the FEBS Journal

Comparison of the substrate specificity of two potyvirus proteases

József Tözsér[1], Joseph E. Tropea[2], Scott Cherry[2], Peter Bagossi[1], Terry D. Copeland[3], Alexander Wlodawer[2] and David S. Waugh[2]

1 Department of Biochemistry and Molecular Biology, Research Center for Molecular Medicine, University of Debrecen, Hungary
2 Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, MD, USA
3 Laboratory of Protein Dynamics and Signaling, Center for Cancer Research, National Cancer Institute at Frederick, MD, USA

Keywords
nuclear inclusion protease; potyvirus protease; substrate specificity; tobacco etch virus protease; tobacco vein mottling virus protease

Correspondence
J. Tözsér, Department of Biochemistry and Molecular Biology, Research Center for Molecular Medicine, University of Debrecen, Debrecen, Hungary
Fax: +1 36 52 314 989
Tel: +1 36 52 416 432
E-mail: tozser@indi.biochem.dote.hu or
D. S. Waugh, Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, PO Box B, Frederick, MD, USA
Fax: +301 846 7148
Tel: +301 846 1842
E-mail: waughd@ncifcrf.gov

(Received 25 August 2004, revised 7 October 2004, accepted 18 November 2004)

doi:10.1111/j.1742-4658.2004.04493.x

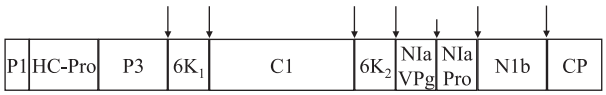
The substrate specificity of the nuclear inclusion protein a (NIa) proteolytic enzymes from two potyviruses, the tobacco etch virus (TEV) and tobacco vein mottling virus (TVMV), was compared using oligopeptide substrates. Mutations were introduced into TEV protease in an effort to identify key determinants of substrate specificity. The specificity of the mutant enzymes was assessed by using peptides with complementary substitutions. The crystal structure of TEV protease and a homology model of TVMV protease were used to interpret the kinetic data. A comparison of the two structures and the experimental data suggested that the differences in the specificity of the two enzymes may be mainly due to the variation in their S4 and S3 binding subsites. Two key residues predicted to be important for these differences were replaced in TEV protease with the corresponding residues of TVMV protease. Kinetic analyses of the mutants confirmed that these residues play a role in the specificity of the two enzymes. Additional residues in the substrate-binding subsites of TEV protease were also mutated in an effort to alter the specificity of the enzyme.

Members of the picornavirus 'super group' are positive-sense RNA viruses with similar genomic organization and replication strategy, which are responsible for a variety of plant and animal diseases [1]. The replication strategy of these viruses includes several proteolytic steps. Consequently, picornaviral proteases are currently used as molecular targets for antiviral therapeutics [2].

Tobacco etch virus (TEV) and tobacco vein mottling virus (TVMV) are members of the family Potyviridae, a subdivision of the picornavirus super group. About 200 potyviruses have been identified to date. Potyvirus RNA genomes are about 10 kb in length, polyadenylated at their 3′ ends, and covalently linked to a viral protein (VPg) at their 5′ ends [3]. The viral genome is translated upon infection into a single polyprotein, which is processed by virally encoded proteases. Most of these cleavages are performed by the nuclear inclusion a (NIa) protease [3–5].

The potyviral NIa protein consists of two domains separated by an inefficiently utilized NIa cleavage site: VPg (22 kDa) at the N-terminus and Pro (27 kDa) at

Abbreviations
TEV, tobacco etch virus; TVMV, tobacco vein mottling virus; NIa, nuclear inclusion protein a.



Site of cleavage	TEV	TVMV
P3/6K1	EDLVEQ↓AKQ	EIVEFQ↓AKN
6K1/C1	EIIYTQ↓SLD	NNVRFQ↓SLD
C1/6K 2	ETIYLQ↓SDS	EAVRFQ↓SES
6K2/NIa-Vpg	EPVYFQ↓GKK	EPVKFQ↓GKS
NIa-Pro/NIb	ELVYSQ↓GEK	DLVRTQ↓GEK
NIb/CP	ENLYFQ↓SGT	ETVRFQ↓SDT
NIa-VPg/NIa-Pro	EDLTFE↓GES	QEVAFE↓SKA
NIa-Pro carboxy	GHKVFM↓SKP	-

Fig. 1. Structure of the potyvirus genome. Locations of the TEV and TVMV NIa protease cleavage sites are indicated by arrows, including the inefficiently utilized cleavage site between NIa-VPg and NIa-Pro. The sequences of the natural TEV and TVMV protease cleavage sites are also indicated below the schematic diagram.

the C-terminus (Fig. 1). The C-terminal domain is a cysteine protease containing His46, Asp81 and Cys151 as the catalytic triad (numbering starts from the VPg/Pro cleavage site). The stringent sequence specificity of the TEV NIa protease has led to its widespread use in the biotechnology industry as a reagent for endoproteolytic removal of affinity tags [6]. The specificity of TEV protease has been analyzed in detail [7–10]. However, much less is known about the substrate specificity of the TVMV protease. The specificity of the latter enzyme has only been studied using oligopeptide substrates that correspond to its naturally occurring cleavage sites [11]. The amino acid sequences of the natural cleavage sites for TEV and TVMV proteases are listed in Fig. 1. A peptide corresponding to the NIb/CP cleavage site (ETVRFQ↓S, where the arrow indicates the site of cleavage) was identified as the best substrate for TVMV protease [11], and the corresponding site in the TEV polyprotein (ENLYFQ↓S) is also utilized by that enzyme with high efficiency [7]. Of the seven natural processing sites for TVMV protease, only peptides representing the NIa-VPg/NIa-Pro and NIa-Pro/NIb sites were not hydrolyzed *in vitro* by recombinant TVMV protease [11].

