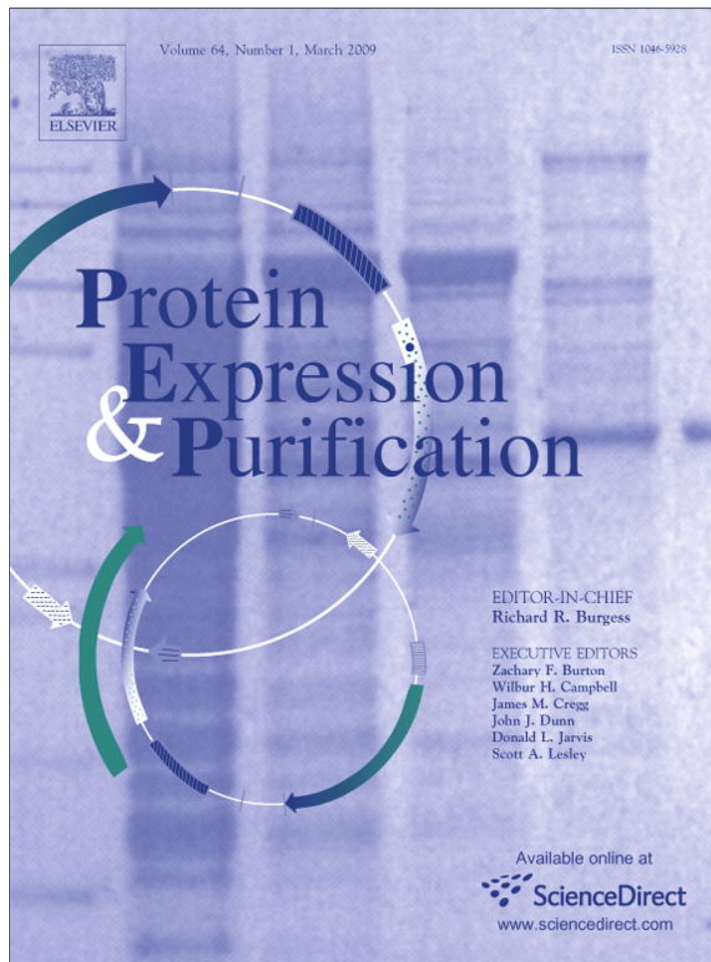
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Protein Expression and Purification 64 (2009) 89–97

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Molecular cloning, overproduction, purification and biochemical characterization of the p39 nsp2 protease domains encoded by three alphaviruses

Di Zhang [a], József Tözsér [b], David S. Waugh [a,*]

[a] Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, P.O. Box B, Frederick, MD 21702-1201, USA
[b] Department of Biochemistry and Molecular Biology, Faculty of Medicine, Medical and Health Science Center, University of Debrecen, P.O. Box 6, Debrecen H-4012, Hungary

ARTICLE INFO

Article history:
Received 20 October 2008
and in revised form 24 October 2008
Available online 30 October 2008

Keywords:
Protease
Affinity tag
Fusion protein
Maltose-binding protein
Venezuelan equine encephalitis virus
Sindbis virus
Semliki forest virus
Nonstructural protein 2

ABSTRACT

Alphaviruses cause serious diseases that pose a potential health threat to both humans and livestock. The nonstructural protein 2 (nsp2) encoded by alphaviruses is a multifunctional enzyme that is essential for viral replication and maturation. Its 39-kDa C-terminal domain (nsp2pro) is a cysteine protease that is responsible for cleaving a viral polyprotein at three sites to generate nonstructural proteins 1, 2, 3 and 4. In the present study, we evaluated nsp2pro domains from the following three sources as reagents for site-specific cleavage of fusion proteins: Venezuelan Equine Encephalitis Virus (VEEV), Semliki Forest Virus (SFV) and Sindbis Virus (SIN). All three alphavirus proteases cleaved model fusion protein substrates with high specificity but they were much less efficient enzymes than potyviral proteases from tobacco etch virus (TEV) and tobacco vein mottling virus (TVMV). Oligopeptide substrates were also cleaved with very low efficiency by the alphavirus proteases. We conclude that, in general, alphavirus nsp2pro proteases are not very useful tools for the removal of affinity tags from recombinant proteins although they do remain promising therapeutic targets for the treatment of a variety of diseases.

Published by Elsevier Inc.

The family of alphaviruses includes 26 known members [1]. They infect a variety of hosts including mosquitoes, birds, rodents and other mammals with worldwide distribution [2]. Alphaviruses also pose a potential threat to human health in many areas. For example, Venezuelan Equine Encephalitis Virus (VEEV) causes encephalitis in humans as well as livestock in Central and South America [3], and some variants of Sindbis Virus (SIN) and Semliki Forest Virus (SFV) have been found to cause fever and arthritis in humans [4,5]. VEEV has also has been weaponized by several nations including the United States [6]. Accordingly, it is classified as a Category B select agent by the Centers for Disease Control and Prevention.

Common laboratory strains of SFV and SIN are avirulent for humans [7] and have therefore served as useful models to study the life cycle and molecular biology of alphaviruses. Alphaviruses possess a single-stranded RNA genome of approximately 12 kb. They can be divided into three groups on the basis of sequence similarity: the VEEV/EEEV group, the SFV group and the SIN group [8]. The genomic RNA of alphaviruses is translated into two polyproteins that, respectively, encode structural proteins and nonstructural proteins. The structural proteins share about 45% amino acid sequence identity and the nonstructural proteins are about 60% identical [8]. The nonstructural proteins may be translated as one or two polyproteins, nsp123 or nsp1234, depending on the virus. These polyproteins are cleaved to generate nsp1, nsp2,[1] nsp3 and nsp4 by a protease activity that resides within nsp2.

The nsp2 protein of alphaviruses has multiple enzymatic activities. Its N-terminal domain has been shown to possess ATPase and GTPase activity [9], RNA helicase activity [10] and RNA 5′-triphosphatase activity [11]. The C-terminal domain of nsp2 is responsible for the regulation of 26S subgenome RNA synthesis [12], switching between negative- and positive-strand RNA synthesis [13], targeting nsp2 for nuclear transport [14] and proteolytic processing of the nonstructural polyprotein. Sequence analysis suggests that the latter function of nsp2 is mediated by a cysteine protease activity of peptidase family C9 and clan CA [15]. Site-directed mutagenesis has indicated that nsp2 utilizes a Cys/His catalytic dyad [16]. Its activity is specific for substrates with certain target sequences such as AGA↓ or AGC↓ [8]. Because nsp2 is essential for alphavirus replication and exhibits some degree of sequence specificity, it is both an attractive target for antiviral therapeutics and a potential biotechnological tool for the removal of affinity tags from recombinant proteins.

* Corresponding author. Fax: +1 301 846 7148.
E-mail address: waughd@ncifcrf.gov (D.S. Waugh).

1046-5928/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.pep.2008.10.013

[1] *Abbreviations used:* VEEV, Venezuelan Equine Encephalitis Virus; SIN, Sindbis Virus; SFV, Semliki Forest Virus; nsp2, nonstructural protein 2; TEV, tobacco etch virus; TVMV, tobacco vein mottling virus; ORF, open reading frame; PCR, polymerase chain reaction; Trx, thioredoxin; IPTG, isopropyl-β-D-thiogalactopyranoside; TCEP, Tris (2-carboxyethyl) phosphine hydrochloride; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; DTT, dithiothreitol.

Author's personal copy

The proteolytic activity of the nsp2pro domains from SIN and SFV has been investigated to some extent. Both enzymes cleaved their cognate nonstructural polyproteins [17,18], but large differences in the cleavage efficiencies of the three sites were observed. For SFV nsp2pro, an artificial substrate containing the p3/p4 site was processed rapidly, while an analogous construct containing the p1/p2 site was processed much less efficiently, and cleavage of the p2/p3 site in the same context was extremely poor [19–21]. However, using native polyproteins translated in vivo as substrates, p2/p3 was cleaved rapidly [18]. For SIN nsp2, cleavage efficiency was only investigated in the context of the native polyprotein. In experiments conducted in vitro, SIN nsp2 could only cleave the p1/p2 and p2/p3 sites but not the p3/p4 site [17,22]. The protease activity of VEEV has not been well characterized, although the crystal structure of the nsp2pro domain of VEEV is available [23].

