Structure of the equine arteritis virus nucleocapsid protein reveals a dimer–dimer arrangement

Equine arteritis virus (EAV) is an enveloped positive-sense RNA virus belonging to the Arteriviridae family, which also includes the porcine pathogen PRRSV and is genetically and structurally related to the coronaviruses. EAV is an important equine pathogen that has caused significant economic losses to the horse-breeding industry and has been difficult to control. The EAV virion consists of a genome-containing nucleocapsid core made of nucleocapsid (N) protein surrounded by a lipid envelope containing several membrane proteins. Here, the crystal structure of the capsid-forming domain of the EAV N protein is presented at 2.0 Å resolution. The dimeric N-protein structure is similar to the previously determined structure of the N protein from PRRSV, with most differences localized to the terminal helices and flexible loops. The N protein is organized as dimers of dimers in the crystal, which may reflect the arrangement of the protein in the viral nucleocapsid.

1. Introduction

The Arteriviridae is a family of enveloped positive-sense RNA viruses that includes two important agricultural pathogens, equine arteritis virus (EAV) and porcine reproductive and respiratory syndrome virus (PRRSV), in addition to lactate dehydrogenase-elevating virus (LDV) of mice and simian haemorrhagic fever virus (SHFV) (Plagemann, 1996; Snijder & Meulenberg, 1998). Arteriviruses are grouped together with coronaviruses in the order Nidovirales based on similarities in genome organization and expression (Cavanagh, 1997; Fauquet et al., 2005; Snijder et al., 1993).

EAV causes respiratory disease and reproductive failure in horses, mules and donkeys (Del Piero, 2000; MacLachlan & Balasuriya, 2006). The virus has increasingly become a problem to the $2 billion horse-breeding industry in the US and has proven difficult to control owing to its pattern of persistent subclinical infection with occasional outbreaks of the disease, in spite of the development of several effective vaccines (Balasuriya & MacLachlan, 2004). The related virus PRRSV is a major porcine pathogen in the US and infects pigs with a similar aetiology to EAV (Neumann et al., 2005; Zimmerman et al., 1997). Although the order Nidovirales includes important human pathogens such as the SARS coronavirus (SARS-CoV), there are no known human pathogens in the Arteriviridae family.

The EAV genome consists of a 12.7 kbp single-stranded positive-sense RNA that is transcribed as a set of subgenomic mRNA transcripts, each coding for a separate open reading frame (ORF; Snijder & Meulenberg, 1998). The first ORF encodes the nonstructural proteins, which include three or
four proteases, an RNA-dependent RNA polymerase, a helicase and an endonuclease, while ORFs 2–5 encode glycosylated membrane proteins (GP2, E and GP3–GP5), ORF6 encodes the nonglycosylated membrane (M) protein and ORF7 encodes the nucleocapsid (N) protein (Snijder & Meulenberg, 1998). Wieringa et al. (2004) described for PRRSV-N the assembly of arterivirus virions appear as smooth roughly spherical particles or oblong particles with a diameter of about 60 nm (Horzinek, 1981; Snijder & Meulenberg, 1998; Wieringa et al., 2004). GP5 may also be involved in receptor binding, but the minor envelope proteins are essential for infectivity and may be the main determinants for tropism (Dobbe et al., 2001). Replication and assembly occur at internal membranes and are probably driven by interactions between N and the envelope proteins (Snijder & Meulenberg, 1998).

The N protein of EAV (EAV-N) consists of 110 residues, while that of PRRSV contains 123 residues, with several additional amino acids in the N-terminal half of the protein (Fig. 1). The amino-terminal half of EAV-N (1–47) is the least conserved part of the sequence; it contains a number of positively charged residues and is involved in RNA binding. The arterivirus N protein is targeted to the nucleus of infected cells, suggesting that N is involved in the modulation of host functions in addition to its role in RNA packaging and assembly (Tijms et al., 2002; Yoo et al., 2003).