The crystal structures of two TEV protease mutants, catalytically inactive C151A and autolysis-resistant S219D, were recently solved as complexes with a substrate and product peptide, respectively [12], revealing the structural basis for its stringent sequence selectivity. In this study, two key residues predicted to be important for the different sequence specificities of the two enzymes were replaced in TEV protease with the corresponding residues of TVMV protease. The specificity of the mutant proteases was evaluated using a series of synthetic oligopeptides as substrates. The high degree of sequence identity (55%) between TEV and TVMV NIa proteases (Fig. 2A) enabled us to build a molecular model of the latter enzyme (Fig. 2B) and to use it, together with the crystal structure of TEV protease, to interpret differences between the specificity of the two enzymes. Additional residues in the substrate-binding subsites of TEV protease were also mutated to investigate their role in providing the specificity of the enzyme.

Results

Potential specificity determinants in TEV and TVMV proteases

Mutational analysis of TEV protease cleavage sites established that the specificity of the enzyme is restricted to the P6–P1′ positions of the substrate [7,13]. The crystal structure of catalytically inactive TEV protease in complex with a peptide substrate [12] revealed which amino acids form the S6–S1′ specificity pockets of the enzyme (Fig. 2A). Using the crystal structure of TEV protease as a starting point, we built a molecular model of TVMV protease. The average RMS deviation between the TEV protease crystal structure and the TVMV protease model was 0.22 Å. The corresponding residues that are predicted to form the specificity pockets in the latter enzyme are also shown in Fig. 2A.

Both TEV and TVMV proteases exhibit a strict requirement for Gln in the P1 position of their substrates and strong preferences for small aliphatic residues (Gly, Ser, Ala) in the P1′, Phe in P2, and Glu in P6 positions, respectively (Fig. 1). It is therefore unlikely that variations in the corresponding subsites of the two enzymes are responsible for their different sequence specificities. The P5 residue is not expected to be a significant specificity determinant either, because its side chain faces the solvent in the enzyme–substrate cocrystal structure [12] and it is not conserved in the natural processing sites for either protease (Fig. 1). It seems likely therefore that the S4 and/or S3 pockets of TEV and TVMV proteases are primarily responsible for their different sequence specificities.

Two residues of the S4 pockets involved in side chain–side chain interactions are different, Ala169(TEV)/Leu169(TVMV) and His214(TEV)/Phe213(TVMV), as shown in in Fig. 3A,B. Having Leu in the TVMV protease in place of Ala169 of the TEV protease decreases the volume of the S4 pocket while maintaining its apolar character. This may explain why all branched-chain aliphatic amino acid residues (Leu, Ile, Val) can be found in the P4 position of TEV protease-processing sites, whereas only Val, the smallest of them, occurs at

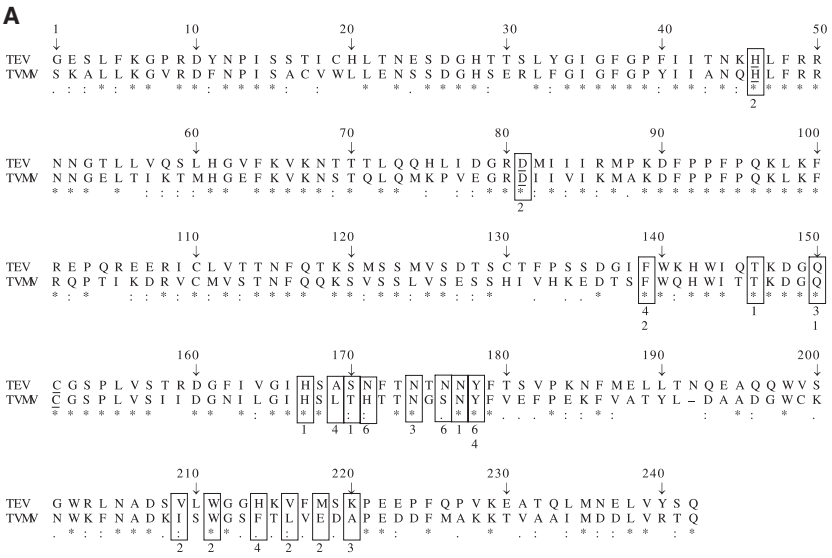


Fig. 2. (A) Sequence alignment of TEV and TVMV NIa proteases. The sequence alignment was made by the program CLUSTALW. Active-site residues are underlined. Conserved (*) and similar residues (: and .) are also indicated below the sequence as given by CLUSTALW. The sequence identity between the two proteases is 55%. Boxed amino acid residues are those involved in substrate binding by side chain–side chain interactions, and part of a given subsite is indicated by the numbers under the boxes (i.e. 1, S_1 binding site; 2, S_2 binding site, etc.). (B) Superimposition of homologous model of TVMV NIa protease (magenta) on the crystal structure of the C151A active site mutant TEV NIa protease (green). The substrate from the crystal structure is shown as a space-filling model. The residues of the catalytic triad are represented by red balls and sticks.