In the present report, we compared the protease activity of the nsp2pro domains from VEEV, SFV and SIN, using MBP-NusG-His_6 fusion proteins with the P6–P5′ residues of the natural polyprotein cleavage sites in the linker between these two domains and oligopeptides as substrates. Because all three protease share similar recognition sites, we also tested their cross-reactivity. Additionally, the optimum reaction conditions for each enzyme were investigated and defined.

Materials and methods

Construction of nsp2 protease expression vectors

Plasmid clones encoding the nonstructural proteins of SIN and VEEV were a gift from Dr. Michael Parker of the United States Army Research Institute of Infectious Diseases. A plasmid clone encoding the nonstructural proteins of SFV was obtained from Dr. Juan Rivera of the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH. The open reading frame (ORF) encoding SIN nsp2pro (residues 462–805) was amplified using PE1941rep and PE1942rep (Table 1) as primers for polymerase chain reaction (PCR). The resulting PCR amplicon was subsequently used as the template for another PCR with primers PE277 and PE1942rep (Table 1), generating a second amplicon that was inserted by recombinational cloning into the destination vector pDEST-HisMBP [24] to yield pDZ1849 (Table 2). The plasmid expression vectors pDZ1830 and pDZ1827, which, respectively, encode HisMBP-VEEV nsp2pro (residues 457–792) and HisMBP-SFV nsp2pro (residues 459–797) fusion proteins, were constructed in the same manner, except primers PE1939 and PE1940 (Table 1) were used to generate pDZ1830 while PE1930 and PE1931 (Table 1) were used to generate pDZ1827. The ORFs of all three proteases were verified by DNA sequencing. SIN nsp2pro has a point mutation at residue 173 (residue 1 in nsp2pro corresponds to residue 462 of full-length nsp2), which changes an amino acid from Lys to Glu, a substitution that is present in several SIN sequences in various databanks and which occurs frequently at this position in many alphavirus nsp2 proteins, including that of SFV. This variable residue, which is a Ser in VEEV nsp2, is solvent-exposed in the crystal structure of VEEV nsp2pro [21] and located very far away from the active site catalytic dyad. Therefore, it is highly unlikely that the K173E mutation has any influence on the activity of SIN nsp2pro. The mutation was present in the plasmid template used for PCR (data not shown).

Construction of MBP-NusG-His_6 fusion protein expression vectors

The construction of pKM625, an MBP-NusG-His_6 fusion protein expression vector with a TEV protease recognition site inserted between the MBP and NusG domains was described previously [25]. The ORF encoding NusG-His_6 was amplified from pKM625, using a series of forward primers containing different nsp2pro protease recognition sites in conjunction with primer PE278 (Table 1). These PCR amplicons were inserted by recombinational cloning into the destination vector pKM596 [26] to generate a series of nine fusion protein expression vectors containing residues P6–P5′ of the nsp1/2, nsp2/3 and nsp3/4 cleavage sites for each alphavirus protease (Table 3) in the linker between MBP and NusG-His_6 domains (Table 2). A schematic representation of an MBP-NusG-His_6 fusion protein substrate is shown in Fig. 1A. All constructs were verified by DNA sequencing.

The fusion protein expression vector containing residues P6–P5′ of the SFV nsp3/4 cleavage site (Table 3) in the linker between MBP and NusG-His_6 domains (Table 2) plus an additional 10 alanine residues between the P5′ site and NusG (pDZ2082) was constructed by QuickChange site-directed mutagenesis (Stratagene, La Jolla, CA) and confirmed by DNA sequencing.

Table 1
Oligonucleotides used for PCR amplification.

Name	Polarity	Sequence (5′ → 3′)
PE277	Sense	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGGAGAACCTGTACTTCCAG
PE278	Antisense	GGGGACCACTTTGTACAAGAAAGCTGGGTTATTAGTGATGATGGTGGTGATG
PE1930	Sense	GAGAACCTGTACTTCCAGGGTATGAAGGTGATTGAAGGACCGGCTG
PE1931	Antisense	GGGGACCACTTTGTACAAGAAAGCTGGGTTATTAGGCCGTGTGCATGGCTTCTCCGGCATAC
PE1939	Sense	GAGAACCTGTACTTCCAGTCTATGAGGCACATCTTGGAGAGACC
PE1940	Antisense	GGGGACCACTTTGTACAAGAAAGCTGGGTTATTAGGCTTCGTGGAGTCTGGAACCTG
PE1941rep	Sense	GAGAACCTGTACTTCCAGTCTGCTGCAATAAACAGCCCCACTC
PE1942rep	Antisense	GGGGACCACTTTGTACAAGAAAGCTGGGTTATTAAACTCCATCTCTTGTACCCTCATAC
PE1978	Sense	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGCTGCAAGAGGCCGGTGCTGGCTCTGTCGAGACTAGCGAGCAACAGGTTCAGGAAC
PE1979	Sense	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGCTGCACGAGGCCGGTTGCGCCCCTTCTTACCATAGCGAGCAACAGGTTCAGGAAC
PE1980	Sense	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGCGTTTCGACGCCGGTGCTTACATCTTCTCTTCTAGCGAGCAACAGGTTCAGGAAC
PE1981	Sense	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGCAAGCCGACGTCGGTGCTGCCCTGGTCGAGACTAGCGAGCAACAGGTTCAGGAAC
PE1982	Sense	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGCGTGACGGTGTCGGTGCTGCCCCTGCTTACCGTAGCGAGCAACAGGTTCAGGAAC
PE1983	Sense	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGCTGACTGGTGTCGGTGGTTACATCTTCTCTACTAGCGAGCAACAGGTTCAGGAAC
PE1996	Sense	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGCTGGGCCGTGCCGGTGCTTACATCTTCTCTTCTAGCGAGCAACAGGTTCAGGAAC
PE2005	Sense	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGGAGTACCACGCCGGTGCTGGCGTCGTCGAGACTAGCGAGCAACAGGTTCAGGAAC
PE2006	Sense	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATGCACACTGCCGGTTGCGCCCCTTCTTACCGTAGCGAGCAACAGGTTCAGGAAC
PE2043	Sense	CGTTGATATCCGGGATTACTTTCGGAGACTTCGACGACGTCCTGCGACTAGGCCGCGCGGGTGCATATATTTTC
PE2044	Antisense	AGATTGGAATTCCTAACGGATTTTTGTTGTAAATGTCCGCTGCCAGTGTCCGAGGAGAAAATATATGCACCCGCGCGG
PE2094	Sense	CGTCCTGCGACTAACCGGCGTGGGTGGATATATTTTCTCCACCGACACTGGCAGCGG
PE2095	Antisense	CCGCTGCCAGTGTCGGTGGAGAAAATATATCCACCCACGCCGGTTAGTCGCAGGACG
PE2136	Sense	GCTTACATCTTCTCTTCTGCTGCTGCCGCCGCTGCCGCCGCTGCTGCCAGCGAGCAACAGGTTCAGG
PE2137	Antisense	CCTGAACCTGTTGCTCGCTGGCAGCAGCGGCGGCAGCGGCGGCAGCAGCAGAAGAGAAGATGTAAGC

Author's personal copy

Table 2
Expression vectors.

