ELSEVIER

Expression, purification and characterization of recombinant severe acute respiratory syndrome coronavirus non-structural protein 1

Kimberly Brucz [a], Zachary J. Miknis [b], L. Wayne Schultz [a,b], Timothy C. Umland [a,b,*]

[a] *Hauptman-Woodward Medical Research Institute, 700 Ellicott Street, Buffalo, NY 14203, USA*
[b] *Department of Structural Biology, SUNY at Buffalo, Buffalo, NY 14203, USA*

Abstract

The coronavirus (CoV) responsible for severe acute respiratory syndrome (SARS), SARS-CoV, encodes two large polyproteins (pp1a and pp1ab) that are processed by two viral proteases to yield mature non-structural proteins (nsps). Many of these nsps have essential roles in viral replication, but several have no assigned function and possess amino acid sequences that are unique to the CoV family. One such protein is SARS-CoV nsp1, which is processed from the N-terminus of both pp1a and pp1ab. The mature SARS-CoV protein is present in cells several hours post-infection and co-localizes to the viral replication complex, but its function in the viral life cycle remains unknown. Furthermore, nsp1 sequences are highly divergent across the CoV family, and it has been suggested that this is due to nsp1 possessing a function specific to viral interactions with its host cell or acting as a host specific virulence factor. In order to initiate structural and biophysical studies of SARS-CoV nsp1, a recombinant expression system and a purification protocol have been developed, yielding milligram quantities of highly purified SARS-CoV nsp1. The purified protein was characterized using circular dichroism, size exclusion chromatography, and multi-angle light scattering.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Severe acute respiratory syndrome; SARS; Coronavirus; Non-structural protein 1; Nspl; Size exclusion chromatography–multi-angle light scattering

The severe acute respiratory syndrome (SARS)[1] outbreak of 2002–2003, followed by a much smaller outbreak in 2004, caused over 8000 illnesses and nearly 800 deaths (World Health Organization; http://www.who.int/csr/sars/country/table2004_04_21/en/index.html). The infectious agent responsible for this disease was quickly identified as a new member of the coronavirus (CoV) family, SARS-coronavirus (SARS-CoV) [1–3], most closely related to the group 2 CoVs [4]. This newly emerged virus prompted a renewed interest in CoV research. Prior to the SARS outbreak, only two CoVs (HCoV-229E and HCoV-OC43) were known to infect humans [5]. These two CoVs have been estimated to cause up to 30% of common colds and mild respiratory illnesses [6]. Other CoVs are widespread in both domestic and wild animals, with several posing significant economic impact on livestock and poultry industries.

Following the emergence of SARS, two additional human CoVs associated with upper and lower respiratory tract diseases were identified. Three groups independently identified in young children what is likely a single CoV species, and this new CoV has been variously designated NL63, NL, and HCoV-NH [7–9]. The second new CoV was discovered in an elderly patient suffering from pneumonia in Hong Kong and has been designated HKU1 [10]. Both of the newly identified human CoVs appear to be

---

* Corresponding author. Fax: +1 716 898 8660.
 E-mail address: umland@hwi.buffalo.edu (T.C. Umland).
[1] *Abbreviations used:* CD, circular dichroism; CoV, coronavirus; DTT, dithiothreitol; GST, glutathione-*S*-transferase; IMAC, immobilized metal affinity chromatography; IPTG, isopropyl-β-D-thiogalactopyranoside; LB, Luria Broth; MALS, multi-angle light scattering; MHV, mouse hepatitis virus; nsp, non-structural protein; ORF, open reading frames; PCR, polymerase chain reaction; PL2^{pro}, papain-like protease 2; RT-PCR, reverse transcriptase-polymerase chain reaction; SARS, severe acute respiratory syndrome; SEC, size exclusion chromatography; Snsp1, SARS-CoV nsp1; TB, Terrific Broth.

widespread, especially in children, and have likely been present in a human host reservoir for an extended time.

SARS-CoV has not re-emerged since 2004, but the natural reservoir of the virus has been putatively identified in several related species of Chinese horseshoe bats [11,12]. These bats are sold in Chinese live animal markets and used in traditional Chinese medicine, and thus re-emergence of SARS is a distinct possibility. Most members of the CoV family exist in a narrow range of host species, specific for each virus. SARS-CoV and HCoV-OC43 are notable examples of CoVs having documented host range expansions from their original animal reservoir (bats [11,12] and bovines [13], respectively), acquiring the ability to infect and be transmitted between humans.

The CoV family possesses the largest RNA genome (28–30 kb) of all RNA viruses. Their positive strand RNA genome exhibits a common organization throughout the family. Two overlapping open reading frames (ORF1a and ORF1ab) are present at the genome's 5′ end, which encode two large polyproteins (pp1a and pp1ab) [14]. Viral proteases process pp1a and pp1ab to yield the mature viral nonstructural proteins (nsps, nsp1–nsp16 in SARS-CoV) [15,16]. Many of these nsps have been associated with viral replication [17], and so are also referred to as replicase proteins. Several of these nsps are unique to the CoV family or to individual CoVs. For example, SARS-CoV nsp1 exhibits weak sequence similarity only to the nsp1s of several other group 2 CoVs [4]. A lack of functional data exists for nsp1 of all CoVs, due at least in part to the absence of homology to a well-characterized protein family and its high variability across the CoV family.