We previously determined the structure of the C-terminal half (residues 58–123) of the PRRSV N protein (PRRSV-Na57), which forms a dimer that is considered to comprise the capsid-forming domain of N (Doan & Dokland, 2003b). Here, we present the structure of the equivalent domain of EAV-N (residues 49–110). In both viruses, the protein forms a tight dimer consisting of a β-sheet that is capped and flanked by α-helices. However, in EAV the N protein forms a dimer of dimers in the crystal, which may reflect the arrangement of the protein in the viral nucleocapsid.

2. Methods

2.1. Cloning, expression, purification and crystallization

Residues 49–110 of EAV-N were cloned into the vector pET16b, incorporating an N-terminal His6 tag (MAHH-HHHH...) immediately preceding the EAV-N sequence, and expressed in Escherichia coli BL21(DE3) as previously described for PRRSV-Na57 (Doan & Dokland, 2003a). To obtain SeMet-derivative crystals, the clones were grown for 15 h at 310 K in M9 media supplemented with glucose, all amino acids except Met, vitamins and SeMet (Gerchman et al., 1994). After lysis by passage through an Avestin Emulsiflex EF-C3 high-pressure disruptor, the protein was purified by affinity on a Hi-Trap Ni-chelating column (Amersham Bio-

![Figure 1](image)

Sequence alignment of EAV-N and PRRSV-N. The alignment was performed with ClustalW (Chenna et al., 2003) and numbered according to EAV-N. Identical residues are shown in black boxes; similar residues are shown in white boxes. The arrow at residue 49 indicates the start of the EAV-Na48 construct used in our study. Secondary-structure elements of the EAV-Na48 structure are shown above the sequence. The grey box above residues 19–31 indicates a predicted α-helix. This figure was created with ESPript (Gouet et al., 1999).
sciences) followed by size-exclusion chromatography on Sephacryl S-200 (Amersham Biosciences). The protein was concentrated to 30 mg ml⁻¹ and crystallized by vapour diffusion against 0.1 M sodium phosphate pH 8.0 and 25% PEG 3350 for 8 d at 291 K. The crystals were harvested, briefly soaked in 10% glycerol, 0.15 M sodium phosphate pH 8.0 and 27% PEG 3350 and frozen by immersion in liquid nitrogen.

2.2. Data collection and structure determination

Data were collected at the UK beamline BM14 at the ESRF using a single frozen crystal that was auto-mounted onto the BM14 beamline goniometer using the SC3 robotic sample changer (Cipriani et al., 2006). An X-ray fluorescence scan was performed to ascertain the Se edge and anomalous scattering factors were calculated using the program CHOOCH (Evans & Pettifer, 2001). A total of 489 frames of data were collected on a MAR Mosaic 225 CCD area detector at the experimentally determined selenium absorption-edge peak energy (12.67587 keV, 0.9781 Å) with an oscillation angle of 1.0°. The data were integrated and scaled using the programs XDS (Kabsch, 1993) and SCALA (Collaborative Computational Project, Number 4, 1994), giving a data set with a redundancy of 19.5 and an overall Rmerge of 7.9% (Table 1). The crystals belonged to space group C2221, with unit-cell parameters a = 68.23, b = 73.23, c = 138.60 Å, corresponding to two dimers in the asymmetric unit with V_M = 2.78 Å³ Da⁻¹. The selenomethionine-derivative crystals showed a high degree of mosaicity (>2°), with usable data extending to about 2.0 Å resolution. Initial phasing was performed with SHELXE (Sheldrick, 2002) followed by fourfold noncrystallographic symmetry averaging, and solvent flattening was performed using DM (Cowtan & Main, 1998). The resulting map could be readily interpreted by comparison with the PRRSV-NΔ57 structure and most of the A, B, C and D subunits were built automatically using ARP/wARP (Morris et al., 2003). Further model building was performed with O (Jones et al., 1991) and refinement was performed using REFMAC (Murshudov et al., 1997). Initial rounds of refinement used NCS restraints, which were relaxed in the final rounds. The final model has R_cry = 0.219 and R_free = 0.248 (Table 1).

