Chapter 7

Expression and purification of recombinant equine arteritis virus nonstructural protein 1

Danny D. Nedialkova, Bruno Coutard, Bruno Canard, and Eric J. Snijder
ABSTRACT

Arteriviruses and coronaviruses, distantly related plus-strand RNA viruses, generate an extensive nested set of subgenomic (sg) mRNAs as templates for the translation of viral structural proteins. These viral RNA molecules are copied from a complementary set of minus-strand templates that are produced via a unique mechanism of discontinuous minus-strand extension that is poorly understood. Non-structural protein 1 (nsp1) from equine arteritis virus (EAV), the prototype arterivirus, is essential for sg mRNA production but dispensable for genome replication. Apart from securing the balance between these key viral processes, nsp1 has also been shown to modulate viral RNA accumulation in an mRNA-specific manner by controlling the levels of templates required for viral mRNA synthesis. The development of in vitro assays to address nsp1 function, combined with solution of the protein's three-dimensional structure, could be instrumental in providing mechanistic details of the multiple roles of this protein in the viral replicative cycle. We therefore explored different expression conditions and fusion tags for the production and purification of soluble recombinant nsp1 in Escherichia coli. We describe the pitfalls we encountered upon purification of GST-tagged nsp1 and subsequent tag removal by proteolytic cleavage, and initial difficulties in purifying hexahistidine-tagged nsp1. We outline a successful approach for nsp1 expression from a construct that allowed its autoproteolytic release from a precursor, and which yielded soluble, stable recombinant nsp1 purified from bacterial cells that can be used for functional studies.
INTRODUCTION

Replication of RNA viruses with single-stranded genomes of messenger RNA polarity (plus-strand [+] RNA) is ensured by an RNA-dependent RNA polymerase (RdRp) that is produced by translation of the viral genome following its uncoating in the infected cell. Plus-strand RNA virus viruses also employ virally-encoded helicases (Hel), presumably to unwind local double-stranded RNA structures that might impede RdRp progress during viral RNA synthesis. The activities of these two essential enzymes need to be modulated to assure the temporal progression of the viral replicative cycle that requires the integral coordination of genome replication, translation, and packaging. +RNA viruses that produce subgenomic RNAs, like arteriviruses and coronaviruses, are challenged with an additional regulatory step. Intricate circuits involving protein-protein and protein-RNA interactions that modulate viral RNA synthesis have therefore evolved. These “molecular switches” can require the recruitment of cellular proteins, but are often mediated by viral domains that can be covalently linked to the RdRp or Hel, or exist as independent polypeptides. An example of such a candidate “molecular switch” is non-structural protein 1 (nsp1) of the equine arteritis virus (EAV), the prototype arterivirus.

The EAV genome is translated to produce a multidomain replicase precursor from the two 5’-proximal open reading frames (ORFs) in the viral genome, resulting in the synthesis of polyprotein (pp) 1a and, after a ribosomal frameshift, pp1ab. The replicase polyproteins are proteolytically processed by virus-encoded (auto)proteinases into 13 nsps that possess enzymatic activities and/or indirect regulatory roles in viral RNA replication. The viral structural genes are encoded in the 3’-end of the genome and expressed from subgenomic (sg) mRNAs. Arterivirus sg mRNAs are 5’- and 3’-coterminal and are copied from sg-length minus-strand templates produced by discontinuous minus-strand extension\(^\text{13}\), an unusual process that is also utilized by the distantly related coronaviruses and remains poorly characterized. Subgenomic mRNA synthesis and genome replication presumably compete for the same template – the genomic plus strand, implying the existence of protein and/or RNA signals that serve as regulatory switch(es).

EAV nsp1 is vital for the quantitative balance between genome replication and sg mRNA production, and is the only arteri- or coronavirus protein specifically implicated in sg mRNA synthesis to date\(^\text{20}\) (Chapters 5 and 6 of this thesis). Nsp1 is also the first protein expressed during infection, by a co-translational release from the nascent replicase polyproteins by a papain-like cysteine proteinase activity (PCPβ) in its C-terminal domain (Fig. 1A); this cleavage is essential for virus viability\(^\text{14}\) (Chapter 5). Two additional conserved subdomains were identified by bioinformatics: a second, proteolytically silent PCP domain that is functional in other arteriviruses (PCPα), and an N-terminal zinc finger (ZF) domain. Interestingly, nsp1 was recently found to modulate the levels at which individual viral mRNAs accumulate relative to each other during EAV infection, likely by controlling the relative abundance of their minus-strand templates (Chapter 6). Unraveling the molecular mechanisms underlying the regulatory functions of nsp1 could therefore provide invaluable insights into the poorly understood process of discontinuous
minus-strand synthesis. For that purpose, expression and purification of recombinant nsp1 and solving its three-dimensional structure would greatly facilitate its detailed characterization. The ability of the protein to form complexes with RNA and specifically recognize viral sequences that regulate sg mRNA synthesis, as well as its influence on the in vitro EAV RdRp activity could be examined\textsuperscript{3,22}. Also, experiments based on isolated active viral RNA-synthesizing complexes from cells transfected with nsp1 mutants and addition of recombinant nsp1 could address the importance of timing of nsp1 expression and whether defects associated with mutations in nsp1 can be complemented by the wild-type protein.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{constructs.png}
\caption{Constructs for expression of recombinant EAV nsp1. (A) Domain organization of EAV nsp1. Three conserved subdomains were identified by comparative sequence analysis: an N-terminal zinc finger (ZF) domain, a papain-like cysteine proteinase activity (PCP\textbeta) in its C-terminal domain which mediates the co-translational release of nsp1 from the nascent replicase polyproteins, and a second, proteolytically silent PCP domain that is functional in other arteriviruses (PCP\textalpha). The position of predicted zinc-coordinating residues and the active-site amino acids of PCP\textalpha and PCP\textbeta are indicated with lines. The lack of PCP\textalpha proteolytic activity is likely due to a natural loss of the active-site Cys. (B) Schematic diagrams of fusion proteins produced from nsp1 expression constructs. Nsp1 was expressed in \textit{E. coli} fused to the C-terminus of GST with a Factor Xa Protease cleavage site (IEGR) and a three-amino-acid spacer. Proteolytic processing of the GST-nsp1 fusion protein by Factor Xa yielded nsp1 containing three foreign amino acids (GIP) at its N-terminus. Numbers indicate amino-acid positions in EAV pp1a. Nsp1 was also produced as an N- or C-terminally His-tagged protein (nsp1-HN and nsp1-HC, respectively). A recombinant protein with an N-terminal His tag followed by the nsp1 sequence and 169 residues from nsp2 was used to generate bacterially expressed autoprocessed nsp1. This recombinant form of nsp1, whose amino acid sequence is identical to that of nsp1-HN, is referred to as nsp1-HN-ap in the text.}
\end{figure}
We therefore set out to design a protocol for the expression and purification of stable and soluble recombinant nsp1 for crystallization and structure determination via X-ray diffraction, as well as functional in vitro studies. Since significant amounts (mg) of protein are usually required for structural studies, and nsp1 is a relatively small protein (30 kDa) that is not expected to be post-translationally modified, we chose E. coli as the expression host. Several terminally extended derivatives of nsp1 were produced and a variety of conditions for their purification were explored. The nsp-coding sequence was fused to the C-terminus of glutathione-S-transferase (GST) via a linker region encoding a Factor Xa Protease cleavage site (Fig. 1B). Alternatively, hexahistidine [(His)₆] tags were fused to the N-terminus or C-terminus of the protein. Lastly, expression of nsp1 with an N-terminal (His)₆ tag from a construct that encoded an amino-acid sequence derived from the N-terminal region of nsp2 was used to assess the effect of autoproteolytic cleavage on the solubility and stability of recombinant nsp1. We outline the technical difficulties encountered upon purification of GST-nsp1 and removal of the affinity tag by proteolysis. Furthermore, we show that autoproteolytic release of nsp1 from a precursor polypeptide is likely key for the successful purification of stable, soluble nsp1 with an N-terminal (His)₆ tag by metal affinity chromatography. Although the current yield of recombinant nsp1 obtained following the protocol described here is likely too low for structural studies, the purified protein can be used for setting up in vitro assays aimed at elucidating one or more of its functions.

