

# Structural determinants of tobacco vein mottling virus protease substrate specificity

Ping Sun,<sup>1</sup> Brian P. Austin,<sup>1</sup> József Tözsér,<sup>2</sup> and David S. Waugh<sup>1\*</sup>

<sup>1</sup>Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, Frederick, Maryland

<sup>2</sup>Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary

Received 6 July 2010; Revised 23 August 2010; Accepted 24 August 2010

DOI: 10.1002/pro.506

Published online 22 September 2010 proteinscience.org

**Abstract:** Tobacco vein mottling virus (TVMV) is a member of the Potyviridae, one of the largest families of plant viruses. The TVMV genome is translated into a single large polyprotein that is subsequently processed by three virally encoded proteases. Seven of the nine cleavage events are carried out by the NIa protease. Its homolog from the tobacco etch virus (TEV) is a widely used reagent for the removal of affinity tags from recombinant proteins. Although TVMV protease is a close relative of TEV protease, they exhibit distinct sequence specificities. We report here the crystal structure of a catalytically inactive mutant TVMV protease (K65A/K67A/C151A) in complex with a canonical peptide substrate (Ac-RETVRFQSD) at 1.7-Å resolution. As observed in several crystal structures of TEV protease, the C-terminus (~20 residues) of TVMV protease is disordered. Unexpectedly, although deleting the disordered residues from TEV protease reduces its catalytic activity by ~10-fold, an analogous truncation mutant of TVMV protease is significantly more active. Comparison of the structures of TEV and TVMV protease in complex with their respective canonical substrate peptides reveals that the S3 and S4 pockets are mainly responsible for the differing substrate specificities. The structure of TVMV protease suggests that it is less tolerant of variation at the P1' position than TEV protease. This conjecture was confirmed experimentally by determining kinetic parameters  $k_{\text{cat}}$  and  $K_{\text{m}}$  for a series of oligopeptide substrates. Also, as predicted by the cocrystal structure, we confirm that substitutions in the P6 position are more readily tolerated by TVMV than TEV protease.

**Keywords:** tobacco vein mottling virus; tobacco etch virus; crystal structure; protease

## Introduction

Tobacco vein mottling virus (TVMV) is a member of the Potyviridae family, which composes one large branch of the Picornaviridae superfamily.<sup>1</sup> Like other positive-sense, single-stranded RNA viruses, the TVMV genome is initially translated into a single large polyprotein that is subsequently processed into individual proteins by three viral proteases: protein 1

(P1), helper component protease (HC-Pro), and nuclear inclusion-a protease (NIa-pro).<sup>2,3</sup> Seven of the nine cleavage events are carried out by the NIa protease.<sup>4</sup> NIa proteases adopt a chymotrypsin-like fold but use a cysteine residue instead of a serine as the active-site nucleophile in the catalytic triad.<sup>5</sup>

The potyviral TEV and rhinoviral 3C proteases are widely used as reagents for endoproteolytic removal of affinity tags from recombinant proteins because of their stringent substrate specificity. TVMV protease, a close relative of TEV protease, has also been used for this purpose.<sup>6–10</sup> Although they share a high degree of sequence identity (52%), these two proteases have distinct substrate specificities and do not cleave each other's canonical

---

Grant sponsor: Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

\*Correspondence to: David S. Waugh, National Cancer Institute at Frederick, P.O. Box B, Frederick, MD.  
E-mail: waughd@mail.nih.gov

recognition sites. Consequently, TVMV protease may be a useful alternative to TEV protease when a recombinant protein happens to contain a sequence that is similar to a TEV protease recognition site or for protein expression strategies that involve the use of more than one protease.<sup>11</sup> Seeking to understand the structural basis for the differing sequence specificities of TEV and TVMV proteases, we have crystallized the latter enzyme in complex with a peptide substrate and determined its structure at a resolution of 1.7 Å. Several intriguing features of the cocrystal structure were investigated in greater detail by characterizing a mutant form of TVMV protease and variants of a canonical TVMV oligopeptide substrate.

## Results and Discussion

### Crystallization and structure determination

Wild-type and mutant forms of the TVMV protease catalytic domain were overproduced in *Escherichia coli* and purified as described.<sup>9</sup> To cocrystallize the enzyme with a canonical peptide substrate, a catalytically inactive mutant was constructed by replacing the nucleophilic active-site cysteine with an alanine (C151A). However, no crystals of this mutant protease were ever obtained. Therefore, we used the technique of surface entropy reduction mutagenesis<sup>12</sup>; two lysine residues were replaced by alanines to create the triple mutant K65A/K67A/C151A. These additional mutations were selected by examining a homology model of the TVMV protease structure that was derived from the structure of TEV protease.<sup>10</sup> Additionally, the C-terminus of the TVMV mutant was trimmed by six residues to remove the P6–P1 sites of the natural polyprotein processing site.

The purified inactive triple mutant TVMV protease (K65A/K67A/C151A) was mixed with a fivefold molar excess of peptide substrate (Ac-RETVRFQSD) before crystallization trials. The crystal used for data collection grown from a solution consisting of 0.2M potassium formate and 20% PEG 3350, belongs to space group  $P2_12_12_1$  and contains two monomers per asymmetric unit. The structure was solved by molecular replacement, using the crystal structure of TEV protease (Protein Data Bank (PDB) code: 1Q31) as a search model. The final model was refined to a resolution of 1.7 Å with an  $R_{\text{work}}$  of 17.5% and an  $R_{\text{free}}$  of 21.0%. It is noteworthy that, as is frequently the case when surface entropy reduction mutants are crystallized,<sup>12</sup> the K65A and K67A mutations in TVMV protease are located at an interface between two symmetry-related molecules in the crystal lattice.

### Overall structure of TVMV protease and comparison with TEV protease

As expected, TVMV protease adopts a typical chymotrypsin-like fold, which consists of two  $\beta$ -barrel

