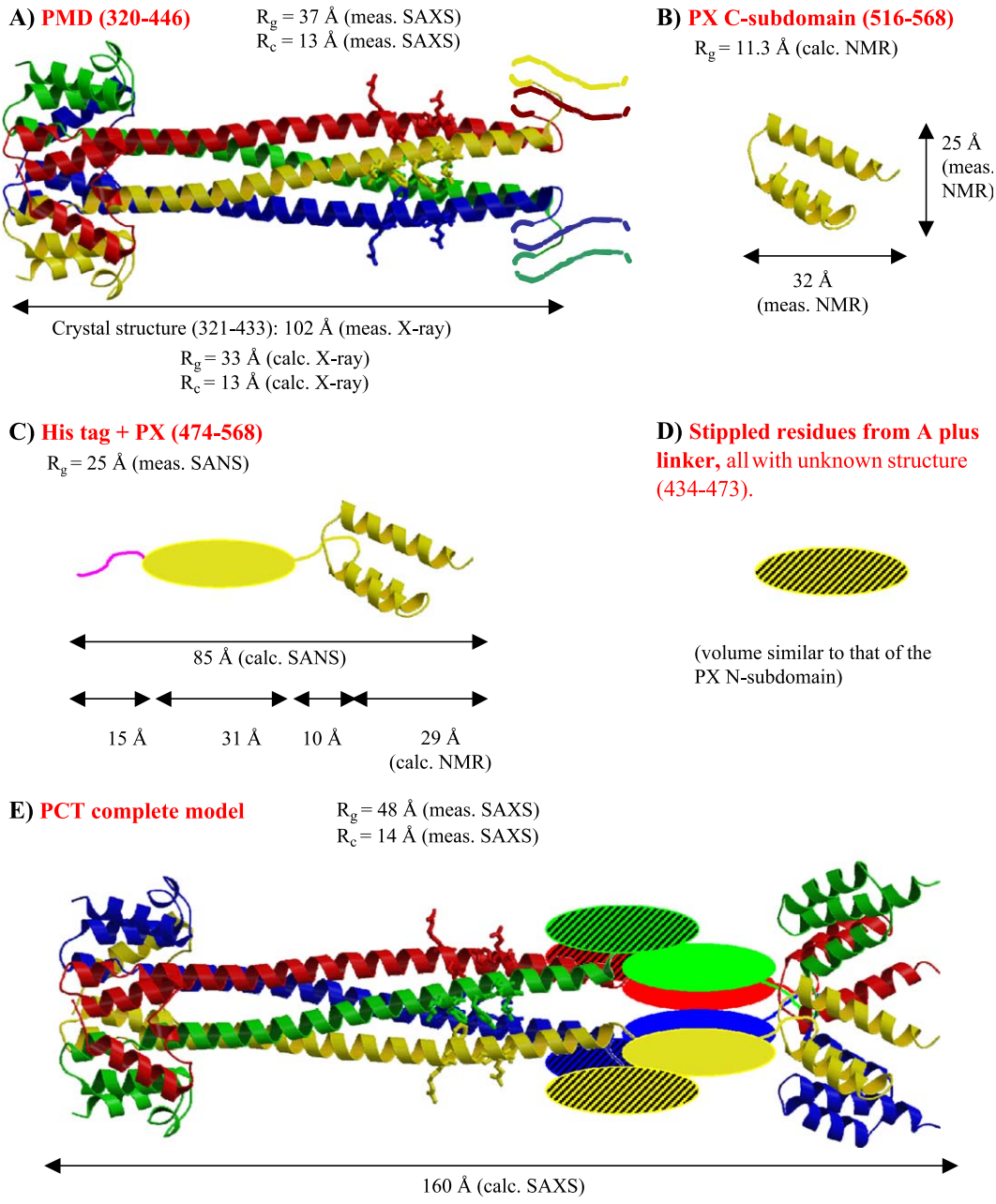


Fig. 6. Structure of the subdomains of PCT and model for the complete PCT. For each domain and for the complete PCT, the values for lengths and widths as measured from the X-ray or NMR model are indicated, as are the values for R_g and R_c as measured (meas.) by SAXS or SANS (small angle X-ray (or neutron) scattering) or as calculated (calc.) from the two actual structures. (A) X-ray structure of PMD, residues present in the structure (321–433) are shown and residues that are present in the molecule used for SAXS but not present in the crystallised protein are stippled. The side chains are from residues that have been shown to interact with L (Bowman et al., 1999). (B) NMR solution structure of the C-subdomain of PX. (C) Model for PX. For this model, the C-subdomain was treated as a sphere with a diameter of 29 Å, whereas N-subdomain and linkers were treated as a rod. Lengths of His-tag and linker were estimated. (D) Domain consisting of residues that are part of PMD but not present in the crystal structure plus residues connecting PMD and PX. There is no structural information on this domain. (E) Model for PCT taking into account the structural information from panels A, B and C plus the SAXS data on PCT.

48 Å and for R_c of 14 Å (Tarbouriech et al., 2000a). In particular, the value for R_c, that is only 1 Å larger than the R_c for PMD, indicates that PCT can be modelled as a cylinder with a width of 40 Å. The value for R_g implies that this cylinder has a length of 160 Å (Pilz, 1982). Because of these constraints, the four PX domains must be positioned close to the axis of the PMD rod and there may be a slight overlap between the lengths of PMD (102 Å) and PX (70 Å) to give a total length of 160 Å. The domain made up by the linker (Fig. 6D) must also be close to the PMD axis and not stick out too much because otherwise it would increase R_c of PCT by more than 1 Å compared to PMD. It cannot contribute itself to the length of PCT on top of the lengths of PMD and PX.

