Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Conformational flexibility in the catalytic triad revealed by the high-resolution crystal structure of *Streptomyces erythraeus* trypsin in an unliganded state

Elise Blankenship,[a]‡ Krishna Vukoti,[a]‡ Masaru Miyagi[a,b,c]* and David T. Lodowski[a,b]*

[a]Case Center for Proteomics and Bioinformatics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA, [b]Department of Pharmacology, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA, and [c]Department of Opthalmology and Visual Sciences, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA

‡ These authors should be considered joint first authors.

Correspondence e-mail: mxm356@cwru.edu, dtl10@cwru.edu

Received 4 October 2013
Accepted 12 December 2013

PDB reference: *Streptomyces erythraeus* trypsin, 4m7g

With more than 500 crystal structures determined, serine proteases make up greater than one-third of all proteases structurally examined to date, making them among the best biochemically and structurally characterized enzymes. Despite the numerous crystallographic and biochemical studies of trypsin and related serine proteases, there are still considerable shortcomings in the understanding of their catalytic mechanism. *Streptomyces erythraeus* trypsin (SET) does not exhibit autolysis and crystallizes readily at physiological pH; hence, it is well suited for structural studies aimed at extending the understanding of the catalytic mechanism of serine proteases. While X-ray crystallographic structures of this enzyme have been reported, no coordinates have ever been made available in the Protein Data Bank. Based on this, and observations on the extreme stability and unique properties of this particular trypsin, it was decided to crystallize it and determine its structure. Here, the first sub-angstrom resolution structure of an unmodified, unliganded trypsin crystallized at physiological pH is reported. Detailed structural analysis reveals the geometry and structural rigidity of the catalytic triad in the unoccupied active site and comparison to related serine proteases provides a context for interpretation of biochemical studies of catalytic mechanism and activity.

1. Introduction

Trypsin, a representative member of the serine protease/hydrolase family, contains a conserved 'catalytic triad' consisting of the eponymous nucleophilic serine, a histidine and an aspartic acid residue. Enzymes with this conserved triad represent over one-third of all known proteases, including trypsins, chymotrypsins, elastases and α-lytic proteases, as well as a number of lipases and hydrolases that act on nonprotein substrates (Hedstrom, 2002*b*; Di Cera, 2009). Their catalytic reactions proceed through an acyl-enzyme mechanism (Hartley & Kilby, 1954; Hedstrom, 2002*a*), in which the hydroxyl group of the catalytic serine attacks the carbonyl C atom of the scissile peptide bond of the substrate, resulting in the generation of a short-lived tetrahedral intermediate, which subsequently collapses upon breaking the carbon–nitrogen peptide bond to form an acyl-enzyme intermediate (Schonbaum *et al.*, 1961). The acyl-enzyme intermediate is then hydrolyzed and the free hydroxyl group of the catalytic serine is regenerated, thus readying the enzyme for the next reaction. Like trypsins from mammalian sources, *Streptomyces erythraeus* trypsin (SET) specifically hydrolyzes both lysyl and arginyl peptide bonds (Yoshida *et al.*, 1971), and follows the generally accepted mechanism of serine protease catalysis (Yoshida *et al.*, 1973; Sakiyama & Kawata, 1983).

Based on sequence comparison of SET with bovine trypsin, the catalytic triad residues of SET were assigned as Ser179, His42 and Asp88 (Yamane *et al.*, 1991; Nagamine-Natsuka *et al.*, 1995), an assignment borne out after the initial structure determination of SET (Yamane *et al.*, 1991). While the structure and activity of serine proteases has been studied in great detail for over 40 years (Di Cera, 2009; Hedstrom, 2002*b*), unanswered questions regarding important details of the catalytic mechanism still remain. In particular, the protonation state of the catalytic residues and the potential for movement of the histidine during catalysis have been probed in numerous structural and biochemical assays (Schmidt *et al.*, 2003; Di Cera, 2009; Topf *et al.*, 2002; Haddad *et al.*, 2005). A more thorough understanding of serine proteases in their unliganded state may help to address these questions by providing a context for the serine protease structure prior to the onset of substrate binding and catalysis.

