Structure of alphacoronavirus TGEV nsp1 has implications for coronavirus nsp1 function and evolution

Anna M. Jansson

Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, SE-751 24 Uppsala, Sweden
E-mail: anna.jansson@icm.uu.se

Abstract
Coronavirus nsp1 have been shown to induce suppression of host mRNA expression and to interfere with host immune response. However, the mechanism is currently unknown. The only available structural information on coronavirus nsp1 is the NMR structure of the N-terminal domain of nsp1 from severe acute respiratory syndrome coronavirus (SARS-CoV) from the genus betacoronavirus. Here we present the first nsp1 structure from an alphacoronavirus, TGEV nsp1. It displays a six-stranded β-barrel fold with a long alpha helix on the rim of the barrel, a fold shared with SARS-CoV nsp113-128. Contrary to previous speculation, the TGEV nsp1 structure suggests that coronavirus nsp1s have a common origin, despite the lack of sequence homology. However, comparisons of surface electrostatics, shape and amino acid conservation between the alpha- and betacoronaviruses lead us to speculate that the mechanism for nsp1 induced suppression of host mRNA expression might be different in these two genera.

Introduction
Coronaviruses (CoVs) cause mainly respiratory and enteric disease (1). In farm animals these viruses cause severe disease and lead to large economic losses. In humans CoVs generally cause mild symptoms like the common cold. However, the emergence of severe acute respiratory syndrome (SARS) in 2003 made it apparent that CoVs could cause serious disease also in the human population. CoVs contain a positive, single-stranded RNA genome of about 30kb, which is the largest among RNA viruses (2, 3). The replicase gene, comprising two thirds of the genome, encodes two large precursor polyproteins that are cleaved into 16 non-structural proteins (nsp), where nsp1 is the first to be expressed (4-6, 2, 7).
CoVs were originally classified into three groups based on antigenic cross-reactivity (8). Subsequent phylogenetic analysis, including analysis of the replicase region, rendered the same three clusters with few exceptions. These were called groups 1, 2 and 3 (9, 10). When SARS coronavirus (SARS-CoV), the etiological agent of SARS, was discovered (11-13) it was placed as the only member in an early split-off from group 2, in sub-group 2b (10). This effectively put the viruses previously established to be members of group 2 in sub-group 2a. These groups have now been recognized as genera, where groups 1, 2 and 3 have become the genera alpha-, beta-, and gammacoronaviruses (α-CoVs, β-CoVs and γ-CoVs), and SARS-CoV is placed in lineage B of the beta genus, β-CoVB. Since then, several SARS-like viruses have been identified, mainly in bats, and placed in β-CoVB (14, 15).

The CoV genome is generally well conserved between the genera. The largest differences in the replicase gene can be found in the 5’-end, and the most N-terminal cleavage product, nsp1, is considered one of the genus specific markers (16, 10). This is based both on sequence comparisons and the fact that nsp1 from α-CoVs, β-CoVA and β-CoVB are different on size, ~ 110, 250 and 180 amino acids respectively. In contrast to the α-CoVs and β-CoVs, the γ-CoVs do not contain an nsp1 protein (17). The fact that no sequence homology could be inferred between the different nsp1s, or any host protein, raised the question of whether these proteins shared similar structure and function (16). However, several studies have shown that nsp1s from both alpha- and betacoronaviruses display both differences and similarities.

It is established that nsp1 suppresses expression of host mRNA. Nsp1s from human CoV-299E, MHV and SARS-CoV significantly reduce reporter gene expression in HEK 293 cells (18-20). In several cell lines, SARS-CoV nsp1 suppresses host gene expression, including that of type I interferon, involved in host immune response (21). SARS-CoV nsp1 also promotes the degradation of host mRNA (22, 23). Like SARS-CoV nsp1, TGEV nsp1, can efficiently suppress host mRNA translation although it seems to lack the ability to modify and degrade host mRNAs. There are indications that SARS-CoV nsp1 also suppresses the expression of the CoV genes (22), but recent experiments on SARS-CoV nsp1 suggest that a short sequence in the 5’-end common
Deletion of nsp1 from infectious clones of murine hepatitis virus (MHV) from β-CoV, abolishes the ability of the virus to infect cultured cells (26). A mutation in the cleavage site between nsp1 and nsp2 in the α-CoV transmissible gastroenteritis virus (TGEV), preventing the release of nsp1, leads to a drastic decrease in viability of the virus (27).