**MATERIALS AND METHODS**

**Construction of plasmids for recombinant nsp1 expression**

In order to generate a plasmid for the expression of EAV nsp1 fused to GST in bacterial hosts, the nsp1- coding sequence (nucleotides 225 to 1004 of the wild-type EAV infectious cDNA clone pEAV211) was amplified by PCR with oligonucleotides E872 (5’- CGCGGATCCCCATGGCAACCTTCTCCGC- 3’) and E386 (5’-GGAATTCGCATGCCTAGCCGTAGTTGCCAG-3’). The fragment was cloned between the BamHI and EcoRI restriction sites of the pGEX-5X-3 expression vector (GE Healthcare, Chalfont St Giles, UK), transcription from which is driven by the chemically inducible tac promoter. The resulting plasmid, pGEX-5X-3-nsp1, encoded EAV nsp1 (amino acids Met-1 to Gly-260 of replicase pp1ab of the EAv-Bucyrus strain, NCBI genome database accession number NC_002532) fused to the C-terminus of GST. The two protein moieties are separated by 7 amino acid spacer representing a Factor Xa cleavage site (NH₂-Ile-Glu-Gly-Arg-COOH), followed by NH₂-Gly-Ile-Pro COOH, originating from translation of sequences in the multiple cloning site of the vector.

For the generation of plasmids for bacterial expression of (His)₆-tagged EAV nsp1, oligonucleotides E777 (5’-CATGCCATGGGCCATCACCATCACCATCACATGGCAACCTTCTCCGCTACTGG-3’) and E386 (5’- GAATTCCATGCCTAGCCGTAGTTGCCAG-3’) were used to
amplification of the nsp1-coding sequence, introducing a start codon followed by sequences specifying a (His)₆ tag, and a stop codon, respectively. Amplification with E385 (5'-TATA-ACGCGTGATACATGGCAACCTTCTCC-3') and E778 (5'-CCGCTCGAGTAGTGATGATGATGATGATGATGATGATGAGGCAAGCTTGGTCCTTGG-3') was used to introduce the (His)₆ tag-coding sequence followed by a stop codon, downstream of nt 1004. The fragments were cloned between the Ncol and XhoI restriction sites of Gateway entry vector pENTR11 (Invitrogen), and, after sequence verification, placed downstream of the T7 promoter in Gateway expression vector pDEST14. The resulting constructs, pDEST14-EAV nsp1-HN and pDEST14- EAV nsp1-HC, mediate the expression of recombinant nsp1 preceded by an N-terminal Met-Gly-(His)₆ tag (nsp1-HN), or followed by a C-terminal (His)₆ tag (nsp1-HC). Lastly, an expression construct encoding aa 1 to 429 of EAV pp1a preceded by a Met-Gly-(His)₆ tag, was generated by PCR amplification of nt 255 – 1511 of pEA211 with E777 and E1064 (5'-CGCTCGAGTTATGTGGCAAGCTTGGTCCTTGG-3'), subsequent cloning of the fragment between the Ncol and XhoI restriction sites of pENTR11 and recombination to pDEST14. The resulting plasmid, pDEST14-nsp1-2N-HN, mediates the expression of EAV nsp1 fused to 169 aa from the N terminus of EAV nsp2, allowing for the generation of autoproteolytically processed nsp1 with an N-terminal (His)₆ tag. This recombinant form of nsp1, whose amino acid sequence is identical to that of nsp1-HN, is referred to as nsp1-HN-ap in the text.

**Bacterial strains for recombinant nsp1 expression**

The BL21(DE3) strain of *Escherichia coli* and its derivatives, C41 (DE3) and C43(DE3) were used as expression hosts as indicated in the text. The BL21(DE3) strain lacks the OmpT and Lon proteases and contains a chromosomal copy of the T7 polymerase gene behind a lacUV5 promoter. Expression of T7 polymerase is induced by addition of the lactose analog isopropyl β-D thiogalactoside (IPTG). The C41 (DE3) and C43(DE3) strains were isolated in a screen for BL21(DE3) derivatives with improved characteristics of toxic protein expression.

**Expression and purification of GST-nsp1**

Plasmid pGEX-5X-3-nsp1 was transformed in an expression strain of *E. coli* and a single colony was used to inoculate 2 ml of LB containing 50 µg/ml of ampicillin (LB-Amp). The culture was grown overnight, diluted 1:100 in fresh LB-Amp, and subsequently grown at 37°C to mid-logarithmic phase (OD₆₀₀ 0.6 – 0.8). Recombinant protein expression was induced by addition of IPTG. The duration of induction, temperature at which bacteria were cultured during induction, and amount of IPTG, are indicated in the text. After induction, cells were harvested by centrifugation at 2,500 x g for 10 min and stored at -20°C. The pellets were resuspended in buffer A (10 mM Tris-HCl pH 7.5, 400 mM NaCl, 10% glycerol; 1 ml of buffer A was used per 10 ml of original culture volume) supplemented with 1 mg/ml lysozyme (Sigma-Aldrich) and 150 U/ml DNase (Invitrogen), and incubated for 1 h
Purification of recombinant EAv nsp1

at 4°C with gentle mixing. The lysates were then supplemented with 0.5 % Triton-X100, incubated for 30 min at 4°C and clarified by high-speed centrifugation. Pellet fractions were resuspended in 1X Laemmli sample buffer (LSB) for SDS-PAGE analysis of insoluble proteins. Supernatant fractions were incubated with GST affinity resin (glutathione sepharose 4B [GS4B] from GE Healthcare, unless otherwise indicated in the text) for 1 – 2 h at 4°C with gentle mixing. The affinity resin was then washed extensively with buffer A supplemented with 0.5 % Triton-X100, followed by washing with buffer A alone. The bound GST-nsp1 was eluted with 10 mM reduced glutathione in 50 mM Tris-HCl, pH 8.0, and analyzed by SDS-PAGE and Coomassie Brilliant Blue staining. Alternatively, the resin-bound fusion protein was equilibrated in Factor Xa cleavage buffer and digested with Factor Xa Protease (see below).