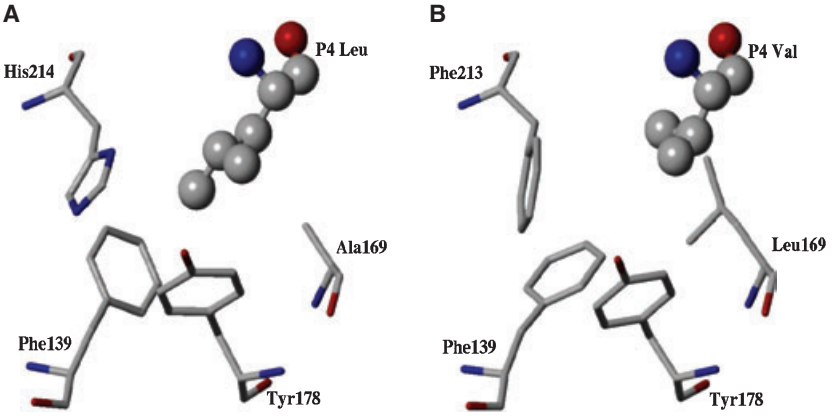


Fig. 3. S4 subsites of TEV (A) and TVMV (B) NIa proteases. Enzyme residues are shown with capped sticks, and the P4 residue of the substrate is shown with ball and stick representation.

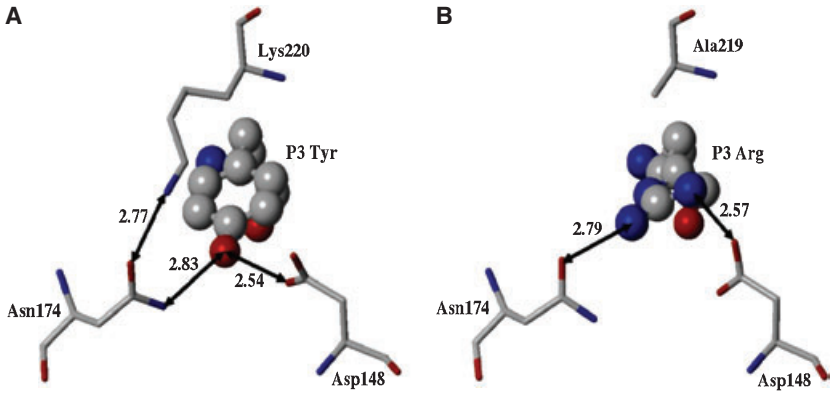


Fig. 4. S3 subsites of TEV (A) and TVMV (B) NIa proteases. Enzyme residues are shown with capped sticks, and the P3 residue of the substrate is shown with ball and stick representation. Hydrogen bonds are indicated by arrows.

the corresponding position in the natural TVMV protease cleavage sites (Fig. 1).

The S3 pockets of the TEV and TVMV proteases with the P3 residues from their NIb/CP cleavage site substrates are shown in Fig. 4A,B. The principal specificity determinant in the S3 pocket of the TEV protease appears to be Lys220 (Fig. 4A). The OH of the P3 Tyr forms hydrogen bonds with the side chains of Asp148 and Asn174, residues that are also present in the TVMV protease. The side chain of Lys220 also forms a hydrogen bond with Asn174 in the TEV protease (Fig. 4A), but this interaction cannot take place in the TVMV protease because the latter enzyme has an Ala residue at position 220 instead (Fig. 4B). In the TVMV protease, the 'missing' positively charged side chain may be supplied by the conserved P3 Arg residue in the substrate (Fig. 4B). It is interesting to note that, with the exception of the inefficiently processed NIa-Vpg/NIa-Pro cleavage site, only charged residues occur at the P3 positions of the natural TVMV protease cleavage sites (Fig. 1).

Comparison of the specificity of the TEV and TVMV proteases by using a peptide series with single mutations in their own cleavage site sequences

The specificity of the TEV and TVMV proteases was compared using a set of oligopeptide substrates based on the NIb/CP natural cleavage sites of these enzymes (peptides 1 and 6 in Table 1). The autolysis-resistant S219V mutant of TEV protease [14] was used as the 'wild-type' enzyme in these experiments. As previously described [14], TEV protease efficiently hydrolyzed the oligopeptide substrate representing its own cleavage site (Table 1). However, substitution of the P4 Leu with Val (peptide 2 in Table 1), the residue found in the equivalent position of the TVMV protease substrate, resulted in a dramatic increase in K_m and a decrease in the specificity constant (k_{cat}/K_m), indicating that the TEV protease strongly prefers Leu in this position even though Val is tolerated and can also be found in naturally occurring cleavage site sequences. The importance of optimum hydrophobic contacts within the P4 pocket of potyviral proteases is further supported by the findings that replacing P4 Val in the TVMV cleavage site peptide with Leu (peptide 7 in Table 1) enabled this peptide to be cleaved by TEV protease, and replacing P4 Leu in the TEV substrate with Ala (peptide 3 in Table 1) reduced the specificity constant to an even greater extent than did Val in this position. Replacement of P4 Val of the TVMV substrate with Ala did not convert the noncleavable sequence to a cleavable one for TEV protease (peptide 8 in Table 1).

When P3 Tyr in the TEV protease substrate was replaced with Arg, the residue found in the corresponding position of the TVMV protease substrate, this also resulted in a very inefficient substrate for TEV protease (peptide 4 in Table 1). Interestingly, even replacing P3 Tyr with Phe (peptide 5 in Table 1) gave rise to a 10-fold increase in K_m and a corresponding decrease in k_{cat}/K_m, underscoring the importance of the interactions between the Tyr OH and the side chains of Asn174 and Asp148 in TEV protease (Fig. 4A). The importance of these interactions is further supported by the results obtained with TVMV substrate substitutions: the replacement of P3 Arg with Tyr (peptide 9 in Table 1) also resulted in a cleavable substrate for the TEV protease (the best one among the singly substituted TVMV cleavage site peptides), whereas a Phe in this position (peptide 10 in Table 1) resulted in a substrate that was also cleavable but substantially less preferred.

