Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease

S. SHANE TAREMI, BRIAN BEYER, MAUREEN MAHER, NANHUA YAO, WINIFRED PROSIDE, PATRICIA C. WEBER, AND BRUCE A. MALCOLM

Department of Structural Chemistry, Schering-Plough Research Institute, Kenilworth, New Jersey 07033

(RECEIVED August 13, 1998; ACCEPTED August 17, 1998)

Abstract

Efficient proteolytic processing of essential junctions of the hepatitis C virus (HCV) polyprotein requires a heterodimeric complex of the NS3 bifunctional protease/helicase and the NS4A accessory protein. A single-chain recombinant form of the protease has been constructed in which NS4A residues 21-32 (GGSVIVGRIILS) were fused in frame to the amino terminus of the NS3 protease domain (residues 3-181) through a tetrapeptide linker. The single-chain recombinant protease has been overexpressed as a soluble protein in E. coli and purified up to homogeneity by a combination of metal chelate and size-exclusion chromatography. The single-chain recombinant protease domain shows full proteolytic activity cleaving the NS5A-5B synthetic peptide substrate, DTEDVVCCSMSYTWTGK with a \( k_a \) and \( k_{cat} \) of 20.0 \( \pm \) 2.0 \( \mu \)M and 9.6 \( \pm \) 2.0 \( \text{min}^{-1} \), respectively, parameters identical to those of the authentic NS31-631/NS4Al~s4 protein complex generated in eukaryotic cells (Sali DL et al., 1998, Biochemistry 37:3392-3401).

Keywords: hepatitis C virus; kinetics; NS3; NS4A; overexpression; protease; recombinant

Hepatitis C virus (HCV) is the major etiologic agent of non-A non-B (NANB) hepatitis with estimated human seroprevalence of 1% globally (Alter & Mast, 1994). The majority of infections result in chronic hepatitis with a significant fraction of those progressing to cirrhosis and hepatocellular carcinoma (Stadhouders & Cooreman, 1997). The current therapeutic regimens are effective in less than 50% of the cases and consequently new anti-HCV agents are actively being sought (Sharara, 1997).

HCV, like all members of the flaviviridae, is a (+)-stranded RNA virus whose genome encodes a single polyprotein of about 3,000 amino acids which undergoes subsequent proteolysis by host and viral enzymes to yield the mature viral proteins (for reviews see Lohmann et al., 1996; Simmonds, 1996; Neddermann et al., 1997; Reed & Rice, 1998). The virally encoded, highly specific NS3 proteolytic activity is responsible for the essential processing of the NS3-4A, NS4A-NS4B, NS4B-NS5A, and NS5A-NS5B junctions (Grakoui et al., 1993; Tomei et al., 1993; D'Souza et al., 1994; Failla et al., 1994; Lin et al., 1994) and consequently is a primary target for antiviral therapeutics. The proteolytic domain, a twin-\( \beta \)-barrel serine protease of the chymotrypsin family, constitutes the amino terminal third of the NS3 protein (residues 1-181) (Hahm et al., 1995; Kim et al., 1996; Love et al., 1996; Yan et al., 1998). A key feature of the NS3 protease is that one strand of its seven stranded N-terminal \( \beta \)-barrel structure is supplied in trans by the NS4A cofactor protein or peptide. Upon binding of the cofactor peptide, the disordered N-terminus of NS3 (residues 1-28), folds to generate a new \( \beta \) strand (A0) and an \( \alpha \) helix (a0) that encase residues 21-33 of NS4A (TANJI et al., 1995; Kim et al., 1996; Yan et al., 1998). This causes the NS4A strand to reposition the D1 strand of the N-terminal \( \beta \) barrel, which in turn stabilizes the a1 helix that bears the second member of the catalytic triad, His57 (Tanji et al., 1995; Kim et al., 1996; Yan et al., 1998). In addition, the refolding of the N-terminus causes the side chain of Asp81 to reorient from the guanidinium group of Arg155 to the iridazole ring of His57 to generate a properly arranged catalytic triad (Tanji et al., 1995; Kim et al., 1996; Yan et al., 1998). In summary, without intercalation of the NS4A strand the N-terminus remains partially disordered resulting in imperfect alignment of the catalytic triad and a corresponding ~950-fold drop in catalytic efficiency (Tanji et al., 1995; Kim et al., 1996; Yan et al., 1998).
Various strategies have been attempted over the last few years to generate milligram quantities of this key target enzyme in a fully active, tractable form for assay and structural analysis. Although modestly successful, expression of the full NS3,NS4AII-5, heterodimeric protein complex in eukaryotic systems yielded relatively small quantities of rather insoluble material insufficient in quantity or quality for structural analysis (Hirowatari et al., 1993; Overton et al., 1995; Suzuki et al., 1995; Steinkuhlcr et al., 1996; Morgenstern et al., 1997; Sali et al., 1998). Examination of the NS4A sequence suggested that the amino-terminal portion (residues 1-21) was primarily responsible for the lack of solubility of the NS3,NS4AII-5 protein complex and subsequently synthetic peptides, based only on the intercalating portion of NS4A (residues 22-33), were used in place of the NS4A protein (Rutkiewicz et al., 1996; Tomei et al., 1996). Proteolytic activity, however, remained an order of magnitude lower than the NS3,NS4AII-5 protein complex produced in the baculovirus system (Sali et al., 1998) and the recombinant NS3 catalytic domain complemented in trans with an excess of the synthetic peptide GSVVIVGRIILS (NS3,181 + NS4A peptide), is shown in Table 1. The kinetic parameters for the