3. Results and discussion

3.1. Structure determination

EAV-NΔ48, comprising residues 49–110 of EAV-N (Fig. 1), was cloned and expressed in E. coli with an N-terminal His₆ tag, purified by nickel-affinity and size-exclusion chromatography and crystallized by the hanging-drop vapour diffusion against 0.1 M sodium phosphate pH 8.0 and 25% polyethylene glycol (PEG) 3350. The crystals belong to the orthorhombic space group C2221, with unit-cell parameters a = 68.23, b = 73.23, c = 138.60 Å, and diffracted X-rays to beyond 2 Å resolution at the UK CRG beamline BM14 at the ESRF (Table 1). The EAV-NΔ48 structure was solved to a resolution of 2.0 Å by single-wavelength anomalous diffraction (SAD) phasing on an SeMet derivative. The resulting model (Fig. 2) was refined with REFMAC against these data to a final R_cry of 21.9% and R_free of 24.8%.

3.2. Structure of EAV-N

EAV-NΔ48 forms a dimer with a similar structure to that of the previously solved PRRSV-NΔ57 structure (Doan & Dokland, 2003b), consisting of a four-stranded antiparallel β-sheet capped by two long α-helices (α2) and flanked by shorter N- and C-terminal helices (Figs. 2b and 3a). Two such dimers are arranged around a noncrystallographic twofold axis into a dimer of dimers that comprises the crystal asymmetric unit (Fig. 3a). Interactions between the two dimers are mediated by ten water molecules that occupy the space between strand β2 of subunits A and C (Fig. 3a). This water

![Figure 2](image-url)

**Figure 2**
Structure of EAV-NΔ48. (a) Stereo diagram of representative electron density from the final 2Fᵣ - Fᵣ map, showing residues Leu88–Lys49 in the β2 strand. (b) Stereo diagram of the Cα backbone of the AB dimer structure. Pertinent residues are labelled for orientation.
hole displays a strongly positively charged surface on the \(\alpha\)-helical side of the tetramer (Fig. 4a). In contrast, the surface of the \(\beta\)-sheet floor has an overall negative charge (Fig. 4b). The structure also showed the N-terminal His\(_6\) tag for subunits \(A\), \(B\) and \(D\), but the last five residues of the C-terminus (PPAGP) were disordered in all four subunits.

The four subunits making up the two dimers in the asymmetric unit are very similar and could be superimposed with pairwise root-mean-square deviations (r.m.s.d.s) ranging from 0.3 to 0.7 Å (Fig. 3b). Likewise, the \(AB\) dimer could be superimposed on the \(CD\) dimer with an r.m.s.d. of 0.7 Å. Nevertheless, there is some variation, primarily localized to residues Ala83–Gly86 which form the loop between \(\beta\)-strands \(\beta1\) and \(\beta2\) (the ‘\(\beta\)-loop’), and to a lesser extent in the N- and C-terminal \(\alpha\)-helices (Fig. 3b). Indeed, the electron density for the \(\beta\)-loop was poor in subunits \(A\) and \(D\), presumably reflecting their exposed location on the outside of the tetramer. Such variation could potentially reflect the conformational variability required for the assembly of an icosahedral shell with more than 60 subunits (Caspar & Klug, 1962; Dokland, 2000).

The flexibility in the subunits is also reflected in the \(B\) factors. The greatest mobility is found in the C-terminal helix \(\alpha3\), as well as in the variable \(\beta\)-loop (Fig. 3c). However, there are clear differences in the degree of mobility in the four subunits. Overall, subunit \(D\) is far more mobile than the other three. The \(B\) factor in \(\alpha3\) of subunit \(D\) is as high as 72.2 Å\(^2\), compared with a mean \(B\) factor of 32.2 Å\(^2\) for the whole structure. In addition, there is a flexible region in the part of subunit \(C\) that forms the outer surface of the tetramer. This mobility is a consequence of the lack of contacts between the \(CD\) dimer and the symmetry-related \(CD\) at this end of the tetramer. In contrast, the \(AB\) dimer, which forms strong contacts with the symmetry-related \(AB\) dimer across the crystal twofold axis at this end of the molecule (see below), displays much lower mobility at the equivalent molecular interface (Fig. 3c).

