Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural genomics of the SARS coronavirus: cloning, expression, crystallization and preliminary crystallographic study of the Nsp9 protein

Valérie Campanacci,[a] Marie-Pierre Egloff,[a] Sonia Longhi,[a] François Ferron,[a] Corinne Rancurel,[a] Aurelia Salomoni,[a] Cécile Durousseau,[a] Fabienne Tocque,[a] Nicolas Brémond,[a] Jessika C. Dobbe,[b] Eric J. Snijder,[b] Bruno Canard[a]* and Christian Cambillau[a]*

[a]Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS and Universités Aix-Marseille I and II, 31 Chemin Joseph Aiguier, 13402 Marseille CEDEX 20, France, and [b]Molecular Virology Laboratory, Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands

Correspondence e-mail: cambillau@afmb.cnrs-mrs.fr, canard@afmb.cnrs-mrs.fr

Received 12 July 2003
Accepted 30 July 2003

The aetiologic agent of the recent epidemics of Severe Acute Respiratory Syndrome (SARS) is a positive-stranded RNA virus (SARS-CoV) belonging to the *Coronaviridae* family and its genome differs substantially from those of other known coronaviruses. SARS-CoV is transmissible mainly by the respiratory route and to date there is no vaccine and no prophylactic or therapeutic treatments against this agent. A SARS-CoV whole-genome approach has been developed aimed at determining the crystal structure of all of its proteins or domains. These studies are expected to greatly facilitate drug design. The genomes of coronaviruses are between 27 and 31.5 kbp in length, the largest of the known RNA viruses, and encode 20–30 mature proteins. The functions of many of these polypeptides, including the Nsp9–Nsp10 replicase-cleavage products, are still unknown. Here, the cloning, *Escherichia coli* expression, purification and crystallization of the SARS-CoV Nsp9 protein, the first SARS-CoV protein to be crystallized, are reported. Nsp9 crystals diffract to 2.8 Å resolution and belong to space group $P6_{1/5}22$, with unit-cell parameters $a = b = 89.7$, $c = 136.7$ Å. With two molecules in the asymmetric unit, the solvent content is 60% ($V_M = 3.1$ Å^3 Da^{-1}).

1. Introduction

The recent epidemics of Severe Acute Respiratory Syndrome (SARS) represent a real paradigm for emerging viral pathogens, as well as an example of worldwide coordinated efforts to control a serious viral outbreak, a test of the reaction time of the scientific community. The first cases of Severe Acute Respiratory Syndrome originated from the Guangdong province in South East China. The number of cases reported and our current knowledge regarding this illness are still currently evolving, but a number of basic facts have been firmly established. The aetiologic agent of SARS is a positive-stranded RNA virus belonging to the *Coronaviridae* family and its genome differs substantially from those of previously identified coronaviruses, including two other human coronaviruses (Peiris *et al.*, 2003; Ksiazek *et al.*, 2003; Drosten *et al.*, 2003; Snijder *et al.*, 2003). The virus, whose name SARS-CoV is now currently accepted, is mainly transmitted by the respiratory route. However, evidence for a secondary faeco–oral route of transmission has also been presented. The viral strain probably primarily infected wild animals traded in Asian markets and crossed the species barrier to infect humans.

There is to date no vaccine and no prophylactic or therapeutic treatments against this agent. A prophylactic treatment would have been useful to combat the epidemics; the only effective measure available to prevent the spread of

© 2003 International Union of Crystallography
Printed in Denmark – all rights reserved

the virus is to quarantine all persons that have been exposed to SARS-CoV. The number of antiviral molecules that can be used to treat patients infected by RNA viruses is incredibly low. Accordingly, it is important to search for efficient antiviral drugs for a large number of RNA viruses, while giving priority to viruses transmitted by the respiratory route because they have the highest potential for causing pandemic outbreaks.

The scientific community has reacted promptly and efficiently to identify and characterize this new infectious agent, as well as to develop methods for SARS-CoV detection and containment protocols. In the meantime, a wide effort is being made to design drugs active against SARS-CoV. Ribavirin has been used in the absence of other candidates, but its intrinsic efficiency against SARS-CoV appears to be low (Koren *et al.*, 2003).