soluble protein and retains full catalytic activity compared with the NS3,NS4AII-5 protein complex produced in eukaryotic cells.

Results

Recombinant single-chain NS4A21-32-GSGS-NS3,181 constituted approximately 5% of the intracellular protein following induction as shown in Figure 1. Following purification, yields approached 16 mg/L of culture. ESI-MS gave a monoisotopic mass of 20,753.9 ± 2.5 Da, which is in good agreement with the calculated value of 20,751.4 Da. Edman sequencing revealed the anticipated sequence beginning with the P' portion of the thrombin cleavage sequence (GSHM) and proceeding to that of the NS4A peptide (GSVVIVGRII).

Kinetic analysis

Kinetic parameters for the recombinant single-chain NS4A21-32-GSGS-NS3,181 protease were obtained from nonlinear regression analysis of Michaelis–Menten plots (Fig. 2). A comparison of Vmax and kcat of the recombinant single-chain NS4A21-32-GSGS-NS3,181, with that of the full length NS3,NS4AII-5 protein complex produced in the baculovirus system (Sali et al., 1998) and the recombinant NS3 catalytic domain complemented in trans with an excess of the synthetic peptide GSVVIVGRIILS (NS3,181 + NS4A peptide), is shown in Table 1. The kinetic parameters for the

\[ V_0 = \frac{[S]}{K_M + [S]} \]

Fig. 2. Dependence of initial velocity on substrate concentration for the recombinant constructs described in Table 1 (○ = NS3,181 + NS4A peptide, △ = NS3,NS4AII-5). Graph shows exemplary data sets for each construct (see Materials and methods).
Table 1. Kinetic parameters of HCV protease constructs

<table>
<thead>
<tr>
<th>Construct</th>
<th>$k_{cat}$ (min$^{-1}$)</th>
<th>$K_m$ (µM)</th>
<th>$k_{cat}/K_m$ (M$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS$3_{1-63}$/NS$4A_{1-54}$</td>
<td>10 ± 2</td>
<td>20 ± 2</td>
<td>(8 ± 2) × 10$^3$</td>
</tr>
<tr>
<td>NS$3_{1-181}$ + NS$4A$ peptide</td>
<td>3 ± 1</td>
<td>80 ± 20</td>
<td>(0.5 ± 0.2) × 10$^3$</td>
</tr>
<tr>
<td>NS$4A_{21-32}$-GS$S$-NS$3_{3-181}$</td>
<td>9 ± 2</td>
<td>19 ± 3</td>
<td>(8 ± 2) × 10$^3$</td>
</tr>
</tbody>
</table>