The observed biochemical effects of nsp1 highlight the importance of this protein in the CoV life cycle, and its potential role as a significant virulence factor, as well as its importance for evasion of host responses. This also indicates that nsp1 is an interesting target in the search for new antiviral drugs. The frequent detection of SARS-like CoV in mammalian hosts indicates a high risk of reintroduction into the human population (14). For development of vaccines and antivirals it is important to understand CoV pathogenicity and its mechanism for avoiding host antiviral systems.

This paper presents the first high-resolution crystal structure of nsp1 from an alphacoronavirus. To date, the only known structure of nsp1 from coronaviruses is that of SARS-CoV nsp1^{13-128} belonging to the betacoronavirus genus, which was determined by NMR (28). The structure of TGEV nsp1 reflects the structural and functional similarities and differences between α-CoV and β-CoV. It also suggests that non-structural protein 1 was not acquired independently by the different coronavirus genera.

**Materials and Methods**

**Cloning, protein expression and purification**

A full length construct of the TGEV nsp1, including an N-terminal (His)₆-tag, was cloned into the expression plasmid pDEST14 (Invitrogen). The protein was expressed in *Escherichia coli* BL21-AI cells (Invitrogen) grown in LB media at 37°C. When OD₆₀₀ reached 0.6, the culture was transferred to 25°C and protein expression was induced with L-arabinose (2g/L). After 3-5 hrs, the cells were harvested by centrifugation. The cells were washed in 1xSSP buffer [150 mM NaCl, 250 mM NaH₂PO₄ pH 7.4] prior to storage at -20°C. For protein purification, the cells from a 1L culture were thawed and re-suspended in 20 mL lysis buffer [50 mM Na₂HPO₄, 50 mM Na₂SO₄, 100mM HEPES, 200 mM NaCl, 10 mM imidazole, 0.5% Triton X-100,
14 mM β-mercaptoethanol, pH 8.0] supplemented with 0.01 mg/mL RNase, 0.02 mg/mL DNase and 0.25 mg/mL lysosome. The cells were subsequently lysed under 2 kBar pressure using a Constant Cell Disruptor (Constant Systems Ltd) and the lysate was centrifuged at 8°C and 45,000 x g (SS-34 rotor, Sorvall) for 20 minutes. The cleared cell lysate was incubated with 0.5 mL Ni Sepharose (6 Fast flow; GE Healthcare) for 30 minutes at 8°C on a shaker. The Ni matrix was washed on a column with 20 mL wash buffer [50 mM Na₂HPO₄, 50 mM Na₂SO₄, 100 mM HEPES, 200 mM NaCl, 20 mM imidazole, 14 mM β-mercaptoethanol, pH 8.0], and the protein was eluted with 2.5 mL elution buffer [same as wash buffer but with 250 mM imidazole]. Directly after elution, the buffer was exchanged on a PD-10 column (BioRad) and eluted with 20 mM Tris-HCl, 300 mM NaCl, pH 8.0, 14 mM β-mercaptoethanol. Ni Sepharose purification and buffer exchange were performed at 8°C. The protein was further purified by size exclusion chromatography (HiLoad 16/60 Superdex-75; GE Healthcare). The fractions from the peak corresponding to a monomer of the TGEV nsp1 protein were pooled and diluted four times with 20 mM Tris-HCl pH 8.0 to a NaCl concentration of 75 mM. The protein was then applied to a 1 mL HiTrapQ anion exchange column (GE Healthcare), which was washed with 20 mL of buffer A [20 mM Tris-HCl, 75 mM NaCl, pH 8.0, 14 mM β-mercaptoethanol] and eluted with a gradient to buffer B [20 mM Tris-HCl, 1 M NaCl, pH 8.0, 14 mM β-mercaptoethanol] over a volume of 20 mL. Both size exclusion chromatography and anion exchange chromatography were carried out at 25°C. The TGEV nsp1 eluted at 500 mM NaCl and the purity of the sample was >98% as judged by analysis on SDS-PAGE. The protein sample was diluted with 20 mM Tris-HCl to a NaCl concentration of 150 mM and thereafter concentrated to between 3-10 μg/μL in a Vivaspin concentrator (Vivascience).