Factor Xa cleavage of GST-nsp1

GS4B resin-bound recombinant GST-nsp1 was equilibrated in FXa buffer (50 mM HEPES pH=6.5, 100 mM NaCl, 10 % glycerol, 1 mM CaCl₂, 50 mM L-Arg pH=9.5; final pH of buffer: 7.5). Subsequently, Factor Xa Protease (Qiagen) was added to the slurry and proteolytic reactions were carried out as specified in the text. Cleaved nsp1, containing three additional amino acids at the N terminus (NH₂-Gly-Ile-Pro COOH) as a result from the cloning strategy, was recovered from the supernatant of the pelleted GS4B resin. After digestion, Factor Xa Protease and traces of GST or uncleaved fusion protein were removed by incubation of the nsp1-containing fraction with Xa Removal Resin (Qiagen; 50 µl per 4 U of enzyme) and GS4B. The nsp-containing fraction was then centrifuged at 100,000 x g for 1 h at 4°C in a Beckman tabletop ultracentrifuge. The supernatant fraction was analyzed by gel filtration on a Superdex 75 column (GE Healthcare) in 50 mM HEPES pH=6.5, 100 mM NaCl, 10 % glycerol, 50 mM L-Arg.

Expression and purification of His-tagged nsp1

Expression of recombinant nsp1 from pDEST14-EAV nsp1-HN and pDEST14-EAV nsp1-HC vectors was performed essentially as described above for GST-nsp1 fusion protein expression. Supernatants from induced bacterial cultures lysed in buffer B (10 mM Tris-HCl pH 8.0, 400 mM NaCl, 10% glycerol supplemented with 0.5 % Triton-X100 were incubated with Ni-NTA agarose (Qiagen) or HIS-Select HF Nickel Affinity Gel (Sigma-Aldrich) for 1 h at 4°C with gentle mixing. Unbound proteins were washed off with buffer B containing 10 mM imidazole, and proteins bound to the resin were eluted in buffer B supplemented with 250 mM imidazole.

For expression and purification of autoproteolytically processed recombinant nsp1-HN-ap, the pDEST14-nsp1-2N-HN expression vector was transformed in C41(DE3) cells and a single colony was used to inoculate 5 ml of LB-Amp. After growth for 4 h at 37°C, the starting culture was diluted in 1 l pre-warmed LB-Amp and cells were grown at 37°C until the OD₆₅₀ equaled 0.6. The culture was cooled on ice and recombinant protein expression
was induced with 1 mM IPTG at 16°C for 16 h. The cells were harvested by centrifugation at 2,500 x g for 10 min, pellets were resuspended in buffer C (50 mM 2-(N-morpholino)-ethanesulphonic acid [MES] pH 7.0, 500 mM NaCl, 10 % glycerol, 5 mM imidazole; 1 ml buffer per 50 ml original culture) supplemented with 1 mg/ml lysozyme, 30 U/ml DNase I, and 0.1 % Triton-X100, and the cell suspension was incubated for 1 h at 4°C with gentle mixing. The Triton-X100 concentration was adjusted to 0.5%, lysates were incubated at 4°C for additional 30 min and subsequently clarified by centrifugation at 20,000 x g for 30 min. Recombinant nsp1 was purified from supernatants using Talon metal affinity resin (Clontech). The resin was washed twice with 50 resin volumes of buffer C containing 0.5 % Triton-X100, followed by three washes with 50 volumes of buffer C containing 10 mM imidazole. Elution was performed in 20 mM MES pH=7.0, 500 mM NaCl, 10 % glycerol, 150 mM imidazole in two steps. Eluates were pooled and immediately loaded onto a Superdex-75 column, and gel filtration was carried out as indicated in the text. Fractions containing nsp1 were pooled, and Ultrafree 10-kDa filter columns (Millipore) were used for protein concentration.

RESULTS AND DISCUSSION

Expression and purification of GST-tagged nsp1

Initial characterization of the proteolytic activity of EAV nsp1 showed that the protein is at least partially active as an autoproteinase when increasingly longer N-terminal fragments of pp1a were expressed in as C-terminal GST fusions in bacterial cells\(^{14}\). Autoproteolytic processing suggested proper folding of nsp1 (and its PCPβ domain in particular) in a fraction of the bacterially produced fusion proteins, although both precursor and self-processed GST-containing proteins could only be recovered from the insoluble fractions of cell lysates under the experimental conditions used\(^{14}\). We therefore set out to explore expression and purification conditions that would allow us to produce soluble GST-nsp1, as well as protocols for the removal of the GST tag to obtain native nsp1. To that end, the nsp1-coding sequence (amino acids 1 – 260 from EAV pp1a) was cloned in frame with the GST gene in plasmid pGEX-5X-3, and the resulting plasmid was used to produce a GST-nsp1 fusion protein under the control of the strong IPTG-inducible tac promoter. The two protein moieties in the fusion are separated by a seven amino-acid linker containing a Factor Xa Protease cleavage site (Ile-Glu-Gly-Arg), followed by three residues (Gly-Ile-Pro) derived from an in-frame vector sequence (see Fig. 1C).

Lowering expression temperature has been shown to improve recombinant protein solubility in E. coli, presumably by allowing enough time for the proper folding of newly produced recombinant proteins due to slower metabolic rates\(^{10,23}\). Expression of GST-nsp1 in BL21(DE3) cells was therefore tested in small-scale bacterial cultures that were grown to mid-log phase, and protein production was induced at 30°C by the addition of 0.5 mM IPTG. Pellet and supernatant fractions of lysates (prepared as described in Materi-
Purification of recombinant EAV nsp1

als and methods) from cells harvested at 2.5 h and 7 h after induction were analyzed by SDS-PAGE. Large quantities of a protein with an apparent molecular mass of 55 kDa, consistent with the expected mass of GST-nsp1, were produced exclusively in the iPTG induced bacterial cultures, and Western blotting with GST- and nsp1-specific antibodies confirmed the identity of the fusion protein (data not shown). The majority of recombinant GST-nsp1, however, was present in the insoluble fraction of cell lysates, especially after longer induction times (Fig. 2A).