Expressed protein		Plasmid vector	Primers used
Protease expression vectors	SIN-nsp2pro	pDZ1849	PE277, PE1941rep and PE1942rep
	SFV nsp2pro	pDZ1827	PE277, PE1930 and PE1931
	VEEV nsp2pro	pDZ1830	PE277, PE1939 and PE1940
MBP-NusG-His$_6$ fusion protein expression vectors (substrates)	SIN p1/p2	pDZ1922	PE1981 and PE278
	SIN p2/p3	pDZ1924	PE1982 and PE278
	SIN p3/p4	pDZ1926	PE1983 and PE278
	SFV p1/p2	pDZ1959	PE2005 and PE278
	SFV p2/p3	pDZ1961	PE2006 and PE278
	SFV p3/p4	pDZ1954	PE1996 and PE278
	SFV p3/p4 + 10	pDZ2082	PE2136 and PE2137
	VEEV p1/p2	pDZ1916	PE1978 and PE278
	VEEV p2/p3	pDZ1918	PE1979 and PE278
	VEEV p3/p4	pDZ1920	PE1980 and PE278
Trx fusion vectors (substrates)	SFV p3/p4	pDZ1979	PE2043 and PE2044
	SIN p3/p4	pDZ2040	PE2094 and PE2095

Table 3
Alphavirus protease recognition sites (P6–P6′).

Virus	p1/p2	p2/p3	p3/p4
SFV	EYHAGA↓GVVETP	MHTAGC↓APSYRV	LGRAGA↓YIFSSD
VEEV	LQEAGA↓GSVETP	LHEAGC↓APSYHV	RFDAGA↓YIFSSD
SIN	QADVGA↓ALVETP	RDGVGA↓APAYRS	LTGVGG↓YIFSTD

Thioredoxin fusion protein substrate expression vectors

Expression vectors encoding thioredoxin (Trx) fusion protein substrates (Table 2) containing either a SFV p3/p4 or SIN p3/p4 site were constructed according to a previous report [19]. Synthetic DNA encoding 19 residues upstream and 18 residues downstream of the SFV p3/p4 site (P19–P18′) was treated with EcoRV and EcoRI, and then ligated in-frame with the Trx gene in pET32c (Novagen, Madison, WI) that had been digested with MscI and EcoRI to create the Trx-SFV p3/p4 fusion protein expression vector pDZ1979. The Trx-SIN p3/p4 fusion protein expression vector, pDZ2040, was constructed by QuickChange site-directed mutagenesis (Stratagene, La Jolla, CA); 6 residues upstream and 5 residues downstream of the SFV p3/p4 site in pDZ1979 were replaced by the corresponding residues from the SIN p3/p4 site (Fig. 1B).

Expression and purification of nsp2pro proteases

Escherichia coli BL21(DE3) CodonPlus-RIL cells (Stratagene) containing a HisMBP-nsp2pro expression vector were grown to mid-log phase (A_{600nm} = 0.5) in Luria Broth supplemented with 100 μg/ml ampicillin and 30 μg/ml chloramphenicol at 37 °C, at which time the temperature was shifted to 25 °C and isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to initiate the production of the fusion protein. After 4 h, the cells were recovered by centrifugation and frozen at −80 °C.

About 30 g of *E. coli* cell paste was suspended in 300 ml of ice-cold 50 mM sodium phosphate (pH 7.5), 150 mM NaCl and 25 mM imidazole (buffer A). The cells were lysed with an APV Gaulin Model G homogenizer (Invensys, Albertslund, Denmark) at 10,000 psi and centrifuged at 30,000g for 30 min at 4 °C. The supernatant was filtered through a 0.45 μm polyethersulfone membrane, applied to a 15 ml HisTrap FF crude affinity column (GE Healthcare, Piscataway, NJ) equilibrated in buffer A, and then eluted with a linear gradient from 25 to 250 mM imidazole in buffer A. Fractions containing the HisMBP-nsp2pro fusion protein were pooled and concentrated using an Amicon YM30 membrane (Millipore, Billerica, MA), then diluted 6-fold with 50 mM sodium phosphate (pH 7.5), 150 mM NaCl to reduce the imidazole concentration to ca. 25 mM. The fusion protein was then digested with a 1 mg/ml stock solution of His-tagged TEV (S219V) protease (100:1 w/w) [27] overnight at 4 °C. Next, the products of the digest were applied to a 15 ml HisTrap FF crude affinity column equilibrated with buffer A. The flow-through fractions now contained the nsp2pro protease. The sample was concentrated to about 10 mg/ml using an Amicon YM30 membrane (Millipore) and applied to a HiPrep 26/60 Sephacryl S-100 column (GE Healthcare) equilibrated with 50 mM Tris–HCl (pH 7.5), 150 mM NaCl and 2 mM Tris (2-carboxyethyl) phosphine hydrochloride (TCEP). The peak fractions containing the nsp2pro protease were pooled and concentrated to 1–5 mg/ml. Aliquots were flash-frozen with liquid nitrogen and stored at −80 °C until further use. The final product was judged to be at least 95% pure by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). The molecular weights of all three proteases were confirmed by electrospray ionization mass spectrometry (data not shown).

Expression and purification of MBP-NusG-His$_6$ and Trx fusion proteins

Escherichia coli BL21 (DE3) CodonPlus-RIL cells containing an MBP-NusG-His$_6$ or Trx fusion protein expression vector were cultivated and harvested as described above for the production of His-

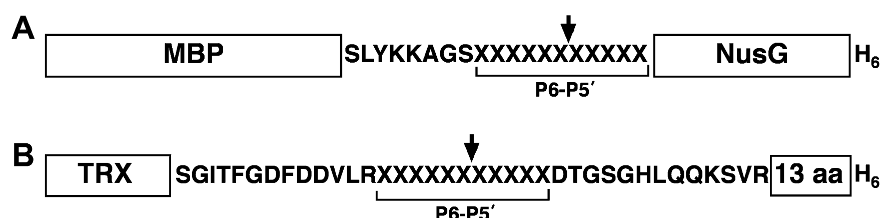


Fig. 1. Schematic diagram (not to scale) of (A) MBP-NusG-His$_6$ and (B) Trx fusion protein substrates. X represents the P6–P5′ amino acid residues of the natural alphavirus protease cleavage sites (see Table 3).

Author's personal copy

MBP-nsp2pro fusion proteins, except that IPTG induction was performed at 30 °C. About 5 g of *E. coli* cell paste was suspended in 50 ml of ice-cold buffer A and processed as described above. The fusion proteins were purified by immobilized metal affinity chromatography, using a 15 ml HisTrap FF crude affinity as described above. Fractions containing the fusion protein were pooled, concentrated using an Amicon YM30 membrane, and then applied to a HiPrep Sephacryl 26/60 S-100 column equilibrated with 50 mM Tris–HCl (pH 7.5), 150 mM NaCl and 2 mM TCEP. The peak fractions containing the fusion protein were pooled and concentrated to 5–10 mg/ml. Aliquots were flash-frozen with liquid nitrogen and stored at −80 °C until further use. The final products were judged to be as least 95% pure by SDS–PAGE (data not shown).

In vitro processing of MBP-NusG fusion proteins

All reactions were performed in 50 mM Tris–HCl (pH 7.5), 150 mM NaCl at 30 °C unless stated otherwise. The concentration of the fusion protein substrates (MBP-NusG-His_6) was 0.9 mg/ml. A sample was removed from each reaction for an undigested control, and then nsp2pro protease was added to the remainder of the reactions to achieve a final concentration of 0.1 mg/ml, unless indicated otherwise. The mixtures were incubated 16 h at 30 °C. The reactions were stopped by adding an equal volume of 2× SDS sample buffer and analyzed by SDS–PAGE. The proteins were visualized by staining with Coomassie brilliant blue.

In vitro processing of Trx fusion proteins

All reactions were performed in 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA and 1 mM DTT at 30 °C unless stated otherwise. The concentration of the fusion protein substrates was 0.9 mg/ml. A sample was removed from each reaction for an undigested control, and then nsp2pro protease was added to the remainder of the reactions to achieve a final concentration of 0.1 mg/ml, unless indicated otherwise. The mixtures were incubated 1 h. The reactions were stopped by adding an equal volume of 2× SDS sample buffer and analyzed by SDS–PAGE. The proteins were visualized by staining with Coomassie brilliant blue.