A second construct with a 5-amino acid C-terminal truncation and an N-terminal (His)₆-tag was cloned into expression vector pEXP5 (Invitrogen). Expression and purification were performed as for the full-length construct.

Crystallization

For crystallization screening, drops containing 0.5 μL protein solution and 0.5 μL reservoir solution were set up as sitting-drop vapor-diffusion experiments using an
Oryx 4 crystallization robot (Douglas Instruments Ltd). Initial crystal hits were obtained at 20°C in two conditions in the JCSG+ suite (Qiagen): A9 [200mM ammonium chloride, 20% (wt/vol) PEG3350] and H7 [200mM ammonium sulfate, 100mM Bis-Tris pH 5.5, 25% (wt/vol) PEG3350]. The crystallization conditions were optimized in terms of precipitant, buffer, pH and protein concentration. Optimal concentration of protein and PEG4000 were batch dependent. Drops, in volumes varying between 3-20 μL, containing protein and reservoir solution in a 2:1 ratio were set up. The drops were seeded with previously obtained crystals 30 minutes after setup. After several rounds of optimization the best crystals were obtained in 5% (wt/vol) PEG4000, 200mM ammonium chloride, 30mM HEPES, 30mM MES, pH6.2 with a protein concentration of 5 μg/μL. Native crystals were dipped for a few seconds in reservoir solution supplemented with 15% glycerol before vitrification in liquid nitrogen. Crystals for phasing were soaked for 2 hours in reservoir solution supplemented with 10mM K₂PtCl₄, and thereafter back-soaked for 30 minutes in the same solution without K₂PtCl₄. The Pt-soaked crystals were cryoprotected and vitrified as described above.

Data collection, phasing and refinement
Crystallographic data were collected at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Native data were collected at beamline ID23eh2, at a wavelength of 0.873Å to 1.6Å resolution. Anomalous data were collected at beamline ID14eh4. Two 360 degree datasets were collected at the Pt edge (λ=1.072Å) at different κ angles, with an oscillation angle of three degrees, to a resolution of 2.5Å. The images were indexed and integrated in MOSFLM (29) and scaled in Scala (30, 31). The space group was determined to be P1 with two molecules in the asymmetric unit, related by a non-crystallographic two-fold axis as revealed by a self rotation function calculated by Molrep (32, 31). The solvent content was estimated to be 40% with a Matthews coefficient of 2.08 (33, 31). Four platinum sites were identified by single isomorphous replacement with anomalous scattering (SIRAS) using ShelxD (34). The sites were further refined in SHARP (35). Subsequent solvent flattening and histogram matching using DM (36) resulted in a significantly improved electron-density map. Buccaneer (37) was used to create a first trace of the polypeptide backbone. This initial model was further improved by alternate cycles of
model rebuilding in O (38) and refinement in Buster-TNT (39, 40). Final refinement was performed using TLS refinement with the two chains as separate groups. In the N-terminal of the A-chain, an additional four residues from the His tag could be modeled. In the B-chain, density for the two first residues in the N-terminal was missing. For full phasing and refinement statistics, see Table 1.

PDB-id

The TGEV nsp1 structure has been deposited to the Protein Data Bank with PDB-id 3ZBD.

Results

The TGEV nsp1 structure exhibits an irregular β-barrel fold

Initial crystallization trials were performed with protein from a construct expressing the full-length TGEV nsp1. Crystals were obtained and tested for diffraction but no data of sufficient quality could be collected. The sequence from TGEV nsp1 was analyzed with the secondary structure prediction software Phyre (41) and I-Tasser (42) and a new construct was produced, containing a five-amino-acid C-terminal truncation. This new construct yielded crystals in the same crystallization conditions as the full-length protein, and native data were collected to 1.5Å. Anomalous data from platinum soaks were collected to 2.4Å. The structure was subsequently solved by Single Anomalous-Isomorphous Replacement (SIRAS) using both datasets. Details of phasing and refinement are in Table 1.