In an effort to increase the fraction of soluble GST-nsp1 produced in bacterial cells, the temperature during protein expression was further decreased to 25°C, and the concentration of IPTG was lowered to 0.3 mM. These conditions resulted in a slightly higher yield of soluble fusion protein, although the bulk of GST-nsp1 still remained in insoluble aggregates (Fig. 2B). The fusion protein was purified from the soluble fraction of bacterial

Figure 2. Expression and purification of GST-nsp1. (A) BL21(DE3) cells transformed with pGEX-5X-3-nsp1 were cultured until \( \text{OD}_{600} \) of 0.7, GST-nsp1 expression was induced by the addition of 0.5 mM IPTG, and the cultures were incubated at 30°C. Lysates were obtained from non-induced or IPTG-induced cells harvested 2.5 or 7 hours post induction. After centrifugation of total cell lysates, obtained as described in Materials and Methods, the insoluble pellet (P) fractions were resuspended in 1x LSB and equal volumes of pellet and supernatant (S) fractions were separated on a 12.5 % polyacrylamide gel by SDS-PAGE. Proteins were visualized by subsequent staining of the gel with Coomassie Brilliant Blue R-250. (B) GST-nsp1 expression was induced with 0.3mM IPTG for 5.5 h at 25°C, and the soluble fraction of the total bacterial lysate was incubated with Glutathione Sepharose 4B (GS4B). The resin was pelleted by centrifugation, and unbound proteins (flow-through, FT) were collected for analysis. Impurities were then removed by washing the resin twice with buffer A with 0.5 % Triton-X100, followed by a single washing step with buffer A alone (for buffer composition, see Materials and Methods). Bound GST-nsp1 was subsequently eluted with 10 mM reduced glutathione in 50 mM Tris-Cl pH 8.0 in three steps. Proteins from the pellet (P) and soluble (S) fractions of the total bacterial lysates, as well as the FT, wash and elution steps were analyzed by SDS-PAGE and Coomassie blue staining. (C) BL21(DE3) cells were transformed with a mutant derivative of pGEX-5X-3-nsp1 containing a C41H substitution in nsp1. Expression and purification of the mutant recombinant protein was performed essentially as described in (B) but protein expression was induced at 20°C.
lysates with glutathione sepharose 4B (GS4B) as described in Materials and methods. The purity of the fusion protein was more than 90%, and the yield of GST-nsp1 obtained under these conditions was ~1 mg/l culture, as estimated by Coomassie Blue staining and comparison to BSA standards (Fig. 2B and data not shown).

We then attempted to purify a mutant form of the fusion protein - GST-nsp1 C41H, carrying a Cys-41 to His substitution of a zinc-coordinating residue in nsp1 that abolishes sg mRNA production when introduced into the infectious cDNA clone of EAV (Chapter 5 of this thesis). The mutant recombinant protein, however, was found exclusively in the insoluble fraction of bacterial lysates when expressed using the protocol described above, as well as when the protein production temperature was lowered even further, to 20°C (Fig. 2C). No significant amounts of GST-nsp1 C41H were obtained upon expression of the protein in a variety of conditions aimed at promoting protein solubility and/or suppressing protein aggregation. The strategies explored included the addition of 3% ethanol to the culture medium during protein production, which was shown to mimic a heat-shock response in *E. coli* and enhance solubility of some recombinant proteins, likely by inducing overexpression of heat-shock proteins, which can act as chaperones and facilitate correct protein folding and assembly17. The increase of osmotic pressure, coupled with addition of “osmolytes” (small organic molecules, e.g. betaine and proline, which have unfavorable interactions with protein surfaces and therefore stabilize their natively folded state) to the culture medium during induction of protein expression has been reported to decrease protein aggregation in *E. coli*4,8. None of these conditions promoted greater solubility of GST-nsp1 C41H (data not shown), and neither did expression of the mutant fusion protein in two different strains of *E. coli* – C41(DE3) and C43(DE3). These results suggest the C41H substitution has a significant destabilizing effect on the overall structure of nsp1 in *E. coli*, which can be difficult to overcome, considering the predicted zinc-binding role of Cys-41 and the structural importance of zinc coordination for the proper folding of ZF domains.

**GST tag removal by Factor Xa proteolysis**

Due to the large size of GST tag (26 kDa) and its potential to influence the biochemical properties of nsp1, tag removal was a prerequisite for subsequent functional studies and attempts to crystallize the protein in order to solve its three-dimensional structure. To this end, GST-nsp1 was expressed and purified as described above, and was subjected to proteolysis with Factor Xa protease while bound to the GST affinity resin (Fig. 3A). Addition of increasing amounts of protease resulted in the release of a 30-kDa protein in the supernatant when cleavage was performed in buffer conditions recommended by the manufacturer (Qiagen FXa buffer; see legend of Fig. 3A) at ambient temperature; this protein was absent in the no-enzyme control reaction. Western blotting confirmed that the 30-kDa cleavage product was indeed nsp1 (data not shown). Centrifugation of cleaved nsp1 at 100,000x *g*, however, revealed that the protein precipitated in the absence of the fusion tag (data not shown) in Qiagen FXa buffer. An optimized protocol for
the Factor Xa-mediated proteolytic reaction that prevented precipitation of cleaved nsp1 was then determined empirically. L-Arg, a known aggregation suppressor, was found to be a crucial additive for stabilizing the protein (Fig. 3B), and proteolysis was carried out in FXa buffer (for composition, see the figure 3 legend) at 4°C to account for potential thermal instability of nsp1. In subsequent experiments, 0.1% CHAPS was added during proteolytic digestion to decrease non-specific association of cleaved nsp1 with the GS4B resin.

**Figure 3.** GST tag removal by Factor Xa proteolysis. (A) The soluble fraction of a total bacterial lysate (S, left panel) was divided in four samples and each sample was incubated with equal volumes of GS4B. Contaminants were washed out as described in the legend to Fig. 2B, and the resin with bound fusion protein was equilibrated in 20 mM Tris pH=6.5, 50 mM NaCl, 1 mM CaCl2 (Qiagen Factor Xa buffer). Proteolytic cleavage of resin-bound GST-nsp1 was carried out at room temperature by addition of different amounts of Factor Xa Protease (1, 0.2, 0.05 or 0 units [U]). Samples were removed from the proteolytic digestion mixture at 3 and 6 hours after the start of the reaction and separated by SDS-PAGE on a 12% polyacrylamide gel that was stained with Coomassie blue in order to estimate the yield and purity of cleaved nsp1. (B) Digestion of GS4B-bound GST-nsp1 with Factor Xa Protease was performed for 16 h at 4°C in 50 mM HEPES pH=6.5, 100 mM NaCl, 10% glycerol, 1 mM CaCl2, supplemented with 50 mM L-Arg. Cleaved nsp1 (input, I) was analyzed by ultracentrifugation at 100,000 x g for 1h at 4°C. The pellet (P100) fractions were resuspended in 1x LSB, and equal volumes of pellet and supernatant (S100) fractions were analyzed by SDS-PAGE and Coomassie blue staining as an indication of cleaved nsp1 solubility.