SARS-CoV nsps 1, 2, and 3 are processed by a papain-like protease (PL2^{pro}) contained within nsp3 [15,16]. Both SARS-CoV pp1a and pp1ab contain nsp1 at their respective N-termini. These two studies demonstrated that viral PL2^{pro} cleaves both polyproteins at $_{180}$G↓A$_{181}$, producing the mature SARS-CoV nsp1 containing 180 amino acid residues with a calculated mass of 19.6 kDa. Furthermore, the mature SARS-CoV nsp1 was shown to be present in Vero cells at 4–6 h post-infection. SARS-CoV nsp1 was not detected as a component of a larger partially processed polyprotein intermediate within lysates from the infected cells, indicating that proteolytic processing at the nsp1–nsp2 cleavage site occurs rapidly following synthesis of pp1a and pp1ab. Immunofluorescence experiments revealed that SARS-CoV nsp1 co-localized with other replicase proteins into discrete cytoplasmic foci that were both perinuclear and dispersed throughout the cytoplasm [15,16]. These cytoplasmic foci likely represent SARS-CoV replication complexes, where viral RNA synthesis occurs. These replication complexes form on double-membrane vesicles, with the vesicles likely constructed through viral manipulation of the cellular autophagy system [18,19]. Later in infection, a more diffuse distribution of SARS-CoV nsp1 was observed, possibly indicative of a change in localization during the viral life cycle, or degradation of previously formed foci. Virus release from infected Vero

cells occurred 3 to 6 h after the initial observation of the presence of nsp1 [16].

SARS-CoV nsp1 possesses weak sequence homology with mouse hepatitis virus (MHV) nsp1 [4], although the mature MHV protein is 8 kDa larger. While comparisons of data regarding nsp1 from MHV and SARS-CoV must be conducted with caution due to significant sequence differences, results from MHV nsp1 studies suggest an essential role for SARS-CoV nsp1. The N-terminal half of MHV nsp1 has been shown to be essential to produce an infectious virus, and point mutations within this region produced virus with altered replication and RNA synthesis [20]. The MHV nsp1 C-terminal half can be deleted or the nsp1–nsp2 cleavage site eliminated, and both mutations will yield viable virus but with delayed replication and lowered peak titers [20,21].

The cellular co-localization of SARS-CoV nsp1 with other viral nsps known to be essential for viral RNA synthesis and viral replication indicates that nsp1 may also have a role in these steps of the viral life cycle. The high sequence variability of nsp1 across the CoV family combined with the tendency for individual members of this family to possess a narrow host species range suggests that nsp1 may have specific host interactions, including suppression of host gene expression [22]. SARS-CoV is known to have expanded its host range from its natural reservoir (bats) to other animals present in live animal markets (e.g., palm civets, raccoon dogs) and to humans. Hence, it may be possible that mutations in nsp1 were involved in the evolution of the SARS-CoV host range. In order to further investigate these possible functions of SARS-CoV nsp1, we have undertaken the expression of recombinant protein in *Escherichia coli*, the purification to homogeneity, and the characterization of this protein. In particular, structural studies require large quantities of highly purified and monodisperse protein samples, and the expression and purification experiments described here were conducted with those goals in mind.

Materials and methods

Cloning of SARS-CoV nsp1

The template for subcloning SARS-CoV nsp1 into an expression vector was a cDNA fragment that encoded nsp1 and a portion of nsp2. This cDNA fragment was generated by reverse transcriptase-polymerase chain reaction (RT-PCR) from SARS-CoV Urbani strain genomic RNA, and was a kind gift from Dr. Mark Denison (Vanderbilt University). The cDNA encoding only nsp1, corresponding to bases 265–804 of the SARS-CoV Urbani strain genome (GenBank Accession No. AY278741), was amplified by polymerase chain reaction (PCR) using the forward primer 5′-ATG GAG AGC CTT GTT CTT GGT G-3′, the reverse primer 5′-TTA ACC TCC ATT GAG CTC ACG AG-3′ and Taq PCR Master Mix (Qiagen). The reverse primer was designed to introduce a STOP codon (TAA) at

the 3′ end of the sense strand of the pcr product. The PCR was conducted in a standard manner, employing 30 cycles and an annealing temperature of 60 °C with an Applied Biosystems Gene Amp PCR System 2400. The PCR amplified fragment was inserted into a modified pET15b expression vector diagramed in Fig. 1a (Yanzhou Wang, unpublished results). The stock pET15b vector (Novagen) encodes a N-terminal (His)$_6$ fusion tag followed by a thrombin cleavage site and a multiple cloning site (MCS) encoding several unique restriction enzyme cleavage sites. The modified vector (Topo-HisGST-YZW) replaced the stock fusion tag, thrombin cleavage site and MCS with a N-terminal (His)$_6$-glutathione-*S*-transferase (GST) fusion tag followed by a tobacco etch virus (TEV) protease cleavage site. This vector was then linearized at the unique *Xho*I site, and adaptors added to both ends to provide *vaccinia* topoisomerase recognition sequences.

The PCR amplified product, with 3′ adenine overhangs due to *Taq* polymerase activity, was incubated with the linear Topo-HisGST-YZW plus topoisomerase at 22 °C for 15 min. DH5α library efficient competent cells (50 μl, Invitrogen) were transformed via heat shock with 3 μl of the ligation reaction, and then plated onto LB agar plates containing 100 μg/mL ampicillin and incubated overnight at 37 °C. Nascent colonies were preliminary screened by PCR with SARS-CoV nsp1 forward primer and T7 terminator primer (Novagen), and the PCR products were analyzed by agarose gel electrophoresis to identify transformants possessing an expression vector with a DNA insert of correct size and directionality. Plasmids from colonies identified by this initial screen were purified using the QIAprep Spin Miniprep kit (Qiagen). Purified plasmids (pHisGST-TEV-Snsp1) were sequenced using T7 promoter and terminator primers on an ABI PRISM 3130XL Genetic Analyzer to verify incorporation of cDNA encoding full-length SARS-CoV nsp1 into the vector.

SARS-CoV nsp1 expression

Several *E. coli* host strains [BL21(DE3), HMS174(DE3), Rosetta(DE3) (Novagen) and BL21 Star(DE3) (Invitrogen)] were transformed with the verified pHisGST-TEV-Snsp1 expression vector for expression and solubility assays. Small scale expression cultures were grown from these transformed cells, testing media (Terrific Broth, TB; Luria Broth, LB), temperature (37 or 22 °C), and time post-induction. All cultures included ampicillin (50 μg/mL), and were induced using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG, Inalco) at an OD$_{600}$ of 0.5–1.0. The Rosetta(DE3) cultures also included chloramphenicol (34 μg/mL). An aliquot of each culture was lysed and then fractionated into supernatant and insoluble pellet. Both fractions from each culture were analyzed by SDS–PAGE.