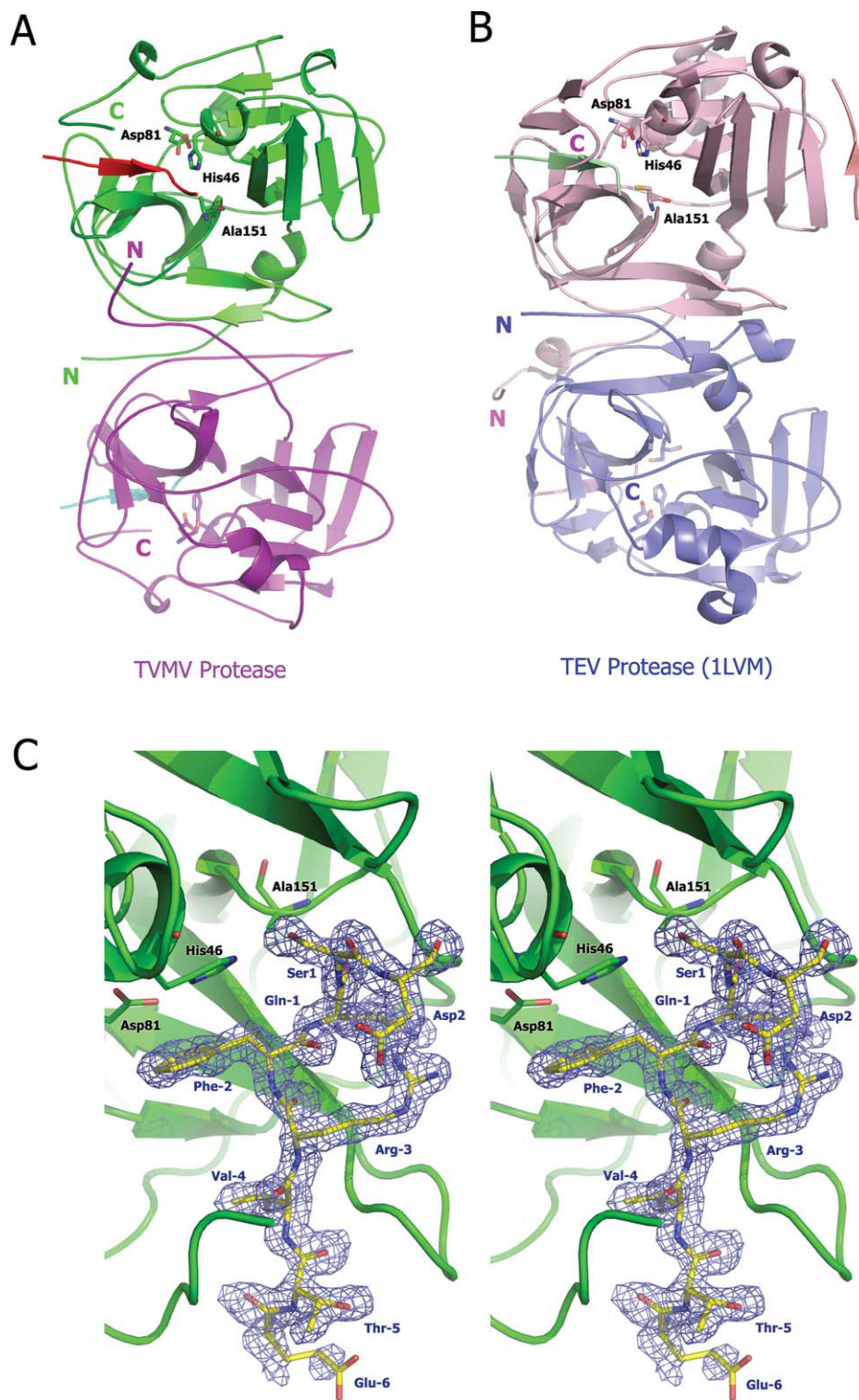
domains that pack together to form a shallow peptide-binding cleft with the catalytic triad residues His<sup>46</sup>, Asp<sup>81</sup>, and Cys<sup>151</sup> located at the interface [Fig. 1(A)]. The two molecules in the asymmetric unit form a dimer that bears a superficial resemblance to the one observed in structure of the S219D mutant of TEV protease<sup>14</sup> [Fig. 1(B)]. However, neither TVMV nor TEV protease has been reported to form dimers in solution, suggesting that the intermolecular interactions observed in the crystals are purely the result of crystal packing.

The two TVMV protease molecules in the asymmetric unit are quite similar, with an overall RMSD of 1.36 Å. The principal differences are located in four loops, which occur between  $\beta 1/\beta 2$ ,  $3_{10}$ -helix-A/ $\beta 4$ ,  $\beta 5/3_{10}$ -helix-B, and  $\beta 8/\beta 9$ . None of these loops are close to the active site of the enzyme. In molecule A, the electron density for the bound substrate (Chain C, Ac-RETVRFQSD) is well defined except for the N-terminal Arg and C-terminal Asp residues. In molecule B, on the other hand, the C-terminal Asp residue of the substrate is clearly visible in the electron density map [Fig. 1(C)]. In both protease molecules, the peptide substrates are bound in an extended conformation in the active site.

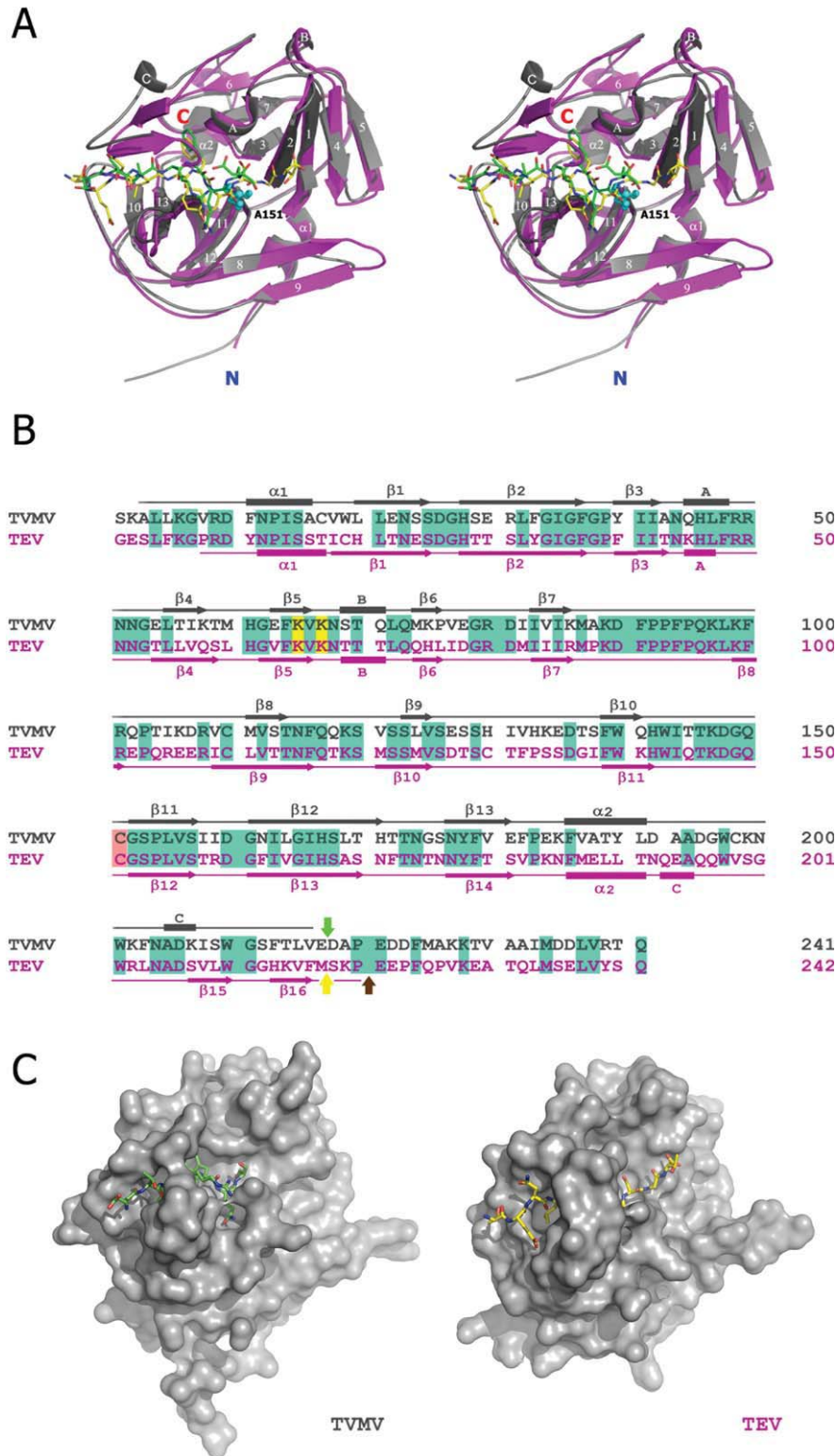
TVMV protease shares 52% amino acid sequence identity with TEV protease. Consequently, as one might expect, the two proteases exhibit a great deal of structural similarity with an overall RMSD of 2.96 Å for 209  $C_{\alpha}$  atoms (Fig. 2). The catalytic triad residues align very well, but as a result of substantive structural differences in the substrate-binding sites of the two enzymes, the backbone of the bound peptide in TVMV protease is shifted about 1 Å from the position that the corresponding peptide occupies in TEV protease [Fig. 2(A)]. The major differences between the two proteases are located in some of the loops, such as the loop between strands  $\beta 8$  and  $\beta 9$  and especially the loop between  $\beta 12$  and  $\beta 13$ , which is involved in the formation of the substrate-binding pocket. In particular, the conformations of the C-termini of the two proteases are extremely different. In TEV protease, the two C-terminal  $\beta$ -strands ( $\beta 15$  and  $\beta 16$ ) form a small antiparallel  $\beta$ -sheet. By contrast, the corresponding region of TVMV protease forms a  $3_{10}$ -helix (helix C) and a short loop.

