Available online at www.sciencedirect.com
ScienceDirect
Protein Expression and Purification 52 (2007) 249–257
Protein Expression & Purification
www.elsevier.com/locate/yprep

# 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*

Received 4 August 2006, and in revised form 6 November 2006
Available online 14 November 2006

**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.


*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.

1046-5928/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2006.11.005

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 $_{180}G{\downarrow}A_{181}$, 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