The side chains visible in Figs. 6A and E are residues that are important for the binding of L as determined by site-directed mutagenesis (Bowman et al., 1999). The L-binding domain was also determined by using deletion mutants (Curran et al., 1994; Smallwood et al., 1994). Thus defined, this domain reaches from residue 411 to residue 445 and includes the "stippled" part of PMD that was not in the crystal structure (Fig. 6A) and which is included in the domain shown in Fig. 6D. Therefore, the exact location of this domain will be of importance for the binding of L to P and the fact that it does not seem to be structurally ordered is probably related to its biological role.

Although PX is monomeric in solution (Tarbouriech et al., 2000a), the C-terminal part of PCT (consisting of the 4 PX domains) must have a compact structure with the four monomers close to each other. In the current model that is shown in Fig. 6E, the C-subdomains of PX are located close to the 4-fold axis. The centres of the C-subdomains of PX (that bind to N-RNA) are about 24 Å apart whereas the distance among symmetry-related surface residues can be up to 45 Å. This should be compared with the distance of 30–40 Å between two neighbouring N molecules of the nucleocapsid of Sendai virus whereas the distance of two N molecules among adjacent turns of the helical coil of the N-RNA coil is 50–60 Å (Egelman et al., 1989, values for the tight coil).

During transcription, PX interacts with the RNA-bound nucleoprotein. Like other paramyxovirus nucleoproteins, Sendai virus N (517aa) is divided into two regions; a well-conserved N-terminal region called N-core (1–400) and a hyper variable C-terminal region called N-tail (401–517). The C-terminal part of the N-tail (460–517) is responsible for binding to PX (Bankamp et al., 1996; Buchholz et al., 1993; Curran et al., 1993; Longhi et al., 2003; Myers et al., 1999). This part of N-tail of SeV N is strongly negatively charged with an extremely negatively charged peptide from residues 468-EEETNDEDVSDIE-480. It is possible that PX binds with its positively charged side to the negatively charged residues in N-tail.