Determination of the optimum pH, temperature and salt concentration

To determine the optimum temperature for each nsp2pro, reactions were performed in 50 mM Tris–HCl (pH 7.5), 150 mM NaCl. The concentration of the fusion protein substrates (MBP-NusG-His_6) was 0.9 mg/ml. A sample was removed from each reaction for an undigested control, and then protease was added to the remainder of the reactions to achieve a final concentration of 0.1 mg/ml in the case of SFV and SIN nsp2pro or 0.025 mg/ml in the case of VEEV nsp2pro. The mixtures were incubated 16 h at a series of temperatures (4, 10, 15, 20, 25, 30, 37 and 42 °C). The reactions were stopped by adding an equal volume of 2× SDS sample buffer and then heating at 90 °C for 5 min. The results were analyzed by SDS–PAGE. The proteins were visualized by staining with Coomassie brilliant blue. Initial velocities were calculated from the relative intensities of Coomassie-stained bands in the gel using Alpha Innotech (San Leandro, CA) Gel Analysis software.

The effect of pH on proteolytic activity was studied in 50 mM of either sodium acetate, sodium phosphate or Tris–HCl and 150 mM NaCl to obtain a range of pH values (5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5 and 9), with the concentration of the substrate and protease being the same as above. Reactions were performed at 30 °C for 16 h, stopped and analyzed as described above.

The effect of monovalent salt concentration on the activity of the nsp2pro proteases was studied by adding NaCl to 50 mM Tris–HCl (pH 7.5) to achieve a range of salt concentrations (0, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 500 mM, 1 and 2 M). Reactions were performed at 30 °C for 16 h, stopped and analyzed as described above.

The effect of glycerol, EDTA, DTT and TCEP on nsp2pro activity

The effect of glycerol on the activity of the nsp2pro proteases was studied by adding glycerol to 50 mM Tris–HCl (pH 7.5) and 150 mM NaCl to achieve a range of glycerol concentrations (0%, 0.5%, 1%, 2%, 3%, 4%, 5% and 10% v/v). The effect of EDTA, DTT and TCEP on the nsp2pro proteases was studied by adding each to 50 mM Tris–HCl (pH 7.5) and 150 mM NaCl to achieve a range of concentrations (0, 0.1, 0.5, 1, 2, 5 and 10 mM). Reactions were performed at 30 °C for 16 h, stopped and analyzed as described above.

Oligopeptide synthesis and characterization

Synthetic oligopeptides having the P6–P6′ residues of alphavirus cleavage site sequences (Table 3) with three additional C-terminal arginines to increase peptide solubility were obtained from Biopeptide Inc. (San Diego, CA). Stock solutions were made in distilled water (or in 2 mM TCEP if the peptide contained a cysteine residue) and the peptide concentrations were determined by amino acid analysis.

Enzyme kinetics

Enzyme assays were initiated by the mixing of 10 µl protease solution (1–50 µM) in 2× reaction buffer (100 mM Tris–HCl pH 7.5, 10 mM DTT, 2 mM EDTA, 300 mM NaCl) and 10 µl substrate solution (0.3 mM). The reaction mixtures were incubated at 30 °C for 1 or 20 h and the reactions were stopped by the addition of 180 µl 1% trifluoroacetic acid. An aliquot was injected onto a Nova-Pak C18 reversed-phase chromatography column (3.9 × 150 mm, Waters Associates, Inc.) using an automatic injector. Substrates and the cleavage products were separated using an increasing water–acetonitrile gradient (0–100%) in the presence of 0.05% TFA. Product peak fractions were collected and analyzed by MALDI TOF mass spectrometry.

In cases when a sufficient degree of substrate cleavage was observed after 1 h of incubation, measurements were performed at six different (0.1–2.0 mM) substrate concentrations to determine the kinetic parameters. Kinetic parameters were determined by fitting the data obtained at less than 20% substrate hydrolysis to the Michaelis–Menten equation by using the Fig. P program (Fig. P Software Corp., Hamilton, Ont., Canada). The catalytic constants (k_{cat} and k_{cat}/K_m) values were calculated by assuming 100% activity for the enzyme. The standard deviations for the k_{cat}/K_M values were calculated as described [28].

Results

Cloning, expression and purification of nsp2pro proteases

The nsp2pro domains from SIN, SFV and VEEV were expressed initially as HisMBP fusion proteins [24] in *E. coli* and purified to homogeneity as described in Materials and methods. All three proteases exhibited a similar expression pattern in *E. coli*. The HisMBP-nsp2pro fusion proteins were only partially soluble when induced at 30 °C, but the solubility improved substantially at 25 °C. After the HisMBP tags were removed by TEV protease, all three nsp2pro domains remained soluble. The enzymes were purified to greater than 95% homogeneity as judged by SDS–PAGE and their molecular weights were confirmed by mass spectrometry (data not shown). The yield of all three proteases was about 1 mg/g of wet cells.

Author's personal copy

Digestion of fusion protein substrates

Initially, nsp2pro protease assays were conducted with purified MBP-NusG-His$_6$ fusion protein substrates. Altogether, nine different substrates were used in these experiments, each including 11 residues from one of the three viral polyproteins (6 residues upstream of the cleavage site and 5 residues downstream, P6–P5′ positions) corresponding to the natural p1/p2, p2/p3 and p3/p4 processing sites. The reaction conditions were similar to those employed in a previous study of SFV nsp2pro [19]; 0.9 mg/ml of substrates and 0.1 mg/ml of protease were incubated at 30 °C overnight in 50 mM Tris–HCl (pH 7.5), 150 mM NaCl. The results are shown in Fig. 2. The VEEV p3/p4 site was cleaved most efficiently by VEEV nsp2pro. Cleavage of VEEV p1/p2 was less efficient, and processing of VEEV p2/p3 was barely detectable. Similar results were obtained with SFV nsp2pro, in accord with an earlier study [19]. The products corresponding to the NusG domains generated by VEEV and SFV nsp2pro were analyzed by N-terminal sequencing to confirm that proteolytic cleavage occurred at the expected sites (data not shown). Curiously, under the same condi-

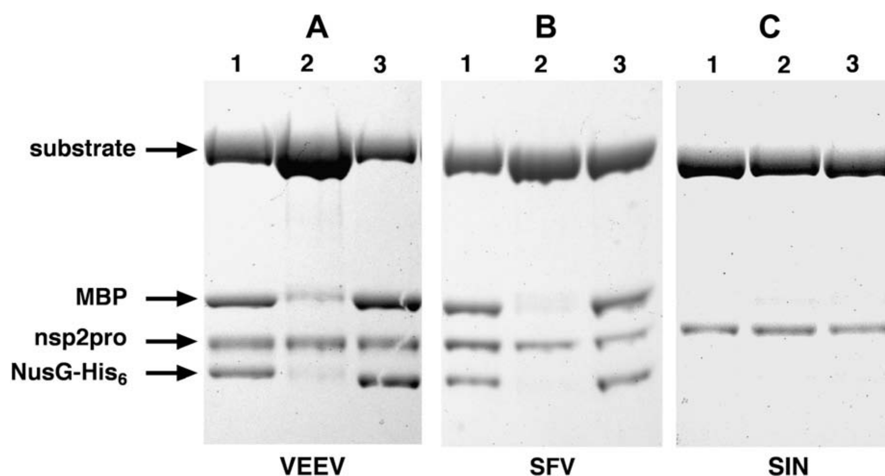


Fig. 2. Digestion of cognate MBP-NusG-His$_6$ fusion protein substrates in vitro by (A) VEEV, (B) SFV and (C) SIN nsp2pro proteases. Reactions were performed with substrates (lane 1, p1/p2; lane 2, p2/p3; lane 3, p3/p4) at a concentration of 0.9 mg/ml in 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, with proteases at a concentration of 0.1 mg/ml. The reactions were incubated 16 h at 30 °C, quenched by the addition of SDS sample buffer and analyzed by SDS–PAGE.