Large scale expression occurred using transformed Rosetta(DE3) cells grown in TB, in the presence of chloramphenicol (34 μg/mL) and ampicillin (50 μg/mL). One liter cultures were grown to an OD$_{600}$ of ∼0.65 and then were induced with 1 mM IPTG. A constant temperature of 37 °C at 260 rpm in a New Brunswick Scientific I250KC incubated shaker was maintained during growth and induction. Cells were harvested 3 h post-induction by centrifugation at 6000g (Beckman Coulter Avanti Centifuge, J-20 XPI). The supernatant was decanted and cell pellets scraped into a sterile 50 mL falcon tube for immediate storage at −80 °C.

Recombinant SARS-CoV nsp1 purification

Sample preparation for recombinant SARS-CoV nsp1 isolation began by thawing frozen cell pellets, harvested from 6 × 1 L cultures, in a 22 °C water bath. The thawed pellets were resuspended in a total of 80 mL of lysis buffer (50 mM Hepes, pH 7.5, 250 mM NaCl, 1 mM β-mercaptoethanol and 10 mM imidazole), supplemented with 3 mL Protease Inhibitor Cocktail (Sigma, P2714) prepared according to the manufacturer's protocol. Cells were lysed using a single 15,000–18,000 psi pass through a Microfluidizer Processor M-110EH (Microfluidics), and the lysate was fractionated by high-speed centrifugation at 97,272g (Beckman L-60 Ultracentrifuge; 45Ti rotor; 30,000 RPM). The soluble fraction containing SARS-CoV nsp1 was
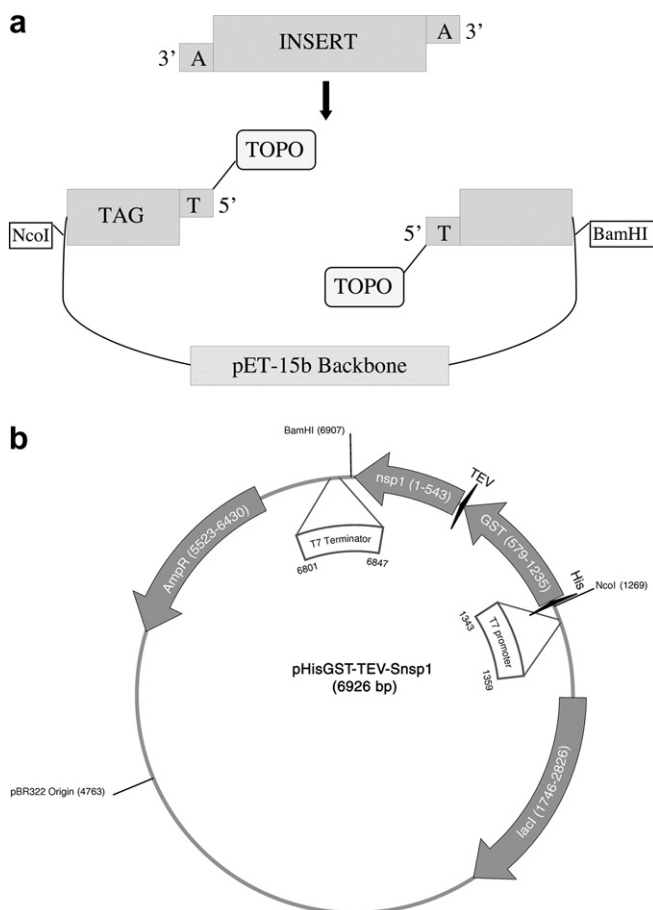

Fig. 1. Expression vector construction. (a) Schematic of the Topo-HisGST-YZW vector. (b) Schematic of the pHisGST-TEV-Snsp1 expression vector coding for SARS-CoV nsp1 with an N-terminal polyhistidine-GST dual affinity tag and a TEV protease cleavage site.

identified by SDS–PAGE and protein immunoblot (Western blot) employing an anti-polyhistidine antibody and the Western Breeze Chromogenic Kit (Invitrogen).

The lysate supernatant was filtered through a 0.45 μM membrane, and then the clarified sample was loaded onto a 5 mL HisTrap immobilized metal affinity chromatography (IMAC) column (Amersham) equilibrated with binding buffer (50 mM Hepes, pH 7.5, 250 mM NaCl, 1 mM β-mercaptoethanol and 15 mM imidazole). Chromatography steps were performed on an AKTA FPLC system (Amersham) unless otherwise noted. Following sample load, the column was subjected to two wash steps to remove weakly bound contaminants using binding buffer in which the imidazole concentration was raised to 25 mM and then to 40 mM. The HisGST-TEV-Snsp1 fusion protein was eluted over a 20 column volume imidazole gradient (15–300 mM) with a final step to 500 mM imidazole to strip the column of any remaining protein. Throughout purification, the purity of the sample was analyzed by Coomassie Brilliant Blue-stained SDS–PAGE and the percentage of the total sample comprised of recombinant SARS-CoV nsp1 was estimated by the densitometry feature of the AlphaImager HP gel imaging system.