$^a$SA-5B cleavage junction substrate DTEDVCC*SMSYTWTG-K (see Materials and methods).

catalytic domain generated in E. coli and complemented with a 1,000-fold excess of the NS$4A$ peptide are measurably different ($k_{cat} = 3$ min$^{-1}$; $K_m = 80$ µM) from those of the single-chain recombinant NS$4A_{21-32}$-GS$S$-NS$3_{3-181}$ catalytic domain and the full length NS$3_{1-63}$/NS$4A_{1-54}$ protein complex generated in the insect cell system ($k_{cat} = 9-10$ min$^{-1}$; $K_m = 19-20$ µM). These differences suggest a greater than 15-fold improvement in the second order rate constant as a consequence of tethering the cofactor to the catalytic domain.

The pH profile (Fig. 3A) and the βOOG profile (Fig. 3D) of the single-chain construct (NS$4A_{21-32}$-GS$S$-NS$3_{3-181}$) suggests a pH optimum of 7.5 and a detergent optimum of 0.05%, which are comparable to that of the full NS$3_{1-63}$/NS$4A_{1-54}$ protein complex from the baculovirus expression system (studies reported in Sali et al., 1998). Unlike the full length complex however (see data in Sali et al., 1998), the single-chain construct shows decidedly higher ionic strength and glycerol optima (Fig. 3B,C). Over the range of the studies (0–2 M NaCl and 0–50%, respectively), there appears to be a simple monotonic increase in activity with increasing ionic strength and concentration of glycerol.

Discussion

Both kinetic and structural studies have shown that the HCV NS3 serine protease requires intercalation of a strand of the NS$4A$ cofactor for proper alignment of the catalytic triad and full proteolytic activity (Kim et al., 1996; Yan et al., 1998). Two approaches have been used to obtain the fully active target enzyme: (1) production of a full length NS$3_{1-63}$/NS$4A_{1-54}$ complex in eukaryotic expression systems (Hirowatari et al., 1993; Overton et al., 1995; Suzuki et al., 1995; Steinkuhler et al., 1996; Morgenstern et al., 1997; Sali et al., 1998) and (2) production of the protease domain alone in E. coli followed by complementation in trans with an excess of synthetic oligopeptide (Shoji et al., 1995; Vishnuvardhan et al., 1997).

The former approach, although feasible, is labor intensive, expensive, and ultimately provides a complex ill-suited to structural analysis by either X-ray diffraction or NMR spectroscopy due to
size and solubility. The latter method, although reasonably fast, inexpensive, and capable of generating material suitable for structural analysis, has been criticized for the low specific activity of the resulting material compared with the full length NS31-631/NS4A1-54 protein complex (Kim et al., 1996; Sali et al., 1998; Yan et al., 1998). Recent studies have shown that the NS4A core peptide's affinity for the protease is several orders of magnitude less than that of the full NS4A protein (Kim et al., 1996; Bianchi et al., 1997; Sali et al., 1998; Yan et al., 1998). Rather than attempt to modify the NS4A peptide cofactor to improve its intrinsic affinity for NS3, a fusion strategy to anchor the cofactor to the protease was devised.

Inspection of the cocrystal structure of the NS31-181/NS4A21-39 complex suggested that it might be possible to fuse the C-terminus of the essential portion of the NS4A peptide (residues 21–32) to the N-terminus of the catalytic domain using an appropriate spacer (Fig. 4). It was reasoned that if an appropriate connection could be engineered, then the tethered cofactor peptide could not undergo complete dissociation and the single-chain construct would more likely remain in the completely ordered state (Kim et al., 1996; Yan et al., 1998). The result would be a small soluble NS3 protease with full catalytic activity requiring no exogenous cofactor. Various linkers, some intentionally unstructured and others incorporating actual turn sequences from known structures (e.g., PKG, PAP, PAG, etc.), were considered to connect the nearest residues of the C-terminal region of the NS4A cofactor and the N-terminal region of the NS3 protease that appeared to retain structural importance—Ser32 and Ile3, respectively (Kim et al., 1996; Yan et al., 1998).