Site-directed mutagenesis was performed on the N-RNA-binding domain of SeV P in which clustered charged or hydrophobic residues were changed to alanine residues (Tuckis et al., 2002, Fig. 5). The P mutants were assayed for viral RNA synthesis and for various protein–protein interactions. Four of the mutants fall into the C-subdomain of PX. Of these mutants, one, JT6 (L524A, V525A, I526A), was completely inactive. The structure shows that two of the mutated residues are involved in the formation of the hydrophobic core of the protein, which suggests that this mutant may be misfolded leading to a complete loss of function. Another mutant, JT9 (E560A, E561A, D562A), was completely inactive in vitro, even though it had some activity in vivo in contrast to JT6. The mutation affects three negatively charged residues at the surface (Figs. 4B and D) of which two are conserved within the parainfluenza virus family. A third mutant, JT7 (R533A, E535A, K536A), is inactive in the N°–P complex formation and in the reconstitution of replication. This lack of function may be due to the mutation of the strictly conserved lysine 536 residue. The fourth mutant JT8 (D549A, E551A, K553A), which mutates the conserved negatively charged residue glutamic acid 551, retains essentially wild-type activity. This mutant is located on the N-terminus of the third α-helix, suggesting that this part of the C-subdomain is probably not involved in the contact with N, whereas mutations of JT9 located on the C-terminal side of α helix 3 have dramatic effects (Fig. 4B). The 3D structure of the PX C-subdomain that we present here cannot explain all the phenotypes. In particular, JT9 that is mutated at its negatively charged side should not have had an effect on activity if PX only interacts with N-tail through its positively charged side. Obviously, the details of the interaction between PX and N-tail are not simple. The structure presented here could be used as a basis for a mutational analysis in which the role of the distribution of charges on the surface of the PX C-subdomain can be tested.

The structural data obtained on PX suggest that flexibility may play an important role in the interactions among P and N and L. The size and the hydrophobic core of the C-subdomain of PX are so small that this domain is at the lowest limit of an ordered domain. The high flexibility in the N-subdomain of PX may also be of biological importance, possibly to allow extra length when necessary, perhaps leading to the coming apart of the PX C-subdomains to span more width. Added to this comes the observations that even the stable four-helical bundle of the oligomerisation domain (Tarbouriech et al., 2000b) is a relatively flexible structure for a coiled coil and that the N-terminal domain of P is also natively unfolded (Karlin et al., 2002). This leads to a structure–function relation of the replication process of nonsegmented negative-strand RNA viruses that is difficult to analyse by traditional structural biology techniques. Of all these techniques, NMR is probably best adapted because it takes movements of proteins and protein domains into account and, because the experiments are performed in solution, there are no (de)stabilising interactions by protein–protein contacts that may take place inside a protein crystal.

Materials and methods

Sample preparation

The SeV X protein spanning amino acids 474–568 of the phosphoprotein (Swiss-prot accession number P04859) was overproduced with six additional histidine residues and a factor Xa cleavage site at its N-terminus and purified as described previously (Tarbouriech et al., 2000a). Uniformly $^{15}N/^{13}C$-labelled X protein was obtained by growing bacteria in M9 minimal medium. The NMR sample used for resonance assignments, nOe collection, $\{^1H\}$-^{15}N nOe, R_1 and $R_{1\rho}$, measurements was prepared at a concentration of ≈1.2–1.5 mM in 50 mM potassium phosphate buffer pH

6.0 with 0.5 M NaCl, 10 mM DTT, protease inhibitor cocktail (Complete, Boehringer Mannheim), 0.02% NaN_3 and 10% D_2O. The NMR tube was then sealed under argon gas.

NMR spectroscopy

All NMR experiments used for ^1H, ^{13}C, ^{15}N resonance assignment were described previously (Marion et al., 2001). Interproton distance restraints were derived from the 3D ^{15}N-edited NOESY and ^{13}C-edited NOESY spectra. These NMR spectra were acquired at 298 K on an 800 MHz Varian Inova spectrometer equipped with a triple resonance (^1H, ^{13}C, ^{15}N) probe including shielded z-gradients. Data processing and peak picking were performed using FELIX program 2000 (Accelrys).

Identification of secondary structure elements

Secondary structure elements of the X protein were characterised using nOe patterns ($H\alpha_i$-Hn_{i+1}, $Hn_i - Hn_{i+1}$, $H\alpha_i - Hn_{i+3}$ and $H\alpha_i - Hn_{i+4}$) and consensus chemical shift index (CSI) for $^1H\alpha$, $^{13}C\alpha$, ^{13}CO and $^{13}C\beta$ resonances of the X protein, calculated using the program CSI (Wishart and Sykes, 1994). α-helices are characterised by short $Hn_i - Hn_{i+1}$ distances and, thus, strong $Hn_i - Hn_{i+1}$ nOe contacts. β-sheets on the contrary are characterised by short $H\alpha_i - Hn_{i+1}$ distances thus strong $H\alpha_i - Hn_{i+1}$ nOe contacts. The chemical shift index (CSI) is an empirical indicator that correlates chemical shift values with secondary structure elements.