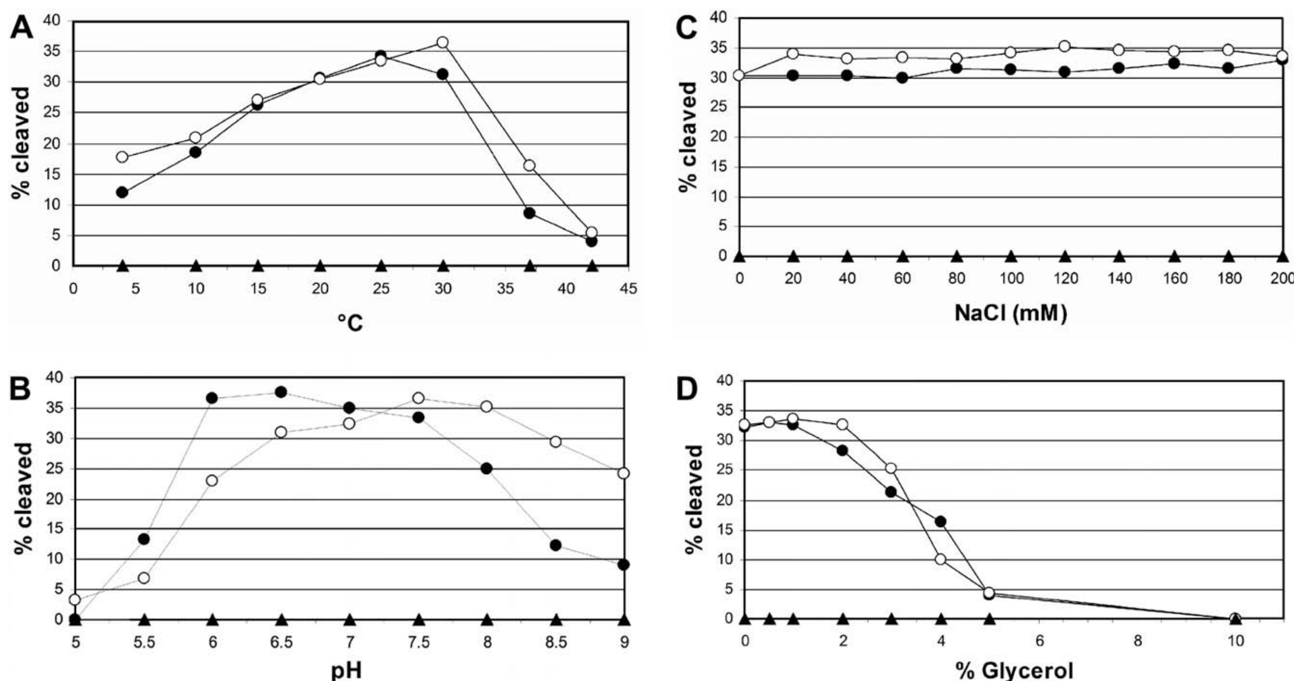


Fig. 3. Determination of optimal reaction conditions for VEEV, SFV and SIN nsp2pro proteases, using MBP-NusG-His$_6$ fusion protein substrates with cognate p3/p4 cleavage sites for each enzyme. Reactions were performed at a substrate concentration of 0.9 mg/ml, with a protease concentration of 0.025 mg/ml for VEEV nsp2pro and 0.1 mg/ml for SFV and SIN nsp2pro, 16 h at 30 °C. The reactions were quenched by the addition of SDS sample buffer and analyzed by SDS–PAGE. Initial velocities were calculated on the basis of data obtained by Alpha Innotech Gel Analysis software of Coomassie-stained gels. Open and closed circles represent data obtained with VEEV and SFV proteases, respectively. Closed triangles represent SIN protease data. Data points represent the average of at least two independent experiments. (A) Optimal temperature. Reactions were performed in 50 mM Tris–HCl (pH 7.5), 150 mM NaCl solution. (B) Optimal pH. Reactions were performed in 150 mM NaCl and either 50 mM sodium acetate, sodium phosphate or Tris–HCl buffer to achieve the indicated pH. (C) Optimal salt concentration. Concentrated NaCl was added to 50 mM Tris–HCl (pH 7.5) to achieve a range of monovalent salt concentrations between 0 and 200 mM. (D) Inhibitory effect of glycerol. Glycerol was added to 50 mM Tris–HCl (pH 7.5), 150 mM NaCl to achieve a range of glycerol concentrations between 0% and 10%.

Author's personal copy

tions, SIN nsp2pro cleaved only about 1% of its p3/p4 substrate and showed no detectable activity on the SIN p1/p2 or p2/p3 substrates. Consequently, its products were not subjected to N-terminal sequencing.

Comparison of pH, temperature and salt dependence

The effect of pH, temperature and monovalent salt concentration on the enzymatic activity of the nsp2pro enzymes was investigated. The MBP-NusG-His_6 fusion proteins VEEV p3/p4, SFV p3/p4 and SIN p3/p4 were used as substrates for these experiments because they were cleaved most efficiently by their cognate enzymes. As shown in Fig. 3A, the temperature dependence of VEEV and SFV nsp2pro is similar. Enzymatic activity increased with increasing temperature between 4 and 30 °C, reaching a maximum at 25 °C for VEEV nsp2pro and 30 °C for SFV nsp2pro. The activity decreased dramatically at temperatures above 30 °C, possibly due to unfolding of the proteases. Both proteases had maximum activity near neutral pH, as shown in Fig. 3B. VEEV nsp2pro was most active in the pH range of 6–7.5, while SFV nsp2pro was maximally active over a somewhat larger range (pH 6.5–8). The concentration of monovalent salt (NaCl) had relatively little effect on enzymatic activity up to 200 mM, as shown in Fig. 3C. However, NaCl concentrations in excess of 500 mM resulted in nonspecific cleavage of the substrates (results not shown). The activity of SIN nsp2pro did not improve under any of the conditions tested.

The effect of glycerol, EDTA and reducing agents

Both VEEV and SFV nsp2pro are very sensitive to glycerol. The activity of the two enzymes is reduced by approximately 20-fold in the presence of 5% glycerol (Fig. 3D). Glycerol was also observed to inhibit the proteolytic activity of Chikungunya virus nsp2 protease, although not as severely [29]. Whereas VEEV nsp2pro could tolerate EDTA up to 10 mM with very little effect, inhibition of SFV nsp2pro was observed at or above 2 mM EDTA. The reducing agents dithiothreitol (DTT) and TCEP had no discernable effect on the activity of VEEV or SFV nsp2pro up to 5 mM, and both enzymes were able to tolerate up to 10 mM of β-mercaptoethanol.

Cross-reactivity of VEEV, SFV and SIN proteases

All three proteases ostensibly recognize similar sites, (A/V)G(A/G/C)↓(A/G/Y) (Table 3). Therefore, it was of interest to investigate whether their substrates are interchangeable. The nine previously described MBP-NusG-His_6 fusion protein substrates were employed for this purpose. As shown in Fig. 4, SIN nsp2pro failed to cleave any of the substrates, including its own, nor were any of its substrates cleaved by either of the other two enzymes. The VEEV p1/p2 and p3/p4 substrates were cleaved by VEEV nsp2pro, as expected. Similarly, the SFV p1/p2 and p3/p4 substrates were cleaved by SFV nsp2pro. The only instance of cross-reactivity was cleavage of the SFV p1/p2 substrate by VEEV nsp2pro. Surprisingly, VEEV nsp2pro cleaved this heterologous substrate more efficiently than any of its own natural processing sites.