After evaluation by visual inspection and energy minimization, four linkers—PAG, PAGG, GSG, and GSGS—were chosen for incorporation into constructs. Although the PAG, PAGG, and GSG containing constructs showed some degree of soluble expression (≤1% of the intracellular protein) and comparable specific activity (data not shown), the flexible GSGS linker was found to yield the highest levels (~5% of the intracellular protein) and consequently was chosen for further study. Kinetic characterization suggests that the catalytic efficiency of the single-chain catalytic domain construct is greater than the NS3 catalytic domain complemented in trans by the equivalent NS4A peptide (NS31-181 + NS4A peptide) and identical to that of the full length NS31-631/NS4A1-54 protein complex. Although identical interactions would be expected between the backbone and side chains of the intercalated portion of the cofactor and the protease domain in all cases, it is reasonable to assume that covalently restraining the peregrinations of the C-terminus of the peptide (not to mention preventing complete dissociation) must reduce disorder in the system with respect to the NS31-181 + NS4A peptide system (whose $K_d$ has been estimated to be 3.8–7.7 μM (Bianchi et al., 1997)), yielding a more stable active site comparable to that of the full length complex (for which the dissociation appears to be sub-nanomolar (Sali et al., 1998)).

The similarities and differences of the single-chain proteolytic domain and the full length NS31-631/NS4A1-54 protein complex with respect to global parameters such as pH, ionic strength, glycerol, and detergent optima can be understood through their common and distinct elements. The fact that both proteases exhibit the same pH optimum is not surprising, as this parameter directly reflects the ionization and structural alignment of their catalytic machineries (Perona & Craik, 1995 and references therein), which appear, from kinetic analyses, to be identical. The comparable effect of detergent and low levels of salt and glycerol on both systems likewise suggests that this is the result of impact on the common elements; either the proteolytic domain (including the intercalated portion of NS4A) and/or peptide substrate binding. The deleterious result of high ionic strength and glycerol on only the NS3/NS4A protein complex suggests that these negative effects are mediated by the distinct elements peculiar to the NS3/NS4A protein complex; specifically the nonintercalated portions of NS4 (residues 1–21 and 33–54) and the helicase domain of NS3 (residues 182–631). The single-chain construct, lacking these pieces, is therefore refractory.

Fig. 4. Ribbon diagram of the proteolytic domain of NS3 showing the tethering of the NS4A cofactor. A: X-ray structure of NS31-181 + NS4A22-35 peptide. B: A model of NS4A21-32-GSGS-NS31-181. The intercalated portion of the construct corresponding to the NS4A cofactor peptide (residues 21–32) is shown in red and the GSGS linker in yellow. The catalytic triad (Ser139, His57, and Asp81) side chains are shown as ball and stick models.
Single-chain HCV protease

Although there may be subtle differences between the full NS3₁₋₆₃₁/NS₄A₁₋₅₄ protein complex and the single-chain proteolytic domain (NS₄A₂₁₋₃₂-GSGS-NS₃₁₈₁) with regard to processing of polyprotein substrates, prima facie kinetic evidence suggests that the single-chain construct contains an optimally organized active site identical to that of the NS₃₋₆₃₁/NS₄A₁₋₅₄ protein complex and should show a similar specificity profile with these substrates. Structural and polyprotein cleavage studies are in progress to confirm these predictions. In the interim, this novel single-chain construct should allow the facile generation of reagent quantities of this important biomedical target and facilitate the identification of medicinally useful inhibitors.