The same series of peptides was also assayed with TVMV protease. The strong preference exhibited by TVMV protease for Val in the P4 position was confirmed by the observation that this enzyme was able to cleave the TEV peptide when the P4 Leu was replaced

Table 1. Comparison of the specificity of TEV and TVMV proteases. The relative specificity constants are given as values relative to that obtained with the respective unmodified substrate of the proteases. Substituted residues in the respective TEV and TVMV cleavage sites are in bold. ND, Not determined.

Peptide no.	Sequence	Enzyme	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}·s^{-1}$)	Rel. k_{cat}/K_m (%)
1	TENLYFQSGTRR	TEV S219V	0.043 ± 0.006	0.194 ± 0.007	4.51 ± 0.65	100
2	TEN**V**YFQSGTRR		> 0.5	ND	0.079 ± 0.001	2
3	TEN**A**YFQSGTRR		> 0.5	ND	0.027 ± 0.001	0.6
4	TENL**R**FQSGTRR		> 0.5	ND	0.027 ± 0.001	0.6
5	TENL**F**FQSGTRR		0.456 ± 0.050	0.161 ± 0.007	0.35 ± 0.041	8
6	TETVRFQSGTRR		–	–	–	
7	TET**L**RFQSGTRR		> 0.5	ND	0.030 ± 0.001	0.7
8	TET**A**RFQSGTRR		–	–	–	
9	TETV**Y**FQSGTRR		> 0.5	ND	0.066 ± 0.005	1.4
10	TETV**F**FQSGTRR		> 0.5	ND	0.007 ± 0.001	0.2
1	TENLYFQSGTRR	TVMV	–	–	–	
2	TEN**V**YFQSGTRR		> 0.5	ND	0.012 ± 0.001	0.3
3	TEN**A**YFQSGTRR		–	–	–	
4	TENL**R**FQSGTRR		–	–	–	
5	TENL**F**FQSGTRR		–	–	–	
6	TETVRFQSGTRR		0.034 ± 0.002	0.064 ± 0.001	1.88 ± 0.12	100
7	TET**L**RFQSGTRR		–	–	–	
8	TET**A**RFQSGTRR		–	–	–	
9	TETV**Y**FQSGTRR		> 0.5	ND	0.022 ± 0.001	1.2
10	TETV**F**FQSGTRR		> 0.5	ND	0.007 ± 0.001	0.4

by Val (peptide 2 in Table 1). None of the other single amino acid substitutions in the TEV peptide yielded peptides that could be cleaved by TVMV protease. Moreover, replacing Val with Leu in the P4 position of the TVMV peptide (peptide 7 in Table 1) prevented cleavage by TVMV protease. The substitution of P3 Arg in the TVMV peptide with either Tyr or Phe (peptides 9 and 10 in Table 1) resulted in a dramatic reduction in catalytic efficiency. As was the case with TEV protease, the substrate with Tyr in this position was cleaved more readily than the peptide with Phe in the P3 position.

Replacement of TEV protease residues by those of TVMV protease

To investigate the structural basis for the different sequence specificities of TEV and TVMV proteases, Ala169 and Lys220 in TEV protease were individually replaced with their counterparts in TVMV protease, which are Leu and Ala, respectively. The same series of substituted peptide substrates was used to assess the specificity of the A169L and K220A mutants (Table 2). In general, the mutant enzymes suffered a substantial loss of catalytic power, but they retained a mainly TEV protease-like specificity. Therefore, the TVMV protease substrate sequences were considered here as mutations in the TEV sequences (Table 2). To quantify the small specificity changes exerted by the mutations, ratios of the relative k_{cat}/K_m values were calculated (Table 2). These values are related to differences in the Gibbs' free energy changes ($\Delta\Delta G‡$) caused by the amino acid change in the substrate for a mutant enzyme relative to the change caused by the same amino acid change for the wild-type enzyme. The A169L mutant still preferred Leu in the P4 position over Val, like wild-type TEV protease. Nevertheless, there was a relative 15-fold decrease in this preference, as evidenced by the relative k_{cat}/K_m values obtained for the mutant and wild-type enzymes, in the TEV substrate sequence background. Ala was also relatively more tolerated by the A169L mutant. Somewhat different results were observed with the modified TVMV substrates: a P4 Leu substitution appeared to be much more favorable in the TVMV substrate sequence background (see peptides 6 and 7 in Table 2), indicating a strong influence of sequence context on enzyme specificity.

The K220A mutant also showed the highest activity on the unmodified TEV substrate, and, as expected, the relative P4 preference was not sensitive to this mutation. Although this mutation did not change the preference for P3 Arg, this residue was eightfold more favorable for this mutant than for the wild-type enzyme (peptide 4 in Table 2). As expected from results of modeling, the Arg side chain of the substrate may partially compensate for the loss of the Lys side

Table 2. Comparison of the specificity of TEV proteases with TVMV residues in their substrate-binding subsites. Residues that are substituted in the TEV substrate sequence are in bold. Because the mutants contained only one amino acid substitution in the TEV protease sequence and retained a predominantly TEV protease-like activity, residues of the TVMV substrates are considered here as mutants of the TEV substrate sequence and are marked differently from in Table 1, but the peptide numbering is the same.