^{15}N relaxation measurements

^{15}N relaxation experiments R_1, $R_{1\rho}$ and $\{^1H\}$-^{15}N nOe were performed using standard 2D pulse sequences (Farrow et al., 1994). For $R_{1\rho}$ relaxation measurements, a spin lock field of 2.5 kHz was applied. The relaxation decay was sampled at the following time points: 20, 100, 250, 400, 550, 700 and 1000 ms for R_1, and 8, 16, 32, 48, 64, 80 104, 132 and 160 ms for $R_{1\rho}$. For the $\{^1H\}$-^{15}N nOe measurements, two spectra with on- and off-resonance ^1H saturation were recorded in an interleaved manner. The saturation time was set to 3.5 s and the recycle delay to 7 s.

Residual dipolar couplings

Residual dipolar couplings (RDC) were collected on a 1 mM uniformly $^{15}N/^{13}C$-labelled sample suspended in a liquid crystalline medium consisting of 5% C12E6/hexanol in 50 mM potassium phosphate buffer pH 6.0 with 0.5 M NaCl, 10 mM DTT and 10% D_2O. The magnitude of the solvent ^2H quadrupolar splitting was 21 Hz. Three different types of dipolar couplings were measured at 600 MHz and 298 K: $^1D_{NH}$, $^1D_{C'C\alpha}$ and $^2D_{C'HN}$ couplings were obtained using 3D HNCO-type experiments (Brutscher, 2001; Yang et al., 1999).

Structure calculations

Structure calculations were performed using a two-step restrained molecular dynamics procedure. The first step, using the program Discover 2.98, utilizes a classical simulated annealing calculation from randomised initial Cartesian coordinates and has been described in detail elsewhere (Blackledge et al., 1995). Structural restraints for this step are composed 17ϕ and 17ψ dihedral angle restraints derived from TALOS (Cornilescu et al., 1999) and 27 hydrogen bonding restraints derived from confidently identifiable helical regions. Interproton distances (280) were introduced as unambiguous distance restraints and calibrated from known intraresidual short-range backbone or methyl–methyl distances from Leu and Val side chains. Ambiguous distance restraints (59) were introduced as previously described (Favier et al., 2002). The 20 best structures, as defined from the total experimental target function, were selected for refinement against RDC restraints.

The second step refines the lowest energy structures calculated from the first step, using all of the restraints mentioned above, combined with 48 $^1D_{NH}$, 48 $^1D_{C'C\alpha}$ and 48 $^2D_{C'HN}$ RDC restraints. This calculation was performed using the program SCULPTOR (Hus et al., 2000) and composes three steps. The molecular coordinates are fixed during the initial step, while the alignment tensor parameters evolve under the influence of the RDCs measured from sites present in secondary structure elements. This step is composed of a sampling period of 4 ps (8000 0.25 fs time steps) at 300 K, followed by 3 ps cooling to 200 K and conjugate gradient minimisation, and defines the most appropriate starting point for the tensor corresponding to each structure. The molecule is then released, and the tensor and molecule evolve under the influence of the same RDC restraints and the complete restraint set from step 1 during 9.3 ps at a temperature of 1000 K. During this period, k_{RDC} is raised from the initial ($k_{RDC,N\text{-}H}$ = 0.005 kcal mol^{-1} Hz^{-2} and $k_{RDC,NH\text{-}C'}$ = 0.01 kcal mol^{-1} Hz^{-2}) to the final values ($k_{RDC,N\text{-}H}$ = 0.50 kcal mol^{-1} Hz^{-2} and $k_{RDC,NH\text{-}C'}$ = 1.0 kcal mol^{-1} Hz^{-2}). This is followed by a 2.7-ps cooling period to 100 K and energy minimisation. This step refines both the local structure and the tensor simultaneously (Sibille et al., 2001). At this point, the remaining RDC restraints are introduced and both tensor and conformation evolve freely during 24 ps at 1000 K, during which time $k_{NOE,h\text{-}bond}$, k_{dihe} and k_{RDC} were again raised from their initial (0.5, 2.0, 0.005, 0.01) to their final values (50.0 kcal mol^{-1} Å^{-2}, 200.00 kcal mol^{-1} deg^{-2}, $k_{RDC,N\text{-}H}$ = 0.50 kcal mol^{-1} Hz^{-2} and $k_{RDC,NH\text{-}C'}$ = 1.0 kcal mol^{-1} Hz^{-2}). Following the sampling period, the system was cooled to 100 K over 2.7 ps and the system again minimised using a conjugate gradient algorithm. This calculation was repeated five times for each initial structure using different initial atomic velocity distributions, and the structure with the lowest total target function comprising nOe, backbone torsion, hydrogen bonding and RDC restraints from each initial structure was selected as a member of the final ensemble.