Materials and methods

Molecular biology

Restriction enzymes, Vent polymerase, and ThermoPol buffer were obtained from New England Biolabs (Beverly, Massachusetts). PCR reactions were carried out in a Perkin Elmer (Foster City, California) Cetus, model 480 DNA thermocycler. pET vectors and BL21(DE3) cells were obtained from Novagen (Madison, Wisconsin). DH5α cells were purchased from Gibco, BRL (Gaithersburg, Maryland). Clone BK 138-1, described in Takamizawa et al. (1991) encoding the entire NS3 region of the 1b/BK strain of HCV, was obtained from the authors. Standard recombinant DNA methods were carried out essentially as described by Sambrook et al. (1989). DNA sequencing was performed according to the Sanger-Dideoxy method by Bioserve Biotechnologies (Laurel, Maryland).

Protein analysis

Protein concentrations were determined by the BIORAD dye method against recombinant HCV standards previously quantified using amino acid analysis. N-terminal sequence analysis was performed on an Applied Biosystems Inc. (Foster City, California) model 480A cation exchange liquid protein sequencer with an on-line PTH analyzer (ABI model 140C). Samples were spotted onto PVDF membranes (ProSorb, ABI, CA) and were subjected to a minimum of 10 cycles of analysis. Electrospray ionization mass spectrometry (ESI-MS) was performed on a Sciex API III triple quadrupole mass spectrometer (Ontario, California) fitted with a pneumatically assisted electrospray (ionspray) device. The sample solution (25–50 μM in 5 mM ammonium acetate, pH 6.5) was infused at 5 μL/min with an orifice voltage of 80–100 V. Multiple scans from 1,650–2,400 Da were acquired for each sample. The standard deviation of the mass estimates ranged from 2–3 Da.

Peptide synthesis

The P8P'8K substrate peptide derived from the 5A–5B junction of the 1a/H strain of HCV was synthesized using standard Fmoc solid phase synthesis protocols on an Applied Biosystems Inc., 431 A Peptide Synthesizer as previously described (Zhang et al., 1997). The substrate was purified using reverse-phase HPLC and the mass confirmed by plasma desorption mass spectrometry using a Bioion 20 California-252 time-of-flight instrument as before.

Assay of protease activity

The activity of protease constructs was determined using the peptide cleavage assay previously described (Sali et al., 1998). In brief, varying amounts of the 5A/SB (P8P'8) substrate peptide DTEDVVC*SMSYTWG-T-K (the * denotes the scissile bond; the C-terminal K was added to facilitate separation of the product) were incubated with appropriate amounts of the recombinant proteases in 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5 mM EDTA, 10 mM DTT, 10% glycerol, and 0.05% βOG at room temperature for 15 min and subsequently quenched with an equal volume of 10% phosphoric acid. Activity profiles varying ionic strength, percent glycerol, and concentration of βOG were conducted in the above buffer. The pH profile study was conducted using the standard solution conditions above with the following buffers: 25 mM MES-HCl, pH 6.0–6.5, 25 mM MOPS-HCl, pH 7.0, and 25 mM Tris-HCl, pH 7.5–9.0.

Cleavage products were quantified by reverse-phase HPLC. Peak areas for the P' product (P'P'8-K) were converted to nanomoles using a standard curve calibrated using authentic material. Initial rates were determined from samples showing ≤10% substrate cleavage. Kinetic parameters were obtained using nonlinear regression analysis of initial rates as a function of substrate concentration (Enzyme Kinetics; Trinity Software, Campton, New Hampshire). All parameters represent the weighted average of three independent determinations.

Recombinant NS3 catalytic domain (NS₃₁₋₁₈₁) without the fused NS4A peptide was precincubated with synthetic NS4A peptide (KKGSVVVTGRVLGSKKAP) for 15 min at 4°C to generate the activated complex (NS₃₁₋₁₈₁+NS₄₄₄₄₄₄peptide) prior to kinetic analysis (NS₃ final concentration 250 nM, NS₄₄₄₄₄₄peptide 20 μM). The mixture was then diluted directly into the proteolytic reaction to give a final concentration of 100 nM NS₃ catalytic domain/8 μM NS₄₄₄₄₄₄peptide. Full length NS₃₋₆₃₁/NS₄A₁₋₅₄ protease was prepared as previously described (Sali et al., 1998).