**Cleaved nsp1 forms dimers in solution**

Using the optimized protocols for GST-nsp1 expression, purification and tag removal outlined above, a large-scale expression culture (1 l) was grown with the aim of producing sufficient recombinant nsp1 for preliminary characterization of the protein preparation’s suitability for crystallization trials. Proteolysis of the resin-bound GST-nsp1 in FXa buffer resulted in the release of soluble native nsp1, and glutathione elution of GS4B-bound proteins following cleavage did not show the presence of remaining unprocessed fusion protein (Fig. 4A). Judging by the intensities of Coomassie-stained protein bands, however, the ratio of cleaved nsp1 and eluted GST protein was not 1:1, suggesting loss of native nsp1, possibly due to aggregation. In addition, impurities with molecular masses consistent with GST-nsp1 and GST remained present in the cleaved nsp1 fraction even after it was incubated with fresh GS4B resin. In an attempt to obtain a native nsp1 sample of higher purity, as well as to analyze the protein’s aggregation state, the S100 fraction of cleaved nsp1 was subjected to size exclusion chromatography. The protein eluted in two peaks with estimated molecular masses of 51 kDa and 27 kDa, and SDS-PAGE analysis
of peak fractions confirmed the presence of nsp1 (Fig. 4B), suggesting that ~90% of the protein form dimers in solution under these experimental conditions. The lack of high-molecular mass peaks in the elution profile argued against the presence of non-specific aggregates in the sample. Traces of GST, however, were still found in the nsp1-containing column fractions (data not shown), likely due to the small difference in molecular mass between the two proteins and GST’s ability to form homodimers. Finally, the yield of native nsp1 obtained in this experiment was estimated to be ~0.2 mg from 1 l of culture.

**Figure 4.** Analysis of cleaved nsp1 by gel filtration. (A) One liter of BL21(DE3) cells transformed with pGEX-5X-3-nsp1 were cultured until OD₆₀₀ of 0.7, expression of GST-nsp1 was induced by addition of 0.3 mM IPTG and the culture was incubated for 5 h at 25°C. Cells were then harvested and lysed, and the fusion protein found in the soluble fraction of the total bacterial lysate was bound to GS4B. After washing out contaminants, 20 U of Factor Xa Protease were added, and proteolytic digestion was carried out in a 600-μl volume for 14 h at 4°C in FXa buffer (see Materials and Methods) supplemented with 0.1 % CHAPS. Cleaved nsp1 was recovered from the resin supernatant, and GST-containing proteins remaining on the resin were eluted with 600μl 10 mM reduced glutathione in 50 mM Tris-HCl, pH 8.0 to check the efficiency of proteolysis (fraction E). Factor Xa and GST-containing contaminants were removed from the cleaved nsp1 fraction by subsequent incubations with Xa Removal Resin and fresh GS4B. The solubility of cleaved nsp1 was determined by ultracentrifugation as in Fig. 3B, and equal volumes of P100, S100 and E fractions were analyzed by SDS-PAGE. Half of the S100 fraction (~250 μl) was loaded onto a Superdex-75 column. (B) Elution profile of cleaved nsp1. Gel filtration chromatography of the S100 sample in (A) was performed in 50 mM HEPES pH=6.5, 100 mM NaCl, 10 % glycerol, 50 mM L-arginine (final pH of buffer: 7.5). Blue dextran (2,000 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), chymotrypsinogen A (25 kDa) and ribonuclease A (13.7 kDa) were used as native molecular mass markers. The elution volumes of the marker proteins filtered through Superdex-75 in FXa buffer were fitted to a linear equation (R²=0.98), and the standard curve was used to estimate the molecular masses of nsp1 elution peaks. (C) Peak fractions from (B) were analyzed by SDS-PAGE and Coomassie Blue staining.
Substantial protein precipitation during Factor Xa proteolysis

The production of soluble GST-nsp1 fusion in bacterial cells and nsp1 aggregation during proteolytic cleavage of the GST tag were two obvious major bottlenecks in obtaining sufficient protein material for nsp1 structure determination. Buffer composition during Factor Xa digestion of the fusion protein, however, was largely constrained by inherent properties of the protease, such as its decreased activity at high pH and salt concentrations above 100 mM. Factor Xa is also active as a disulfide-linked heterodimer, and therefore sensitive to the presence of reducing reagents during proteolytic cleavage. Addition of non ionic detergents such as Triton X-100 or Nonident P-40 reportedly does not interfere with Factor Xa proteolytic activity, although their presence even in trace amounts is poorly compatible with crystallization trials, due to their low critical micelle concentration (CMC). We therefore attempted to increase the fraction of soluble GST-nsp1 produced in expression cultures by using a number of the approaches outlined above that do not employ detergents with low CMC. A combination of expression in C43(DE3) cells and addition of 330 mM sorbitol and 2.5 mM betaine to the culture during protein production was the only condition which resulted in an increase in purified GST-nsp1, although the effects were modest at best (data not shown). In addition, a number of independent experiments for purification of GST-nsp1 from a large-scale culture and subsequent cleavage of the GST tag resulted in variable yields of cleaved nsp1 that could be as low as ~0.025 mg per liter of bacterial culture, despite high levels of GST-nsp1 in the soluble fractions of bacterial lysates.