To select drugs active against a viral pathogen, one usually relies on screening candidate drugs for their efficacy in virus-infected cell cultures and/or animal models. However, during the current research on drugs for treating hepatitis C virus (HCV) infections, a novel and promising approach has been introduced. The RNA-dependent RNA polymerase of HCV has been purified and crystallized and enzymatic tests have been used to find potent nucleoside and non-nucleoside inhibitors of the virus, the structure–activity relationships of which allow further testing and clinical developments (de Francesco *et al.*, 2003). This approach is gaining momentum owing to a concomitant increase in the power of new technologies and technological developments. Among those, genomics approaches are being conducted to solve the crystal structures of large sets of clinically relevant proteins, which will become the subjects of future structure–function relationship studies.

A crystal structure has not yet been determined for any of the 28 predicted mature SARS-CoV proteins. The crystal structure of the main (or 3CL) protease of transmissible gastroenteritis virus, a related coronavirus, has been determined and was used to construct a model of the SARS-CoV 3CL protease, facilitating future drug design against this important target (Anand *et al.*, 2003). The putative coronavirus RNA-dependent RNA polymerase has been purified, but is inactive *in vitro* (Grotzinger *et al.*, 1996).

In this context, we have developed a SARS-CoV whole-genome approach aimed at determining the crystal structure of all SARS-CoV proteins. We anticipate that this will greatly facilitate drug design as well as the study of many other aspects related to the biology of these complex viruses. Coronaviruses are enveloped viruses with a single-stranded RNA genome of positive polarity (Lai & Holmes, 2001). Their genome is between 27 and 31.5 kbp in length, the largest of the known RNA viruses. Like other coronaviruses, the SARS-CoV genome is known to encode two large replicase polyproteins (the ORF1a and ORF1ab proteins), which are processed into a set of mature non-structural proteins (Nsps) by internal viral proteases (Snijder *et al.*, 2003). The functions of many of these products, such as the Nsp9–Nsp10 polypeptides produced from the C-terminal domain of the ORF1a-encoded polyprotein, are still unknown. In the related mouse hepatitis virus, which is a group 2 coronavirus, the SARS-CoV Nsp9 corresponds to a 12 kDa cleavage product (P1a-12) that is found preferentially in the perinuclear region of infected cells, where it co-localizes with other components of the viral replication complex (Bost *et al.*, 2000). No clues to the function of the Nsp9 equivalent of any coronavirus have been obtained thus far. Here, we report the cloning, expression, purification and crystallization of the SARS-CoV Nsp9 protein, a 113-residue protein (Fig. 1), which is the first SARS-CoV protein to be crystallized.

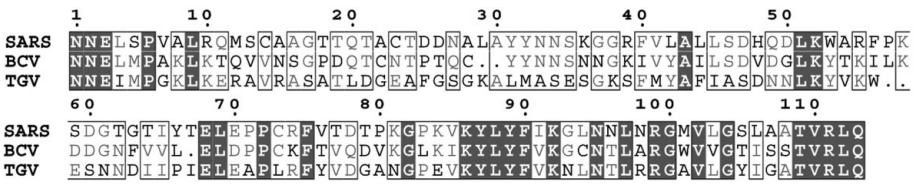

Figure 1
Alignment of the sequence of the SARS-CoV Nsp9 protein with that of bovine coronavirus (BCV) and of the transmissible gastroenterisis virus (TGV). Conserved residues are identified with a black background. Homologous residues are boxed.

2. Material and methods

2.1. Infection and RNA isolation

Vero cells were infected with SARS-CoV (Frankfurt-1 strain; NCBI Accession No. AY291315; Drosten *et al.*, 2003) at a multiplicity of infection of 0.01. At the onset of the cytopathogenic effect (approximately 40 h post-infection), intracellular RNA was isolated by cell lysis for 10 min at room temperature with 5% lithium dodecyl sulfate in LET buffer (100 m*M* LiCl, 1 m*M* EDTA, 10 m*M* Tris–HCl pH 7.4) containing 20 µg ml^{-1} of proteinase K. After shearing of the cellular DNA using a syringe, lysates were incubated at 315 K for 15 min, extracted with phenol (pH 4.0) and chloroform and the RNA was ethanol-precipitated. cDNA was obtained by reverse transcription using primer SAV009 (5′-GGACAG-CAACCGCTGGACAATC-3′), complementary to nucleotides 13644–13665 of the Frankfurt-1 genome, using Thermoscript reverse transcriptase (Invitrogen).