Digestion of thioredoxin fusion protein substrates

Concern about the relatively weak catalytic power of the nsp2pro enzymes prompted us to compare the activity of our preparation of SFV nsp2pro with SFV nsp2pro obtained from a commercial source (Quattromed, Tartu, Estonia), using both the MBP-NusG-His_6 substrates that we designed and a previously described Trx fusion protein substrate for SFV nsp2pro (Fig. 1B) [19]. Rather than consisting of 2-folded domains connected by an unstructured linker (like the MBP-NusG-His_6 substrates), the Trx domain is fol-

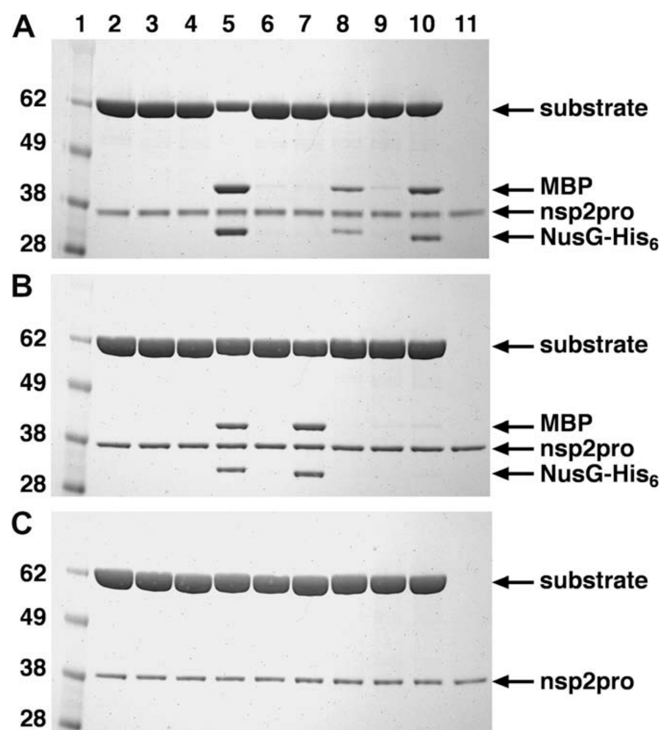


Fig. 4. Analysis of cross-reactivity between VEEV, SFV and SIN proteases. Reactions were performed with MBP-NusG-His_6 fusion protein substrates at a concentration of 0.9 mg/ml in 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, with proteases at concentration of 0.1 mg/ml. The reactions were incubated at 30 °C overnight. Lane 1, molecular weight standards (SeeBlue Plus2 Prestained; Invitrogen, Carlsbad, CA, USA). The approximate sizes of the markers (in kDa) are indicated to the left of lane 1. Lanes 2–10 correspond to MBP-NusG-His_6 substrates SIN p1/p2, SIN p2/p3, SIN p3/p4, SFV p1/p2, SFV p2/p3, SFV p3/p4, VEEV p1/p2, VEEV p2/p3 and VEEV p3/p4, respectively. Lane 11 contains the indicated nsp2pro protease only. (A) Digestion of substrates by VEEV nsp2pro. (B) Digestion of substrates by SFV nsp2pro. (C) Digestion of substrates by SIN nsp2pro.

lowed by 50 ostensibly unstructured residues, 37 of which are derived from the SFV p3/p4 processing site (P19–P18′). The results of these experiments confirmed that the specific activity of the two enzyme preparations is comparable (Fig. 5), but the Trx fusion protein was cleaved far more efficiently by SFV nsp2pro than the MBP-NusG-His_6 p3/p4 substrate; the former substrate was completely digested after only 1 h. In view of the particularly low catalytic activity of our preparation of SIN nsp2pro (see above), we also constructed a second Trx fusion protein substrate, otherwise identical to the first, in which the P6–P5′ residues were replaced by the corresponding residues from the SIN p3/p4 processing site. This substrate was cleaved approximately 10-fold more efficiently than the SIN p3/p4 MBP-NusG-His6 fusion protein (data not shown).

The effect of linker length on the efficiency of cleavage

In contrast to the MBP-NusG-His_6 fusion proteins containing alphavirus recognition sites, virtually identical fusion protein substrates with recognition sites for TEV (ENLYFQG) or TVMV (ETVRFQS) proteases between the MBP and NusG-His_6 domains are processed very efficiently in vitro by their cognate enzymes [25,30]. On the other hand, the Trx fusion protein substrates for the alphavirus proteases, which have longer linkers and lack a structured C-terminal domain, are cleaved more efficiently than their MBP-NusG-His_6 counterparts. Two features distinguish the latter substrates from the former. First, the Trx fusion proteins have longer segments of the viral polyprotein sequence on either side of the cleavage site than do the MBP-NusG-His_6 fusion pro-

Author's personal copy

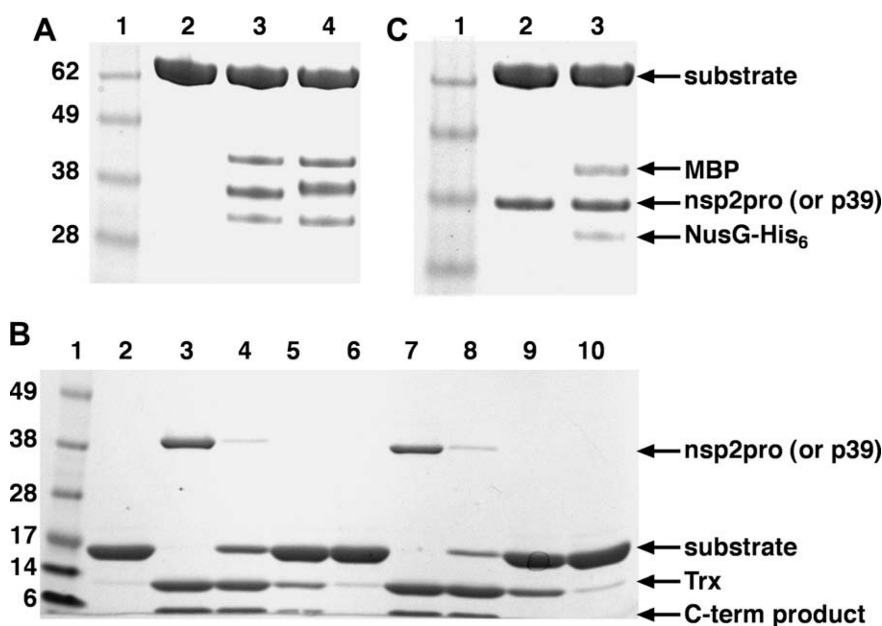


Fig. 5. Effect of enzyme source and type of substrate on the activity of SFV nsp2pro. (A) Digestion of MBP-NusG-His$_6$ fusion proteins by SFV nsp2pro and Pro39. Reactions were performed with substrates at a concentration of 0.9 mg/ml in 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 1 mM DTT, 1 mM EDTA with proteases at a concentration of 0.1 mg/ml. The reactions were incubated at 30 °C for 16 h, quenched by the addition of SDS sample buffer and analyzed by SDS–PAGE. Lane 1, molecular weight standards (SeeBlue Plus2 Prestained). The approximate sizes of the markers (in kDa) are indicated to the left of lane 1. Lane 2, undigested fusion protein. Lane 3, digestion with SFV nsp2pro. Lane 4, digestion with Pro39. (B) Digestion of TRX fusion proteins by SFV nsp2pro and Pro39. Reaction conditions were the same as in panel A, but reactions were terminated after 1 h. Lane 1, molecular weight standards (SeeBlue Plus2 Prestained). The approximate sizes of the markers (in kDa) are indicated to the left of lane 1. Lane 2, undigested fusion protein. Lanes 3–6, SFV nsp2pro at 1 mg/ml, 0.1 mg/ml, 10 μg/ml and 1 μg/ml, respectively. Lanes 7–10, Pro39 at 1 mg/ml, 0.1 mg/ml, 10 μg/ml and 1 μg/ml, respectively. (C) Digestion of MBP-NusG-His$_6$ (SFV p3/p4) and MBP-NusG-His$_6$ with 10 additional alanine residues between the P5′ site and NusG (SFV p3/p4 + 10) by SFV nsp2pro. Reaction conditions were the same as in panel B (1 h). Lane 1, molecular weight standards (SeeBlue Plus2 Prestained). Lane 2, MBP-NusG-His$_6$ SFV p3/p4. Lane 3, MBP-NusG-His$_6$ SFV p3/p4 + 10.

teins. Second, the distance between the cleavage site and the beginning of ordered structure in the C-terminal domain of the MBP-NusG-His$_6$ fusion proteins is shorter than in the Trx fusion protein substrates. In crystal structures of *Aquifex aeolicus* NusG, between 4 and 8 N-terminal residues are disordered [31,32]. Together with the 5 alphavirus residues distal to the cleavage site, this amounts to a total of 9–13 disordered residues between the cleavage site and the folded NusG domain of the fusion protein. To investigate whether the length of the linker between the alphavirus cleavage site and NusG has an impact on the processing efficiency, an MBP-NusG-His$_6$ fusion protein substrate was constructed with 10 consecutive Ala residues between the P5′ residue of the SFV p3/p4 cleavage site and the N-terminus of NusG. As shown in Fig. 5C, this fusion protein was cleaved with greater efficiency by SFV nsp2pro than the original MBP-NusG-His$_6$ substrate with the shorter linker. While this experiment does not rule out the possibility that the sequence specificity of alphavirus proteases extends beyond the P5′ site, it does reveal a clear dependency on linker length.

