doi:10.1016/j.jmb.2005.04.013 J. Mol. Biol. (2005) 350, 145–155

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Crystal Structure of Tobacco Etch Virus Protease Shows the Protein C Terminus Bound within the Active Site

Christine M. Nunn, Mark Jeeves, Matthew J. Cliff, Gillian T. Urquhart Roger R. George, Luke H. Chao, Yugo Tscuchia and Snezana Djordjevic*

Department of Biochemistry and Molecular Biology University College London Gower Street, London, WC1E 6BT, UK

*Corresponding author

Tobacco etch virus (TEV) protease is a cysteine protease exhibiting stringent sequence specificity. The enzyme is widely used in biotechnology for the removal of the affinity tags from recombinant fusion proteins. Crystal structures of two TEV protease mutants as complexes with a substrate and a product peptide provided the first insight into the mechanism of substrate specificity of this enzyme. We now report a 2.7 Å crystal structure of a full-length inactive C151A mutant protein crystallised in the absence of peptide. The structure reveals the C terminus of the protease bound to the active site. In addition, we determined dissociation constants of TEV protease substrate and product peptides using isothermal titration calorimetry for various forms of this enzyme. Data suggest that TEV protease could be inhibited by the peptide product of autolysis. Separate modes of recognition for native substrates and the site of TEV protease self-cleavage are proposed.

Keywords: protease; crystal structure; dissociation constants

Introduction

Tobacco etch virus (TEV) belongs to the Potyviridae family of viruses.[1–3] As with other positive-strand RNA viruses, the TEV genome is translated into a single large polyprotein that is subsequently processed by virally encoded proteins with proteolytic activities. The majority of cleaving events in TEV are carried out by the nuclear inclusion *a* (NIa) protease. The 27 kDa NIa protease is functionally and structurally homologous to picornaviral 3C proteases.[4,5] A distinctive feature of these enzymes is that their biological activity requires very stringent sequence specificities that would ensure appropriate viral maturation without damage to the host cell's protein. High specificity and good activity under a range of conditions renders TEV NIa protease a popular choice of enzyme for cleaving genetically engineered fusion proteins.[6–8]

Within the viral polyprotein, or in artificial substrates, the TEV NIa protease (or simply TEV protease) recognises an extended seven amino acid residue long consensus sequence E-X-X-Y-X-Q-S/G (where X is any residue) that is present at the protein junctions.[9,10] The sequence is cleaved after the glutamine residue and as a result the serine or glycine residue becomes the new N terminus. Residues at the P6, P3, P1 and P1′ positions are essential for optimal cleavage; however, the rate of the cleavage is effected by the type of amino acids occupying other positions.[11,12] Crystal structures of the catalytically inactive and catalytically active TEV protease in complex with a peptide substrate and product, respectively, provided the first insight into the mechanism of substrate specificity of this enzyme, at atomic resolution.[5] Previously, although the structures of the nascent 3C proteases from hepatitis A virus,[13] rhinovirus-14[14] and poliovirus[15,16] were available, no crystal structure of a 3C-type protease had been determined in the presence of the peptide. The TEV protease crystal structures confirmed that, as expected, this enzyme adopts the characteristic two-domain antiparallel β-barrel fold that typifies the trypsin-like serine proteases with the catalytic triad residues His46, Asp81 and

Present address: C. M. Nunn, Department of Health and Human Sciences, London Metropolitan University, 166-220 Holloway Road, London, N7 8DB, UK.

Abbreviations used: TEV, tobacco etch virus; NIa, nuclear inclusion *a* protease; NIb, nuclear inclusion protein b; eIF2α, α-subunit of the eukaryotic translation initiation factor 2; ITC, isothermal titration calorimetry.

E-mail address of the corresponding author: snezana@biochem.ucl.ac.uk

Cys151 located at the interface between the domains. Further, structural descriptions of the S6-S1 and S1′ TEV substrate-binding pockets were obtained.

It has been shown, however, that TEV protease undergoes self-cleavage generating a truncated protein with reportedly diminished activity.[17] The cleavage occurs near the C terminus of TEV protease after residue 218, at a site that does not follow the canonical sequence. Using site-directed mutagenesis, efforts were made to produce a non-cleavable form of the TEV protease that would retain the wild-type enzymatic activity.[18,19] These studies were also trying to understand how and why an enzyme that otherwise is so highly specific undergoes this self-cleavage. The previously reported crystallographic investigation of both the catalytically active and catalytically inactive TEV protease structures revealed little information about the C-terminal region of TEV protease.[5] In that report, the protein product that was used for crystallisation ended at residue 236, and in three of the four protomers observed in the two crystal structures, residues 222–236 were not located within the electron density maps. In one of the protomers the seven residue long C-terminal fragment (residues 230–236) was located in a position that appeared to be a crystallisation artefact.

We have crystallised the C151A inactive mutant of TEV protease containing all 242 residues of the mature enzyme. The structure reveals a potential mechanism of auto-inhibition for the TEV protease in which the C-terminal tail of the protein binds within the active site region. Based on the crystal structure and the available biochemical data we are able to postulate different substrate recognition modes for the proteolysis and autolysis events and to address the previously observed changes in the apparent activity of the truncated protease form.

Results and Discussion

Protein purification and crystallisation

The protein used for crystallisation spans the full length of 27 kDa NIa protease corresponding to the enzyme obtained upon processing of the viral polyprotein (Figure 1). This includes residues 237–242 that were not present in the previously crystallised domain and that correspond to a site of an internal viral polyprotein cleavage. During the purification of the recombinant protein from *Escherichia coli*, we noticed that the C151A mutation changed the properties of the protein compared to the wild-type. During the gel filtration step, the protein appeared not to be completely monomeric as indicated by a much broader and less symmetric elution peak. Furthermore, upon concentrating the purified protein sample to above 4 mg/ml, the bands on SDS/polyacrylamide gels appeared smeary even in the presence of high concentrations of 2-mercaptoethanol, also suggesting some aggregation of the protein. Despite possible aggregation we subjected the protein to standard crystallisation screens and several conditions yielded crystals. The protein crystallised in space group *C*2 unlike the C151A–substrate and S219D–product complexes previously studied, which crystallised in space groups $P4_22_12$ and $P4_32_12$, respectively. The Matthews coefficient of 2.68 Å^3/Da is similar to that observed previously for the S219D mutant protein which had a Matthews' coefficient of 2.25 Å^3/Da, whereas the crystals of the C151A–substrate complex exhibited a higher solvent content of 71% and corresponding Matthews coefficient of 4.35 Å^3/Da. The crystals of full-length C151A diffracted weakly and although the data were collected to 2.5 Å resolution we were able to process and refine the structure only to 2.7 Å resolution (Table 1).