Cleavage of fusion protein

TEV protease with an N-terminal polyhistidine affinity tag was added to the IMAC purified HisGST-TEV-Snsp1 sample in a 1:50 mass ratio to cleave the fusion protein and simultaneously the sample was dialyzed into Buffer A (50 mM Hepes, pH 7.5, 250 mM NaCl and 1 mM β-mercaptoethanol) overnight at 4 °C. The TEV protease treated SARS-CoV nsp1 was separated from the protease, the cleaved HisGST affinity tag and any remaining uncleaved HisGST-Snsp1 fusion protein by loading the sample onto a second 5 mL HisTrap IMAC column equilibrated with Buffer A. Imidazole gradients were created using Buffer B, which was identical to Buffer A with the addition of 500 mM imidazole. Following sample loading onto the column, a step gradient to was applied to raise the imidazole concentration to 25 mM, and the column was washed to elute the SARS-CoV now lacking the affinity tag. A final sharp linear gradient to raise the imidazole concentration to 500 mM was performed to elute the species possessing a polyhistidine tag (e.g., TEV protease and the cleaved HisGST tag). The SARS-CoV nsp1 containing fractions were pooled and concentrated to 18.3 mg/mL by an Amicon Ultra-15 Centrifugal Filter Unit with a 5000 MWCO membrane (Millipore) using centrifugation at 2000*g* at 4 °C. The protein was prepared for size exclusion chromatography (SEC) by dialyzing overnight at 4 °C against SEC buffer [25 mM Hepes, pH 7.5, 150 mM NaCl, 1 mM EDTA and 5 mM dithiothreitol (DTT)].

Size exclusion chromatography and multi-angle light scattering

A Superdex 200 HL 16/60 SEC column (Amersham) was employed as a final polishing purification step to remove aggregated protein and low molecular weight contaminants. The column was equilibrated against SEC buffer. Preparative SEC was run at 1.0 mL/min at 4 °C. The resulting fractions were analyzed by SDS–PAGE and those containing pure SARS-CoV nsp1 were pooled. Concentration was performed as required for additional experiments.

A Superdex 200 HR 10/30 SEC column (Amersham) was used to estimate the molecular mass and oligomeric state of the purified SARS-CoV nsp1. This column was equilibrated with SEC buffer and run at 0.5 mL/min at 4 °C. A calibration curve for molecular size estimation was generated by individually loading blue dextran 2000, bovine serum albumin (BSA), chymotrypsinogen A, and aprotinin onto this analytical SEC column and eluting under similar conditions. These data were input into Unicorn v.5.0.1 (Amersham) to calculate a retention volume vs. molecular weight calibration curve.

Size exclusion chromatography coupled with multi-angle light scattering (SEC–MALS) experiments employed the same analytical SEC column installed on an AKTA Purifier modified to include differential refractive index and multi-angle light scattering (MALS) detectors (Optilab DSP (Wyatt) and miniDAWN (Wyatt), respectively) downstream of the Purifier's standard UV flow cell. The system was extensively equilibrated with SEC buffer at 0.5 mL/min at 4 °C. Purified SARS-CoV nsp1 (200 ul @ 3 mg/mL) was injected onto the column and eluted at 0.5 mL/min at 4 °C. ASTRA software (Wyatt) was used to evaluate the MALS data.

Circular dichroism

SARS-CoV nsp1 was concentrated to 5 mg/mL and dialyzed against 10 mM phosphate buffer, pH 7.5, composed of 0.26 g monosodium phosphate monohydrate and 2.2 g disodium phosphate heptahydrate per 1 L in preparation for circular dichroism (CD) analysis on a Jasco Spectropolarimeter Model J-715. SARS-CoV nsp1 dilutions of 1:3, 1:4, 1:5, 1:10, 1:50, 1:100, and 1:150 were measured in one of three reference cells (1.0 cm, 1.0 mm and 0.1 mm) at 20 °C to determine optimal conditions. Standard Analysis (Jasco) program files were extracted and further analyzed using the k2d web server (www.embl-heidelberg.de/~andrade/k2d/) to estimate secondary structure composition [23]. Mean residue ellipticity ($[\theta]$ expressed in deg × cm^2/dmol) was calculated using $[\theta] = \theta \times 100 \times M_r/(c \times l \times N_A)$, where θ is the experimental ellipticity in mdeg, M_r is the protein's molecular weight in Daltons, c is protein concentration in mg/mL, l is the cuvette path length in cm and N_A is the number of residues in the protein.

Secondary structure was predicted based on the recombinantly expressed SARS-CoV nsp1 amino acid sequence (post-affinity tag cleavage) using SCRATCH [24], PSI-PREP [25], PROFsec [26], Sable-2 [27], and Predator [28]. A consensus secondary structure prediction was made based upon these individual prediction results and used to compare to the secondary structure content measured experimentally by CD.

Results

Construction of SARS-CoV nsp1 bacterial expression plasmid

An *E. coli* vector (Topo-HisGST-YZW, Yanzhou Wang, unpublished) was employed to construct an expression vector to produce full-length SARS-CoV nsp1. The backbone of the plasmid is based on pET-15b, retaining the advantages of this pET vector (i.e., the powerful but stringent T7*lac* promoter, ampicillin resistance) while introducing a dual N-terminal affinity tag (polyhistidine and GST), a highly specific TEV-protease cleavage site, and topoisomerase ligation (Fig. 1b). All modifications to the pET-15b plasmid are between the unique *Bam*HI and the *Nco*I sites.

The SARS-CoV nsp1 cDNA fragment was produced by PCR, using a template that was generated by RT-PCR from the 5′ end of the viral RNA genome. A STOP codon was introduced during the PCR step, as the template cDNA did not contain a STOP codon immediately 3′ to the nsp1 coding sequence because the wild-type SARS-CoV nsp1 is proteolytically processed from the N-terminal end of the large pp1a and pp1ab polyproteins (486 and 790 kDa, respectively). The PCR amplified cDNA was 543-nt long, plus 3′ adenine overhangs due to the use of *Taq* polymerase in the PCR. The overhangs are required for topoisomerase TA cloning. The completeness of the pHis-GST-TEV-Snsp1 expression vector was confirmed by DNA sequencing.

Expression and purification of the fusion protein

The optimal expression condition for the SARS-CoV nsp1 fusion protein in 5 mL cultures was determined to employ the Rosetta(DE3) *E. coli* strain in TB media containing ampicillin and chloramphenicol, with growth and expression occurring at 37 °C, and harvesting occurring 3 h post-induction. Sufficient soluble SARS-CoV nsp1 expressed as a His-GST fusion protein was present for the desired structural and biophysical studies that additional expression optimization was not required.