A major complication in the crystallographic study of serine proteases is attributed to their instability at physiological pH owing to autolysis, resulting in inactivation or complete digestion of the enzyme (Várallyay *et al.*, 1998). Thus, most of the previous crystallization studies of serine proteases have required inactivation of the protease by chemical modification of active-site residues or nonphysiological acidic conditions to inhibit the enzyme activity, or have utilized engineered substrates or inhibitors designed to trap the enzyme in particular states akin to presumed catalytic intermediates. While these studies have provided a large amount of information on the structure of serine proteases, few structural studies have revealed a native trypsin enzyme in its unliganded state; indeed, the majority of trypsin structures determined to date contain an inhibitor or are chemically modified at the catalytic residues. Unlike mammalian trypsins, SET is resistant to autolysis at physiological pH, making it an ideal protease for structural studies (Kiser *et al.*, 2009). SET also exhibits the prototypical trypsin fold as found in many mammalian trypsins, with 204 of 223 residues aligned between SET and bovine trypsin, and contains similar S1 and S2 binding pockets (Yamane *et al.*, 1995). In this study, SET was crystallized at a pH (7.9) at which the enzyme is active and does not exhibit appreciable autolysis even after days at 37°C, making it an ideal protease for structural studies at a physiologically relevant pH and without the use of artificial substrates or inhibitors.

Here, we present the structure of SET in its unliganded state determined to subatomic resolution (0.81 Å from the $CC_{1/2}$; Diederichs & Karplus, 2013; Karplus & Diederichs, 2012) and compare it with representative structures of other serine hydrolases at comparable resolution. The inclusion of data beyond an $I/\sigma(I)$ level of 2.0 (which in the case of this crystal is 0.85 Å) better enabled the visualization of $F_o - F_c$ densities attributable to H atoms (see Supplementary Fig. S1[1]). Although the structure of SET has been determined previously at 2.7 Å resolution in 1991 (Yamane *et al.*, 1991) and was improved to 1.9 Å resolution in 1995 (Yamane *et al.*, 1995), the crystallographic coordinates were never made available in the Protein Data Bank. In this study, we extend the previous analysis to subatomic resolution, and compare the geometry of the active site of SET with structures of other serine hydrolases at comparable resolution.

2. Experimental procedures

2.1. Protein expression, preparation and crystallization

2.1.1. Cloning and protein expression. The *Escherichia coli* codon-optimized full-length SET gene containing a hexahistidine tag at the N-terminus was inserted into the PQE-80L vector as its pro-protease form (Kiser *et al.*, 2009). The DNA sequence of the resultant plasmid was verified by DNA sequencing. *E. coli* BL21(DE3) cells (Invitrogen, Carlsbad, California, USA) transformed with the PQE80-SET plasmid were grown to mid-log phase in 2×YT medium at 37°C in the presence of 0.1 mg ml^{-1} ampicillin. Induction of the culture was carried out with 0.1 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) for 6 h and the cells were then harvested by centrifugation at 2500*g* for 15 min and stored at −80°C until use in protein purification.

2.1.2. Protein purification. In a typical purification procedure, 10 g cell pellet was resuspended in 100 ml buffer *A* [10 m*M* Tris–HCl pH 8.0, 300 m*M* NaCl, 10%(*v*/*v*) glycerol, 0.05%(*v*/*v*) Tween 20] and lysed by incubation with 1.5 mg ml^{-1} lysozyme at room temperature for 1 h. The viscous DNA was then sheared by a brief sonication on ice. Cell debris was removed by centrifugation at 20 000*g* for 1 h at 4°C. The cleared supernatant was loaded onto a 5 ml Ni–NTA resin column (Thermo Scientific, Rockford, Illinois, USA) pre-equilibrated with buffer *A*, washed with 100 ml buffer *A*, and the bound His_6-Pro-SET was eluted with 25 ml buffer *A* containing 250 m*M* imidazole. The fractions containing His_6-pro-SET were pooled and dialyzed against 3 l buffer *B* (10 m*M* Tris–HCl, 150 m*M* NaCl, pH 8.0) to remove imidazole. The fusion protein was then treated with chymotrypsin (0.1 mg ml^{-1}) for 3 h at 25°C to produce mature SET and was loaded onto a 5 ml *p*-aminobenzamidine–agarose column (Sigma–Aldrich, St Louis, Missouri, USA) equilibrated with buffer *C* (50 m*M* Tris–HCl, 300 m*M* NaCl, 20 m*M* $CaCl_2$, pH 8.0). After washing the column with 100 ml buffer *C*, SET was eluted with 0.5% formic acid, and 2 ml fractions were collected into glass tubes containing 2 ml 1 *M* ammonium bicarbonate (NH_4HCO_3) to immediately neutralize the formic acid in the eluant. Fractions containing the mature SET were assessed by SDS–PAGE, pooled, concentrated and the buffer was exchanged in a 10 000 MWCO Amicon Ultra-4 centrifugal filter device (Millipore, Billerica, Massachusetts, USA). The removal of formate and excess ammonium bicarbonate was accomplished by repeated concentration and dilution with 10 m*M* ammonium bicarbonate. The concentration of the protein was determined using the molar absorption coefficient (ε = 14 773 M^{-1} cm^{-1}) at 280 nm. The typical yield was 5 mg of purified protein per 10 g of bacterial pellet.