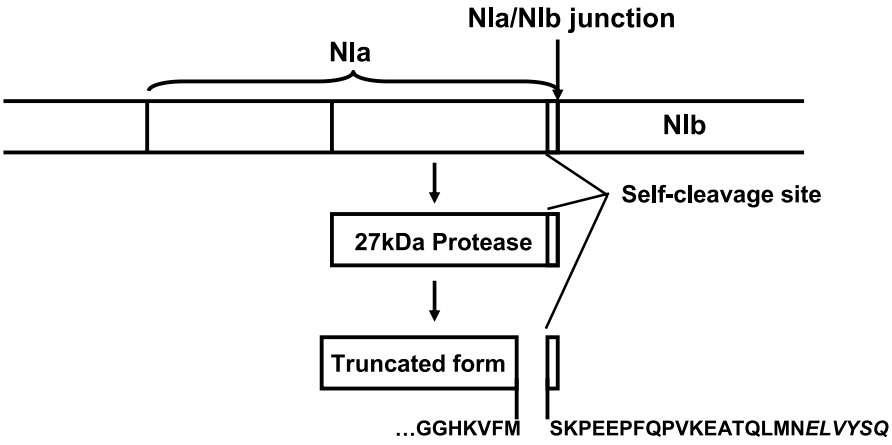

Figure 1. Schematic diagram of TEV polyprotein processing resulting in release of the 27 kDa NIa. The diagram also shows the site of the self-cleavage. The last six residues in the sequence of the 27 kDa protease are shown in italics to emphasise that these six residues, present in the current structure, correspond to the internal proteolytic site.

Table 1. Diffraction data and final refinement statistics

Space group	*C*2
Cell dimensions (Å,°)	146.587×41.544×96.063, β=108.95
Wavelength (Å)	0.933
Temperature (K)	110
Resolution (Å)	91–2.7
Total number of reflections collected	128,457
Unique reflections	13,436
Completeness (%)[a]	98.7 (93.5)
$I/\sigma(I)$	17.2 (3.5)
R_{merge}[b]	0.075 (0.32)
No. of protein atoms	Protein A (residues 4–222 and 235–242) 1822 Protein B (residues 4–221 and 235–242) 1813
No. of beta-mercaptoethanol molecules	2
Number of solvent water molecules	37
Resolution used in refinement	500–2.7
RF-function—correlation coefficient for cross-rotation solution	0.174
Weight for NCS restraints	300
No. of reflections: working set/test set	12,072/1364
R_{cryst}	0.240
R_{free}	0.302
RMS deviation bonds (Å)	0.0115
RMS deviation angles (°)	1.72
Ramachandran plot	
Most favoured: additionally allowed: generously allowed (%)	77.4: 20.5: 2.1

[a] Numbers in parentheses are for the last shell.
[b] $R_{merge} = \left[\sum_{hkl}\sum_i |I_{hkl} - \langle I_{hkl}\rangle| / \sum_{hkl}\sum_i |I|\right]$.

Description of the structure and comparison to other structures of TEV protease

TEV protease adopts the characteristic two-domain antiparallel β-barrel fold that is the hallmark of trypsin-like serine proteases[20,5] (Figure 2). The structure comprises two protein molecules within the crystallographic repeat unit, which are related by non-crystallographic 2-fold symmetry. In the *C*2 space group there are two possible non-crystallographic dimer relationships. The one, for which the coordinates have been deposited, has a small, buried protein surface area at the interface between protomer A and B, measuring 688 Å^2. The second non-crystallographic 2-fold symmetry axis generates a dimer that is very similar to that in the previously described S219D mutant structure of TEV.[5] In this dimer the contacts between two protomers are extensive with N-terminal residues 4–10 from one subunit crossing over the hairpin 110–127 of another subunit and forming main-chain antiparallel β-strand interactions as well as side-chain interactions. The root-mean-square (RMS) deviation for all atoms between protomer A and protomer B is 1.79 Å. The main difference between the two protomers occurs within the β-hairpin formed by residues 114–122. This was similarly observed in the two previously determined structures of TEV protease, and was attributed to both the effects of crystal packing and conformational flexibility.[5]

The structure reported here and that of the S219D/product complex differ in both their crystallographic space groups (*C*2 and $P4_32_12$, respectively) and crystallographic packing of their protein molecules, whilst exhibiting the same type of intra-dimer interactions. In the structure of the C151A–substrate complex the two molecules forming a dimer were oriented very differently, with the two crystallographic symmetry-related molecules

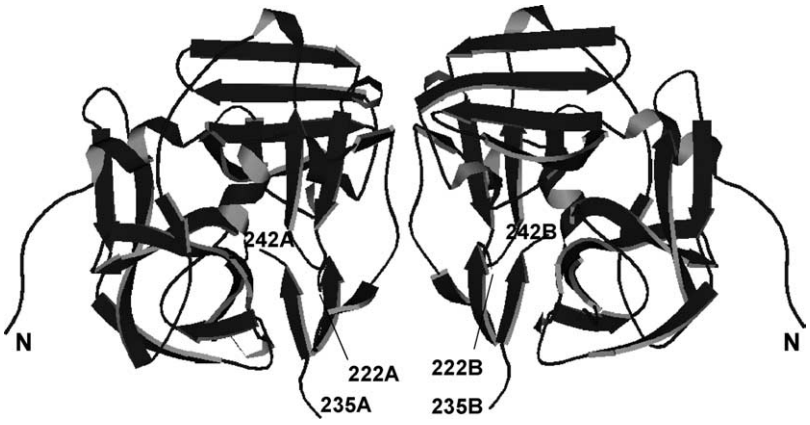

Figure 2. Ribbon diagram of the TEV protease dimer found in the asymmetric unit of the *C*2 crystal form described here. The molecules are oriented such that the peptide-binding site is positioned at the bottom of the Figure. The C-terminal residues 235–242 forming a β-strand are shown binding to the active site.