Expression was easily scaled up to 1 L cultures grown in 2.8 L Fernbach flasks. The initial step of fusion protein purification was IMAC employing a nickel-charged col-

umn. The thawed and resuspended pellets were lysed using a Microfluidizer. The Microfluidizer not only efficiently lyses the cells in a single run, but the resulting supernatant's viscosity was lower than that obtained by other methods (e.g., sonication) allowing for easier sample loading onto the IMAC column. The SARS-CoV nsp1 fusion protein was eluted as a single but somewhat broad peak by a linear gradient of increasing imidazole concentration. Fractions were pooled based upon protein purity, as judged by Coomassie Brilliant Blue stained SDS-PAGE (Fig. 2a). The purity of the pooled fusion protein, post-IMAC, was 80% (Table 1), with a single major band running at a molecular weight of ~50 kDa, as expected for the fusion protein (SARS-CoV nsp1 at 19.6 kDa + HisGST-TEV fusion tag at 28.1 kDa). Two major contaminants running at ~30 kDa
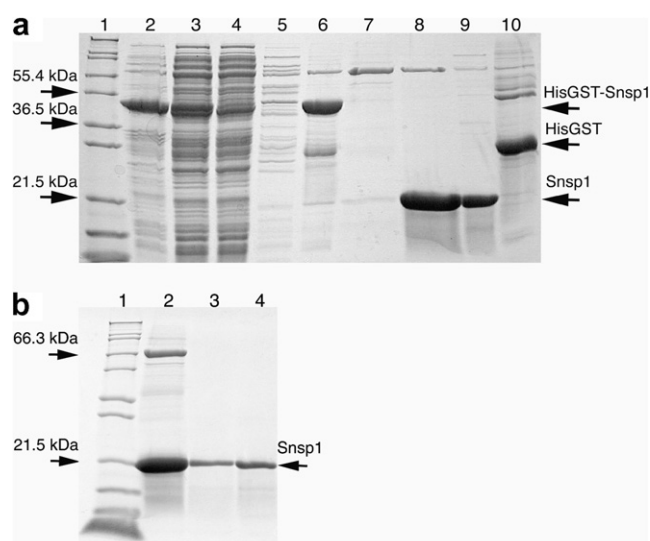

Fig. 2. Coomassie stained SDS–PAGE analysis of SARS-CoV nsp1 at various stages of purification. (a) Gel displaying fractions from the initial IMAC purification (IMAC #1) and the second IMAC following cleavage of the dual affinity tag (IMAC #2). Lane 1, Mark12 molecular weight marker (Invitrogen); lane 2, total cell lysate; lane 3, lysate supernatant; lane 4, IMAC #1 flow-through; lane 5, IMAC #1 wash; lane 6, IMAC #1 HisGST-Snsp1 elution; lanes 7–9, IMAC #2 flow-through fractions containing SARS-CoV nsp1; lane 10, IMAC #2 elution of cleaved dual affinity tag. (b) Gel analysis of preparative SEC. Lane 1, Mark12 molecular weight marker (Invitrogen); lane 2, sample loaded onto SEC column; lanes 3–4, pooled SARS-CoV nsp1 eluted from SEC loaded onto the gel at 3 and 6 μg, respectively.

Table 1
Yield of recombinant SARS-CoV nsp1 purified from *E. coli*

Step	Total protein (mg)[a]	Purity (%)[c]	His-GST-nsp1 (mg)	SARS-CoV nsp1 (mg)
Lysate (soluble)[b]	1080	28[d]	300	–
IMAC #1	75	80[d]	60	–
Tag cleavage + IMAC #2	30	80[e]	0	24
SEC	21	99[e]	0	21

[a] Estimated by Bradford assay; fraction containing SARS-CoV nsp1.
[b] From 6 L culture.
[c] Estimated from densitometry on Coomassie-stained SDS–PAGE gels.
[d] Purity of the His-GST-nsp1 fusion protein.
[e] Purity of SARS-CoV nsp1 post-affinity tag cleavage.

Table 2
Molecular weight estimation by SEC

Sample	Type[a]	Mol. weight (kDa)[b]	Retention vol. (mL)	Est. mol. weight (kDa)
Blue Dextran 2000	S	~2000	7.33	–
BSA (dimer)	S	132.6	11.44	–
BSA (monomer)	S	66.3	13.12	–
Chymotrypsinogen A	S	25.0	15.79	–
Aprotinin	S	6.5	17.70	–
SARS-CoV nsp1	U	20.2	14.45	37.2

Superdex 200 HR 10/30 on AKTA purifier calibration: retention vol. (mL) $= A \times \log(\text{MW}) + B$; MW in kDa $A = -4.291$, $B = 21.19$, correlation $= -0.9928$.

[a] Type: standard (S) or unknown (U).
[b] MW calculated from amino acid sequence.

and at ~65 kDa plus several minor contaminants were also observed.

TEV protease cleavage of fusion protein

The HisGST dual affinity fusion tag was cleaved from SARS-CoV nsp1 using TEV protease possessing its own $(\text{His})_6$ affinity tag. This protease retains a useful level of activity over a wide range of buffer conditions and temperature. Thus, it is possible to perform the TEV protease cleavage in conjunction with a dialysis step to remove the imidazole, rather than performing cleavage and dialysis separately. Cleavage was nearly complete following incubation overnight at 4 °C. The cleaved fusion tag, TEV protease, and any remaining uncleaved fusion protein was separated from the now tagless SARS-CoV nsp1 using a second IMAC column. Several of the minor contaminants, presumably *E. coli* proteins that co-eluted with the fusion protein on the initial IMAC run, were resolved from the cleaved SARS-CoV nsp1 during this step. A single major contaminant running at ~65 kDa on SDS–PAGE remained.