In an effort to identify the basis of the observed variability, GST-nsp1 produced in 1 l of bacterial culture was bound to GS4B resin and subjected to Factor Xa digestion essentially as in Fig. 4. Cleaved nsp1 was recovered in the supernatant, and proteins remaining associated with the GS4B resin were eluted by boiling in an equal volume of 1x LSB. Analysis of the two fractions by SDS-PAGE revealed that by far the majority of cleaved nsp1 remained associated with the glutathione resin, as – surprisingly – did GST and a large fraction of uncleaved fusion protein (Fig. 5 A). The fact that these proteins could only be dissociated from the resin by a harsh elution step is indicative of substantial protein aggregation that occurred irrespective of whether the GST-nsp1 fusion protein was produced in BL21(DE3) or C43(DE3) cells (Fig. 5B). This phenomenon could be due to the tendency of recombinant GST to undergo oxidation-induced aggregation, probably via disulfide bond formation among surface-exposed cysteine residues found close to the C-terminus of the protein. Aggregated GST, however, was shown to retain its ability to bind glutathione affinity matrix and therefore should still be eluted efficiently in the presence of excess glutathione, which was not the case for the protein aggregates we observed (Fig. 5B). Furthermore, when recombinant GST was expressed on its own in BL21(DE3) cells and subjected to Factor Xa proteolysis under conditions identical to those used for digestion of GST-nsp1, a large fraction of GST could still be eluted from the affinity resin with glutathione (Fig. 5B). The nsp1 moiety is therefore the most likely cause of the observed aggregation, which was not related to the presence of the affinity resin, since it was also detected upon Factor Xa digestion of GST-nsp1 in solution (data
not shown). Taken together, these results prompted us to better characterize the properties of purified GST-nsp1. The fusion protein was eluted from GS4B and dialyzed against a wide range of buffers with different pH and ionic strength, all of which contained 1 mM CaCl₂ and 50 mM L-Arg (essential components of Factor Xa buffer). The aggregation state of the fusion protein was then assessed by dynamic light scattering, a method for determination of particle size in solution. The hydrodynamic radius of GST-nsp1 particles indicated that the fusion protein is largely present as “soluble aggregates” in the majority of buffers tested, or completely precipitated under certain conditions (data not shown). In view of the limitations on buffer additives compatible with Factor Xa activity outlined above, further optimization aimed at improving GST-nsp1 solubility was not pursued.

Figure 5. Protein precipitation during Factor Xa cleavage of GST-nsp1. (A) Expression and purification of GST-nsp1, followed by on-resin cleavage of the GST tag by Factor Xa Protease were performed essentially as in Fig. 4A, apart from the absence of CHAPS from the cleavage buffer. Cleaved nsp1 was recovered from the buffer fraction after pelleting of GS4B by centrifugation, and unprocessed and/or resin-associated proteins were eluted from GS4B with 1x LSB. Equal volumes of LSB eluate and cleaved (C) protein fractions were analyzed by SDS-PAGE and Coomassie blue staining. (B) GST-nsp1 was expressed in BL21(DE3) or C43(DE3) cells. GST was expressed in BL21(DE3) cells that were transformed with empty pGEX-5X-3 vector. GS4B-bound GST-nsp1 and GST only were subjected to proteolysis with Factor Xa Protease in FXa buffer and cleaved proteins (designated with C) were recovered from the buffer fraction. Proteins that remained bound to GS4B after the proteolytic digestion were eluted once with 10 mM reduced glutathione in 20 mM Tris pH=7.5, 150 mM NaCl (E), followed by elution with 1x LSB. Equal volumes of the C, E and LSB fractions were analyzed by SDS-PAGE followed by Coomassie Blue staining.

Expression and purification of hexahistidine-tagged nsp1

The difficulties encountered during expression and purification of recombinant GST-nsp1 were seemingly related to the inherent instability and/or improper folding of nsp1 when fused to the C-terminus of GST. Taking this into consideration, fusion to another large protein tag with solubility-enhancing properties was not tested. Instead, a (His)₆ tag was fused to the N- or C-terminus of nsp1. A (His)₆ tag allows for convenient one-step protein purification by immobilized metal affinity chromatography. In general, this small tag has little effect on protein folding or stability, and consequently, it rarely needs to be removed from the target protein after purification. However, it does not alter the solubility properties of the target protein. It was therefore surprising that, upon IPTG induction
of (His)$_6$-tagged nsp1 expression in BL21(DE3) cells transformed with pDEST14-nsp1-HN or pDEST14-nsp1-HC, robust expression of a 30-kDa protein was detected only in the induced cultures, and was present predominantly in the soluble fractions of bacterial lysates (Fig. 6A). Western blotting with an nsp1-specific antibody confirmed the identity of the recombinant proteins produced in induced cultures (data not shown). Both nsp1-HN and nsp1-HC bound very poorly to Ni-NTA agarose, however, and the majority of both recombinant proteins were found in the flow-through fractions (Fig. 6, B and C). Similar results were obtained when a different metal affinity matrix (HIS-Select HF Nickel Affinity Gel) was used, and washing out unbound proteins in the absence of imidazole did not improve the recovery of nsp1-HN or nsp1-HC neither, but increased the amount of contaminants in the elution fractions (data not shown). The PCPβ domain of nsp1 serves to autoproteolytically release the C-terminus of the protein from the rest of pp1a during EAV infection, cleaving a Gly-Gly sequence, and since the protease domain is active in E. coli, it might have cleaved off the C-terminal (His)$_6$ tag in the nsp1-HC protein. Expression of an active-site protease mutant (C164S) from pDEST12-nsp1-HC, however, did not improve the binding of nsp-HC to metal affinity matrices (data not shown). These results suggest that neither the N-terminal nor the C-terminal (His)$_6$ tag is exposed for binding metal ligands under the experimental conditions used, while a control (His)$_6$-tagged protein could be purified efficiently under identical conditions (data not shown). Nevertheless, both nsp1-HN and nsp1-HC seemed to be stably folded in bacterial cells in the absence of solubility-enhancing fusion tags indicating to importance of authentic terminal sequences for folding. The protein’s N-terminal subdomain is a ZF (Fig. 1B), with the first of predicted zinc-coordinating residues located relatively close (24 amino acids)

**Figure 6.** Expression and purification of His-tagged nsp1. (A) BL21(DE3) cells transformed with pDEST14-nsp1-HN or pDEST14-nsp1-HC were cultured until OD$_{600}$ of 0.7, recombinant protein expression was induced by the addition of 0.5 mM IPTG, and the cultures were incubated at 30°C. Lysates were obtained from non-induced or IPTG-induced cells harvested 18 hours postinduction. After centrifugation of total cell lysates, the insoluble pellet (P) fractions were resuspended in 1x LSB, and equal volumes of pellet and supernatant (S) fractions were separated on a 12.5 % polyacrylamide gel by SDS-PAGE. Proteins were visualized by subsequent staining of the gel with Coomassie blue. (B, C) The soluble fraction of the total bacterial lysates from induced cells in (A) was incubated with Ni-NTA Agarose. The beads were pelleted by centrifugation, and the flow-through (FT) was collected for analysis of unbound proteins. Contaminants were then removed by washing the beads twice with buffer B containing 10 mM imidazole. Bound nsp1-HN (B) or nsp1-HC (C) was subsequently eluted with buffer B supplemented with 250 mM imidazole in two steps. Proteins from the pellet and soluble fractions of the total bacterial lysates, as well as the FT, wash (W) and elution (E) steps were analyzed by SDS-PAGE and Coomassie Blue staining.
Another viral autoprotease, the alphavirus Sindbis virus core protein, is rendered proteolytically inactive following autocatalytic cis cleavage in this manner. A C-terminal (His)_6 tag would thus also be inaccessible for binding to affinity resins. To overcome these intrinsic problems with tag accessibility, attempts were made to purify individual (His)_6-tagged nsp1 subdomains of nsp1. Constructs encoding only the ZF and PCPα domains (amino-acids 1 to 140 or 1 to 156 from EAV pp1a) fused to a C-terminal (His)_6 tag expressed very poorly in bacterial cells and were mostly insoluble, while expression of an N-terminally (His)_6-tagged PCPβ (amino-acids 151 to 260 from EAV pp1a) was not detected at all in induced cultures (data not shown).