The TGEV nsp1 structure is characterized by an irregular six-stranded β-barrel, flanked by a small β-sheet connected to a short 310-helix (Fig. 1). A 15 amino acid long α-helix is placed on the rim of the barrel. Four anti-parallel strands, β3, β7, β5 and β6, make up one side of the barrel, with β3 and β6 loosely connected to strands β1 and β8 that create the other side of the barrel. β2 and β4 form a small parallel sheet flanking the barrel adjacent to β3 and β7. Some of the strands are irregular, and have breaks in the hydrogen bonding pattern. This is due to a β-bulge in β7 involving the carbonyl oxygen from Val64 and amide nitrogens from Gln85 and Gly86, as well as Pro74 that makes a kink in the end of the strand β6.
A small cavity is located at the top of the barrel between $\beta_5$ and $\beta_7$, next to the $\alpha$-helix. The cavity is lined by residues Val18, Pro19, Leu21, Val26, Glu26, Tyr41, Val61, Ile62, Val89 and the aliphatic stem of Arg90. Glu29 has a different conformation in the A-chain compared to that in the B-chain and the position of the side chain determines the size of the opening of the cavity to the solvent.

The TGEV nsp1 sequence shares no significant similarity to any known structures in the Protein Data Bank. However, a search with the crystal structure in the PDBeFold database resulted in a single significant hit, the NMR structure of the N-terminal domain of nsp1 from SARS-CoV (residues 13-128, pdb-id 2HSX/2GDT), with a q-score of 0.37 and a z-score of 6.3. The q-score indicates the quality of the alignment where 1 corresponds to an identical hit. The z-score measures statistical significance of the match, where a higher number corresponds to a higher statistical significance (43).

To explore the relationship between the nsp1 proteins in $\alpha$-CoV and $\beta$-CoVB, sequences from both groups were gathered and aligned separately with Clustal W (44). A careful structure based sequence alignment between TGEV nsp1 and SARS-CoV nsp1$^{13-128}$ was used to merge the alignments of $\alpha$-CoV and $\beta$-CoVB.

Conservation within the alphacoronavirus genus

The $\alpha$-CoV alignment shows a number of highly conserved areas (Fig. 2). A large portion of the conserved residues in $\alpha$-CoV make up the hydrophobic core of the $\beta$-barrel fold: these include Val 44, Val52, Val61, Leu77, Leu 84, Phe87, Ile88 and Val89. This cluster of residues is connected to a conserved solvent-exposed hydrophobic patch consisting of Phe43 and Phe100, via Gly86, which is highly conserved due to space restraints. The Ile23-Arg35 helix shows little conservation. However, it is anchored to the hydrophobic core by the highly conserved Gly37 and Phe38.

The TGEV nsp1 structure contains two salt bridges: Lys7 - Asp99, connecting $\beta_1$ to $\beta_8$ and Lys103 - Asp71, connecting $\beta_8$ to $\beta_6$. Although Lys7 - Asp99 can be found in
mink-CoV in the α-CoV group, none of these electrostatic interactions seem to be well retained throughout the CoV family.

The surface of the TGEV nsp1 structure exhibits two highly conserved areas. The first is located on a ridge formed by the loops between stands β1-β2 and β7-β8 together with strand β2. The second comprises residues from strand β8, where the beginning of the strand shows the highest level of conservation (Fig. 3 and 4). On the ridge, Asp13, Gln15, Asn92 and Asn94, all conserved, are positioned in a ring around Tyr14, which is not. Gln15 is consistently replaced by a Glu in the other α-CoVs. The second conserved patch is mainly made up from side chains from strand β8 where the highest level of conservation is found in the N-terminal part. The cluster also includes one residue from β3, Phe43. Together with Phe100, this residue forms a small exposed hydrophobic patch. In this conserved area, two more hydrophobic surface exposed residues are found: Val96 and Leu97.

Overall comparison between α-CoV and β-CoVB

A comparison of TGEV nsp1 with the structure of SARS-CoV nsp113-128 clearly shows that the two structures share the same fold with a characteristic six-stranded β-barrel with a long alpha helix on one side of the barrel. However, a three-dimensional alignment of TGEV nsp1 with SARS-CoV nsp113-128 reveals that there are large differences between the structures. The location of the strands in the barrel is shifted, along with an outward shift of the α-helix in TGEV nsp1 compared to SARS-CoV nsp113-128 where the helix is positioned closer to the barrel. The loop between β5 and β6 is significantly shorter in the TGEV nsp1 structure. In addition, the small β-sheet, comprising β2 and β4, flanking the barrel next to strands β3 and β7, is only found in the TGEV nsp1 structure.