### Enzymatic activity of TVMV<sup>WT</sup> and TVMV<sup>1-217</sup>

TEV protease readily undergoes autolysis at a specific site near its C-terminus, giving rise to a truncated enzyme with greatly reduced activity.<sup>15,16</sup> Curiously, although they clearly influence the enzymatic activity of the protease, the C-terminal residues that are removed by this truncation are not visible in electron density maps of the full-length enzyme in complex with a peptide substrate or product.<sup>14</sup> The closely related TVMV protease does not undergo autolysis.<sup>7,9</sup> Even so, those TVMV residues



**Figure 1.** Three-dimensional structure of TVMV protease. (A) Overall structure of TVMV protease in one asymmetric unit. Chain A (1–217, purple) and Chain B (3–216, green) bound to peptide substrates Chain C (residues –6 to 1, cyan) and Chain D (residues –6 to 2, red), respectively. The N- and C-termini are labeled with the letters N and C. The catalytic triad residues are shown in ball-and-stick representation. (B) Crystal structure of TEV protease (PDB ID: 1LVM) viewed from the same perspective. Chains A to E are colored in pink, blue, green, pale pink, and salmon, respectively. (C) Stereoview of the peptide substrate (Chain D, yellow) bound to inactive TVMV protease (Chain B, green). The catalytic triad residues [H46, D81, and A151 (normally C151 in wild-type TVMV protease)] are shown in ball-and-stick representation. The peptide substrate is also displayed in a ball-and-stick format and covered by an omit map contoured at  $1.0\sigma$ . The composite omit map was calculated at 1.7 Å resolution by CNS,<sup>13</sup> with an omit ratios of 7.5%.



**Figure 2.** Comparison between TVMV and TEV proteases. (A) Stereoview of superposition of the crystal structures of TVMV (Chain B, gray) and TEV (PDB ID: 1LVB, Chain A, magenta) proteases with their peptide substrates. Secondary structure elements are labeled according to the TVMV structure. Canonical peptide substrates bound to TVMV and TEV proteases are shown in green and yellow, respectively. The N- and C-termini of the proteases are labeled in blue and red, respectively. The side chains of the mutated catalytic residue (C151A in TVMV protease) from both structures are shown as cyan spheres. (B) Sequence alignment of TVMV (gray) and TEV (magenta) proteases. The secondary structures were determined by the results from iMolTalk server (<http://i.moltalk.org/>).  $\beta$ -strands and  $\alpha$ -helices are numbered and  $3_{10}$ -helices are labeled A–C. Identical residues are shaded in cyan. Two surface entropy reduction mutation sites (K65&K67) are shaded in yellow. The active-site cysteine residues are shaded in salmon. The self-cleavage site within TEV protease is indicated by the yellow arrow. The C-termini of available crystal structure models of TEV protease (PDB ID: 1LVB and 1LVM) are denoted by the brown arrow. The C-terminus of the truncated TVMV protease (TVMV<sup>1–217</sup> protease) investigated in this study is indicated by the green arrow. (C) Surface representation of TVMV (left) and TEV (right) proteases viewed at the same angle.

**Table I.** Kinetic Parameters for the Wild-Type (TVMV<sup>WT</sup>) and Truncated (TVMV<sup>1-217</sup>) Proteases with the Peptide Substrate TETVRFQSGTRR-NH<sub>2</sub>

Enzyme	$K_m$ (mM)	$k_{cat}$ (s <sup>-1</sup> )	$k_{cat}/K_m$ (mM <sup>-1</sup> s <sup>-1</sup> )	rel. $k_{cat}/K_m$
TVMV <sup>WT</sup>	0.082 ± 0.020	0.092 ± 0.010	1.12 ± 0.30	1.00
TVMV <sup>1-217</sup>	0.108 ± 0.016	0.094 ± 0.006	0.87 ± 0.14	0.78