Preparative size exclusion chromatography

Preparative scale SEC was used as the final purification step. Prior to SEC, the SARS-nsp1 sample was concentrated to minimize the volume applied to the SEC column in order to enhance resolution. The concentrated sample was stable in the SEC buffer and could be stored at 4 °C for several days with no observed precipitation or degradation. The SEC elution profile included a small early eluting peak corresponding to the high molecular weight contaminant, and a single large well-formed peak corresponding to SARS-CoV nsp1. Coomassie Brilliant Blue stained SDS–PAGE analysis (Fig. 2b) indicated that the SEC purified SARS-CoV nsp1 was 99% pure, and the major contaminant at ~65 kDa was removed. The protocol yielded 21 mg of purified protein (3.5 mg per 1 L culture).

Estimating molecular weight and oligomerization state

The molecular weight and oligomeric state of the SEC purified SARS-nsp1 in its native, soluble state was esti-

mated by two methods: standard analytical SEC with a calibration curve derived from well-behaved protein standards and SEC–MALS. The same Superdex 200 HR 10/30 column and the same SEC buffer was used in both techniques, and the protein eluted as a single well-formed peak in all runs. For the standard SEC size estimation, the purified SARS-CoV nsp1 reproducibly eluted at 14.45 mL, corresponding to a molecular weight estimate of 37.2 kDa (Table 2). The calculated molecular weight of the recombinantly expressed SARS-CoV nsp1 is 20.2 kDa, including six vector-derived N-terminal amino acid residues (GSLDAL) remaining post-cleavage. Thus, the molecular weight of 37.2 kDa estimated by SEC implies that the SARS-CoV nsp1 is present as a dimer (37.2/20.2 kDa = 1.84) in solution. The preparative scale SEC column was also calibrated using protein standards, and the molecular weight estimated from the results of this larger SEC column confirmed the analytical SEC results (data not shown). The molecular weight estimated by SEC–MALS was 19.6 kDa. However, the protein eluted from the column at a similar elution volume as in the standard SEC run. These data implies that the SARS-CoV nsp1 are present as a monomer (19.6/20.2 kDa = 0.97) in solution, in contrast to the standard SEC results. The discrepancy in these results will be discussed below.

Circular dichroism

The purified SARS-CoV nsp1 was subjected to CD analysis to experimentally determine the protein's secondary structure composition. A 0.1 mm path length cell and a minimal phosphate buffer were used to minimize buffer effects upon the measured spectrum. The secondary structure composition estimated from the CD spectrum was 28% α-helix, 33% β-strand, and 39% random coil. By comparison, the consensus predicted secondary structure composition based upon the amino acid sequence alone was 26% α-helix, 24% β-strands, and 50% random coil. See Fig. 3.

Discussion

SARS-CoV nsp1 has been successfully produced in a recombinant *E. coli* expression system, meeting the goal of producing milligram quantities of highly purified protein
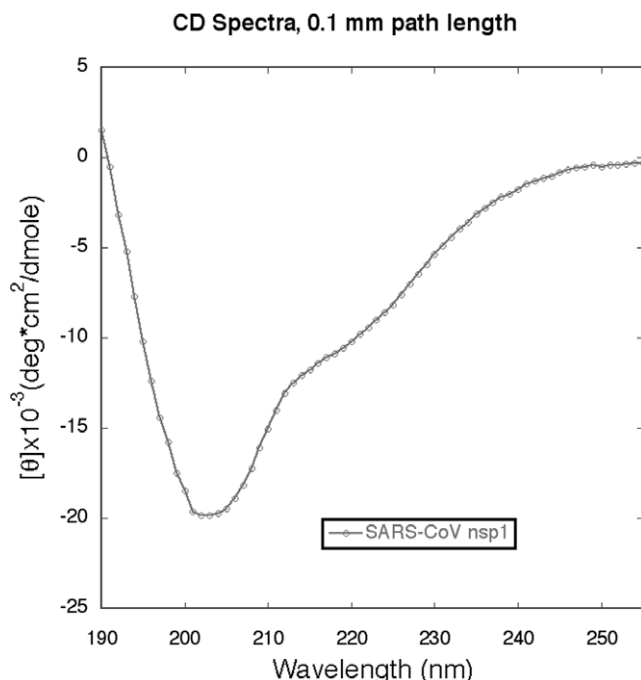
Fig. 3. CD spectra of SARS-CoV nsp1 at 1.0 mg/mL using a 0.1 mm path length cell.

for structural and biophysical study. The purified sample was stable in solution at concentrations of 1 mg/mL to ⩾ 18 mg/mL, was present as a single well-defined oligomeric state, and possessed a significant amount of secondary structure.

The use of an expression vector encoding a dual affinity tag (polyhistidine and GST) allows for the possibility of purification by two orthogonal affinity methods. It is not unusual for a small number of native *E. coli* proteins to co-elute with a recombinantly expressed fusion protein containing a single affinity tag purified on the appropriate affinity resin; whereas, it is unusual for a native *E. coli* protein to effectively bind to both IMAC and glutathione resins. We have successfully used this dual tag/dual affinity column procedure to highly purify a number of proteins, and in at least one instance (mouse HoxA5 homeodomain) the presence of the dual affinity tag dramatically increased expression compared to the comparable fusion protein possessing only a polyhistidine tag (Umland, unpublished data). The Topo-HisGST-pET15bTEV was chosen for expression of recombinant SARS-CoV nsp1 for these reasons. Upon development of the purification protocol, it was found that sufficiently high purity was obtained by IMAC, followed by removal of high molecular weight aggregates by SEC. However, the presence of the GST portion of the affinity tag provides options for future purifications, if required.