The small cavity with Glu29 as a gatekeeper is not conserved. Instead there is a narrow tunnel in the SARS-CoV nsp113-128 structure, not found in TGEV nsp1, that stretches through the center of the barrel. It appears too large to be an artifact from poor packing of the protein core. However, it is not likely to be conserved throughout the β-CoVB lineage, given the low conservation of the neighboring side chains.
Thus far, the viruses that belong to β-CoVB show lower diversity than those in α-CoV. An alignment of nsp1 from four viruses in β-CoVB, including SARS-CoV, shows three conserved areas. The mapping of these onto the SARS-CoV nsp113-128 structure is illustrated in Fig. 4. The two separate alignments of α-CoV and β-CoVB were merged using the structure based alignment of TGEV nsp1 and SARS-CoV nsp113-128 (Fig. 2). Interestingly, the conserved regions within each group show very little overlap in the combined α-CoV and β-CoVB alignment. For example, the β-CoVB nsp1 proteins show a high level of conservation in helix α1, absent in α-CoV. The conservation pattern in the barrel is also different between the two groups. The consensus sequence LRKxGxKG, referred to by Almeida et al. (28), is roughly conserved within β-CoVB. When comparing with the α-CoV family, only the two glycines are conserved, both of which seem to be located in the linker region between nsp1 and nsp2. However, a few residues seem to be retained across the α-CoV genus and the β-CoVB lineage. Most of these, like Ile8, Ile88, Phe43, Val44, Val52, Ile88 and Val 89, are part of the conserved hydrophobic cluster, extending from the core of the barrel to the surface. Val44 is the center of a less well-conserved cluster on the other side of β3. The β-bulge located by Gly86 seems to be absolutely conserved throughout α-CoV and β-CoVB and might be a characteristic feature of the nsp1 β-barrel fold.

The poor sequence conservation between α-CoV and β-CoVB is also reflected in the surface electrostatics. The open side of the TGEV nsp1 barrel exhibits a strong negative electrostatic potential whereas the long helix features mainly positive electrostatics (Fig. 4C and 4I). The SARS-CoV nsp113-128 structure reveals a significantly different pattern. The electrostatic potential also seems to be slightly more conserved in β-CoVB than in α-CoV (Fig. 4).

Discussion

The high-resolution crystal structure of TGEV nsp1 reveals that nsp1s from α-CoV share a common fold with the N-terminal domain of the nsp1s in β-CoVB, despite their lack of sequence similarity. At the same time the structure also highlights that there are important structural differences between the two lineages, potentially
explaining their differences in function. SARS-CoV nsp1 inhibits IFN expression in infected cells (19) and interferes with antiviral signaling pathways of the host (21). TGEV nsp1, together with SARS-CoV nsp1 and several other CoV nsp1s, can also efficiently inhibit expression of host mRNA. However, little is known about the mechanism behind this function.

The structure of TGEV nsp1 is characterized by an irregular six-stranded β-barrel flanked by an α-helix. In order to identify the evolutionarily conserved areas, an alignment was made from various nsp1 sequences from viruses in the alpha genus. The conserved residues were plotted onto the surface of the TGEV nsp1 protein. The conservation pattern within α-CoV does not give any immediate clues about the function or mechanism of the α-CoV nsp1. A large portion of the conserved residues, centered around the highly conserved strand β7, make up the hydrophobic core of the protein and are more likely to be involved in the structural stability of the protein than being important for the function. However, the TGEV nsp1 surface features two highly conserved patches. From these, a few residues stand out as candidates for potential interaction with a partner or target molecule. The patch made up from the two loops between strands β1-β2 and β7-β8 together with strand β2 has four residues that are of special interest. These are Asp13, which is completely conserved, Gln15 which is a conserved Glu in all α-CoVs except TGEV, and two asparagines on the neighboring loop, Asn92 and Asn94. These conserved residues are all placed on a protruding, ridge formation. The highest conservation of the second patch is mainly found on the edge of the ridge and going down on one side (Fig. 3B), including residues Leu97, Glu98 and Asp99. Both these patches are potential surfaces for interaction with another molecule. The protruding shape of the ridge, as opposed to a cavity or a bowl shape, suggest that the partner molecule may be another protein. There are indications that TGEV nsp1 may need a host factor for its function. Experiments performed in cell-free HeLa extracts and rabbit reticulocyte lysate (RRL) reveal that TGEV nsp1 suppresses protein translation in the first experiment but not the second, suggesting that a host factor that exist in the HeLa extracts, but not in RRL, is needed for TGEV nsp1 function (22).
Intriguingly, the combined alignment of α-CoV and β-CoVB nsp1s shows that there is not much overlap between the conservation patterns of the two groups. The lack of conservation is also reflected in the shape and the electrostatics of the TGEV and SARS-CoV nsp1\textsuperscript{13-128} structures, resulting in different three dimensional volumes despite the similar β-barrel fold (Fig. 4).