2.2. Subcloning, *Escherichia coli* protein expression and purification

The SARS-CoV Nsp9-coding sequence was amplified by PCR from the cDNA prepared above using two primers containing the attB sites of the Gateway recombination system (Invitrogen). At the 5′ end of the gene, a sequence encoding a hexahistidine tag was attached. The cDNA was then subcloned in the pDest14 plasmid (Invitrogen). The open reading frame of the final construct (referred to as pDest14/Nsp9-HN and encoding an N-terminally His-tagged version of SARS-CoV orf1a polyprotein residues

4118–4230) was checked by sequencing (MilleGen, Toulouse, France). Expression was performed in *E. coli* strain C41(DE3) (Avidis SA, France) transformed with the pLysS plasmid (Novagen). This plasmid carries the lysozyme gene, allowing tight regulation of the expression, and supplies the tRNAs for six rare codons used with a very low frequency in *E. coli*. Cultures were grown at 310 K until OD_{600} reached 0.6 and were then stored for 2 h on ice; 2% ethanol was added for the induction of stress chaperones (Gong & Shuman, 2002). Expression was induced by adding 50 µ*M* IPTG and cells were incubated for 16 h at 290 K. Cells were collected by centrifugation and the bacterial pellets were resuspended and frozen in 50 m*M* Tris–HCl, 150 m*M* NaCl, 10 m*M* imidazole pH 8.0.

Cellular suspensions were thawed with 0.25 mg ml^{-1} lysozyme, 0.1 µg ml^{-1} DNase and 20 m*M* $MgSO_4$ and were centrifuged at 12 000*g*. The supernatant was applied onto an Ni-affinity column connected to an FPLC system (Amersham Pharmacia Biotech). The protein was eluted with 50 m*M* Tris–HCl, 150 m*M* NaCl, 250 m*M* imidazole pH 8.0 and then applied onto a preparative Superdex 200 gel-filtration column pre-equilibrated in 10 m*M* Tris–HCl, 300 m*M* NaCl pH 8.0. The recombinant protein was characterized by N-terminal sequencing, mass spectroscopy, dynamic light scattering (DLS) and circular dichroism (CD).

2.3. Protein characterization

DLS was performed with a Dynapro Microsampler (Protein Solutions) using a protein solution at 5.8 mg ml^{-1} in 10 m*M* Tris–HCl, 300 m*M* NaCl pH 8.0. The CD spectrum of the final purified product was recorded between 185 and 260 nm on a JASCO J810 spectrometer using a protein solution at 0.1 mg ml^{-1} in sodium phosphate buffer pH 7.0 containing 25 m*M* NaCl.

2.4. Crystallization

Crystallization screening was performed by vapour diffusion with nanodrops using a Cartesian robot as described previously (Sulzenbacher *et al.*, 2002; Vincentelli *et al.*, 2003). Briefly, three commercial kits were used: Wizard Screens 1 and 2 (Emerald BioStructures), Structure Screens 1 and 2 and Stura Footprint screen (Molecular Dimensions Ltd). The crystals were obtained in 2.0 *M* ammonium sulfate, 0.1 *M* phosphate–citrate pH 4.2 and with a protein concentration of 5.8 mg ml^{-1} in the gel-filtration buffer. The optimization of the crystallogenesis was performed with nanodrops in a two-dimensional matrix (Lartigue *et al.*, 2003) with a precipitant range of 1.8–2.2 *M* ammonium sulfate and a pH range of 4.0–4.5 (0.1 *M* phosphate–citrate), leading to a crystal size of ~100 × 100 × 80 µm (Fig. 2).