Experiments with oligopeptide substrates

To corroborate the results obtained with fusion protein substrates and facilitate the measurement of kinetic parameters, activity assays were also conducted with a series of oligopeptide substrates encompassing residues P6–P6′ of the viral cleavage sites, using each enzyme on all of the substrates. When substrates were incubated overnight with 0.1 mg/ml (2.5 μM) enzymes, no hydrolysis was observed for most of the enzyme/substrate combinations. A low level (~5%) of hydrolysis was observed for the VEEV p1/p2 and p3/p4 peptides with VEEV protease (data not shown). Using substantially higher enzyme concentrations, a low level of hydrolysis of the SIN p1/p2 and p3/p4 peptides by SIN protease

Table 4
Kinetic parameters determined for oligopeptide substrates.

Site	Protease	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1} s^{-1})
SFV p1/p2	VEEV	0.58 ± 0.09	0.016 ± 0.001	0.028 ± 0.005
SFV p1/p2	SFV	1.20 ± 0.22	0.043 ± 0.005	0.036 ± 0.008
SFV p3/p4	SFV	1.27 ± 0.20	0.352 ± 0.032	0.277 ± 0.050

was also observed at the correct site, as verified by MALDI MS. In three cases, the rate of hydrolysis at 1 h of incubation was sufficient to allow the determination K_m, k_{cat} and k_{cat}/K_m values using the HPLC-based assay, as described previously for TEV and TVMV proteases [30,33]. These include the cleavage of SFV p1/p2 site peptide by VEEV protease, and the cleavage of SFV p1/p2 and p3/p4 site peptides by SVF protease. The results of these experiments are presented in Table 4. In general, the hierarchy of processing efficiency is in good agreement with the results obtained using the MBP-NusG-His$_6$ fusion protein substrates, although the SFV nsp2pro appeared to be somewhat more active than the VEEV nsp2pro on the peptide substrates. The specificity constants (a measure of catalytic efficiency) for those combinations of alphaviral proteases and oligopeptides that could be estimated experimentally are approximately 10- to 100-fold lower than those obtained with TEV and TVMV proteases and their specific substrates [30,33].

Discussion

In addition to their obvious utility for protein purification, affinity tags can improve the yield of recombinant proteins, protect them from intracellular proteolysis, and enhance their solubility [34–38]. However, it is ordinarily desirable to remove an affinity

Author's personal copy

tag from a protein prior to functional and structural studies. Enzymatic methods are most commonly employed to remove affinity tags, yet not all proteases perform this task equally well.

The objective of this study was to compare the ability of three alphavirus nsp2 proteases to cleave genetically engineered fusion proteins at a specific site. The methods we used here are similar to those that were previously used to characterize TEV and TVMV proteases [30]. To this end, we constructed a series of nine MBP-NusG-His$_6$ fusion proteins, each including the P6–P5′ residues of a different viral polyprotein cleavage site (Table 3) in between the MBP and NusG domains. Based on a previous study of SFV nsp2 [39], we anticipated that 11 viral residues surrounding the cleavage site would be more than sufficient for efficient processing. Unexpectedly, however, in contrast to previous studies of SFV nsp2pro [19,21,39], we discovered that the alphavirus nsp2pro enzymes exhibited relatively poor catalytic activity in vitro, being much less efficient at processing MBP-NusG-His$_6$ fusion proteins than TEV or TVMV proteases [30]. One possible explanation for this observation is that the alphavirus proteases are inherently weaker catalysts than the potyviral proteases, which is supported by the data obtained with oligopeptide substrates.

We also found that the TRX fusion proteins were cleaved much more efficiently by the alphavirus nsp2pro proteases than were the MBP-NusG-His$_6$ substrates. This may indicate that the specificity determinants for the nsp2pro enzymes extend beyond the P6–P5′ residues of their substrates. Consistent with this hypothesis, Lulla et al. reported a decrease in processing efficiency when the length of the SFV sequence on the C-terminal side of the p3/p4 site was reduced from 10 to 5 residues [39]. On the other hand, as few as five viral residues on the N-terminal side of the SFV p3/4 site appeared to be enough for efficient processing. Another possible explanation for this observation is that the larger size of the nsp2pro enzymes (ca. 39 kDa compared to 25 kDa for TEV and TVMV proteases) may inhibit their access to short interdomain linkers. Although the crystal structure of VEEV nsp2pro has been published [23], no nsp2pro/peptide co-crystal structures are currently available, and so it is not known how many residues are occluded by the enzyme when it binds to its processing sites. However, the more efficient processing of an MBP-NusG-His$_6$ fusion with an extended linker following the cleavage site is consistent with this hypothesis. A third possibility is that alphavirus nsp2 proteases utilize a form of facilitated diffusion to locate their target sequences, akin to the manner in which restriction endonucleases "scan" DNA via nonspecific interactions in search of their cognate binding sites [40]. From a practical standpoint, it does not matter whether steric hindrance is to blame, if the sequence specificity of alphavirus proteases extends beyond the P5′ site in the C-terminal direction, or if facilitated diffusion is involved in the recognition of target sites; any of these circumstances would reduce the utility of these enzymes as reagents for removing N-terminal affinity tags because efficient processing would require that quite a few unstructured, non-native residues remain on the N-terminus of the target protein after cleavage.

We conclude from this study that although these alphavirus proteases exhibit sufficient sequence specificity to be useful reagents for removing affinity tags, in practice their ability to cleave genetically engineered fusion protein substrates is far less impressive than that of proteases encoded by potyviruses such as TEV and TVMV, which makes them comparatively unattractive from a biotechnological standpoint. These proteases may require a longer linker between domains in a fusion protein substrate for efficient processing to occur. Additionally, experiments with oligopeptide substrates suggest that they are inherently weaker catalysts than the potyviral proteases. In either case, they do not offer any significant advantages over existing proteolytic reagents. On the other hand, alphavirus proteases remain promising therapeutic targets for the treatment of a variety of diseases, which is ample justification to continue studying them.

Acknowledgments

This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. We thank Scott Cherry for technical assistance, the SAIC-Frederick Protein Chemistry Laboratory for assistance with N-terminal amino acid sequencing and Beáta Bozóki for help in peptide assays. Electrospray mass spectrometry experiments were conducted on the LC/ESMS instrument maintained by the Biophysics Resource in the Structural Biophysics Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, while amino acid analysis and MALDI MS measurements for peptides were performed in the Food Science and Quality Assurance Laboratory and Proteomics Core Facility of the Debrecen University, respectively.