It has been previously speculated that the SARS-CoV nsp1 might be a unique SARS protein, and that its ability to suppress expression of host mRNA potentially could account for its elevated pathogenicity relative to other coronaviruses (16). It is now established that nsp1 from α-CoV as well as that from β-CoVA and β-CoVB can induce suppression of host mRNA expression (22, 23, 18, 19, 45, 20). It is also established that SARS-CoV nsp1 binds the 40S subunit of the ribosome to make it translationally inactive. The nsp1-40S complex can modify the 5' end of capped mRNA and induce cleavage in certain mRNAs containing the internal ribosome entry site (IRES). However, this activity cannot alone account for the substantially reduced expression of reporter protein, suggesting that there is an additional mechanism for the suppression of host gene expression (23).

In contrast to these results, although TGEV nsp1 has been shown to effectively suppress host gene expression, no binding to the 40S ribosomal subunit has been observed (22). TGEV nsp1 also failed to promote host mRNA degradation (22). In SARS-CoV nsp1 it seems like the ability to bind to the 40S subunit is related to the second domain (residues 129-180) as SARS-CoV nsp1 carrying K164A H165A mutations was inactive in terms of 40S binding, and consequently unable to degrade mRNA (23). However, nsp1 proteins from two other alphacoronaviruses, HCoV-229E and HCoV-NL63, have been shown to immunoprecipitate together with the S6 protein, which is part of the 40S subunit (46). Interestingly, these two nsp1 proteins share all of the conserved regions in the α-CoV group (Fig. 2). Thus, it cannot be ruled out that the α-CoV TGEV nsp1 might interact with parts of the 40S ribosomal subunit under certain conditions.

It is tempting to speculate that the β-barrel domain of SARS-CoV nsp1\textsuperscript{13-128} and TGEV nsp1 share a similar mechanism for the additional suppression of host mRNA,
not accounted for by the SARS-CoV nsp1 induced modification of mRNA. However, there is no experimental data to support this. On the contrary, the K164A H165A double mutation harbored in the second domain renders the SARS-CoV nsp1 completely inactive in experiments where TGEV nsp1 effectively suppressed host mRNA expression (22).

The structural differences in the TGEV nsp1 and SARS-CoV nsp1, together with the available biochemical data, lead us to speculate that the nsp1 protein from α-CoV and β-CoV have different mechanisms for 40S-independent suppression of host mRNA. However, since the TGEV nsp1 structure has the same fold as SARS-CoV nsp1\textsuperscript{13-128} it is very unlikely that the nsp1s were acquired independently by the different genera. This suggests that the coronavirus nsp1s are evolutionarily related and that the different mechanisms are a result of divergent evolution.

Acknowledgements

I thank Linda Boomaars-van der Zanden and Eric J. Snijder (Leiden University Medical Center, The Netherlands) for providing the initial TGEV nsp1 full length expression clone, and Luis Enjuanes (Centro Nacional de Biotecnologia, Madrid, Spain) for providing the cloned cDNA that was used as a template. I thank Bruno Coutard and Bruno Canard (AFMB, Marseille, France) for assistance. Terese Bergfors is thanked for critical reading of the manuscript.

This work was supported by the EU IP Project VIZIER (CT 2004–511960) and Uppsala University Faculty Support to T. Alwyn Jones.

References


Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ,


Figure legends:

FIG 1 (A) Overall structure of TGEV nsp1 colored in rainbow from blue in the N-terminal to red in the C-terminal. (B) Topology diagram with coloring corresponding to that in A. Strands β1, β3, β5, β6, β7 and β8 make up the barrel. Strands β2 and β4 are unique for TGEV nsp1 and are not found in SARS-CoV nsp113-128.