Both the fusion protein and SARS-CoV nsp1 following removal of the affinity tag were stable in solution under the conditions described for purification and characterization. The pH was maintained near neutrality, but ionic strength was varied significantly during the experiments, ranging from only 10 mM phosphate buffer up to 250 mM NaCl. The protein remained in solution in monomeric form following storage at 4 °C for one week. For long term storage, the protein was flash frozen in small aliquots using liquid nitrogen, and then stored at −80 °C. The preparation of a stable protein sample was an important goal, and is required prior to placing significant efforts into structural and biophysical studies. The protein was also resistant to proteolytic degradation. While no explicit proteolytic digestion assays were performed on the sample, there was no indication that native *E. coli* proteases caused any observable degradations either pre- or post-lysis. The lack of proteolytic degradation is important for easily obtaining a homogeneous sample. It is an indication that the protein maintains a globular fold, hindering proteolysis.

Circular dichroism was used to determine the secondary structure composition of the purified SARS-CoV nsp1 (Fig. 3). The experimentally derived composition displayed reasonable agreement with the consensus prediction based on amino acid sequence alone. The major deviation between experiment and prediction was the experimental data indicated a higher than expected amount of β-strand, resulting in a less than expected amount of random coil. Having more residues in a regular secondary structure conformation likely aids the stabilization of the protein, and is an indication of a well folded protein. However, it should be noted that the program k2d, used to analyze the CD data, considers *random coil* to include all residues that do not participate in an α-helix or a β-strand, and this term does not imply that such residues lack a defined and stable structure within a given protein. The experimentally determined composition of 28% α-helix, 33% β-strand, and 39% random coil is consistent with values observed for other proteins having a globular fold. For example, using the same CD protocol, we have determined the secondary structure composition of another SARS-CoV protein (nsp9) as being 10% α-helix, 39% β-strand, and 51% random coil (unpublished results). These values compare extremely well with the values (13% α-helix, 35% β-strand, and 52% random coil) calculated from its crystal structure (PDB: 1UW7). Hen egg white lysozyme (PDB: 1HEW) and bovine trypsin (PDB: 1GBT) contain approximately 50% and 54%, respectively, of their residues in other than α-helical or β-strand conformations, based upon their crystal structures.

Molecular weight estimation by SEC calibrated against the elution volumes of several well-behaved protein standards is a well-established procedure. SEC is also capable of providing an estimation of the oligomeric state of the protein in solution under the chosen buffer conditions. This method provided an estimated molecular weight for the purified recombinant SARS-CoV nsp1 of ∼37 kDa and indicated that it was present predominantly as a single species. These data can be interpreted as the protein being present as a dimer in solution, as the calculated mass of a monomer is 20.2 kDa. However, molecular weight estimation by traditional SEC is limited by the assumptions that

the protein sample interacts with the column resin in an ideal fashion (e.g., no electrostatic or hydrophobic interactions), and the individual protein particles (monomers or complexes) are approximately spherical, as the elution profile is influenced not only by molecular weight but also by molecular shape.

SEC–MALS employs a combination of light scattering and refractive index detectors to continuously monitor the SEC eluant, providing molecular weight estimates unaffected by a sample's non-ideal interaction with the SEC resin. The sole role of SEC in a SEC–MALS experiment is to maximize the homogeneity of the sample being analyzed by MALS at any given instant, as MALS provides a weighted average of the molecular weight of all species in the aliquot under analysis. Molecular weight estimation by MALS is largely independent of molecular shape, and so the SEC–MALS results are influenced substantially less by non-ideal sample behavior then when using SEC alone. Analysis of 14 protein standards showed that the SEC–MALS method can routinely estimate molecular weights of native proteins within 5% [29]. SEC–MALS indicates that purified SARS-CoV nsp1 is present as a monodisperse monomeric population weighing 19.6 kDa in solution. The disagreement between the SEC and the SEC–MALS molecular weight estimations may be due to non-ideal interactions between the protein and the SEC media. However, it is likely an indication that the protein's shape deviates significantly from spherical (e.g., oblate or prolate). SARS-CoV nsp1 is likely present as a monomer in solution as molecular weights estimated by SEC–MALS are more accurate than those estimated from SEC alone.

Acknowledgments

References

[1] C. Drosten, S. Gunther, W. Preiser, S. van der Werf, H.R. Brodt, S. Becker, H. Rabenau, M. Panning, L. Kolesnikova, R.A. Fouchier, A. Berger, A.M. Burguiere, J. Cinatl, M. Eickmann, N. Escriou, K. Grywna, S. Kramme, J.C. Manuguerra, S. Muller, V. Rickerts, M. Sturmer, S. Vieth, H.D. Klenk, A.D. Osterhaus, H. Schmitz, H.W. Doerr, Identification of a novel coronavirus in patients with severe acute respiratory syndrome, N. Engl. J. Med. 348 (2003) 1967–1976.

[2] T.G. Ksiazek, D. Erdman, C.S. Goldsmith, S.R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J.A. Comer, W. Lim, P.E. Rollin, S.F. Dowell, A.E. Ling, C.D. Humphrey, W.J. Shieh, J. Guarner, C.D. Paddock, P. Rota, B. Fields, J. DeRisi, J.Y. Yang, N. Cox, J.M. Hughes, J.W. LeDuc, W.J. Bellini, L.J. Anderson, A novel coronavirus associated with severe acute respiratory syndrome, N. Engl. J. Med. 348 (2003) 1953–1966.

[3] J.S. Peiris, S.T. Lai, L.L. Poon, Y. Guan, L.Y. Yam, W. Lim, J. Nicholls, W.K. Yee, W.W. Yan, M.T. Cheung, V.C. Cheng, K.H. Chan, D.N. Tsang, R.W. Yung, T.K. Ng, K.Y. Yuen, Coronavirus as a possible cause of severe acute respiratory syndrome, Lancet 361 (2003) 1319–1325.

[4] E.J. Snijder, P.J. Bredenbeek, J.C. Dobbe, V. Thiel, J. Ziebuhr, L.L. Poon, Y. Guan, M. Rozanov, W.J. Spaan, A.E. Gorbalenya, Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage, J. Mol. Biol. 331 (2003) 991–1004.