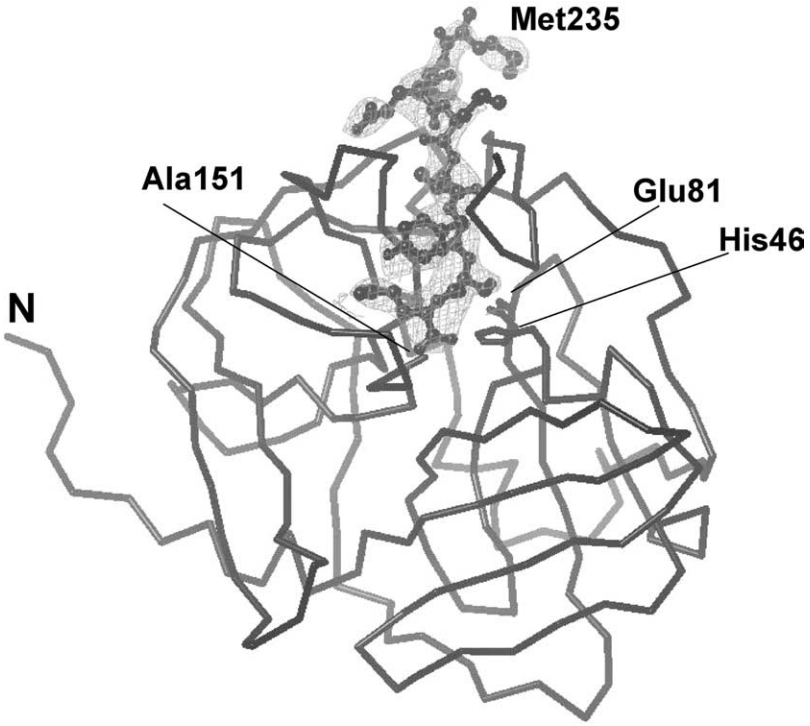

Figure 3. Electron density corresponding to the C-terminal residues 235–242. A ball-and-stick model of the C-terminal residues is overlaid on difference electron density contoured at 1.2σ that was unaccounted for after the molecular replacement solution was obtained. The active site residues are shown as a stick-only model and labelled. For clarity, only a C^{α} model of the TEV residues 4–222 is shown.

linked through a disulphide bridge formed by oxidation of the surface exposed residue Cys130. The equivalent cysteine residue in the current structure was bound to a molecule of 2-mercaptoethanol in both protomers. Overall the three structures of TEV protease are very similar; the positional root-mean-square deviations of the superimposed α-carbon atoms between the title TEV protease structure and that of the S219D–product and C151A–substrate complexes are 0.55 Å and 0.73 Å, respectively.

We were unable to locate all the residues from the amino acid sequence of the protein used for the crystallisation within the electron density maps. Those residues between Glu222 and Met235 in monomer A and Pro221 and Met235 in monomer B were not visible and were not included in the structure refinement. This region was also not observed in the previously determined C151A–substrate and S219D–product structures, further confirming that this region exhibits intrinsic flexibility and displays molecular disorder irrespective of the crystallographic packing constraints. This C-terminal region of the TEV protease follows immediately after the site of the autolysis and has no counterpart in other trypsin-like proteases.

Table 2. H-bonding interactions between the C-terminal tail (residues 235–242) and TEV protein (residues 4–221)

C-terminal tail	TEV Protein atom	Distance (Å)
Gln242 OE1	His167 NE2	2.7
	Thr146 OG1	2.6
Gln242 NE2	Asp148 OD1	3.1
Gln242 OXT	His46 NE2	3.0
Gln242 OT1	Gly149 N	2.7
Gln242 OT1	Glu150 N	3.2
Ser241 OG	His46 ND1	3.1
Tyr240 OH	Asp148 OD1	2.7
	Asn174 ND2	2.8
Glu237 OE1	Asn176 ND2	2.9
	Tyr178 OH	2.7

The C terminus binds to the active site

The previously crystallised protein product spanned residues 1–236 of the mature NIa protease. Here we have crystallised the protein that contained an additional six residues at the C terminus up to the NIa/NIb cleavage point of the viral polypeptide. Following structure solution using only the enzyme portion of C151A–substrate complex as a search model, additional electron density was clearly identified within the active site (Figure 3). This density was easily interpreted as that of the C-terminal residues 235–242 bound tightly within the active site in a similar fashion to that in which the peptide product bound to the S219D–protein complex. In addition to hydrogen bonds formed by the side-chains of the C terminus (Table 2), the main-chain atoms are also involved in binding to the substrate-binding site (Figure 4(a)). Furthermore, the C-terminal carboxylic group itself is tightly bound to the active site. One of the terminal oxygen atoms is hydrogen-bonded to the main-chain amide nitrogen atoms of residues 149–151 identifying this region as an oxyanion hole involved in stabilising the negative charge generated on the carbonyl oxygen atom of the putative tetrahedral intermediate during the peptide bond hydrolysis (Figure 4(b)). This is a somewhat