References

[1] C.H. Calisher, N. Karabatsos, Arbovirus serogroups: definition and geographic distribution, in: T.P. Monath (Ed.), The Arboviruses: Epidemiology and Ecology, vol. 1, CRC Press Inc., Baca Raton, FL, 1988, pp. 19–57.
[2] K. Lundstrom, Biology and application of alphaviruses in gene therapy, Gene Ther. 12 (Suppl. 1) (2005) S92–S97.
[3] S.C. Weaver, R. Salas, R. Rico-Hesse, G.V. Ludwig, M.S. Oberste, J. Boshell, R.B. Tesh, Re-emergence of epidemic Venezuelan equine encephalomyelitis in South America. VEE Study Group, Lancet 348 (1996) 436–440.
[4] B. Niklasson, A. Espmark, J.W. LeDuc, T.P. Gargan, W.A. Ennis, R.B. Tesh, A.J. Main Jr., Association of a Sindbis-like virus with Ockelbo disease in Sweden, Am. J. Trop. Med. Hyg. 33 (1984) 1212–1217.
[5] C.C. Mathiot, G. Grimaud, P. Garry, J.C. Bouquety, A. Mada, A.M. Daguisy, A.J. Georges, An outbreak of human Semliki Forest virus infections in Central African Republic, Am. J. Trop. Med. Hyg. 42 (1990) 386–393.
[6] M.S. Bronze, M.M. Huycke, L.J. Machado, G.W. Voskuhl, R.A. Greenfield, Viral agents as biological weapons and agents of bioterrorism, Am. J. Med. Sci. 323 (2002) 316–325.
[7] S. Schlesinger, Alphavirus vectors: development and potential therapeutic applications, Expert Opin. Biol. Ther. 1 (2001) 177–191.
[8] J.H. Strauss, E.G. Strauss, The alphaviruses: gene expression, replication, and evolution, Microbiol. Rev. 58 (1994) 491–562.
[9] M. Rikkonen, J. Peranen, L. Kaariainen, ATPase and GTPase activities associated with Semliki Forest virus nonstructural protein nsP2, J. Virol. 68 (1994) 5804–5810.
[10] M. Gomez de Cedron, N. Ehsani, M.L. Mikkola, J.A. Garcia, L. Kaariainen, RNA helicase activity of Semliki Forest virus replicase protein NSP2, FEBS Lett. 448 (1999) 19–22.
[11] L. Vasiljeva, A. Merits, P. Auvinen, L. Kaariainen, Identification of a novel function of the alphavirus capping apparatus. RNA 5′-triphosphatase activity of Nsp2, J. Biol. Chem. 275 (2000) 17281–17287.
[12] J. Suopanki, D.L. Sawicki, S.G. Sawicki, L. Kaariainen, Regulation of alphavirus 26S mRNA transcription by replicase component nsP2, J. Gen. Virol. 79 (Pt. 2) (1998) 309–319.
[13] D.L. Sawicki, S.G. Sawicki, A second nonstructural protein functions in the regulation of alphavirus negative-strand RNA synthesis, J. Virol. 67 (1993) 3605–3610.
[14] J. Peranen, M. Rikkonen, P. Liljestrom, L. Kaariainen, Nuclear localization of Semliki Forest virus-specific nonstructural protein nsP2, J. Virol. 64 (1990) 1888–1896.
[15] N.D. Rawlings, A.J. Barrett, Evolutionary families of peptidases, Biochem. J. 290 (Pt. 1) (1993) 205–218.
[16] E.G. Strauss, R.J. De Groot, R. Levinson, J.H. Strauss, Identification of the active site residues in the nsP2 proteinase of Sindbis virus, Virology 191 (1992) 932–940.
[17] W.R. Hardy, J.H. Strauss, Processing the nonstructural polyproteins of sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in *cis* and in *trans*, J. Virol. 63 (1989) 4653–4664.
[18] K. Takkinen, J. Peranen, L. Kaariainen, Proteolytic processing of Semliki Forest virus-specific non-structural polyprotein, J. Gen. Virol. 72 (Pt. 7) (1991) 1627–1633.
[19] L. Vasiljeva, L. Valmu, L. Kaariainen, A. Merits, Site-specific protease activity of the carboxyl-terminal domain of Semliki Forest virus replicase protein nsP2, J. Biol. Chem. 276 (2001) 30786–30793.
[20] L. Vasiljeva, A. Merits, A. Golubtsov, V. Sizemskaja, L. Kaariainen, T. Ahola, Regulation of the sequential processing of Semliki Forest virus replicase polyprotein, J. Biol. Chem. 278 (2003) 41636–41645.
[21] A. Golubtsov, L. Kaariainen, J. Caldentey, Characterization of the cysteine protease domain of Semliki Forest virus replicase protein nsP2 by in vitro mutagenesis, FEBS Lett. 580 (2006) 1502–1508.

Author's personal copy

[22] R.J. de Groot, W.R. Hardy, Y. Shirako, J.H. Strauss, Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing in vivo, EMBO J. 9 (1990) 2631–2638.
[23] A.T. Russo, M.A. White, S.J. Watowich, The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease, Structure 14 (2006) 1449–1458.
[24] J.E. Tropea, S. Cherry, S. Nallamsetty, C. Bignon, D.S. Waugh, A generic method for the production of recombinant proteins in *Escherichia coli* using a dual hexahistidine-maltose-binding protein affinity tag, Methods Mol. Biol. 363 (2007) 1–19.
[25] R.B. Kapust, J. Tozser, T.D. Copeland, D.S. Waugh, The P1′ specificity of tobacco etch virus protease, Biochem. Biophys. Res. Commun. 294 (2002) 949–955.
[26] J.D. Fox, D.S. Waugh, Maltose-binding protein as a solubility enhancer, Methods Mol. Biol. 205 (2003) 99–117.
[27] R.B. Kapust, J. Tozser, J.D. Fox, D.E. Anderson, S. Cherry, T.D. Copeland, D.S. Waugh, Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency, Protein Eng. 14 (2001) 993–1000.
[28] P. Boross, P. Bagossi, T.D. Copeland, S. Oroszlan, J.M. Louis, J. Tozser, Effect of substrate residues on the P2′ preference of retroviral proteinases, Eur. J. Biochem. 264 (1999) 921–929.
[29] B.A. Pastorino, C.N. Peyrefitte, L. Almeras, M. Grandadam, D. Rolland, H.J. Tolou, M. Bessaud, Expression and biochemical characterization of nsP2 cysteine protease of Chikungunya virus, Virus Res. 131 (2008) 293–298.
[30] J. Tozser, J.E. Tropea, S. Cherry, P. Bagossi, T.D. Copeland, A. Wlodawer, D.S. Waugh, Comparison of the substrate specificity of two potyvirus proteases, FEBS J. 272 (2005) 514–523.
[31] J.R. Knowlton, M. Bubunenko, M. Andrykovitch, W. Guo, K.M. Routzahn, D.S. Waugh, D.L. Court, X. Ji, A spring-loaded state of NusG in its functional cycle is suggested by X-ray crystallography and supported by site-directed mutants, Biochemistry 42 (2003) 2275–2281.
[32] T. Steiner, J.T. Kaiser, S. Marinkovic, R. Huber, M.C. Wahl, Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities, EMBO J. 21 (2002) 4641–4653.
[33] S. Nallamsetty, R.B. Kapust, J. Tozser, S. Cherry, J.E. Tropea, T.D. Copeland, D.S. Waugh, Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro, Protein Expr. Purif. 38 (2004) 108–115.
[34] R.B. Kapust, D.S. Waugh, *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused, Protein Sci. 8 (1999) 1668–1674.
[35] A. Jacquet, V. Daminet, M. Haumont, L. Garcia, S. Chaudoir, A. Bollen, R. Biemans, Expression of a recombinant *Toxoplasma gondii* ROP2 fragment as a fusion protein in bacteria circumvents insolubility and proteolytic degradation, Protein Expr. Purif. 17 (1999) 392–400.
[36] D.B. Smith, K.S. Johnson, Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase, Gene 67 (1988) 31–40.
[37] C.V. Maina, P.D. Riggs, A.G. Grandea 3rd, B.E. Slatko, L.S. Moran, J.A. Tagliamonte, L.A. McReynolds, C.D. Guan, An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein, Gene 74 (1988) 365–373.
[38] D.S. Waugh, Making the most of affinity tags, Trends Biotechnol. 23 (2005) 316–320.
[39] A. Lulla, V. Lulla, K. Tints, T. Ahola, A. Merits, Molecular determinants of substrate specificity for Semliki Forest virus nonstructural protease, J. Virol. 80 (2006) 5413–5422.
[40] A. Pingoud, W. Wende, A sliding restriction enzyme pauses, Structure 15 (2007) 391–393.