[5] J.S. Kahn, The widening scope of coronaviruses, Curr. Opin. Pediatr. 18 (2006) 42–47.

[6] K.V. Holmes, Coronaviruses, in: D.M. Knipe, B. Roizman, P.M. Howley, S.E. Straus, B.N. Fields, D.E. Griffin, M.A. Martin, R.A. Lamb (Eds.), Fields Virology, Lippincott Williams & Wilkins, Philadelphia, 2001, pp. 1187–1203.

[7] F. Esper, C. Weibel, D. Ferguson, M.L. Landry, J.S. Kahn, Evidence of a novel human coronavirus that is associated with respiratory tract disease in infants and young children, J. Infect. Dis. 191 (2005) 492–498.

[8] R.A. Fouchier, N.G. Hartwig, T.M. Bestebroer, B. Niemeyer, J.C. de Jong, J.H. Simon, A.D. Osterhaus, A previously undescribed coronavirus associated with respiratory disease in humans, Proc. Natl. Acad. Sci. USA 101 (2004) 6212–6216.

[9] L. van der Hoek, K. Pyrc, M.F. Jebbink, W. Vermeulen-Oost, R.J. Berkhout, K.C. Wolthers, P.M. Wertheim-van Dillen, J. Kaandorp, J. Spaargaren, B. Berkhout, Identification of a new human coronavirus, Nat. Med. 10 (2004) 368–373.

[10] P.C. Woo, S.K. Lau, C.M. Chu, K.H. Chan, H.W. Tsoi, Y. Huang, B.H. Wong, R.W. Poon, J.J. Cai, W.K. Luk, L.L. Poon, S.S. Wong, Y. Guan, J.S. Peiris, K.Y. Yuen, Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia, J. Virol. 79 (2005) 884–895.

[11] S.K. Lau, P.C. Woo, K.S. Li, Y. Huang, H.W. Tsoi, B.H. Wong, S.S. Wong, S.Y. Leung, K.H. Chan, K.Y. Yuen, Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats, Proc. Natl. Acad. Sci. USA 102 (2005) 14040–14045.

[12] W. Li, Z. Shi, M. Yu, W. Ren, C. Smith, J.H. Epstein, H. Wang, G. Crameri, Z. Hu, H. Zhang, J. Zhang, J. McEachern, H. Field, P. Daszak, B.T. Eaton, S. Zhang, L.F. Wang, Bats are natural reservoirs of SARS-like coronaviruses, Science 310 (2005) 676–679.

[13] L. Vijgen, E. Keyaerts, E. Moes, I. Thoelen, E. Wollants, P. Lemey, A.M. Vandamme, M. Van Ranst, Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event, J. Virol. 79 (2005) 1595–1604.

[14] A.E. Gorbalenya, L. Enjuanes, J. Ziebuhr, E.J. Snijder, Nidovirales: evolving the largest RNA virus genome, Virus Res. 117 (2006) 17–37.

[15] B.H. Harcourt, D. Jukneliene, A. Kanjanahaluethai, J. Bechill, K.M. Severson, C.M. Smith, P.A. Rota, S.C. Baker, Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity, J. Virol. 78 (2004) 13600–13612.

[16] E. Prentice, J. McAuliffe, X. Lu, K. Subbarao, M.R. Denison, Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins, J. Virol. 78 (2004) 9977–9986.

[17] Y.J. Tan, S.G. Lim, W. Hong, Characterization of viral proteins encoded by the SARS-coronavirus genome, Antiviral Res. 65 (2005) 69–78.

[18] R. Gosert, A. Kanjanahaluethai, D. Egger, K. Bienz, S.C. Baker, RNA replication of mouse hepatitis virus takes place at double-membrane vesicles, J. Virol. 76 (2002) 3697–3708.

[19] E. Prentice, W.G. Jerome, T. Yoshimori, N. Mizushima, M.R. Denison, Coronavirus replication complex formation utilizes components of cellular autophagy, J. Biol. Chem. 279 (2004) 10136–10141.

[20] S.M. Brockway, M.R. Denison, Mutagenesis of the murine hepatitis virus nsp1-coding region identifies residues important for protein processing, viral RNA synthesis, and viral replication, Virology 340 (2005) 209–223.

[21] M.R. Denison, B. Yount, S.M. Brockway, R.L. Graham, A.C. Sims, X. Lu, R.S. Baric, Cleavage between replicase proteins p28 and p65 of mouse hepatitis virus is not required for virus replication, J. Virol. 78 (2004) 5957–5965.

[22] W. Kamitani, K. Narayanan, C. Huang, K. Lokugamage, T. Ikegami, N. Ito, H. Kubo, S. Makino, Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation, Proc. Natl. Acad. Sci. USA 103 (2006) 12885–12890.

[23] M.A. Andrade, P. Chacon, J.J. Merelo, F. Moran, Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network, Protein Eng. 6 (1993) 383–390.

[24] J. Cheng, A.Z. Randall, M.J. Sweredoski, P. Baldi, SCRATCH: a protein structure and structural feature prediction server, Nucleic Acids Res. 33 (2005) W72–W76.

[25] K. Bryson, L.J. McGuffin, R.L. Marsden, J.J. Ward, J.S. Sodhi, D.T. Jones, Protein structure prediction servers at University College London, Nucleic Acids Res. 33 (2005) W36–W38.

[26] K. Lin, V.A. Simossis, W.R. Taylor, J. Heringa, A simple and fast secondary structure prediction method using hidden neural networks, Bioinformatics 21 (2005) 152–159.

[27] R. Adamczak, A. Porollo, J. Meller, Combining prediction of secondary structure and solvent accessibility in proteins, Proteins 59 (2005) 467–475.

[28] D. Frishman, P. Argos, Seventy-five percent accuracy in protein secondary structure prediction, Proteins 27 (1997) 329–335.

[29] E. Folta-Stogniew, K.R. Williams, Determination of molecular masses of proteins in solution: Implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory, J. Biomol. Tech. 10 (1999) 51–63.