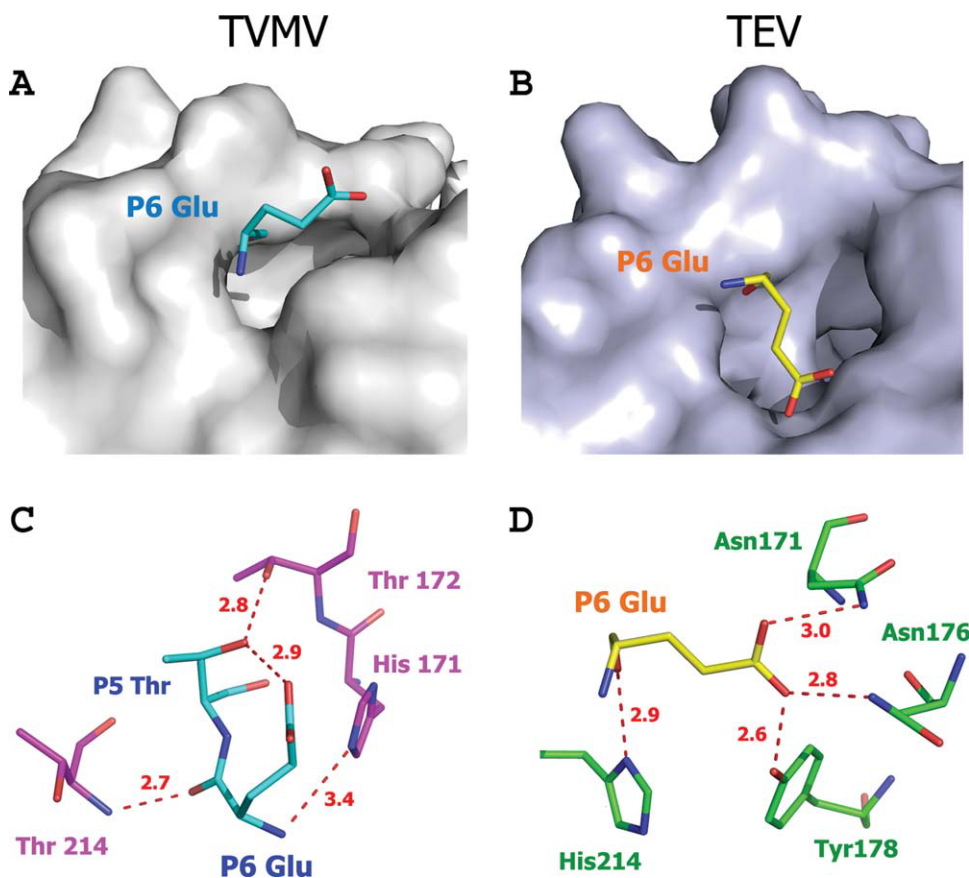
corresponding to the disordered C-terminal residues of TEV protease are also disordered in the cocrystal structure of the TVMV/substrate complex.

To ascertain whether there is any difference between the enzymatic activity of wild-type (TVMV<sup>WT</sup>) and truncated TVMV proteases, the latter protein (TVMV<sup>1-217</sup>) was overproduced and purified in the same manner as the other forms of TVMV protease. Kinetic parameters  $K_m$  and  $k_{cat}$  were determined for the full-length and truncated TVMV enzymes using an oligopeptide substrate (Table I). Surprisingly, the truncated TVMV enzyme was substantially more active than the analogous TEV protease construct, exhibiting nearly as much activity as the full-length enzyme. These results were confirmed by conducting assays with an MBP-NusG fusion protein substrate<sup>9</sup> *in vitro* (data not shown).

### Structural determinants of substrate specificity

Previous studies have shown that seven residues surrounding their cleavage sites (positions P6–P1') comprise the specificity determinants for TEV and TVMV proteases.<sup>8,17</sup> The most efficient substrates for TEV and TVMV proteases are ENLYFQS and ETVRFQS, respectively.<sup>8,17</sup> The availability of crystal structures of both proteases in complex with oligopeptide substrates containing their optimal (canonical) recognition sequences provides substantial insight into the structural determinants of substrate specificity for both enzymes.

The P6 glutamic acid is an important specificity determinant for TEV protease.<sup>10,18</sup> As shown in Figure 3(A,B), P6 Glu interacts with the loop between  $\beta$ 12 and  $\beta$ 13A in TEV protease, whereas the side chain of P6 Glu rotates 90° and points into the



**Figure 3.** Comparison between the P6 positions of TVMV and TEV protease (PDB ID: 1LVB, Chain A) substrates. (A, B) P6 Glu binds to the surface of its corresponding protease, viewed at the same angle. (C, D) Hydrogen-bond interactions between P6 Glu and TVMV protease (C) and TEV protease (D). Residues are shown in ball-and-stick representation. The hydrogen bonds are shown as dashed lines colored red. Residues from TVMV and TEV protease are colored in purple and green, respectively. The P6 substrate residues for TVMV and TEV proteases are colored cyan and yellow, respectively.

solvent in the TVMV protease/peptide cocrystal structure. As shown in Figure 3(C), the only interaction made by the side chain of P6 Glu in the TVMV/peptide structure is a hydrogen bond between its O- $\epsilon$ 1 atom and the O- $\gamma$  atom of P5 Thr. Protease residues make only main-chain contacts with P6 Glu, which should have little or no impact on sequence specificity. In TEV protease, the interactions with P6 Glu are far more extensive. As shown in Figure 3(D), atom O- $\epsilon$ 1 and atom O- $\epsilon$ 2 of P6 Glu are within hydrogen-bonding distance of N- $\delta$ 2 of Asn171, N- $\delta$ 2 of Asn176, and O- $\eta$  of Tyr178. In addition, its main-chain O atom makes a hydrogen bond with N- $\delta$ 1 of His214. These observations suggest that the P6 position is a more important specificity determinant for TEV protease than TVMV protease. Consistent with this notion, the P6 position in all of the naturally occurring TEV polyprotein processing sites is occupied by glutamic acid, whereas variations occur at the P6 sites in the TVMV polyprotein.<sup>10</sup>

The side chains of P5 Thr in the TVMV peptide and P5 Asn in TEV peptide both project away from the proteins into solvent. Consequently, consistent with the high diversity of residue types that occur at the P5 positions of the natural polyprotein processing sites,<sup>8</sup> this position should not be a significant specificity determinant for either enzyme. Furthermore, biochemical studies have shown that almost any residue can occupy the P5 position with little or no impact on the cleavage efficiency of TEV protease.<sup>19</sup>