3.3. Comparison with other nidoviral nucleocapsid proteins

The \(N\) protein in EAV consists of 110 residues, compared with 123 residues in PRRSV (Fig. 1). The largest sequence difference between the EAV and PRRSV \(N\) proteins exists in the N-terminal RNA-binding domain, in which there are ten fewer residues in EAV-N and the sequence identity is less than 16%. Notably, many of the positively charged Lys residues involved in RNA binding and the Cys residues considered important for assembly in PRRSV (Wootton & Yoo, 2003) are absent in EAV.

**Figure 3**
The EAV-N\(\Delta48\) structure. (a) Ribbon representation of the two dimers in the asymmetric unit, coloured in a spectrum from blue (subunit \(A\)) through green (subunit \(B\)) and yellow (subunit \(C\)) to red (subunit \(D\)). The ten water molecules between the \(AB\) and \(CD\) dimers are shown as grey spheres. Secondary-structure elements in the \(A\) and \(B\) subunits are labelled. (b) Superposition of the \(C\) backbone of all four subunits, coloured as in (a). (c) Worm trace of the \(AB\)-\(CD\) tetramer backbone. The tube thickness and colour reflect the average \(B\) factors (thin/blue, low; thick/red, high). (d) Superposition of the \(C\) backbone of the EAV-\(AB\) dimer (blue) on the PRRSV-\(N\) dimer (red). \(N\)- and \(C\)-termini are indicated. (e) Ribbon representation of the \(AB\)-\(CD\) tetramer and the \(A\primeB\prime-C\primeD\prime\) tetramer related to \(AB\)-\(CD\) by the crystallographic twofold symmetry axis parallel to \(b\) in the \(ab\) plane (filled oval). The unit cell is indicated with balls at each corner. (f) Superposition of the \(AB\)-\(CD\) tetramer (blue) with the \(AB\)-\(A\primeB\prime\) tetramer (red). (For this experiment, only one of the two dimers in the each tetramer was superimposed in order to emphasize the difference in angle between the two tetramers.) All figures were produced using PyMOL (DeLano, 2002).
In contrast, the C-terminal capsid-forming domains represented by the present structure share about 35% sequence identity. The greatest sequence divergence is in the α2 helix, which is exposed on the outer surface of the protein. These residues are presumed to form virus-specific interactions with other viral proteins and do not directly affect the three-dimensional conformation of the N protein. Consequently, the two structures are very similar, with an overall r.m.s.d. of 2.6 Å for the A subunit. The largest difference is found in the C-terminal helix α3 and in the β-sheet, which are 3.9 and 4.0 Å apart, respectively, in the A subunits of EAV-N and PRRSV-N. These also correspond to the regions of highest flexibility in EAV-N. The r.m.s.d. between EAV-N and PRRSV-N increases to 2.8 Å when the whole dimer is considered, indicating that there are additional differences between EAV-N and PRRSV-N in the way the monomers are organized into a dimer; EAV-N forms a tighter dimer than PRRSV-N (Fig. 3d).

Coronaviruses have much larger nucleocapsid proteins than arteriviruses: up to 422 residues in the case of SARS-CoV. An arterivirus-like nucleocapsid domain was recently discovered in the C-terminal dimerization domains of the nucleocapsid proteins from the coronaviruses IBV and SARS-CoV (Chang et al., 2005; Jayaram et al., 2006; Yu et al., 2006). In these structures, the orientation of the α-helices relative to the β-sheet differs from that in EAV and PRRSV, but the topology is the same and could be classified as the ‘nidoviral nucleocapsid fold’. Thus, the evolutionary relationships between arteriviruses and coronaviruses evidenced by the similar organization of the genomes is also reflected in the nucleocapsid protein structures and will probably also be reflected in other viral protein structures. The nidoviral nucleocapsid fold differs from the nucleocapsid proteins of both the flaviviruses (Dokland et al., 2004) and the alphaviruses (Choi et al., 1997).