Small angle neutron scattering experiment (SANS)

SANS was performed on the D22 instrument at the Institut Laue-Langevin (Grenoble, France) using a wavelength of 6 Å and a sample-detector distance of 3 m. The protein concentration was 10 mg ml^{-1} in 200 mM NaCl, 20 mM Tris–HCl pH 7.6 in H_2O. The global radius of gyration is derived from the Guinier approximation: $I(Q) = I_o\exp(-R_g^2Q^2/3)$, where I_o is the zero angle scattered intensity, R_g the overall radius of gyration and $Q = 4\pi\sin\theta/\lambda$, with 2θ being the scattering angle and λ the wavelength. The radius of gyration can be calculated from the initial slope of the plot $\log(I)$ vs. Q^2.

Model calculations

Calculations of the radius of gyration of simple geometric bodies were done using formulae from Glatter (1982) and Steiner's law (Gerthsen et al., 1982) using EXCEL (Microsoft). Molecular modelling was done using O (Jones et al., 1991) and PDBSET from the CCP4 program suite (CCP4, 1994).

(a) PX including the His-tag has 109 residues. Using an average volume of 134 $Å^3$ per residue for proteins, the radius of a sphere with that volume $V = \frac{4}{3}\pi r^3$ would be r = 15.2 Å. Its corresponding radius of gyration is 11.7 Å with $R_g^2 = 3/5r^2$ for a compact sphere. This value is very different from the measured radius of gyration of 25 Å, which means that PX is not globular.

(b) The radius of gyration for a thin rod with a length l is given by $R_g^2 = \frac{l^2}{12}$. The radius of gyration of an object composed of two objects with individual radii of gyration R_{g1} and R_{g2} and distances of their centres of gravity of the common centre of gravity x_1 and x_2 and respective scattering m_1 and m_2 is given by:

$$R_g^2 = \frac{(m_1x_1^2 + m_2x_2^2 + m_1R_{g1}^2 + m_2R_{g2}^2)}{m_2 + m_2} \quad (1)$$

according to Steiner's law. Here, the C-terminal domain (53 aa) has been approximated by a sphere with a radius r (14.6 Å) corresponding to the radius of gyration R_{g2} calculated from the NMR structure (11.3 Å). The N-terminal domain (56 aa) comprising the His-tag has been modelled as a thin rod with a length l. For simplicity, the mass of the two domains can be approximated as equal. The distance between the centre of the global centre of gravity and the individual centre of gravity of the rod and the sphere becomes $\frac{l}{4} + \frac{r}{2}$. Eq. (1) gives $R_g^2 = \left(\left(\frac{l}{4} + \frac{r}{2}\right)^2 + \frac{1}{2}\frac{1}{12}l^2 + \frac{1}{2}R_{g2}^2\right)$. From this expression, l can be calculated as 54 Å. Adding $2r$ gives the total length L of 84 Å. However, if we model PX more simply as a thin rod, then from $R_g^2 = \frac{L^2}{12}$ follows L = 87 Å. The average value of these two calculations, 85 Å, was used for the model in Fig. 6.

(c) Estimation of the contribution of the 14 residues of the His-tag is difficult. For simplicity, we used the value of the diameter of a sphere (15 Å), which contains the same volume as the 14 residues. This dimension has been subtracted to obtain the length of PX without His-tag.

Protein data bank accession number

The coordinates are deposited at the RCSB Protein data Bank under the accession number 1R4G.