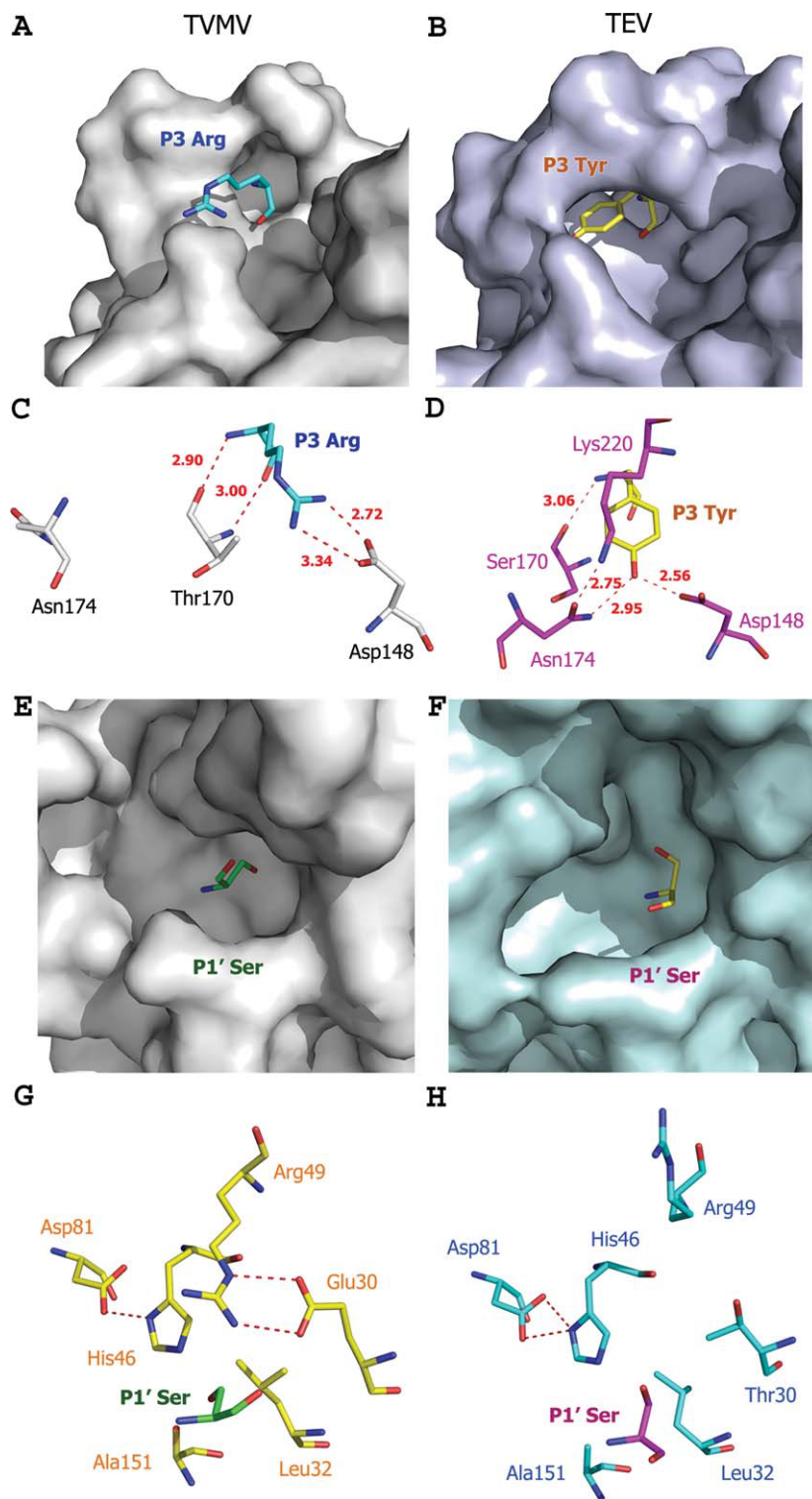
The S4 pockets are hydrophobic in both TVMV and TEV proteases. However, the S4 pocket of TVMV protease is shallower than that of TEV protease. The Van der Waals cavity volumes of the S4 pockets in TVMV protease and TEV protease are 137 and 241 Å<sup>3</sup>, respectively, calculated using *VOIDOO*<sup>20</sup> with a probe radius of 1.4 Å. This may explain why the P4 position is invariably occupied by valine in the natural TVMV processing sites whereas, although leucine is the residue most frequently found at this position in TEV processing sites, valine and isoleucine also occur in some cases.<sup>21</sup> Additionally, experiments conducted with oligopeptide substrates *in vitro* demonstrated that when the P4 valine in an otherwise optimal substrate for TVMV protease was replaced by leucine, it can no longer be cleaved by the enzyme.<sup>10</sup> This is understandable because the much smaller S4 pocket of TVMV protease cannot accommodate the larger side chain of leucine. On the other hand, when the P4 leucine is replaced with valine in the TEV protease substrate, the cleavage efficiency is dramatically decreased to only 2% of the canonical substrate,<sup>10</sup> indicating TEV protease strongly prefers leucine instead of valine in the P4 position because of its larger S4 pocket. Smaller residues in this position, like alanine, result in even less efficient cleavage.<sup>10</sup>

Unlike TEV protease, TVMV protease has no S3 pocket [Fig. 4(A–D)]. Rather, the side chain of P3 Arg in the TVMV protease substrate is fully exposed to solvent. Nevertheless, it forms two salt bridges with the side chain of Asp148. The precise geometry required to achieve these strong ionic interactions may be the reason why arginine is found in the P3 positions of all natural TVMV polyprotein processing sites except one, which contains a lysine residue instead. Recognition of the P3 Tyr by TEV protease is totally different and has been described in detail.<sup>14</sup> Consequently, the S3/P3 interactions appear to be major discriminators of specificity by the two proteases.

The S2 pockets in both TVMV and TEV protease are very hydrophobic, except that in TEV protease, it is more closed from the exposure to solvent. In TEV protease this pocket is formed by four hydrophobic residues (Val 209, Trp 211, Val 216, and Met218) and a face of His46. However, in TVMV protease, because of the different conformation of its C-terminus, the S2 pocket is not covered by a  $\beta$ -strand ( $\beta$ 16) as it is in TEV protease, leaving the pocket partially open. In addition to three conserved residues (Trp210, Leu215, and His46), Leu169 from  $\beta$ 12 and Phe203 from the loop between  $\alpha$ 2 and  $3_{10}$ -helix C form the S2 pocket in TVMV protease.

The P1 sites in both TVMV and TEV protease substrates are occupied by glutamine. The hydrogen bonds between atoms N- $\epsilon$ 2 and O- $\epsilon$ 1 of P1 Gln and the side chains of His167 and Thr146 are conserved in the two proteases. However, there are also some differences between the conformations of the S1 pockets. As shown in Figure 2(C), the S1 pocket in TVMV protease is partially open, whereas in TEV protease this pocket is fully closed with the P1 residue of the substrate buried inside. The hydrogen-bond network among the main-chain and side-chains atoms of Asp148, Ser170, Asn174, Lys220, and P3 Tyr covers the side chain of P1 Gln in TEV protease. This also explains why in the TEV N1a-Pro carboxy cleavage site, the aliphatic P1 Met is also acceptable. In TVMV protease, this hydrogen-bonding interaction no longer exists because of the movement of Asn174 from the loop between strands  $\beta$ 12/ $\beta$ 13 and the missing counterpart of Lys220 from its flexible C-terminus.

Unlike the S1' pocket of TEV protease, which is a shallow and narrow groove on its surface,<sup>14</sup> the S1' pocket of TVMV protease is round and small with the side chain of the P1' serine residue pointing into its inner surface [Fig. 4(E–H)]. The P1' positions of both TEV and TVMV polyprotein processing sites are typically occupied by Ser, Gly, or Ala. However, experimental data indicated that TEV protease exhibits a wide tolerance for variation in this position, including Met and Cys residues.<sup>22</sup> The fact that TVMV protease has a completely different S1' pocket raises the question of its sensitivity to variation in



**Figure 4.** Comparison of the S1' and S3 subsites in TVMV and TEV proteases. Left and right panels refer to the TVMV and TEV protease, respectively. (A, B) Surface representation of the S3 subsites. P3 residues are colored cyan (TVMV) and yellow (TEV). (C, D) Hydrogen-bond interaction at the S3 subsites. Residues from TVMV and TEV protease are colored gray and magenta, respectively. (E, F) Surface representation of the S1' pockets. P1' Ser residues of the peptide substrates are colored green (TVMV) and yellow (TEV). (G, H) Hydrogen-bond interactions in the S1' pockets. Residues are shown in ball-and-stick representation. The hydrogen bonds are shown as dashed lines colored red. Residues from TVMV and TEV proteases are colored yellow and cyan, respectively. P1' Ser residues of the peptide substrates are colored in green (TVMV) and magenta (TEV).