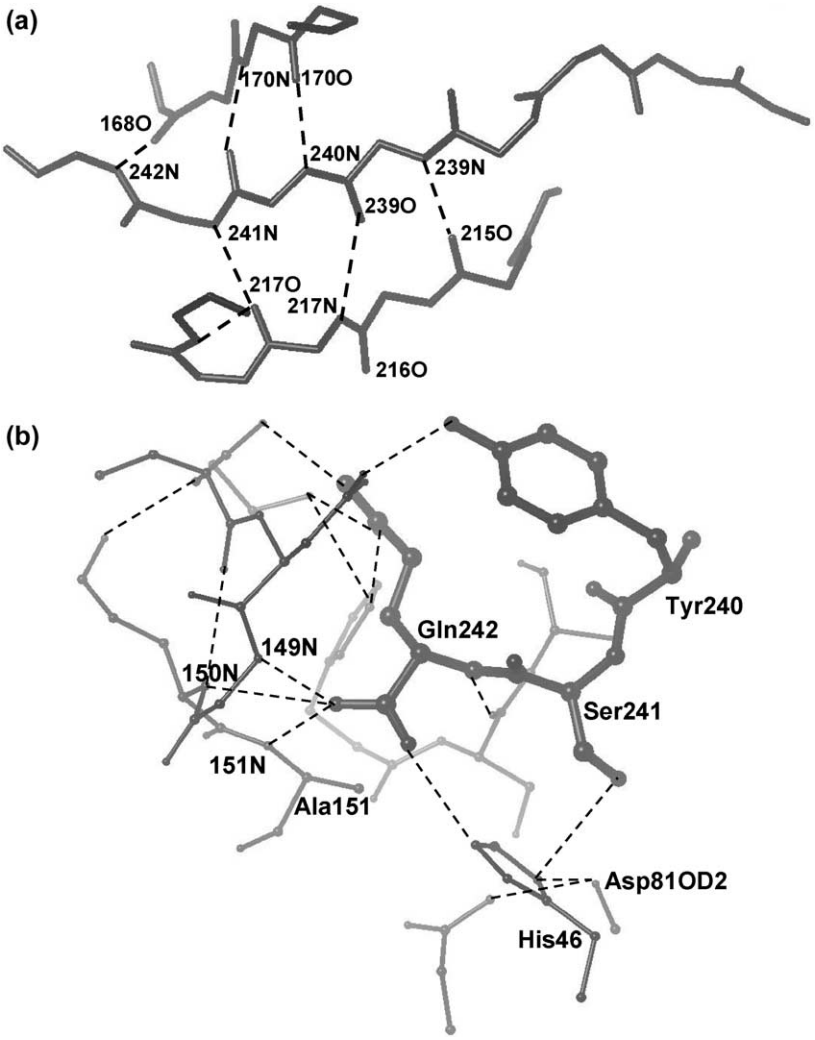

Figure 4. Hydrogen bonds involved in binding and recognition of the C terminus to the active site. (a) Main-chain hydrogen bonds involved in binding of the C terminus. Only the main-chain atoms are shown for the two strands forming the peptide-binding groove and for the C-terminal residues bound at the site. The β-type main-chain hydrogen bonds are denoted explicitly by broken lines and the atoms involved in forming these bonds are labelled. (b) The complex network of hydrogen bonds involved in binding and recognition of the C-terminal residues at the protease active site. The potential hydrogen bonds are shown as broken lines. The Figure emphasises the specific interactions of the Gln242 side-chain, the positioning of the C-terminal carboxylate and the nitrogen residues that were postulated to be involved in the formation of an oxyanion hole. The C-terminal residues are represented in a thicker ball-and stick model relative to the rest of the structure. For clarity, only selected residues and atoms of the active site are shown.

unusual observation. According to the model of peptide hydrolysis, peptide substrate and product should not be in a conformation agreeable with binding to the oxyanion hole, as stabilisation of the tetrahedral intermediate exclusively, through interaction with the oxyanion hole, is a significant component of the catalytic mechanism.[21] The second terminal oxygen atom is hydrogen-bonded to the NE2 atom of the catalytic residue His46. The ND1 atom of the imidazole ring is hydrogen-bonded to both the carboxylate oxygen atom of Asp81 and the hydroxyl oxygen atom of the Ser241 residue. This network of hydrogen bonds implies that the C terminus of the protease binds to the active site with high specificity. While the structure shown here is that of the active site mutant C151A there is no structural reason to suppose that these interactions could not exist in the presence of the cysteine side-chain. Unfortunately, the previously reported structure of the S219D mutant with the peptide product does not give the full model for its terminal carboxylate group and comparison between the two structures cannot be made with confidence.

As residues 222–235 were not identified in the electron density maps there remains the possibility that swapping of the C termini forms the dimer. Within the single globular unit, the distance between the α carbon of the residue 222 and the α carbon of the residue 235 is around 20 Å, sufficient to accommodate 12 residues. However, a similar distance separates residue 222 to residue 235 of the symmetry-related molecule. This molecular arrangement observed in our structure further confirms that release of the protease from the polyprotein could be carried out in an intramolecular as well as in an intermolecular fashion.

Implications for the substrate recognition and the catalytic mechanism

The TEV protease recognition site comprises the seven amino acid consensus sequence Glu-X-X-Tyr-X-Gln/Ser. The importance of the specific residues at positions P6, P3, P1 and P1′ in creating an optimal substrate cleavage site was characterised through biochemical and mutational analysis.[9,11,22] These residues are engaged in interactions with the TEV

protease, primarily *via* hydrogen bonding interactions (Table 2). Within the active site for the three TEV protease structures solved to date the conformation of the bound peptide is very similar, with the positions of amino acids P1, P3 and P6 almost superimposed. The only exception is the Tyr residue at position P3, the side-chain of which can acquire two alternative conformations. These different conformations were seen in the C151A–substrate and S219D–product structures and one of these conformations was observed in the title structure. The peptides used in the previous studies contained the following sequences: **E**NL**Y**F**QG** and TT**E**NL**Y**F**QS**GT (where the residues underlined are those for which no electron density was identified and the residues in bold are conserved amino acids). The C-terminal sequence bound to the active site of the structure reported here is MN**E**LV**Y**S**Q**, thus differing from the sequences described above at positions P2, P4 and P5. The most striking difference is at position P2, where a Phe in the two peptides is substituted with a Ser within the C terminus. At the P5 position Asn in the two peptide structures is substituted by a Leu in the C terminus. While the two sites are of lesser importance for the substrate specificity, the polarity of the two residues within the C-terminal end is the inverse of those for the equivalent residues in the two peptides and this might have an impact on the strength of the interaction with the binding site. Specifically, a Ser in the P2 position is located within the S2 binding pocket, which is lined by hydrophobic residues Val209, Trp211, Val216, Met218, the only exception being His46, which in this structure is oriented such that it forms a hydrogen bond to the Ser side-chain. Trp211 moves into the cavity towards the space previously occupied by the Phe side-chain present in the S219D–product and C151A–substrate structures. Furthermore, as there is no S5 pocket in TEV protease, the side-chain of substrate residue P5 protrudes away from the protein and is exposed to solvent. When the C-terminal end is bound to the substrate-binding site the corresponding P5 residue is Leu and exposure of this hydrophobic residue to solvent would not be favourable thermodynamically. The implication is that the strength of the interactions of the P1, P3 and P6 sites could be overriding the unfavourable interactions of the C-terminal end. Additionally Val at position P4 replaced a Leu residue in the peptides studied here; however, we could not observe any structural implications for this substitution.