FIG 2 Eight nsp1 sequences from α-CoV and four nsp1 sequences from β-CoV were aligned separately. The two alignments were subsequently merged by using the three dimensional structure alignment of TGEV nsp1 and SARS-CoV nsp113-128. The level of sequence conservation within each genus is highlighted with dark background color and white letters. The darker background color, the higher conservation. Residues conserved between the genera are marked with boxes. Residues likely to be important for α-CoV function is marked by stars. These residues are further highlighted in Figs 3 and 4. The β-CoVα consensus sequence suggested by Almeida et al (28) is marked by circles. Secondary structure elements from the TGEV nsp1 structure are displayed above the sequence and colored according to the scheme in Fig. 1. The figure was prepared using Aline (47).

FIG 3 The surface of TGEV nsp1 display two areas with high sequence conservation. Two loops make up the first area (A) where Asp13, Gln15, Asn92 and Asn94 make up a conserved circle. The second area (B), centered around strand β8, display both exposed hydrophobic residues as well as charged residues that potentially could interact with a partner molecule or partner protein.
FIG 4 Superposed structures of TGEV nsp1 and SARS-CoV nsp1\textsuperscript{13-128} are presented separately in two different rotations – rotation 1 A-F and rotation 2 G-L. Conserved residues within each genus are shown in green, where darker green corresponds to a higher level of conservation. In each rotation the TGEV nsp1 and SARS-CoV nsp1\textsuperscript{13-128} are shown as cartoons and surface representation displaying the conserved areas. Subfigures C, F, I and L show the electrostatic surface potential of the two structures from -4/+4 mV. The electrostatic potential was calculated with APBS and PDB2PQR using the PARSE force field (48-50). The area within the dotted oval in H corresponds to the two conserved areas shown in FIG 3.
Table 1: Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection statistics</th>
<th>TGEV nsp1 native</th>
<th>TGEV nsp1 Pt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beam line</td>
<td>ID23eh2 (ESRF)</td>
<td>ID14eh4 (ESRF)</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.873</td>
<td>1.072</td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td>Cell axial lengths (Å)</td>
<td>35.4 36.0 42.2</td>
<td>35.6 36.1 42.7</td>
</tr>
<tr>
<td>Cell angles (°)</td>
<td>91.3 109.1 94.2</td>
<td>90.9 109.4 93.9</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>27.4 - 1.5 (1.58-1.50)</td>
<td>40.3- 2.5 (2.64-2.5)</td>
</tr>
<tr>
<td>Number of reflections measured</td>
<td>122299 (17500)</td>
<td>167992 (24887)</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>30811 (4441)</td>
<td>6888 (998)</td>
</tr>
<tr>
<td>Average multiplicity</td>
<td>4.0 (3.9)</td>
<td>24.4 (24.9)</td>
</tr>
<tr>
<td>Anomalous average multiplicity</td>
<td>12.0 (12.3)</td>
<td>99.4 (99.8)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.6 (95.4)</td>
<td>99.6 (99.8)</td>
</tr>
<tr>
<td>Anomalous completeness (%)</td>
<td>99.4 (99.8)</td>
<td></td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.056 (0.476)</td>
<td>0.079 (0.263)</td>
</tr>
<tr>
<td>(I/σI)</td>
<td>14.3 (3.0)</td>
<td>45.5 (19.7)</td>
</tr>
</tbody>
</table>

Refinement statistics

| Resolution range (Å)      | 26.1-1.5        |
| Number of reflections used in working set | 31262       |
| Number of reflections for Rfree calculation | 1535        |
| R (%)                     | 17.5            |
| Rfree (%)                 | 20.9            |
| Number of non-hydrogen atoms | 1718        |
| Number of solvent waters  | 207             |
| Mean B-factor (Å²)        | 23.3            |
| Ramachandran plot outliers (%)\textsuperscript{a} | 0.0          |
| RMSD from ideal bond length (Å)\textsuperscript{b} | 0.01         |
| RMSD from ideal bond angle (°)\textsuperscript{b} | 1.06         |

Values in parentheses refer to the outer resolution shell. Data collection statistics as calculated by Scala, part of the CCP4 program suite (30, 31).

\textsuperscript{a}Refinement statistics, except for Ramachandran outliers, were calculated by Buster-TNT (39).

\textsuperscript{b}Calculated using a strict-boundary Ramachandran plot definition (51).

\textsuperscript{1}Ideal values from Engh & Huber (52).