Construction of recombinant single-chain NS₄₄₄₄₄₄₂₁₋₃₂-GSGS-NS₃₁₋₁₈₁

Initial DNA sequencing of the BK 138-1 plasmid harboring the entire NS3 bifunctional protein revealed four nonsynonymous differences with the published sequence (Takamizawa et al., 1991), at positions aa 66/nt 3606 (A → G; ggc → gga), aa 86/nt 3666 (P → Q; ccc → cag), aa 87/nt 3669 (K → A; aag → gcc), and aa 147/ nt 3879 (F → S; ttc → tcc) of the protease catalytic domain. A 5' PCR primer encoding: (1) an Nde 1 site, (2) residues 21–32 of NS4A, and (3) a tetrapeptide linker, Gly-Ser-Gly-Ser along with residues 3–9 of NS3, in frame with the pET 28b+ cistron, 5'GATAATCATAATGGTCTCGTTGTTATTGTTGGTGAATATTTTATCTGGTAGTGGTAGTAA ACTCC-3' was used along with a 3' PCR primer complementary to amino acid residues 175–181 of NS3 as well as encoding an in frame opal termination codon flanked by an EcoRI site 5'CTCAG CGAATTCTTACAGGCCGATATGTGTTCCAT3' to amplify (in frame) residues 3–181 of plasmid BK 138-1. To improve fidelity, Vent DNA polymerase was used following the manufacturer's recommendations to amplify the DNA. Final primer/template/dNTP concentrations were: primers 0.1 μM; template 0.1 μg/mL; dNTPs 2.5 mM each. The PCR conditions were as follows: 95°C for 45 s (pre-cycle); 95°C for 30 s, 55°C for 1 min, 72°C for 2 min (25 cycles). The amplified 616 base pair fragment was purified and double digested with EcoRI and Ndel, as were aliquots of the expression vector pET-28b+. The digested fragment and vector were then gel purified, ligated overnight at 15°C, and used to transform compe-
tent DH5α cells according to the manufacturer’s protocol. Transformed cells were selected on LB agar plates with kanamycin (25 μg/mL). Recombinant clones were identified by restriction analysis and three were selected for complete sequence verification. The resulting recombinant single-chain constructs fused the NS4A cofactor (residues 21–32), the GSGS linker, and the NS3 protease (residues 3–181) followed by an opal stop codon, in frame to the histidine leader sequence (MGSSHHHHHHSSGLVPRGSHM) of pET-28b+. An identical construct, but lacking the NS4A cofactor and linker, was assembled to generate the catalytic domain alone (NS31–181).

Expression and purification

Both the single-chain proteolytic domain (NS4A21–32–GSGS–NS31–181) and the catalytic domain alone (NS31–181) were expressed and purified following the procedure described below. Single colonies from plates of freshly transformed BL21 (DE3) cells were used to initiate growth of all recombinant protease constructs. Terrific Broth medium was supplemented with 25 μg/mL kanamycin. When the cell density reached an OD600 of 2–3, the fermenter was stopped and purified following the procedure described below. Single colonies from plates of freshly transformed BL21 (DE3) cells were used to initiate growth of all recombinant protease constructs. Terrific Broth medium was supplemented with 25 μg/mL kanamycin. When the cell density reached an OD600 of 2–3, the fermenter was stopped

Protease, as judged by SDS-PAGE, were pooled and flash-frozen at –80°C.

Acknowledgments

The authors would like to thank Drs. Yan-hui Liu, Charles McNemar, and Runmin Zhang for their support with mass spectrometry, protein sequencing, and peptide synthesis, respectively, and acknowledge the expert technical assistance of Hernando de Fex, James Durkin, and Richard Ingram in the preparation of peptide substrates and protein complexes.

References