Based on the structure we hypothesise that upon self-cleavage, the C-terminal peptide comprising residues 219–242 associates with the truncated enzyme and remains bound acting as an auto-inhibitor. In the presence of the appropriate substrate, however, it is expected that the C terminus will be displaced from the active site.

Determination of dissociation constants

Isothermal calorimetry measurements were carried out in order to determine apparent dissociation constants for the peptides corresponding to the substrate and the product of the proteolytic reaction. We carried out the measurements on the wild-type enzyme, C151A mutant protein, and on the C-terminally truncated version of the protease (Table 3; Figure 5). The dissociation constant of the substrate peptide was determined only for the active site mutant form of the enzyme, as proteolysis by the catalytically active forms would have interfered with the measurements. Although, no direct comparison could be made, the K_D value of 47 μM for the C151A mutant–substrate pair is within the range of values that could be anticipated for the affinity of the enzyme towards the substrate based on reported K_m values for the wild-type enzyme of ∼60 μM.[17,18] Both catalytically active and inactive enzymes bind the product of the reaction with similar affinity of 165 μM and 187 μM, respectively, three times lower affinity than that of substrate. However, the C-terminally truncated form of the enzyme binds the peptide product with affinity of 90 μM, several-fold higher

Table 3. ITC-derived dissociation constants for substrate and product peptides

Enzyme	K_D product (mM)	K_D substrate (mM)	K_m (mM) Kapust *et al.*[18]
Wild-type TEV	0.165 ± 0.019	n/a	0.061 ± 0.010
Truncated TEV	0.090 ± 0.018	n/a	0.448 ± 0.049
C151A	0.187 ± 0.019	0.047 ± 0.024	n/a

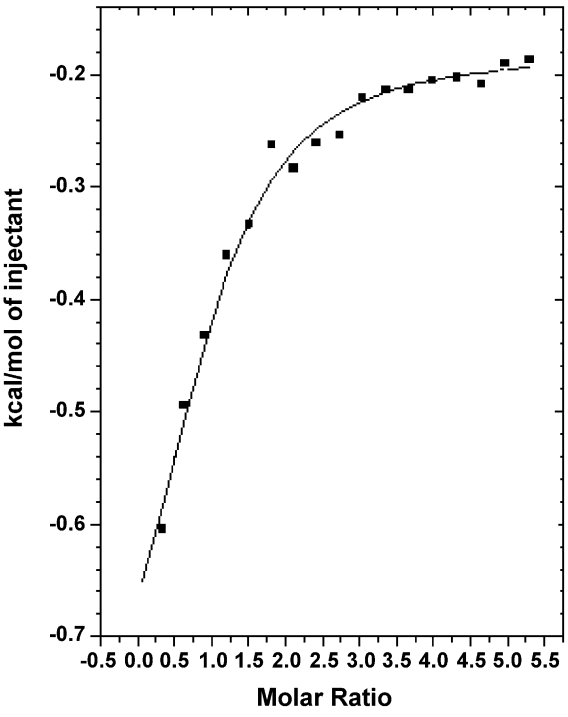

Figure 5. Isothermal titration calorimetry plot for the proteolytic product peptide interaction with the wild-type TEV protease.

than the anticipated affinity for the substrate considering the previously determined 448 μM K_m of substrate peptide for this enzyme variant.[18] Although we were unable to determine a dissociation constant for the truncated form-substrate peptide pair, a good correlation between our ITC data and previously reported K_m values for the substrate-wild-type enzyme would suggest that indeed the truncated form of the protein is inhibited by the proteolytic product.

The mechanisms of self-cleavage site recognition and substrate recognition are separated

Self-cleavage of TEV protease occurs between residues 218 and 219 producing a truncated enzyme with diminished activity.[17,18] Interestingly, the amino acid sequence at this site does not resemble the consensus cleavage sequence. When compared to the consensus recognition sequence (E-X-X-Y-X-Q-S/G), among the residues surrounding the site of autolysis, it is only the Ser residue at the P_a1' site that agrees with the consensus sequence. (Here, subscript a denotes the P1 position of the autolysis site.) In addition, the Phe residue at P_a2 corresponds to Phe at P2 of many substrates; however, this site tolerates substitutions and is not required for the recognition. Furthermore, extensive mutagenesis studies have shown that even at the Ser $P1'$ site a number of substitutions can be made without having a significant effect on the autolytic or proteolytic activity. Few substitutions (Val, Asp, Pro) resulted in diminished autolysis without affecting the main enzymatic activity of TEV protease. We have used site-directed mutagenesis to study the importance of residues at the P_a3 position surrounding the autolytic site. When the Val residue at this position was mutated to Asp we found that while autolysis still occurred, the enzyme lost apparent activity towards the substrate fusion protein under conditions that are used to carry out the reaction with the wild-type enzyme (Figure 6). Even when the mutant protein was incubated with the substrate for 24 hours there was no detectable cleavage of substrate as judged by SDS/PAGE. The crystal structures clearly show that this region of the protein is involved in substrate binding. We would like to point out that residues 216–218 are not involved in the formation of a substrate binding site through their side-chains, but that these three residues contribute towards the formation of a distinct, relatively long and narrow groove that provides the space for substrate binding. Binding of the recognition sequence to the enzyme is, in part, stabilised through a set of β-type main-chain hydrogen bonds (Figure 4(a)). Since Val216 does not participate in substrate recognition through its side-chain, the change to Asp must affect the conformation of that region and interfere with consensus substrate binding while the catalytic triad remains active, as is evident from the fact that the enzyme is able to perform a self-cleavage