3.4. Assembly and structure of the virion

As previously pointed out, the arterivirus N protein resembles the coat protein of RNA bacteriophage MS2, although the MS2 protein has a wider eight-stranded β-sheet and α-helices in a different orientation (Doan & Dokland, 2003b; Valegård et al., 1990). The organization of MS2 thus provides a reasonable starting point for under-
standing the organization of N in the EAV nucleocapsid. In this model, the β-sheet of the dimer would face the inside of the shell, where the N-terminal domains would be positioned to interact with the RNA (Fig. 5a). The α-helical outer surface might form a docking site for the envelope proteins, most likely the endodomains of the major envelope proteins M and/or GP5 (Fig. 5a; Doan & Dokland, 2003b). Assembly of the shell is probably mediated through interactions between the predicted α-helix in residues 19–31 as well as with the RNA.

In order to build an icosahedral shell with more than 60 subunits (or any nonsymmetric large shell) there has to be some conformational variability and/or differences in subunit interactions between the capsid-protein subunits (Caspar & Klug, 1962; Dokland, 2000). In the T = 3 plant viruses, for example, such conformational variation is generated through the formation of two different types of dimers: a flat CD dimer and a curved AB dimer (Abad-Zapatero et al., 1980; Harrison et al., 1978). The differences in dimer organization are reflected in conformational differences and order–disorder transitions in the N-terminal part of the protein.

In the EAV-N crystal structure, the two dimers (AB and CD) that comprise the tetramer in the crystal asymmetric unit are organized via a quasi-twofold axis so that the β-sheets of both dimers are more or less facing the same way (Fig. 3a). This tetramer is almost identical to that formed by the AB dimer and the A′B′ dimer related by the crystallographic twofold axis (Figs. 3e and 3f). It is tempting to suggest that the tetramer comprises the building block for the shell and that the quasi-equivalence between these two types of tetramers reflects the organization of the N protein in the virus. For example, we might build a shell with a diameter of around 31–38 nm from 360 copies of N (Fig. 5b). However, there is currently no experimental evidence to support this model and crystal contacts are not necessarily an indicator of in vivo interactions. Confirmation of this model would have to await characterization of the arterivirus structure by cryo-electron microscopy and three-dimensional reconstruction.

We are grateful to Dr Eric Snijder at Leiden University for suggesting this project, for providing us with the original EAV-N clone and for his continued support of this work.