**Table II.** Comparison of the P6 and P1' Specificity of TEV and TVMV Proteases

Enzyme	Substrate	$K_m$ (mM)	$k_{cat}$ ( $s^{-1}$ )	$k_{cat}/K_m$ ( $mM^{-1} s^{-1}$ )	rel. $k_{cat}/K_m$
TVMV protease	TETVRFQ↓SGTRR	0.082 ± 0.020	0.092 ± 0.010	1.12 ± 0.30	1.00
	TETVRFQ↓AGTRR	0.460 ± 0.087	0.112 ± 0.015	0.24 ± 0.06	0.22
	TETVRFQ↓MGTRR	0.671 ± 0.100	0.037 ± 0.003	0.06 ± 0.01	0.05
	TETVRFQ↓YGTRR	0.425 ± 0.087	0.062 ± 0.008	0.15 ± 0.04	0.13
	TETVRFQ↓HGTRR	0.372 ± 0.074	0.012 ± 0.001	0.03 ± 0.01	0.03
	TATVRFQ↓SGTRR	0.244 ± 0.038	0.032 ± 0.002	0.13 ± 0.02	0.12
	TNTVRFQ↓SGTRR	0.142 ± 0.035	0.074 ± 0.008	0.52 ± 0.14	0.46
	TQTVRFQ↓SGTRR	0.169 ± 0.034	0.076 ± 0.008	0.45 ± 0.10	0.40
	TENLYFQ↓SGTRR	0.093 ± 0.009	0.345 ± 0.009	3.71 ± 0.37	1.00
TEV protease (S219V)	TENLYFQ↓AGTRR	0.109 ± 0.011	0.367 ± 0.016	3.36 ± 0.38	0.91
	TENLYFQ↓MGTRR	0.088 ± 0.008	0.198 ± 0.006	2.25 ± 0.22	0.61
	TENLYFQ↓YGTRR	0.060 ± 0.009	0.062 ± 0.004	1.03 ± 0.17	0.28
	TENLYFQ↓HGTRR	0.215 ± 0.056	0.300 ± 0.030	1.40 ± 0.38	0.31
	TANLYFQ↓SGTRR	>1.0	N.D.	0.15 ± 0.01	0.04
	TNNLYFQ↓SGTRR	0.648 ± 0.130	0.244 ± 0.023	0.38 ± 0.08	0.10
	TQNLVYFQ↓SGTRR	0.535 ± 0.090	0.109 ± 0.011	0.20 ± 0.04	0.05

ND, not determined.

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. Cleavage site is indicated in arrow.

the P1' position. As shown in Figure 4(C), it is the side chains of Arg49 and Glu30 that form the lid of the S1' pocket of TVMV protease. Further, this conformation is maintained through hydrogen bonds involving N- $\eta$ 1 of Arg49 and O- $\epsilon$ 2 of Glu30, N- $\epsilon$  of Arg49, and O- $\epsilon$ 1 of Glu30. Although Arg49 is conserved between TEV and TVMV proteases, the hydrogen-bonding partner Glu30 in this interaction is not. The equivalent residue in TEV protease is a threonine. Although threonine is also a hydrophilic residue with a hydroxyl group that might contribute to a hydrogen-bonding interaction, its comparatively shorter side chain eliminates the possibility in this case. Instead, Arg49 in TVMV protease points away from the molecule into the solvent and forms a hydrogen bond with the side chain of Asp207 from the loop between  $3_{10}$ -helix C and  $\beta$ -15 [Fig. 2(B)]. It is interesting to note that the conserved equivalent residue to Asp207 in TVMV protease and Asp206 in TEV protease actually is located in the  $3_{10}$ -helix C with its  $C_\alpha$  atom located 11.3 Å away from that of Asp207 in TVMV protease. The different conformations of their C-termini are the main reason that TVMV and TEV proteases possess such distinct S1' pockets.

### Experimental tests of insights derived from the structure of TVMV protease

Comparison of the structures of the two enzyme/substrate complexes led to a number of insights into the structural basis for their distinct sequence specificities, as described above. Among these, two conjectures were selected to be tested experimentally: (1) that the P6 specificity of TEV protease should be more stringent than that of TVMV protease, and (2) that the P1' specificity of TVMV protease should be

more stringent than that of TEV protease. To this end, a series of oligopeptide substrates with variations in the P6 or P1' sites were prepared, and the kinetic parameters  $K_m$  and  $k_{cat}$  were determined for reactions with TVMV and TEV proteases. The results are presented in Table II. As predicted, the kinetic data confirm that TVMV protease is much less tolerant than TEV protease of amino acid substitutions in the P1' position of its substrates. Hence, relaxed specificity in the S1'pocket is not a general feature of potyvirus proteases, but a fortuitous property of TEV protease. Also as predicted, substitutions in the P6 position are more readily tolerated by TVMV than TEV protease. Hence, these results are consistent with the cocrystal structure of the enzyme/substrate complex.

### Conclusions

Although it is a common practice to analyze and compare substrate specificity determinants of related proteases in a pocket-by-pocket and residue-by-residue manner, one must bear in mind that the substrate-binding site of TVMV protease is not, strictly speaking, composed of independent binding pockets for each residue. Rather, the properties of one pocket may very well influence those of the adjacent ones. Nevertheless, the structure of TVMV protease in complex with its canonical substrate peptide has provided considerable insight into the structural basis of its substrate specificity, leading us to conclude that the S3 and S4 pockets appear to function together as the main specificity determinants. Although structurally similar to TEV protease, TVMV protease has a distinct sequence specificity. In some cases, it may be advantageous to use more than one protease during the production of a



recombinant protein or domain. For example, the Midwest Center for Structural Genomics has developed a strategy for protein expression and purification in *E. coli* that uses both TEV and TVMV proteases.<sup>23</sup> Using this approach, they successfully “rescued” several targets that proved very difficult to purify with the conventional single protease method.

## Materials and Methods

### Protein expression and purification

The catalytically inactive C151A TVMV protease mutant was constructed by polymerase chain reaction (PCR) amplification of the open reading frame (ORF) from pRK1035<sup>9</sup> ([www.addgene.org](http://www.addgene.org), plasmid #8832) using the forward primer PE-1449 (5'-GAG AAC CTG TAC TTC CAG TCT AAA GCT TTG CTG AAG GGC GTG-3'), which annealed to the 5' end of the TVMV ORF and added an in-frame tobacco etch virus (TEV) protease recognition site upstream, of the ORF, and the reverse primer PE-639 (5'-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT ATT AGT CCA TGA TGG CGG CAA CAG-3'), which annealed to the 3' end of TVMV ORF and added an attB2 Gateway recombination site (Invitrogen, Carlsbad, CA) to the end of the amplicon. An attB1 Gateway recombination site was subsequently added to the 5' end of the resulting PCR amplicon in a second round of PCR, using the attB1-TEV primer PE-277<sup>24</sup> and the same reverse primer. The final PCR product was recombined into pDONR201 via the Gateway BP reaction to generate pKP1485, and the DNA sequence was confirmed. The C151A mutation was subsequently introduced by QuikChange mutagenesis (Stratagene, La Jolla, CA). The C151A mutant TVMV protease ORF, preceded by an in-frame TEV protease recognition site, was moved into the destination vector pDEST-HisMBP<sup>25</sup> via the Gateway LR reaction to generate pKP1487. The surface entropy reduction mutations (K65A/K67A) were introduced into pKP1487 by overlap extension PCR<sup>26</sup> to create pBA1675.