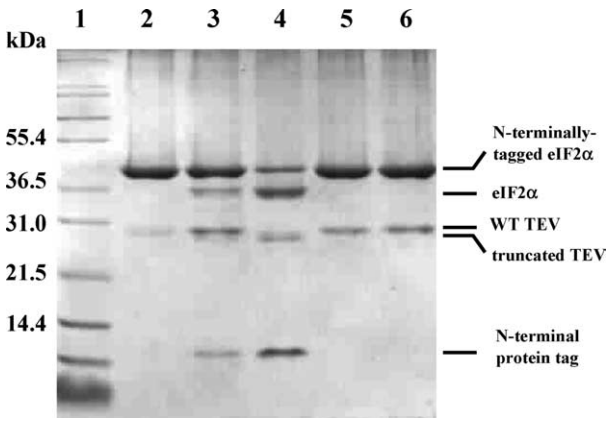

Figure 6. SDS/PAGE showing the proteolytic activities of the different forms of TEV protease. Equal amounts (7 μg) of the partially purified fusion protein substrate (eIF2α containing the 5 kDa N-terminal affinity tag) were digested with 0.7 μg of a protease. Subsequently the reaction mixtures were loaded on the SDS/12.5% polyacrylamide gel. The reaction mixtures contained: no enzyme, lane 2; wild-type TEV, lane 3; truncated TEV, lane 4; V216D mutant, lane 5; C151A mutant, lane 6. Lane 1 contains molecular mass markers. In lane 2 a faint band of approximately 30 kDa comes from a purification contaminant.

reaction. A further implication of this mutation is that differential mechanisms are employed for consensus substrate recognition and for autolysis site recognition. Thus, it is not surprising that the two sites have very little in common. During the autolysis, which occurs intramolecularly,[18] residues 216–218, must move closer to the active site residues and at that point the intramolecular substrate recognition and binding is probably carried out only through side-chains of the residues near to the site of cleavage, i.e. Met and Ser. In contrast, binding of the seven residues long consensus substrate sequence relies on the formation of a narrow groove, one wall of which comprises residues 216–219. The specifically shaped binding groove allows for the appropriate positioning of side-chains of the consensus residues P6, P3 and P1 to their binding sites S6, S3 and S1,[5] in addition to stabilising the substrate binding through main-chain hydrogen bonds.

It was reported that the F217K mutation also produces an enzyme with greatly diminished catalytic activity.[18] Additionally, in this case, the self-cleavage was greatly reduced compared to the wild-type enzyme. In concert with the specific involvement of residues 216–218 in the formation of the overall shape of the substrate-binding groove, the F217K mutation resulted in an increased K_m value for the canonical substrate and had no effect on the k_{cat} value of the enzyme. However, with respect to the self-cleavage site, Phe217 corresponds to the P2 position, which does not play a significant role in recognition of the canonical substrate and thus the observed reduction of self-cleavage for that

particular mutation again supports the hypothesis that the two substrates of hydrolysis reaction are recognised by separate molecular mechanisms.

Activity of a truncated form of protease

Reports by Parks *et al.*[17] and Kapust *et al.*[18] showed that the truncated form of TEV protease resulting from self-cleavage has significantly reduced activity. While Parks *et al.*[17] showed that the effect of truncation is manifested in both the K_m and the k_{cat} values of reaction, Kapust *et al.*[18] showed no effect on k_{cat} and obtained a sevenfold increased K_m value compared to the full-length enzyme. Since the residues preceding the self-cleavage site constitute part of the substrate-binding site, one can speculate that some structural perturbations upon C-terminal removal may influence substrate binding, although the crystal structures provide no clear explanation for this phenomenon. Furthermore, there is no convincing structural interpretation of how the k_{cat} value would be affected by the C-terminal truncation. The two reports, discussed above,[17,18] tested the activity of the truncated enzyme against peptide substrates, and hence, we decided to generate a truncated enzyme and test its activity against fusion proteins containing the specific TEV cleavage site.

When the truncated enzyme was tested under buffer conditions and enzyme/substrate ratios commonly used in overnight reactions with wild-type protease, we observed no significant difference in activity between the full-length and the truncated enzyme (Figure 6). In some instances, and depending on the fusion protein substrate, efficiency of overnight cleavage appeared somewhat higher with truncated enzyme than with wild-type protease. We attempted to determine kinetic parameters for the truncated enzyme with respect to the fusion proteins by carrying out proteolytic reactions with fixed concentrations of enzyme while varying the concentrations of substrate protein. From the preliminary experiments it was apparent that we could not prepare sufficiently high concentrations of substrate in order to saturate the enzyme and that the K_m value for the particular substrate was at least 500 μM (data not shown). This value is comparable to previously reported K_m values for the peptide substrate of 448 μM.[18] The same experiment resulted in a K_m value of the wild-type enzyme of > 200 μM which is higher than that reported for the peptide (61 μM) and the dissociation constant obtained in our ITC experiments (47 μM). This discrepancy may be due to poor reliability of our kinetic experiments or an effect of the target protein on the binding affinity for the canonical sequence within the context of the fusion protein. It is worth noting that due to the purification protocols and in order to avoid protein precipitation, fusion proteins were usually subjected to TEV proteolysis at concentrations of 20–100 μM and thus the reactions were carried out under less than optimal conditions.

Conclusions

We have crystallised and solved the structure of the active-site mutant of the full-length TEV protease including all residues up to the naturally occurring NIa/NIb cleavage junction in the TEV polyprotein. In this structure, the C-terminal end of the protein was found bound to the active site in a fashion similar to that previously observed for the peptide substrate and peptide product. A distinct network of hydrogen bonds, involved in stabilising the carboxyl end, suggested that the product of the proteolysis could act as an inhibitor. ITC measurements demonstrated that the inhibitory effect of the peptide product is particularly significant in the C-terminally truncated form of the enzyme. The auto-inhibitory properties of the C-terminal end might be somehow involved in regulation of the proteolytic activity during the viral maturation. A recombinant protease lacking the last ~20 residues is much more soluble than the wild-type enzyme, probably because the C-terminal region is nearly completely disordered, and can remove an affinity tag from the fusion proteins although with greatly reduced affinity for the substrate.