**Importance of nsp1 autopropeolytic processing for recombinant protein stability**

In the course of this project, the crystal structure of nsp1α from another arterivirus, porcine reproductive and respiratory syndrome virus (PRRSV; strain XH-GD) was reported by Sun and co-workers. The N-terminal part of PRRSV pp1a/pp1ab also contains a ZF domain followed by two PCP domains. Unlike its ortholog in EAV, PCPα is proteolytically active, mediating the self-release of nsp1α from the viral replicase polyproteins, while the PCPβ domain cleaves nsp1β off. Homology between EAV nsp1 and PRRSV nsp1α is thus limited to the first two subdomains of EAV nsp1. Nevertheless, PRRSV nsp1α is also essential for sg mRNA production in virus-infected cells, suggesting similar functions of the two distantly related proteins in the replicative cycle of arteriviruses. In the report by Sun and co-authors, PRRSV nsp1α was purified after expression of N-terminally (His)_6-tagged nsp1α-nsp1β sequence in *E. coli* in order to obtain nsp1α in its self-processed form. Interestingly, a number of nsp1α C-terminal residues were indeed found to reside in the substrate-binding pocket of the protease domain in a well-stabilized form that would preclude further proteolytic reactions. Thus, autopropeolytic processing of PRRSV nsp1α might very well be a prerequisite for correct folding and stability of the protein – a notion further supported by the insolubility of a protease active-site mutant, C76S, when expressed in *E. coli* under a variety of conditions.

Although the problems we encountered while attempting to purify (His)_6-tagged EAV nsp1 did not seem related to the protein’s solubility but rather to tag accessibility, we could not exclude that nsp1-HN and nsp1-HC were also present as soluble aggregates in lysates fractions, which could also account for poor binding to metal affinity resins. We therefore set out to determine whether bacterially expressed self-cleaved nsp1 could be purified with more success. To this end, the N-terminal 429 amino-acids from EAV pp1a, fused to an N-terminal (His)_6 tag (see Fig. 1C), were cloned in pDEST14, resulting in expression plasmid pDEST14-nsp1-2N-HN. Nsp1 expressed from similar constructs was previously shown to efficiently mediate its autopropeolysis both in vitro and in vivo.
and the difference in molecular mass between the precursor protein (47 kDa) and native nsp1 (30 kDa) is sufficiently large to be easily seen by SDS-PAGE. Recombinant (His)\textsubscript{6}–tagged nsp1 obtained from this construct is referred to as nsp1-HN-ap to distinguish it from nsp1-HN, although the primary structure of the two proteins is identical.

The wild-type pDEST14-nsp1-2N-HN and a mutant derivative carrying a C44H substitution in nsp1 (a mutation of a zinc-coordinating residue in nsp1 that renders sg mRNA production undetectable when introduced into the infectious cDNA clone of EAV) were transformed in C41(DE3) cells, and protein expression and purification were performed following a protocol similar to that used by Sun and co-workers for the purification of PRRSV nsp1\textalpha. Briefly, transformed cells were grown at 37°C to mid-log phase and the culture temperature was lowered to 16°C during protein production, which was induced by addition of 1 mM IPTG. The levels of recombinant protein in the induced cultures were not sufficiently high to be easily detected by Coomassie staining (Fig. 7A, top panels), but Western blotting with an nsp1-specific antibody revealed robust expression of both nsp1-HN-ap and its mutant derivative, nsp1-HN-ap C44H (Fig. 7A, bottom panels). Both proteins were found predominantly in the supernatant fractions of bacterial lysates and exclusively in self-processed form, since virtually no precursor proteins were detected. These results suggest that nsp1 autoproteolysis with this combination of expression construct and expression conditions is very efficient in \textit{E. coli}, despite the low temperature of protein production. The C44H mutation also does not seem to influence the efficiency of nsp1 self-processing, in line with previous results showing that mutations of zinc-coordinating residues do not affect nsp1 autoprocessing \textit{in vitro} (Chapter 5 of this thesis).

Wild-type nsp1-HN-ap could be efficiently purified by metal affinity chromatography with Talon resin (charged with cobalt ions), albeit in the presence of a number of abundant protein contaminants (Fig. 6A, top left panel). Also, some of the protein was still found in the fraction of proteins that did not bind to the affinity resin. Nevertheless, this protocol resulted in the purification of \textasciitilde{}0.2 mg recombinant nsp1-HN-ap from 1 l of bacterial culture (data not shown). The yield of nsp1-HN-ap C44H, however, was much lower (Fig. 7A, top right panel), and immunoblot analysis of fractions obtained during purification revealed that most of the mutant recombinant protein did not bind the Talon affinity resin (Fig. 7A, bottom right panel). The C44H substitution might interfere with zinc binding and thus adversely affect the conformational stability of the N-terminal part of nsp1, and possibly of the whole protein, if interdomain interactions are required for its correct folding. This might not impede the PCP\textbeta domain-mediated autoproteolytic release of nsp1, which most probably occurs co-translationally\textsuperscript{14,15}, but could result in protein aggregation that could, in turn, render the (His)\textsubscript{6} inaccessible for metal binding.