Peptide no.	Sequence	Enzyme	k_{cat}/K_m ($mM^{-1}·s^{-1}$)	Rel. k_{cat}/K_m (%)	Rel. k_{cat}/K_m ratio (mut/wt E)
1	TENLYFQSGTRR	S219V	4.51	100	
2	TEN**V**YFQSGTRR		0.079	2	
3	TEN**A**YFQSGTRR		0.027	0.6	
4	TENL**R**FQSGTRR		0.027	0.6	
5	TENL**F**FQSGTRR		0.35	8	
6	TE**TVR**FQSGTRR		–	0	
7	TE**T**L**R**FQSGTRR		0.030	0.7	
8	TE**TAR**FQSGTRR		–	0	
9	TE**TV**YFQSGTRR		0.066	1.4	
10	TE**TVF**FQSGTRR		0.007	0.2	
1	TENLYFQSGTRR	S219V A169L	0.011	100	1
2	TEN**V**YFQSGTRR	(S4 mutant)	0.0033	30	15
3	TEN**A**YFQSGTRR		0.0011	10	17
4	TENL**R**FQSGTRR		0.0003	3	5
5	TENL**F**FQSGTRR		0.0012	11	1.4
6	TE**TVR**FQSGTRR		–	0	0
7	TE**T**L**R**FQSGTRR		0.0007	6	9
8	TE**TAR**FQSGTRR		–	0	0
9	TE**TV**YFQSGTRR		0.0015	13	9
10	TE**TVF**FQSGTRR		–	0	
1	TENLYFQSGTRR	S219V K220A	0.40	100	1
2	TEN**V**YFQSGTRR	(S3 mutant)	0.013	3	1.5
3	TEN**A**YFQSGTRR		0.003	1	1.7
4	TENL**R**FQSGTRR		0.021	5	8
5	TENL**F**FQSGTRR		0.038	10	1.3
6	TE**TVR**FQSGTRR		–	0	0
7	TE**T**L**R**FQSGTRR		0.021	5	7
8	TE**TAR**FQSGTRR		–	0	0
9	TE**TV**YFQSGTRR		0.008	2	1.4
10	TE**TVF**FQSGTRR		0.001	0.2	1

chain in the enzyme. However, the better tolerance for Arg at P3 is not observed in the TVMV substrate series (peptides 6 and 9 in Table 2). This is probably due to the altered sequence context, while the relative preference for P3 Tyr over Phe remained conserved (peptides 9 and 10 in Table 2).

Mutational analysis of other putative specificity determinants in TEV protease

The Phe in the P2 position of the canonical TEV protease substrate engages in hydrophobic interactions with Phe139, Val209, Trp211, Val216 and Met218 in the S2 pocket of the enzyme (Fig. 5A). Tyr is unfavorable in this position because of steric hindrance and disturbance of the hydrophobicity in the pocket. Inspection of the co-crystal structure suggested that specificity of the enzyme might be altered so that it would prefer Tyr instead of Phe in its S2 pocket by replacing Val209 with Ser. In principle, this mutation would increase the size of the S2 pocket, enabling it to accommodate the OH group of Tyr, while simultaneously creating an opportunity for a hydrogen bond to form between the OH of P2 Tyr in the substrate and that of Ser209 in the enzyme. In practice, however, the V209S mutant still preferred Phe over Tyr, although this preference was reduced $\approx$16-fold relative to the wild-type enzyme, while the relative preferences for peptides with substitutions at other positions (P4 and P6) did not change substantially (Table 3).

The P4 Leu in the canonical TEV protease substrate makes very favorable hydrophobic interactions with the side chains of Phe139, Ala169, Tyr178 and His214 in the enzyme. Owing to its small size, the S4 pocket cannot accommodate larger hydrophobic side chains such as that of Phe. Tyr178 forms the bottom of the

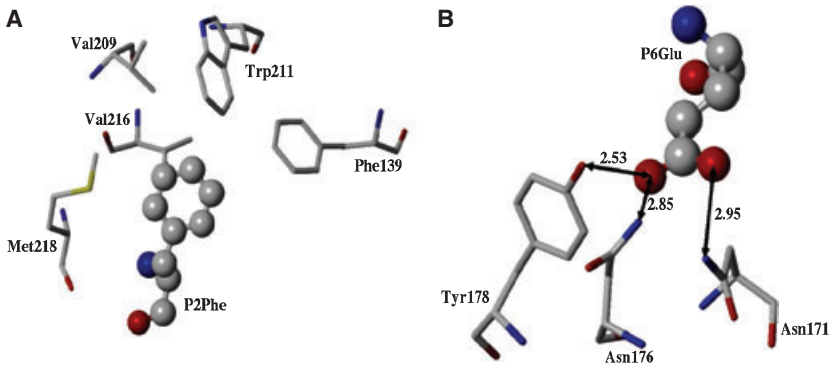


Fig. 5. S2 (A) and S6 (B) subsites of TEV NIa protease. Enzyme residues are shown with capped sticks, and the P2 and P6 residues of the substrate are shown with ball and stick representations. Hydrogen bonds are indicated by arrows.

Table 3. Comparison of the specificity of TEV protease with mutations of key residues of the substrate-binding subsites. Substituted residues in the TEV substrate sequence are in bold. ND, Not determined.