Materials and Methods

Protein expression and purification

The TEV(C151A) and TEV(V216D) protease expression vectors were created by generating single-site mutations in the original plasmid pRET3aTEV that contained a coding sequence for the full-length TEV NIa (residues 1–242 of the naturally occurring mature 27 kDa TEV NIa) with a six histidine affinity tag at the N terminus. This plasmid was kindly given to us by Dr Daniela Rhodes, LMB-MRC, Cambridge UK.[23] The mutations were carried out with the PCR-based QuikChange site-directed mutagenesis kit (Stratagene). Oligonucleotide primers that were used to introduce the specific changes in coding sequences were synthesised by Amersham Pharmacia Biotech. The sequences of the new expression vectors were confirmed by sequencing (OSWEL Sequencing). Wild-type and mutant proteins were commonly expressed in C41(DE3) cells[24] and grown in Luria broth supplemented with 100 μg/ml ampicillin. The cells were grown at 37 °C until absorbance at 600 nm reached 0.7. At that point the growth temperature was reduced to 22 °C and the expression was induced by the addition of 0.5–1.0 mM isopropyl-β-D-thiogalactopyranoside. The cells were then incubated for a further three hours at 22 °C before being harvested by centrifugation. Wild-type, V216D and C151A proteases were purified by Ni^{2+} affinity chromatography. The C151A protease variant was additionally purified by gel filtration on an S100-HR column (Pharmacia). The final purity of the protein sample was evaluated by SDS/PAGE.

A truncated version of the enzyme was created by changing the coding sequence for residue Ser219 to the stop codon TGA, thus obtaining the gene coding for the protein corresponding to the product of the self-cleavage reaction. Compared to the wild-type protein much higher yields of the truncated enzyme were obtained in the soluble fraction of cell extracts and 40 mg of purified

protein per litre of cell culture were produced routinely. In general, the purification procedure was similar to that used for the full-length enzyme except that to obtain maximum solubility the buffers had to be kept at pH 7.5.

Activity assay

Protease activity was assessed by incubating a protein fusion substrate with TEV protease at room temperature for 12 hours. Typically the reaction mixture was prepared in 50 mM Tris–HCl (pH 8.0) (or 50 mM Hepes (pH 7.5)), 500 mM KCl, 10% (v/v) glycerol, 10 mM 2-mercaptoethanol with a final volume of 15–20 μl. The ratio of the protease and the substrate in the reaction was 1 μg of TEV: 10 μg of substrate protein. The substrate was recombinant yeast eIF2α or human PP2A methyl esterase (PME-1) containing an N-terminal affinity tag of 5 kDa that was linked through the TEV recognition sequence E-N-L-Y-F-Q-G. Products of the reactions were analysed by SDS/PAGE on 12.5% (w/v) polyacrylamide gels.

Kinetic measurements

Wild-type TEV, truncated TEV enzyme and tagged-PME-1 were purified in parallel and stored at −20 °C in aliquots. The protein concentration was estimated using Bio-Rad protein assay. All proteins were stored in buffer containing 50 mM Hepes (pH 7.5), 250 mM NaCl, 20% glycerol and 10 mM 2-mercaptoethanol. The reactions were initiated by mixing 10 μl of protease solution (0.6 or 1.2 μM depending on the enzyme) with 10 μl of 6.6–660 μM fusion protein substrate (affinity tagged PME-1). The reaction tubes were incubated in a 30 °C water bath for 30 minutes. Reactions were stopped by adding 5 μl of Laemmli buffer.[25] Individual reactions were then analysed by SDS/PAGE on 12.5% polyacrylamide gels. Coomassie blue-stained gels were scanned and analysed by the GeneScan system from SYNGENE and individual bands were quantified using GeneTools software. The reactions were run in duplicate.

Isothermal titration calorimetry (ITC) measurements

The peptides used in the experiments were purchased from Peptide Protein Research UK. The sequences of the peptides were as follows: PTTENLYFQSGTVDRR, substrate; and PTTENLYFQ, product. ITC experiments were conducted on a VP ITC (Microcal Inc., Northampton, MA) as described.[26,27] Experiments were performed at 30 °C in a buffer containing 50 mM Hepes (pH 7.5), 250 mM NaCl, 10% glycerol and 5 mM DTT. The data were analysed using the ORIGIN software supplied with the calorimeter. In a typical titration 1.8 ml of 35–120 μM enzyme variant was placed in the calorimeter cell, and 1.5 mM peptide was injected in 4×8 μl injections followed by five injections of 12 μl and finally ten injections of 18 μl. The reported data for K_D are based on two repetitions of the titrations.

Crystallisation and data collection

The C151A TEV protein was screened for crystallisation using the range of crystallisation conditions from Hampton Research Screen II and Molecular Dimension Structure Screen I.[28,29] Crystallisation was carried out at 20 °C by vapour diffusion using the hanging drop method. The drops were made by combining 2 μl of the enzyme and 2 μl of each of the crystallisation solutions. For the crystallisation experiments the protein was concentrated to 5 mg/ml in a buffer solution containing 50 mM Tris–HCl (pH 8.0), 500 mM KCl, 10% glycerol and 20 mM 2-mercaptoethanol. Within a week several conditions from both screens yielded diamond-shaped crystals with the longest dimension of 0.2–0.3 mm. The crystal that was used for data collection grew in 0.2 M lithium sulphate, 0.1 M Tris–HCl (pH 8.5), 30% (w/v) polyethylene glycol 4000. Crystals belong to space group $C2$ with cell dimensions $a=146.59$ Å, $b=41.54$ Å, $c=96.06$ Å, $\beta=108.95°$. The Matthews coefficient[30] is 2.68 Å^3/Da, corresponding to a solvent content within the unit cell of 53.8%. Diffraction data were collected at 100 K using a single crystal on beamline ID14-2 at the European Synchrotron Radiation Facilities in Grenoble, France. The data set consisted of 205 frames, each with an oscillation of 1° and exposure time of 10 s/frame. Data were processed using the HKL2000[31] program suite.

Structure solution and refinement

Structure solution was carried out by molecular replacement using, as the search model, coordinates for a monomer of the C151A mutant structure (pdb code 1LVB), excluding the substrate peptide. Molecular replacement was carried out using the programs in CNS.[32] After the initial molecular replacement solution was obtained for the first monomer of the asymmetric unit, a translation search was performed with the second highest peak in the rotation function and the complete asymmetric unit was generated. The protein dimer was subjected to rigid body refinement to give an R-factor of 0.49 and R_{free} of 0.49. Manual fitting of the structure was then carried out with the graphics program TURBO FRODO[33] until all of the electron density for the two protein monomers was assigned. For each monomer the three N-terminal residues of the TEV protease were not visible within electron density maps, furthermore residues 223–234 of chain A, and residues 222–234 of chain B were also not located. This indicates conformational flexibility for these regions of the protein structure. In addition, no affinity tag (Gly-His$_6$) was located within the structure.