In order to improve sample purity, gel-filtration chromatography of pooled eluates containing recombinant nsp1-HN-ap (Fig. 6A) was performed in 20 mM MES pH=7.0, 500 mM NaCl, 20 % glycerol, 1 mM DTT. The bulk of wild-type nsp1-HN-ap eluted in a single peak with an estimated molecular mass of 63 kDa (Fig. 7B), consistent with a dimer. Column fractions from the single elution peak preceding that of dimeric nsp1 contained largely a \textasciitilde{}75-kDa contaminant (Fig. 7B, lower panel, a-b), and higher oligomeric forms
Figure 7. Expression and purification of autoproteolytically processed His-tagged nsp1. (A) C41(DE3) cells were transformed with wild-type pDEST14-nsp1-2N-HN or a mutant derivative of the plasmid carrying a C44H substitution in nsp1. Induction of recombinant protein expression and purification of His-tagged proteins was performed as described in Materials and Methods. Samples obtained from total lysate of non-induced (-) cells, pellet (P) and supernatant (S) fractions of total lysates from IPTG-induced cells, and fractions obtained during purification (FT, flow-through) were resolved by SDS-PAGE. The gels were stained with Coomassie blue (top panels) or blotted onto Hybond-P PVDF membrane (GE Healthcare) by semi-dry Western blotting (WB, bottom panels). The membranes were subsequently incubated with an nsp1-specific monoclonal antibody (12A4), followed by an HRP-conjugated secondary antibody (DAKO) and an ECL-Plus kit (GE Healthcare) for detection. The expected position of the unprocessed nsp1-2N-HN protein (predicted molecular mass: 46.6 kDa) is indicated with an asterisk. (B, C) Top: Gel-filtration profile of wild-type nsp1-HN-ap (B) or nsp1-HN-ap C44H (C) obtained upon size exclusion chromatography of the pooled
197

Purification of recombinant EAV nsp1

of nsp1 were thus unlikely to be present in solution. The resolution of the gel filtration column was, however, insufficient to achieve a substantially improved purity of recombinant nsp1. Despite the low protein yield and purity, the fractions eluted from the Talon resin after purification of nsp1-HN-ap C44H (Fig. 7A, left panel) were also analyzed by gel filtration, to assess whether the C44H mutation is associated with a change in oligomeric state. The elution profile of nsp1-HN-ap C44H was much less symmetric due to the complexity of the sample, but the mutant protein was only detected in fractions from two elution peaks with approximately equal areas and estimated molecular masses of 57 kDa and 34 kDa (Fig. 7C). The C44H thus seems to destabilize, but not completely prevent nsp1 self-association. It is difficult to establish, however, whether this may be attributable to an importance of the ZF domain for dimerization, or to a more general defect in the overall fold of nsp1 (see above). Nevertheless, recombinant nsp1 purified by two different protocols forms dimers in solution (Fig. 4B and 7B), even in the presence of 1 mM DTT and high salt concentrations (500 mM NaCl, Fig. 7B). It is therefore unlikely that the observed dimerization is caused by tag sequences in the recombinant protein. Also, in view of nsp1’s homo-oligomerization in infected cells19, the ability of recombinant nsp1 to form dimers might be of functional significance for its role(s) in the EAV replicative cycle.

Though nsp1-HN-ap was efficiently resolved by size-exclusion chromatography in 20 mM MES pH=7.0, 500 mM NaCl, 20 % glycerol, 1 mM DTT, the protein was unstable in this buffer, since precipitation was observed during concentration of nsp1-containing chromatography fractions. The loss of protein was even more pronounced after overnight storage at 4°C, or when the sample buffer was exchanged with one of a lower ionic strength (20 mM MES pH 7.0, 200 mM NaCl, 50% glycerol). In an attempt to obtain a purer and more stable preparation of recombinant nsp1-HN-ap, the protein was expressed and purified as above, with minor adjustments to the original protocol (described in detail in Materials and Methods). Beta-mercaptoethanol was added at 5 mM to buffer C prior to bacterial lysis and Talon purification, and a single-step elution with 100 mM imidazole, followed by three elution steps with 250 mM imidazole in buffer C. Fractions from cell lysis and Talon purification analyzed by SDS-PAGE and Coomassie blue staining are denoted as in (A). (E) Eluates 2 – 4 from (D) were pooled and subjected to size-exclusion chromatography on a Superdex-75 column in 20 mM MES pH=7.0, 1M NaCl, mM DTT. Nsp1-containing peak fractions were concentrated in the same buffer, or in the presence of 0.1 % β-octylglucoside (β-oct). Equal sample volumes obtained immediately after concentration (1, 3) or following overnight incubation of the concentrated samples at 4°C (2, 4) were resolved by SDS-PAGE and visualized by Coomassie blue staining.
of proteins bound to the Talon resin was performed as follows: first, proteins that were bound with low affinity were eluted with 100 mM imidazole in buffer C, followed by three elution steps with 250 mM imidazole in buffer C. Most of the contaminants were eluted in the 100 mM imidazole step, together with a relatively low proportion of recombinant nsp1-HN-ap. The nsp1-containing eluate fractions obtained in this way were of higher purity (Fig. 7D). The pooled eluates were loaded onto a Superdex-75 column, gel-filtration chromatography was carried out in 20 mM MES pH=7.0, 1M NaCl, 20 % glycerol, 5 mM DTT. The peak column fractions that contained nsp1 were split in two samples that were subsequently concentrated at 4°C by ultrafiltration in the absence of additives, or in the presence of 0.1 % β-octylglucoside. There was no detectable protein precipitation during ultrafiltration in both buffers, and SDS-PAGE showed similar levels of recombinant nsp1 present in the two samples immediately after protein concentration, as well as following overnight incubation at 4°C (Fig. 7E). The yield of nsp1-HN-ap after Talon affinity purification and after gel filtration and was estimated at 0.3 mg and 0.25 mg/ l culture, with the minor loss of protein likely due to exclusion of gel-filtration fractions that contained a low level of nsp1 and high levels of contaminating proteins (data not shown). The inclusion of an additional purification step, i.e. ion exchange chromatography, might remove co-purifying proteins more efficiently to obtain a sample that is sufficiently pure for structure determination.

CONCLUDING REMARKS

Despite recent insights into successful strategies for the expression and purification of recombinant proteins emerging from large-scale structural genomics studies (see e.g. reference7), optimal conditions for each protein still have to be determined empirically. Our experience greatly underscores the importance of the choice of fusion tag that can be facilitated by at least some prior knowledge about the domain structure and/or biochemical properties of the target protein. The extensive aggregation we observed upon expression of nsp1 as a C-terminal GST fusion might be due to a number of factors, such as e.g. improper folding of the ZF domain when the N-terminus of nsp1 was blocked by its fusion to GST, or the formation of GST dimers forcing nsp1 dimerization before the protein has reached its stably folded state. Ultimately, successful purification of soluble recombinant nsp1 was achieved with nsp1 which was autoproteolytically released from a precursor polypeptide in E. coli, an observation that is in line with results of other studies in which soluble viral proteins that are derived from proteolytic processing of longer polypeptides were obtained5,16. It is thus tempting to suggest that mimicking the natural way of protein production may be a generally applicable approach for obtaining soluble recombinant viral proteins. This strategy could also constitute an important addition to the widely-used approaches of expressing subdomains and/or truncated proteins for structural and functional studies.
ACKNOWLEDGEMENTS

We thank Martijn van Hemert and Sjoerd van den Worm for helpful discussions, and Linda Struijk and Els van der Meijden for the pGEX-5X-3 plasmid.
REFERENCES