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: MN5042).

2.1.3. Protein crystallization and cryoprotection. Initial crystallization screening was performed utilizing 200 nl crystallization drops set up with a Crystal Phoenix crystallization robot (Art Robbins Instruments, Sunnyvale, California, USA) and various commercial sparse-matrix crystallization screens from Qiagen (Venlo, Limburg, Germany), Molecular Dimensions (Altamonte Springs, Florida, USA) and Hampton Research (Aliso Viejo, California, USA). Once initial hits had been analysed at the synchrotron, the best conditions were optimized by hand with larger drop volumes using sitting-drop setups in ChrysChem plates (Hampton Research) containing 500 µl reservoir solution. Crystals exhibiting optimal diffraction resolution grew from crystallization drops consisting of 1.5 µl SET protein at 39 mg ml^{-1} and 1.5 µl well solution (10 m*M* calcium chloride, 100 m*M* HEPES pH 7.0, 2.72 *M* ammonium sulfate; the measured pH of the well solution was 7.9). Crystals appeared after 4–5 d at 23°C and reached maximum dimensions of 700 × 500 × 300 µm over a period of three weeks. Crystals were harvested into appropriately sized Dual-Thickness MicroLoops (MiTeGen, Ithaca, New York, USA), cryoprotected by removal of the surrounding mother liquor in perfluoropolyether oil PFO-X175/08 (Hampton Research), cooled by plunging into liquid nitrogen and stored under liquid nitrogen until analysis. An image of a representative crystal is shown in Fig. 1.

2.2. Data collection, structure determination and refinement

2.2.1. Data collection. Initial crystals obtained from robotic screening were screened on NE-CAT beamline 24-ID-C at the Advanced Photon Source (APS) at Argonne National Laboratory and on beamline X29c at the National Synchrotron Light Source at Brookhaven National Laboratory, and exhibited diffraction to *d* spacings of ∼1.1 Å. After optimization of the crystallization conditions, diffraction data were measured on NE-CAT beamline 24-ID-C at the Advanced Photon Source at Argonne National Laboratory utilizing the MD2 microfocus end station and a 100 × 40 nm X-ray beam. Data were collected in shutterless mode on a Pilatus detector with 0.2° 'images' at 0.6888 Å wavelength and 150 mm crystal-to-detector distance. Owing to the large size of the crystals, we were able to take advantage of the continuous vector scan data-collection scheme implemented at NE-CAT to distribute dose over the entire crystal and maximize resolution.

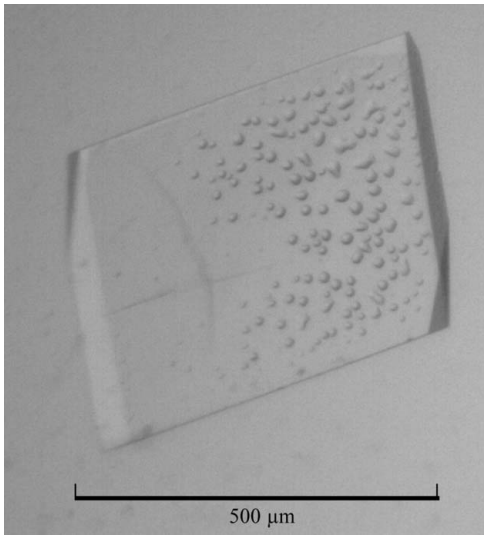

Figure 1
Typical crystal of SET. Purified SET protein at 39 mg ml^{-1} was mixed in a 1:1 ratio with a well solution consisting of 10 m*M* calcium chloride, 100 m*M* HEPES pH 7.0, 2.72 *M* ammonium sulfate (the measured pH of the final well solution was 7.9) and incubated at 23°C. Crystals appeared after 3–5 d and attained maximum dimensions of 700 × 500 × 300 µm over the course of three weeks. The scale bar corresponds to 500 µm.