Enzyme	Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}s^{-1}$)	Rel. k_{cat}/K_m (%)	Rel. k_{cat}/K_m ratio (mut/wt E)
S219V	TENLYFQSGTRR	0.043 ± 0.006	0.194 ± 0.007	4.51 ± 0.65	100	
	TENLY**Y**QSGTRR	0.400 ± 0.031	0.022 ± 0.001	0.056 ± 0.005	1.2	
	TEN**F**YFQSGTRR	> 0.5	ND	0.024 ± 0.001	0.5	
	T**Q**NLYFQSGTRR	0.535 ± 0.090	0.109 ± 0.011	0.20 ± 0.04	4	
S219V/V209S	TENLYFQSGTRR	0.143 ± 0.025	0.036 ± 0.002	0.25 ± 0.05	100	1
	TENLY**Y**QSGTRR	>0.5	ND	0.048 ± 0.002	19	16
	TEN**F**YFQSGTRR	> 0.5	ND	0.001 ± 0.0001	0.4	0.8
	T**Q**NLYFQSGTRR	> 0.5	ND	0.013 ± 0.0004	5	1.2
S219V/Y178V	TENLYFQSGTRR	> 0.5	ND	0.017 ± 0.001	100	1
	TENLY**Y**QSGTRR	N.D.	ND	< 0.001	–	–
	TEN**F**YFQSGTRR	> 0.5	ND	0.012 ± 0.001	71	42
	T**Q**NLYFQSGTRR	> 0.5	ND	0.003 ± 0.0002	18	5
S219V/N171D	TENLYFQSGTRR	0.246 ± 0.028	0.049 ± 0.003	0.20 ± 0.03	100	1
	TENLY**Y**QSGTRR	> 0.5	ND	0.007 ± 0.0001	4	3
	TEN**F**YFQSGTRR	> 0.5	ND	0.013 ± 0.001	6	12
	T**Q**NLYFQSGTRR	0.610 ± 0.130	0.090 ± 0.012	0.15 ± 0.04	75	19

S4 pocket (Fig. 3A). The structure of the enzyme–substrate complex suggested that the depth of this pocket might be increased by replacing Tyr178 with Val, enabling it to tolerate Phe in the P4 position of the substrate. Indeed, the Y178V mutant exhibits only a slight preference for Leu over Phe in this position, whereas the wild-type enzyme is nearly 200-fold more selective (Table 3). However, the Y178V mutation causes a vast reduction in the general catalytic efficiency of the enzyme, which may be due to the loss of a hydrogen bond between Tyr178 and P6 Glu. In good agreement with this prediction, Gln in the P6 position of the substrate is also much better tolerated by this mutant than the wild-type enzyme.

Glu is highly conserved in the P6 position of the natural TEV protease cleavage sites (Fig. 1). This residue is involved in an intricate network of hydrogen bonds in the crystal structure of the enzyme–substrate complex (Fig. 5B). All of these hydrogen bonds can be formed only if the P6 side chain is Glu because any other residue would interrupt this co-operative network. For instance, Gln in the P6 position would place two nitrogens in close proximity to one another, resulting in unfavorable interactions. At the same time, the remaining hydrogen bonds in the network would prevent the side chain of P6 Gln from rotating 180 ° to alleviate the electrostatic repulsion between the two side-chain amide nitrogens. The Oε2 atom of P6 Glu forms a hydrogen bond with Nδ2 of Asn171. We reasoned that replacing Asn171 with Asp might create a more favorable environment for Gln than Glu in the S6 pocket of the enzyme. The N171D mutant still exhibits a slight preference for Glu over Gln, yet it tolerates Gln in the P6 position much more readily than does the wild-type enzyme, resulting in a 19-fold loss of selectivity (Table 3).

Discussion

The principal objectives of this study were to identify amino acid residues in both the enzymes and the substrates that are responsible for the different sequence specificities of TEV and TVMV proteases, in order to create enzymes with altered specificity by site-directed mutagenesis of putative specificity determinants. A comparison of the natural cleavage sites for the two enzymes, together with the results of kinetic analyses reported here, indicate that the residues in the P3 and P4 positions of the substrate are the most crucial specificity discriminators. Similarly, comparison of the crystal structure of TEV protease in complex with a peptide substrate with a homology model of TVMV protease suggested that the major differences between the active sites of the two enzymes involves their S3 and S4 pockets.

Two parallel strategies were pursued in an effort to alter the sequence specificity of TEV protease. In one approach, the homology model of TVMV protease was compared with the experimentally determined crystal structure of TEV protease in complex with a canonical peptide substrate in order to identify residues that are likely to be responsible for the different sequence specificities of the two enzymes. The leading candidates, Ala169 and Lys220 in TEV protease, were mutated to Leu and Ala, their respective counterparts in TVMV protease, in an effort to create a chimeric enzyme of intermediate sequence specificity. In a complementary approach, based purely on a close inspection of the crystal structure of TEV protease in complex with a canonical peptide substrate, presumptive specificity determinants were mutated in an effort to elicit specific effects. Collectively, these experiments probed specificity determinants in the S2, S3, S4 and S6 pockets of TEV protease.