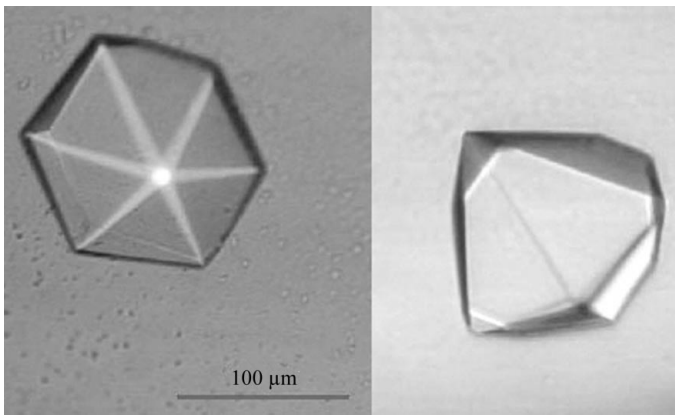

Figure 2
Optimized crystals of the SARS-CoV Nsp9 protein. The scale bar is 100 µm.

Table 1
Crystal parameters and data-reduction statistics of the Nsp9 protein crystals.

Values in parentheses are for the last resolution shell.

Space group	$P6_{1/5}22$
Unit-cell parameters (Å)	$a = b = 89.7$, $c = 136.7$
Beamline	ID14-EH1 at ESRF (λ = 0.934 Å)
Resolution (Å)	26.0–2.8 (2.94–2.8)
R_{sym} (%)	5.3 (28.1)
$I/\sigma(I)$	9.9 (2.5)
No. reflections	90899 (11486)
No. unique reflections	8395 (1166)
Completeness	98.7 (98.7)
Multiplicity	10.8 (9.9)

2.5. Data collection

The crystals were cryocooled in a pure solution of silicone oil DC200. They were exposed at beamline ID14-EH1, ESRF, Grenoble using a Quantum ADSC Q4R detector. A total of 110 1° oscillations were recorded with a crystal-to-detector distance of 180 mm and a collection time of 9 s per frame. Diffraction data were integrated with *DENZO* (Otwinowski & Minor, 1997) and were reduced with *SCALA* (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

3.1. *E. coli* protein expression and purification

We have subcloned 35 SARS-CoV targets in the Gateway system, including 20 full-length proteins and 15 protein domains. To date, 70 constructs have been generated, of which 28 were expressed, 14 were soluble and five were purified. Four of them led to small crystals, among which were those of the Nsp9 protein described in this report. Expression of selenomethionine-substituted Nsp9 was performed using the method of methionine-biosynthesis pathway inhibition (Doublié, 1997). Purification of the selenomethionine protein was performed as described above and crystal optimization is under way.

3.2. Data collection and reduction

Nsp9 crystals diffract to 2.8 Å at ID14-EH1 (ESRF, Grenoble). Data integration and reduction indicate that they belong to the *P*622 space group. R_{sym} is 5.3%, an excellent value considering the redundancy of the data (Table 1). Reflections are observed at multiples of six along the *c* axis

(00*l*), indicating that the space group is either $P6_122$ or its enantiomorph $P6_522$. The unit-cell parameters are $a = b = 89.7$, $c = 136.7$ Å, which lead to a V_M value of 3.1 Å^3 Da^{-1} (60% solvent) with two molecules in the asymmetric unit (Matthews, 1968). The observed distribution of centric or acentric intensities overlaps with the theoretical curve, an indication that merohedral twinning, a feature that is often observed in trigonal or hexagonal crystals, is not present.

3.3. Characterization

SARS-CoV Nsp9 has been purified to homogeneity in two steps. The identity of the final product has been confirmed by N-terminal sequencing. The oligomeric status of Nsp9 has been checked using gel filtration and DLS. The former technique indicates that the protein is monomeric, while the DLS analysis is consistent with a monodisperse species with an apparent Stokes radius of 26 Å and an equivalent mass of 31 kDa, which corresponds to a dimer. This discrepancy might be related to the concentration differences between the two techniques.