Structure refinement was carried out using the program CNS[32] with 2-fold non-crystallographic symmetry restraints for all amino acid residues except 28, 113, 172, 207, 222 and 114–122 inclusive. The structure was subjected to torsion angle simulated annealing, positional refinement and group temperature factor refinement against a maximum likelihood target. In addition to the two protomers, two molecules of 2-mercaptoethanol and 37 water molecules were located during structure refinement and included in the model. Diffraction data and the final refinement statistics are shown in Table 1.

Protein Data Bank accession codes

The atomic coordinates and structure factors (code 1Q31) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University Piscataway NJ†.

† http://www.rcsb.org/

Acknowledgements

This work was supported by HEFCE UK. We thank Dr. Ajit Basak for his help during data collection, Professor John Ladbury for the use of isothermal titration calorimetry equipment and Dr Mark Williams for constructive criticism.

References

1. Dougherty, W. G. (1983). Analysis of viral RNA isolated from tobacco leaf tissue infected with tobacco etch virus. *Virology*, **131**, 473–481.
2. Allison, R. F., Johnston, R. E. & Dougherty, W. G. (1986). The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for the synthesis of a single polyprotein. *Virology*, **154**, 9–20.
3. Ryan, M. D. & Flint, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *J. Gen. Virol.* **78**, 699–723.
4. Seipelt, J., Guarne, A., Bergmann, E., James, M., Sommergruber, W., Fita, I. & Skern, T. (1999). The structure of picornaviral proteinases. *Virus Res.* **62**, 159–168.
5. Phan, J., Zdanov, A., Evdokimov, A. G., Tropea, J. E., Peters, H. K., III, Kapust, R. B. *et al*. (2002). Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* **277**, 50564–50572.
6. Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. & Dougherty, W. G. (1994). Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal. Biochem.* **216**, 413–417.
7. Lucast, L. J., Baley, R. T. & Doudna, J. A. (2001). Large-scale purification of a stable form of recombinant tobacco etch virus protease. *BioTechniques*, **30**, 544–554.
8. Polayes, D. A., Ward, G. & Hughes, A. J, Jr (1994). TEV protease, recombinant: A site-specific protease for efficient cleavage of affinity tags from expressed proteins. *Focus*, **16**, 2–5.
9. Carrington, J. C. & Dougherty, W. G. (1988). A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proc. Natl Acad. Sci. USA*, **85**, 3391–3395.
10. Dougherty, W. G. & Parks, T. D. (1991). Post-translational processing of the tobacco etch virus 49-kDa small nuclear inclusion polyprotein: Identification of an internal cleavage site and delimitation of VPg and proteinase domains. *Virology*, **183**, 449–456.
11. Dougherty, W. G. & Parks, T. D. (1989). Molecular genetic and biochemical evidence for the involvement of the heptapeptide cleavage sequence in determining the reaction profile at two tobacco etch virus cleavage sites in cell-free assays. *Virology*, **172**, 145–155.
12. Dougherty, W. G., Cary, S. M. & Parks, T. D. (1989). Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. *Virology*, **171**, 356–364.
13. Allaire, M., Chernaia, M. M., Malcolm, B. A. & James, M. N. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature*, **369**, 72–76.
14. Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W. *et al*. (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell*, **77**, 761–771.
15. Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S. & James, M. N. (1997). Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.* **273**, 1032–1047.
16. Bergmann, E. M., Mosimann, S. C., Chernaia, M. M., Malcolm, B. A. & James, M. N. (1997). The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. *J. Virol.* **71**, 2436–2448.
17. Parks, T. D., Howard, E. D., Wolpert, T. J., Arp, D. J. & Dougherty, W. G. (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.
18. Kapust, R. B., Tözsér, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D. & Waugh, D. S. (2001). Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14**, 993–1000.
19. Kapust, R. B., Tözsér, J., Copeland, T. D. & Waugh, D. S. (2002). The P1′ specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* **294**, 949–955.
20. Bazan, J. F. & Fletterick, R. J. (1988). Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proc. Natl Acad. Sci. USA*, **85**, 7872–7876.
21. Ishida, T. & Kato, S. (2003). Theoretical perspectives on the reaction mechanism of serine proteases: the reaction free energy profiles of the acylation process. *J. Am. Chem. Soc.* **125**, 12035–12048.
22. Dougherty, W. G., Carrington, J. C., Cary, S. M. & Parks, T. D. (1988). Biochemical and mutational analysis of a plant virus polyprotein cleavage site. *EMBO J.* **7**, 1281–1287.
23. Fairall, L., Chapman, L., Moss, H., de Lange, T. & Rhodes, D. (2001). Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. *Mol. Cell.* **8**, 351–361.
24. Miroux, B. & Walker, J. E. (1996). Over-expression of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298.
25. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
26. Renzoni, D. A., Pugh, D. J., Siligardi, G., Das, P., Morton, C. J., Rossi, C. *et al*. (1996). Structural and thermodynamic characterization of the interaction of the SH3 domain from Fyn with the proline-rich binding site on the p85 subunit of P13-kinase. *Biochemistry*, **35**, 15646–15653.
27. O'Brien, R., DeDecker, B., Fleming, K. G., Sigler, P. B. & Ladbury, J. E. (1998). The effects of salt on the TATA binding protein–DNA interaction from a hyperthermophilic archaeon. *J. Mol. Biol.* **279**, 117–125.
28. Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallog. sect. D*, **50**, 414–423.
29. Saridakis, E. & Chayen, N. E. (2000). Improving protein crystal quality by decoupling nucleation and growth in vapor diffusion. *Protein Sci.* **9**, 755–757.
30. Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

31. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
32. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
33. Roussel, A., Inisan, A. G., Knoops-Mouthuy, E. & Cambillau, C. (2000) *Turbo-Frodo, Version OpenGL. 1*, AFMB/CNRS, Marseille, France.

Edited by M. Guss

(Received 12 January 2005; received in revised form 6 April 2005; accepted 7 April 2005)