The catalytic activity and specificity of the mutant TEV proteases were compared with the wild-type TEV and TVMV enzymes. All of the mutants examined in this study were much less active than the wild-type enzyme. Moreover, all of them still cleaved the canonical peptide substrate more efficiently than the substrates that they were designed or predicted to recognize, although in some cases the difference was slight. Nevertheless, they all exhibited differences in specificity that are consistent with the predicted effects of the mutations. Hence, the results are consistent with the predicted role for these residues (based on crystal structure) in the interaction with substrate. The loss of activity of the mutants could be the result of less efficient folding compared with the wild-type protease, due to the local conformational/electrostatic changes exerted by the mutations at the active site, or by a combination of these effects. Because no tight-binding inhibitor of TEV protease is available, it is difficult to address the folding efficiency, which would only be expected to influence the k_{cat} values calculated from the total protein content. The changes in K_m for the mutants, together with the specificity alterations suggest that at least part of the effect of mutations was directly due to conformational/electrostatic changes of the substrate binding sites. At the same time, the results of this study also indicate that it will probably be very difficult to generate potyviral proteases with unique sequence specificities and acceptable catalytic power using either of the approaches taken here. As observed for various other proteases including papain [15] and HIV protease [16], the intertwined network of interactions that form the specificity pockets in potyviral proteases does not appear to be well suited for protein engineering.

Experimental procedures

Protein expression and purification

A mutant form of TEV protease, harboring an S219V substitution, was used as the 'wild-type' enzyme in this study. This mutation prevents autodigestion of TEV protease, but does not affect its catalytic efficiency [14]. Ser219 is located near the side of the S3 pocket, but its side chain points away from the enzyme. Consequently, it is not expected to influence the specificity of the protease. The vector used to produce the S219V TEV protease mutant, pRK793, was described previously [14]. Additional mutations (A169L, N171D, Y178V, V209S or K220A) were introduced into the ORF encoding the S219V TEV protease by overlap extension PCR [17], using pRK793 as the template. AttB recombination sites were added to the ends of the PCR amplicons, which were subsequently recombined into the Gateway® destination vector pKM596 [18] to create the protease expression vectors. The nucleotide sequences of the inserts in all of the expression vectors were confirmed experimentally. All of the mutant proteases were produced in the form of maltose-binding protein fusion proteins which cleaved themselves *in vivo* at a canonical TEV protease-recognition site (ENLYFQ↓G) to yield TEV protease catalytic domains with N-terminal His tags and C-terminal polyarginine tags [14].

Wild-type and mutant forms of TEV protease were overproduced and purified as follows. BL21(DE3) CodonPlus RIL cells (Stratagene, La Jolla, CA, USA) containing a TEV protease expression vector were grown in shake flasks at 37 °C in Luria broth containing 100 $\mu g \cdot mL^{-1}$ ampicillin and 30 $\mu g \cdot mL^{-1}$ chloramphenicol. When the cells reached mid-exponential phase ($A_{600} \approx 0.5$), isopropyl thio-β-D-galacto-

side was added to a final concentration of 1 mM, and the temperature was reduced to 30 °C. After 4 h of induction, the cells were collected by centrifugation and stored at −70 °C.

All purification procedures were carried out at 4 °C. Cell pellets were suspended in ice-cold lysis buffer [50 mM Hepes (pH 8.0), 100 mM NaCl, 10% glycerol, and 25 mM imidazole] containing Complete® protease inhibitor cocktail (Roche, Mannheim, Germany) and 1 mM benzamidine, and disrupted by three passes through an APV Gaulin model G1000 homogenizer at 70–76 MPa. Polyetheleneimine from a 5% stock solution (adjusted to pH 7.9 with HCl) was added to a final concentration of 0.1%, and the homogenate was centrifuged at 30 000 ***g*** for 30 min. The supernatant fractions were filtered through a 0.2-μm polyethersulfone membrane and applied to a Ni/nitrilotriacetate/agarose column (Qiagen, Valencia, CA, USA) equilibrated with lysis buffer. The column was washed extensively and eluted with a linear gradient to 200 mM imidazole over 10 column volumes. Fractions containing recombinant protease were pooled, and EDTA and dithiothreitol were added to a final concentration of 1 mM and 5 mM, respectively. The samples were diluted fourfold with 50 mM Hepes (pH 8)/1 mM EDTA, and then applied to a HiTrap SP FF column equilibrated with this buffer. Proteins were eluted with a linear gradient to 1 M NaCl over 30 column volumes. Relevant fractions were pooled and concentrated using an Amicon YM-10 membrane. The samples were fractionated on a HiPrep 26/60 Sephacryl S100 HP column (Amersham Biosciences, Piscataway, NJ, USA) equilibrated with buffer [25 mM Hepes (pH 7.5), 100 mM NaCl, 5% glycerol, 2 mM dithiothreitol]. Purified recombinant proteases (> 95% pure as assessed by SDS/PAGE) were concentrated to 1–2 $mg{\cdot}mL^{-1}$, flash-frozen with liquid nitrogen, and stored at −70 °C until use. The molecular masses were confirmed by electrospray ionization MS.

Expression and purification of the wild-type TVMV protease catalytic domain with an N-terminal His tag has been described elsewhere [19].

Oligopeptide synthesis and characterization

Oligopeptides were synthesized by standard 9-fluorenylmethyloxycarbonyl chemistry on a model 430A automated peptide synthesizer (Applied Biosystems, Inc., Foster City, CA, USA) with amide C-terminus. Stock solutions were made in distilled water and the peptide concentrations were determined by amino acid analysis after peptide hydrolysis using a Beckman 6300 amino acid analyzer (Beckman Coulter Inc, Fullerton, CA, USA).