References
Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D. & Tsukihara, T. (1980). Nature (London), 286, 33–39.
Balasuriya, U. B. & MacLachlan, N. J. (2004). Vet. Immunol. Immunopathol. 102, 107–129.
Caspar, D. L. D. & Klug, A. (1962). Cold Spring Harbor Symp. Quant. Biol. 27, 1–24.
Cavanagh, D. (1997). Arch. Virol. 142, 629–633.
Chang, C. K., Sue, S. C., Yu, T. H., Hsieh, C. M., Tsai, C. K., Chiang, Y. C., Lee, S. J., Hsiao, H. H., Wu, W. J., Chang, C. F. & Huang, T. H. (2005). FEBS Lett. 579, 5563–5568.
Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003). Nucleic Acids Res. 31, 3497–3500.
Choi, H. K., Lu, G., Lee, S., Wengler, G. & Rossmann, M. G. (1997). Proteins, 27, 345–359.
Cipriani, F. et al. (2006). Acta Cryst. D62, 1251–1259.
Collaborative Computational Project, Number 4 (1994). Acta Cryst. D50, 760–763.
Cowan, K. & Main, P. (1998). Acta Cryst. D54, 487–493.
Del Piero, F. (2000). Vet. Pathol. 37, 287–296.
DeLano, W. L. (2002). The PyMOL Molecular Graphics System. San Carlos, CA, USA: DeLano Scientific.
Doan, D. & Dokland, T. (2003a). Acta Cryst. D59, 1504–1506.
Doan, D. & Dokland, T. (2003b). Structure, 11, 1445–1451.
Dobbe, J. C., van der Meer, Y., Spaan, W. J. M. & Snijder, E. J. (2001). Virology, 286, 283–294.
Dokland, T. (2000). Structure, 8, R157–R167.
Dokland, T., Walsh, M., MacKenzie, J. M., Khromykh, A. A., Ee, K.-H. & Wang, S. (2004). Structure, 12, 1157–1163.
Evans, G. & Pettifer, R. F. (2001). J. Appl. Cryst. 34, 82–86.
Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. & Ball, L. A. (2005). Editors. Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses. San Diego: Elsevier/Academic Press.
Gerchman, S. E., Graziano, V. & Ramakrishnan, V. (1994). Protein Expr. Purif. 5, 242–251.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). Bioinformatics, 15, 305–308.
Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). Nature (London), 276, 368–373.
Horzinek, M. C. (1981). Non-arthropod-borne Togaviruses. London: Academic Press.
Jayaram, H., Fan, H., Bowman, B. R., Ooi, A., Jayaram, J., Collisson, E. W., Lescar, J. & Prasad, B. V. V. (2006). J. Virol. 80, 6612–6620.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Acta Cryst. A47, 110–119.
Kabsch, W. J. (1993). J. Appl. Cryst. 26, 795–800.
MacLachlan, N. J. & Balasuriya, U. B. (2006). Adv. Exp. Med. Biol. 581, 429–433.
Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). Methods Enzymol. 374, 229–244.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Acta Cryst. D53, 240–255.
Neumann, E. J., Kliebenstein, J. B., Johnson, C. D., Mabry, J. W., Bush, E. J., Seitzinger, A. H., Green, A. L. & Zimmerman, J. J. (2005). J. Am. Vet. Med. Assoc. 3, 385–392.
Plagemann, P. G. W. (1996). Fields Virology, 3rd ed., edited by B. N. Fields, D. M. Knipe & P. M. Howley, pp. 1105–1120. Philadelphia: Lippincott–Raven.
Sheldrick, G. M. (2002). Z. Kristallogr. 217, 644–650.
Snijder, E. J., Dobbe, J. C. & Spaan, W. J. M. (2003). J. Virol. 77, 97–104.
Snijder, E. J., Horzinek, M. C. & Spaan, W. J. M. (1993). Adv. Exp. Med. Biol. 342, 235–244.
Snijder, E. J. & Meulenberg, J. J. M. (1998). J. Gen. Virol. 79, 961–979.
Tijms, M. A., van der Meer, Y. & Snijder, E. J. (2002). J. Gen. Virol. 83, 795–800.
Valkenberg, K. L., Liljas, L., Fridborg, K. & Unge, T. (1990). Nature (London), 345, 36–41.
Wierenga, R., de Vries, A. A., van der Meulen, J., Godeke, G. J., Onderwater, J. J., van Tol, H., Koerten, H. K., Moommaas, A. M., Snijder, E. J. & Rottier, P. J. M. (2004). J. Virol. 78, 13019–13027.
Wootton, S. & Yoo, D. (2005). J. Virol. 77, 5456–5457.
Yoo, D., Wootton, S., Li, G., Song, S. & Rowland, R. R. (2003). J. Virol. 77, 12173–12183.
Yoo, I. M., Oldham, M. L., Zhang, J. & Chen, J. (2006). J. Biol. Chem. 281, 17134–17139.
Zimmerman, J. J., Yoon, K. J., Wills, R. W. & Swenson, S. L. (1997). Vet. Microbiol. 55, 187–196.