The expression vector for C-terminally truncated TVMV protease (TVMV<sup>1-217</sup>) was constructed by (PCR) amplification of the ORF from pRK1035 using the forward primer PE-727 (5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC GGA AAC CGT GCG TTT CCA GTC TC-3'), which annealed to the 5' end of the TVMV ORF and the upstream TVMV protease recognition site while adding an attB1 recombination site to this end of the amplicon, and the reverse primer PE-2070 (5'-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT ATT ATT CAA CCA GGG TAA AGG AAC-3'), which introduced a termination codon after residue 217 of the TVMV ORF while adding an attB2 Gateway recombination site to the other end of the amplicon. The PCR product was recombined into pDONR201 via the Gateway

BP reaction to generate pPS2021, and the DNA sequence was confirmed. The truncated TVMV protease ORF, now preceded by an in-frame TVMV protease recognition site, was moved into the destination vector pDEST-HisMBP<sup>25</sup> via the Gateway LR reaction to generate pPS2022. The construction of vectors for the production of full-length TEV (pRK793) and TVMV (pRK1035) proteases has been described previously.<sup>9,16</sup>

All recombinant proteins were expressed in *E. coli* BL21(DE3) CodonPlus-RIL cells (Stratagene, La Jolla, CA), which were induced at mid-log phase with 1 mM IPTG for 4 h at 30°C. The cells were harvested by centrifugation at 4°C and frozen at -80°C until use. Purification of full-length TEV and TVMV proteases was carried out as described.<sup>9,27</sup> Truncated TVMV protease (TVMV<sup>1-217</sup>) was purified in the same manner as full-length TVMV protease.<sup>9</sup> The catalytically inactive mutant TVMV protease (K65A/K67A/C151A) was purified as follows. The cell pellet was resuspended in 50 mM sodium phosphate, pH 8.0, 150 mM NaCl, and 25 mM imidazole, and the cells were disrupted using a APV Model G1000 homogenizer (Invensys, Røhølmvej, Denmark). The lysate was centrifuged at 15,000 rpm at 4°C using an SA-600 rotor, filtered, and then the HisMBP-TVMV fusion protein was purified by immobilized metal affinity chromatography (IMAC) as described.<sup>25</sup> Fractions containing the fusion protein were pooled, cleaved overnight with hexahistidine-tagged TEV protease,<sup>27</sup> and then subjected to another round of IMAC as described.<sup>25</sup> The flow-through fractions were pooled and concentrated to 5 mL, using an Amicon stirred cell with a YM10 membrane (Millipore, Billerica, MA) and applied to a 26/60 Superdex-75 preparative size exclusion column (GE Healthcare, Piscataway, NJ) equilibrated in 25 mM Tris, pH 7.5, and 2 mM tris(2-carboxyethyl) phosphine hydrochloride, and the peak fractions corresponding to TVMV protease were pooled and concentrated to ~15 mg/mL with an Amicon stirred cell (Millipore) as above.

### Crystallization and data collection

The inactive enzyme-substrate complex was prepared by mixing the protein solution (10 mg/mL) with a fivefold molar excess of the peptide substrate (Ac-RETVRFQSD). The complex was then subjected to crystallization trials with various kits from Hampton Research, Qiagen, and Emerald Biosystems. The Hydra II Plus One crystallization robot (Matrix Technologies, Hudson, NH) was used to setup the screens in a sitting drop vapor diffusion format at 18°C. The crystal used for data collection was grown from a solution consisting of 0.2M potassium formate, 20% PEG 3350 with a ratio of protein to reservoir solution of 1:3. Crystals of mutant 1 (K65A/K67A/C151A) appeared within 4 days. The

crystals belong to the space group  $P2_12_12_1$  and contain two monomers per asymmetric unit. The solvent content of the crystal was estimated to be ~43.4% (v/v) with a Matthews coefficient ( $V_M$ ) of  $2.13 \text{ \AA}^3 \text{ Da}^{-1}$  for two monomers in the asymmetric unit. The unit cell has dimensions of  $a = 76.5 \text{ \AA}$ ,  $b = 77.6 \text{ \AA}$ , and  $c = 78.5 \text{ \AA}$ . Crystals were equilibrated in a cryoprotectant composed of reservoir buffer plus 5% (v/v) ethylene glycerol and were flash frozen in a cold nitrogen stream at  $-170^\circ\text{C}$ . One dataset was collected at the SER-CAT beamline 22-ID (Advanced Photon Source, Argonne National Laboratory). Data were processed and scaled using the HKL2000 program suite.<sup>28</sup> Data collection and processing statistics are summarized in Table III.

### Structure determination and refinement

The structure of TVMV protease mutant (K65A/K67A/C151A) in complex with the peptide substrate was solved by molecular replacement method using the monomer of TEV protease (PDB code: 1Q31<sup>29</sup>) as the search model and the program MOLREP.<sup>30</sup> The substrate peptide and the loops were deleted from the search model. Five percent of the reflections were set aside for cross validation ( $R_{\text{free}}$ ). After initial rigid body refinement, the molecular replacement solution yielded an  $R_{\text{factor}}$  of 0.45 and  $R_{\text{free}}$  of 0.48. The resulting model was manually corrected and finished in O.<sup>31</sup> Refinement was carried out with REFMAC5.<sup>32</sup> The final model consists of residues 1–217 of Chain A, 3–216 of Chain B, 2–8 of Chain C, and 2–9 of Chain D. In addition, one molecule of formic acid and 553 water molecules were located during structure refinement and included in the model. Model quality was assessed with PROCHECK<sup>33</sup> and MolProbity.<sup>34</sup> All nonglycine residues reside either in the most favorable or in the allowed regions of the Ramachandran plot. Model refinement statistics are listed in Table III. The atomic coordinates and structure factors for the TVMV protease/peptide structure have been deposited in the PDB<sup>35</sup> with accession code 3MMG. All figures were generated by the graphics program PyMOL.<sup>36</sup>

### Enzyme kinetics

Enzyme assays were initiated by mixing 20  $\mu\text{L}$  of protease (50–5700 nM) in 50 mM sodium phosphate (pH 7.0), 5 mM dithiothreitol, 800 mM NaCl, 10% glycerol, with 20  $\mu\text{L}$  of substrate (0.04–1.1 mM; the actual range was selected on the basis of approximate  $K_m$  values) in the same buffer. The enzyme concentrations were determined by amino acid analysis. Measurements were performed at six different substrate concentrations. The reaction mixtures were incubated at  $30^\circ\text{C}$  for 30 min and then stopped by the addition of 160  $\mu\text{L}$  4.5M guanidine hydrochloride containing 1% trifluoroacetic acid. Aliquots were injected on to a Nova-Pak C18 reversed-phase