Enzyme kinetics

The protease assays were initiated by the mixing of 20 μL protease solution of S219V TEV protease, S219V/A169L, S219V/N171D, S219V/Y178V, S219V/V209S, S219V/K220A double mutant TEV proteases or TVMV protease (50–5700 nM) in 50 mM sodium phosphate, pH 7.0, containing 5 mM dithiothreitol, 800 mM NaCl, 10% glycerol, and 20 μL substrate solution (0.04–1.1 mM, actual range was selected on the basis of approximate K_m values). The enzyme concentrations were determined by amino acid analysis. Measurements were performed at six different substrate concentrations. The reaction mixture was incubated at 30 °C for 30 min, and the reaction was stopped by the addition of 160 μL 4.5 M guanidine hydrochloride containing 1% trifluoroacetic acid. An aliquot was injected on to a Nova-Pak C_{18} reversed-phase chromatography column (3.9 × 150 mm; Waters Corporation, Milford, MA, USA) 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% trifluoroacetic acid. To determine the correlation between peak areas of the cleavage products and their amount, fractions were collected and analyzed by amino acid analysis. The k_{cat} values were calculated by assuming 100% activity for the enzyme. 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., Durham, NC, USA). The standard deviations for the k_{cat}/K_m values were calculated as described [20]. If no saturation was obtained in the studied concentration range, the k_{cat}/K_m value was determined from the linear part of the rate vs. concentration profile.

Molecular modeling of TVMV protease

A molecular model of TVMV protease was built by MODELLER 3 [21], based on the structure of C151A mutant TEV protease (PDB code: 1LVB [12]). A sequence alignment of the two proteases was made by the CLUSTALW 1.74 program [22]. Structures were examined on Silicon Graphics O2 workstation using SYBYL (Tripos, St Louis, MO, USA).

Acknowledgements

We thank Karen Routzahn and Howard Peters for expert technical assistance, and Suzanne Specht for peptide synthesis and amino acid analyses. Electrospray ionization MS experiments were conducted using the LC/ES-MS instrument maintained by the Biophysics Resource in the Structural Biophysics Laboratory, Center for Cancer Research, National Cancer Institute at Frederick.

References

1 Ryan MD & Flint M (1997) Virus-encoded proteinases of the picornavirus super-group. *J Gen Virol* **78**, 699–723.

2 Tong L (2002) Viral proteases. *Chem Rev* **102**, 4609–4626.
3 Urcuqui-Inchima S, Haenni AL & Bernardi F (2001) Potyvirus proteins: a wealth of functions. *Virus Res* **74**, 157–175.
4 Riechmann JL, Lain S & Garcia JA (1992) Highlights and prospects of potyvirus molecular biology. *J Gen Virol* **73**, 1–16.
5 Dougherty WG & Semler BL (1993) Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiol Rev* **57**, 781–822.
6 Melcher K (2000) A modular set of prokaryotic and eukaryotic expression vectors. *Anal Biochem* **277**, 109–120.
7 Dougherty WG, Carrington JC, Cary SM & Parks TD (1988) Biochemical and mutational analysis of a plant virus polyprotein cleavage site. *EMBO J* **7**, 1281–1287.
8 Carrington JC, Haldeman R, Dolja VV & Restrepo-Hartwig MA (1993) Internal cleavage and trans-proteolytic activities of the VPg-proteinase (NIa) of tobacco etch potyvirus *in vivo*. *J Virol* **67**, 6995–7000.
9 Schaad MC, Haldeman-Cahill R, Cronin S & Carrington JC (1996) Analysis of the VPg-proteinase (NIa) encoded by tobacco etch potyvirus: effects of mutations on subcellular transport, proteolytic processing, and genome amplification. *J Virol* **70**, 7039–7048.
10 Kapust RB, Tözsér J, Copeland TD & Waugh DS (2002) The P1′ specificity of tobacco etch virus protease. *Biochem Biophys Res Commun* **294**, 949–955.
11 Yoon HY, Hwang DC, Choi KY & Song BD (2000) Proteolytic processing of oligopeptides containing the target sequences by the recombinant tobacco vein mottling virus NIa proteinase. *Mol Cells* **10**, 213–219.
12 Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HPK III, Kapust RB, Li M, Wlodawer A & Waugh DS (2002) Structural basis for the substrate specificity of tobacco etch virus protease. *J Biol Chem* **277**, 50564–50572.
13 Dougherty WG, Cary SM & Parks TD (1989) Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. *Virology* **171**, 356–364.
14 Kapust RB, Tözsér J, Fox JD, Anderson DE, Cherry S, Copeland TD & Waugh DS (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* **14**, 993–1000.
15 Berti PJ, Faerman CH & Storer AC (1991) Cooperativity of papain–substrate interaction energies in the S2 to S2′ subsites. *Biochemistry* **30**, 1394–1402.
16 Tözsér J, Bagossi P, Weber IT, Louis JM, Copeland TD & Oroszlan S (1997) Studies on the symmetry and sequence context dependence of the HIV-1 proteinase specificity. *J Biol Chem* **272**, 16807–16814.
17 Ho SN, Hunt HD, Horton RM, Pullen JK & Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59.
18 Fox JD & Waugh DS (2003) Maltose-binding protein as a solubility enhancer. *Methods Mol Biol* **205**, 99–117.
19 Nallamsetty S, Kapust RB, Tözsér J, Cherry S, Tropea JE, Copeland TD & Waugh DS (2004) Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease *in vivo* and *in vitro*. *Protein Expr Purif* **38**, 108–115.
20 Boross P, Bagossi P, Copeland TD, Oroszlan S, Louis JM & Tözsér J (1999) Effect of substrate residues on the P2′ preference of retroviral proteinases. *Eur J Biochem* **264**, 921–929.
21 Sali A & Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779–815.
22 Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.