A PSI-Blast search retrieved seven homologous sequences, all belonging to members of the *Coronaviridae* family. They were aligned using *MULTALIGN* (Corpet, 1988) with standard options. The consensus of the secondary-structure predictions obtained with *JPRED* (Cuff *et al.*, 1998), *PSI-PRED* (McGuffin *et al.*, 2000) and *PREDICT PROTEIN* (Rost, 1996) converges to a fold of seven β-strands. A fold-recognition analysis was performed with the threading programs *3D-PSSM* (Kelley *et al.*, 2000) and *INBGU* (Fischer, 2000). Both programs fail to detect any protein homologue to Nsp9, but converge to a fold of two seven-stranded β-sheets. In agreement, the CD spectrum of purified Nsp9 reveals a structured protein formed by a majority of β-strands (35%) and β-turns (18%), but which also contains 15% α-helix. Random-coil segments account for 32% of the total.

4. Conclusion

The SARC-CoV Nsp9 protein expressed in *E. coli* was readily crystallized using the nanodrop screening (Sulzenbacher *et al.*, 2002) and optimization (Lartigue *et al.*, 2003) approaches. Crystals diffract to 2.8 Å resolution and are amenable to structure determination using SeMet substitution and MAD methods (Hendrickson, 1991) at synchrotrons.

This study was funded by the SPINE project of the European Union 6th PCRDT (QLRT-2001-00988), by the French Genopole programme and by the Conseil General of the Bouches-du-Rhone. We thank H. W. Doerr and H. Rabenau (Institute for Medical Virology, Johan Wolfgang Goethe University, Frankfurt-am-Main, Germany) for providing us with the virus and P. Bredenbeek, S. Gorbalenya and W. Spaan for technical assistance and helpful discussions/suggestions.

References

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. (2003). *Science*, **300**, 1763–1767.
Bost, A. G., Carnahan, R. H., Lu, X. T. & Denison, M. R. (2000). *J. Virol.* **74**, 3379–3387.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Corpet, F. (1988). *Nucleic Acids Res.* **16**, 10881–10890.
Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. & Barton, G. J. (1998). *Bioinformatics*, **14,** 892–893.
De Francesco, R., Tomei, L., Altamura, S., Summa, V. & Migliaccio, G. (2003). *Antivir. Res.* **58**, 1–16.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Drosten, C. *et al.* (2003). *N. Engl. J. Med.* **348**, 1967–1976.
Fischer, D. (2000). *Pac. Symp. Biocomput.* **5**, 119–130.
Gong, C. & Shuman, S. (2002). *J. Biol. Chem.* **277**, 15317–24.
Grotzinger, C., Heusipp, G., Ziebuhr, J., Harms, U., Suss, J. & Siddell, S. G. (1996). *Virology*, **222**, 227–235.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. (2000). *J. Mol. Biol.* **299**, 499–520.
Koren, G., King, S., Knowles, S. & Phillips, E. (2003). *CMAJ*, **168**, 1289–1292.
Ksiazek, T. G. *et al.* (2003). *N. Engl. J. Med.* **348**, 1953–1966.
Lai, M. M.C. & Holmes, K. V. (2001). *Fields Virology*, 4th ed., edited by D. M. Knipe & P. M. Howley, pp. 1163–1185. Philadelphia: Lippincott Williams & Wilkins.
Lartigue, A., Rivière, S., Brossut, R., Tegoni, M. & Cambillau, C. (2003). *Acta Cryst.* D**59**, 916–918.
McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). *Bioinformatics*, **16**, 404–405.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Peiris, J. S., Lai, S. T., Poon, L. L., Guan, Y., Yam, L. Y., Lim, W., Nicholls, J., Yee, W. K., Yan, W. W., Cheung, M. T., Cheng, V. C., Chan, K. H., Tsang, D. N., Yung, R. W., Ng, T. K. & Yuen, K. Y. (2003). *Lancet*, **361**, 1319–1325.
Rost, B. (1996). *Methods Enzymol.* **266**, 525–539.
Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L. M., Guan, Y., Rozanov, M., Spaan, W. J. M. & Gorbalenya, A. E. (2003). In the press.
Sulzenbacher, G. *et al.* (2002). *Acta Cryst.* D**58**, 2109–2115.
Vincentelli, R., Bignon, C., Gruez, A., Sulzenbacher, G., Canaan, S., Tegoni, M., Campanacci, V. & Cambillau, C. (2003). *Acc. Chem. Res.* **36**, 165–172.