Note added in proof

Since the submission of this paper, Johansson et al. (2003) have published the crystal structure of the C-subdomain of the phosphoprotein of measles virus. The structure of the measles virus protein is similar to that of the Sendai virus C-subdomain determined here by NMR (rms deviation 1.6 Å for equivalent C atoms).

Acknowledgments

We thank Bernhard Brutscher for pulse sequences and Christophe Deprez for discussions.

References

Abraham, G., Rhodes, D.P., Banerjee, A.K., 1975. The 5′ terminal structure of the methylated mRNA synthesized in vitro by vesicular stomatitis virus. Cell 5, 51–58.

Bankamp, B., Horikami, S.M., Thompson, P.D., Huber, M., Billeter, M., Moyer, S.A., 1996. Domains of the measles virus N protein required for binding to P protein and self-assembly. Virology 216, 272–277.

Blackledge, M.J., Medvedeva, S., Poncin, M., Guerlesquin, F., Bruschi, M., Marion, D., 1995. Structure and dynamics of ferrocytochrome c_{553} from *Desulfovibrio vulgaris* studied by NMR spectroscopy and restrained molecular dynamics. J. Mol. Biol. 245, 661–681.

Bohm, G., Muhr, R., Jaenicke, R., 1992. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. Protein Eng. 6, 383–390.

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.

Brutscher, B., 2001. Accurate measurement of small spin–spin couplings in partially aligned molecules using a novel J-mismatch compensated spin-state-selection filter. J. Magn. Reson. 151, 332–338.

Brutscher, B., Bruschweiler, R., Ernst, R.R., 1997. Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by 1H, 13C, and 15N nuclear magnetic resonance spectroscopy. Biochemistry 36, 13043–13053.

Buchholz, C.J., Spehner, D., Drillien, R., Neubert, W.J., Homann, H.E., 1993. The conserved N-terminal region of Sendai virus nucleocapsid protein NP is required for nucleocapsid assembly. J. Virol. 67, 5803–5812.

Calain, P., Roux, L., 1993. The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. J. Virol. 67, 4822–4830.

Collaborative Computational Project, Number 4, 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr., Sect. D: Biol. Crystallogr. 50, 760–763.

Cornilescu, G., Delaglio, F., Bax, A., 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR 13, 289–302.

Curran, J., Kolakofsky, D., 1988. Scanning independent ribosomal initiation of the Sendai virus X protein. EMBO J. 7, 2869–2874.

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.

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.

Curran, J., Boeck, R., Lin-Marq, N., Lupas, A., Kolakofsky, D., 1995a. Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils. Virology 214, 139–149.

Curran, J., Marq, J.B., Kolakofsky, D., 1995b. An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. J. Virol. 69, 849–855.

DeLano, W.L., 2003. The PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org.

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.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., Kay, L.E., 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry 33, 5984–6003.

Favier, A., Brutscher, B., Blackledge, M., Galinier, A., Deutscher, J., Penin, F., Marion, D., 2002. Solution structure and dynamics of Crh, the Bacillus subtilis catabolite repression HPr. J. Mol. Biol. 317, 131–144.

Gerthsen, C., Kneser, H.O., Vogel, H., 1982. Physik, 14 ed. Springer, Berlin, pp. 65–66.

Glatter, O., 1982. Data treatment. In: Kratky, O., Glatter, O. (Eds.), Small Angle X-ray Scattering. Academic Press, London, pp. 119–165.

Horikami, S.M., Moyer, S.A., 1995. Structure, transcription, and replication of measles virus. Curr. Top. Microbiol. Immunol. 191, 35–50.

Hunt, D.M., Smith, E.F., Buckley, D.W., 1984. Aberrant polyadenylation by a vesicular stomatitis virus mutant is due to an altered L protein. J. Virol. 52, 515–521.

Hus, J.-C., Marion, D., Blackledge, M., 2000. De novo determination of protein structure by NMR using orientational and long-range order restraints. J. Mol. Biol. 298, 927–936.

Iseni, F., Baudin, F., Garcin, D., Marq, J.B., Ruigrok, R.W., Kolakofsky, D., 2002. Chemical modification of nucleotide bases and mRNA editing depend on hexamer or nucleoprotein phase in Sendai virus nucleocapsids. RNA 8, 1056–1067.