**Table III.** Summary of the TVMV Protease Crystallographic Data

Parameter	TVMV
Data collection	
Space group	$P2_12_12_1$
Unit cell dimensions	
$a, b, c$ ( $\text{\AA}$ )	76.5, 77.6, 78.5
$\alpha, \beta, \gamma$ ( $^\circ$ )	90, 90, 90
Protein molecules/A.U.	2
Wavelength ( $\text{\AA}$ )	0.97939
Data processing	
Resolution range ( $\text{\AA}$ ) <sup>a</sup>	35–1.7 (1.76–1.70)
No. of reflections	50064
$R_{\text{merge}}$ <sup>a</sup>	0.08 (0.431)
$I/\sigma_I$ <sup>a</sup>	17.4 (2.7)
Completeness (%) <sup>a</sup>	96.1 (92.7)
Redundancy <sup>a</sup>	5.2 (3.4)
Refinement statistics	
Resolution range ( $\text{\AA}$ )	35–1.7
$R_{\text{work}}/R_{\text{free}}$ (%)	17.5/21.0
No. atoms/molecules	
Protein	3591
Water	553
Formic acid	1
Average $B$ -factor ( $\text{\AA}^2$ )	
Protein	17.4
Water	32.3
Formic acid	36.5
R.m.s deviations	
Bond lengths ( $\text{\AA}$ )	0.01
Bond angles ( $^\circ$ )	1.33
Ramachandran plot statistics	
Residues in most favored regions	351
Residues in additional allowed regions	38
Residues in generously allowed regions	1
Residues in disallowed regions	0

A.U., asymmetric unit.

<sup>a</sup> Values in parentheses refer to the highest resolution shell.

chromatography column (3.9 mm  $\times$  150 mm; Waters Corporation, Milford, MA) 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  $k_{\text{cat}}$  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). The standard deviations for the  $k_{\text{cat}}/K_m$  values were calculated as described.<sup>37</sup> If no saturation was obtained in the studied concentration range, the  $k_{\text{cat}}/K_m$  value was determined from the linear part of the rate versus concentration profile. Because of slight variations in the activity of different protease preparations, all measurements were performed with the same batch of enzymes.

## Acknowledgments

The authors thank Scott Cherry and Kerri Penrose for technical assistance. X-ray diffraction data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at <http://www.ser-cat.org/members.html>. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. W-31-109-Eng-38. This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

## References

- Ryan MD, Flint M (1997) Virus-encoded proteinases of the picornavirus super-group. *J Gen Virol* 78 (Part 4): 699–723.
- Stanway G (1990) Structure, function and evolution of picornaviruses. *J Gen Virol* 71 (Part 11):2483–2501.
- Domier LL, Franklin KM, Shahabuddin M, Hellmann GM, Overmeyer JH, Hiremath ST, Siaw MF, Lomonosoff GP, Shaw JG, Rhoads RE (1986) The nucleotide sequence of tobacco vein mottling virus RNA. *Nucleic Acids Res* 14:5417–5430.
- Hellmann GM, Shaw JG, Rhoads RE (1988) In vitro analysis of tobacco vein mottling virus NIa cistron: evidence for a virus-encoded protease. *Virology* 163: 554–562.
- Seipelt J, Guarne A, Bergmann E, James M, Sommergruber W, Fita I, Skern T (1999) The structures of picornaviral proteinases. *Virus Res* 62:159–168.
- Hwang DC, Kim DH, Kang BH, Song BD, Choi KY (2000) Molecular cloning, expression, and purification of nuclear inclusion A protease from tobacco vein mottling virus. *Mol Cells* 10:148–155.
- Hwang DC, Kim DH, Lee JS, Kang BH, Han J, Kim W, Song BD, Choi KY (2000) Characterization of active-site residues of the NIa protease from tobacco vein mottling virus. *Mol Cells* 10:505–511.
- 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.
- Nallamsetty S, Kapust RB, Tozser 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.
- Tozser J, Tropea JE, Cherry S, Bagossi P, Copeland TD, Wlodawer A, Waugh DS (2005) Comparison of the substrate specificity of two potyvirus proteases. *FEBS J* 272:514–523.
- Donnelly MI, Zhou M, Millard CS, Clancy S, Stols L, Eschenfeldt WH, Collart FR, Joachimiak A (2006) An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. *Protein Expr Purif* 47:446–454.
- Cooper DR, Boczek T, Grelewska K, Pinkowska M, Sikorska M, Zawadzki M, Derewenda Z (2007) Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallogr D Biol Crystallogr* 63:636–645.
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921.
- Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HKIII, 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.
- Parks TD, Howard ED, Wolpert TJ, Arp DJ, Dougherty WG (1995) Expression and purification of a recombinant tobacco etch virus NIa proteinase: biochemical analyses of the full-length and a naturally occurring truncated proteinase form. *Virology* 210:194–201.
- Kapust RB, Tozser 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.
- 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.
- 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.
- Dougherty WG, Cary SM, Parks TD (1989) Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. *Virology* 171:356–364.
- Kleywegt GJ, Jones TA (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D Biol Crystallogr* 50: 178–185.
- Carrington JC, Dougherty WG (1987) Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. *J Virol* 61:2540–2548.
- Kapust RB, Tozser J, Copeland TD, Waugh DS (2002) The P1' specificity of tobacco etch virus protease. *Biochem Biophys Res Commun* 294:949–955.
- Eschenfeldt WH, Maltseva N, Stols L, Donnelly MI, Gu M, Nocek B, Tan K, Kim Y, Joachimiak A (2010) Cleavable C-terminal His-tag vectors for structure determination. *J Struct Funct Genomics* 11:31–39.
- Evdokimov AG, Tropea JE, Routzahn KM, Waugh DS (2002) Three-dimensional structure of the type III secretion chaperone SycE from *Yersinia pestis*. *Acta Crystallogr D Biol Crystallogr* 58:398–406.
- Tropea JE, Cherry S, Nallamsetty S, Bignon C, Waugh DS (2007) A generic method for the production of recombinant proteins in *Escherichia coli* using a dual hexahistidine-maltose-binding protein affinity tag. *Methods Mol Biol* 363:1–19.
- 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.
- Tropea JE, Cherry S, Waugh DS (2009) Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol Biol* 498:297–307.
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
- Nunn CM, Jeeves M, Cliff MJ, Urquhart GT, George RR, Chao LH, Tsuchia Y, Djordjevic S (2005) Crystal structure of tobacco etch virus protease shows the protein C terminus bound within the active site. *J Mol Biol* 350:145–155.

30. Vagin AA, Isupov MN (2001) Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr D Biol Crystallogr* 57:1451–1456.
31. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47 (Part 2):110–119.
32. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255.
33. Laskowski RA, Moss DS, Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231:1049–1067.
34. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB,III, Snoeyink J, Richardson JS, et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–W383.
35. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235–242.
36. DeLano WL (2002) The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific.
37. Boross P, Bagossi P, Copeland TD, Oroszlan S, Louis JM, Tozser J (1999) Effect of substrate residues on the P2' preference of retroviral proteinases. *Eur J Biochem* 264:921–929.