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.

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 Crystallogr., A 47, 110–119.

Karlin, D., Longhi, S., Receveur, V., Canard, B., 2002. The N-terminal domain of the phosphoprotein of Morbilliviruses belongs to the natively unfolded class of proteins. Virology 296, 251–262.

Kolakofsky, D., Pelet, T., Garcin, D., Hausmann, S., Curran, J., Roux, L., 1998. Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited. J. Virol. 72, 891–899.

Longhi, S., Receveur-Brechot, V., Karlin, D., Johansson, K., Darbon, H., Bhella, D., Yeo, R., Finet, S., Canard, B., 2003. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. J. Biol. Chem. 278, 18638–18648.

Lopez Garcia, F., Zahn, R., Riek, R., Wüthrich, K., 2000. NMR structure of the bovine prion protein. Proc. Natl. Acad. Sci. U.S.A. 97, 8334–8339.

Marion, D., Tarbouriech, N., Ruigrok, R.W., Burmeister, W.P., Blanchard, L., 2001. Assignment of the ^{1}H, ^{15}N and ^{13}C resonances of the nucleocapsid-binding domain of the Sendai virus phosphoprotein. J. Biomol. NMR 21, 75–76.

Myers, T.M., Smallwood, S., Moyer, S.A., 1999. Identification of nucleocapsid protein residues required for Sendai virus nucleocapsid formation and genome replication. J. Gen. Virol. 80, 1383–1391.

Nagadoi, A., Nakazawa, K., Uda, H., Okuno, K., Maekawa, T., Ishii, S., Nishimura, Y., 1999. Solution structure of the transactivation domain of ATF-2 comprising a zinc finger-like subdomain and a flexible subdomain. J. Mol. Biol. 287, 593–607.

Nicholls, A., Sharp, K.A., Honig, B., 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 11, 281–296.

Pilz, I., 1982. Proteins. In: Kratky, O., Glatter, O. (Eds.), Small Angle X-ray Scattering. Academic Press, London, pp. 239–293.

Richarz, R., Wüthrich, K., 1978. Carbon-13 NMR chemical shifts of the common amino acid residues measured in aquaeous solutions of the linear tetrapeptides H-Gly-Gly-X-$_{L}$ala-OH. Biopolymers 17, 2133–2141.

Sayers, E.W., Gerstner, R.B., Draper, D.E., Torchia, D.A., 2000. Structural preordering in the N-terminal region of ribosomal protein S4 revealed by heteronuclear NMR spectroscopy. Biochemistry 39, 13602–13613.

Sibille, N., Pardi, A., Simorre, J.-P., Blackledge, M., 2001. Refinement of local and long range structural order in theophylline-binding RNA using ^{13}C-^{1}H residual dipolar couplings and restrained molecular dynamics. J. Am. Chem. Soc. 123, 12135–12146.

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.

Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R.W., Burmeister, W.P., 2000a. On the domain structure and the polymerization state of the Sendai virus P protein. Virology 266, 99–109.

Tarbouriech, N., Curran, J., Ruigrok, R.W., Burmeister, W.P., 2000b. Tetrameric coiled coil domain of Sendai virus phosphoprotein. Nat. Struct. Biol. 7, 777–781.

Testa, D., Banerjee, A.K., 1977. Two methyltransferase activities in the purified virions of vesicular stomatitis virus. J. Virol. 24, 786–793.

Tuckis, J., Smallwood, S., Feller, J.A., Moyer, S.A., 2002. The C-terminal 88 amino acids of the Sendai virus P protein have multiple functions separable by mutation. J. Virol. 76, 68–77.

Wishart, D.S., Sykes, B.D., 1994. The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. J. Biomol. NMR 4, 171–180.

Yang, D., Venters, R.A., Mueller, G.A., Choy, W.Y., Kay, L.E., 1999. TROSY-based HNCO pulse sequences for the measurement of ^{1}HN-^{15}N, ^{15}N-^{13}CO, ^{1}HN-^{13}CO, ^{13}CO-^{13}Cα, and ^{1}HN-^{13}Cα dipolar couplings in ^{15}N, ^{13}C, ^{2}H-labeled proteins. J. Biomol. NMR 14, 333–343.