Table 1
Data-collection and refinement statistics for the described structure of SET.

Values in parentheses are for the outer shell.

Data-collection statistics	
Beamline	NE-CAT 23-ID-C
Space group	$P2_12_12_1$
Unit-cell parameters (°)	
a	46.96
b	61.25
c	74.14
Resolution range† (Å)	61.380–0.82 (0.82–0.81)
$R_{p.i.m.}$‡	0.031 (1.053)
Mean $I/\sigma(I)$	13.1 (1.0)
No. of observations	234158
No. of unique observations	208803
Completeness (%)	99.80 (96.9)
Multiplicity	5.40
Average mosaicity (°)	0.06
Refinement statistics	
Resolution (Å)	0.81
R_{cryst}§ (%)	11.16
R_{free}¶ (%)	12.19
R_{total} (%)	11.22
No. of atoms	3909
No. of solvent atoms	353
No. of H atoms	1771
R.m.s. deviations	
Bonds (Å)	0.02
Angles (°)	2.032
Ramachandran analysis (%)	
Favored	98.40
Allowed	1.60

† The resolution range used in refinement was based on $CC_{1/2}$ for the data set, which was 0.81 Å along the *h* axis, 0.83 along the *l* axis and 0.88 along the *k* axis owing to mild anisotropy. When an $I/\sigma(I) = 2.0$ criterion was used, the data extended to 0.87 Å along *h*, 0.88 Å along *k* and 0.85 Å along *l*. ‡ $R_{p.i.m.}$ (within I^+/I^-). § R_{cryst} =

2.2.2. Data processing, model building, refinement and analysis. Data from a single crystal were integrated and scaled using *XDS* (Kabsch, 2010) and *AIMLESS*, respectively (Winn *et al.*, 2011). The resolution cutoff for subsequent refinement was determined using the $CC_{1/2}$ values reported by *AIMLESS* (Winn *et al.*, 2011; Evans, 2011). An initial model was generated using the structure of *S. griseus* trypsin (SET shares ∼40% identity and 50% homology with *S. griseus* trypsin; PDB entry 1sgt; Read & James, 1988) as a molecular-replacement model in *Phaser*, and the initial phases were used directly from *Phaser* as input into *ARP*/*wARP*, which was able to build a preliminary model consisting of 221 out of a total of 227 amino acids (McCoy *et al.*, 2007; Perrakis *et al.*, 1999). Models were built in *Coot*, including adding alternative conformations

where appropriate, and were refined using *PHENIX*, utilizing a riding hydrogen model, individual anisotropic temperature factors and occupancy refinement (Adams *et al.*, 2010; Emsley *et al.*, 2010). Allowing individual hydrogen refinement did not result in better geometry statistics or better *R* factors. For the final round of refinement, we elected to refine against all data using *REFMAC*5 (Murshudov *et al.*, 2011; $R_{\rm total}$ in Table 1) to better fit the disordered surface side chains and alternative conformations. Analysis of the structures was performed in *PyMOL* (Schrödinger), *MolProbity* (Chen *et al.*, 2010) and *Coot* (Emsley *et al.*, 2010; Krissinel & Henrick, 2004). The final model contained 223 of 227 residues, lacking the extreme C-terminus owing to disorder. At a contour level of 2.0σ, ~60% of amide H atoms exhibit observable $F_{\rm o} - F_{\rm c}$ map peaks, with ~70% of H atoms visible in the structure overall (see representative residual hydrogen density in Supporting Fig. S1). The final atomic coordinates and structure factors have been deposited in the PDB as entry 4m7g. The data-collection and refinement statistics are summarized in Table 1.

3. Results

3.1. Description of the overall structure

Although we attempted to replicate the previous crystallization conditions as reported, we were unable to grow crystals in the original $P3_221$ space group (Yamane *et al.*, 1995) and rather produced protein crystals in space group $P2_12_12_1$. While the crystal lattice differed between the previous SET structures and our current model, both models represent the unliganded form of SET. SET contains the two α-helices and 13 β-sheets of the prototypical trypsin fold exhibited by bovine trypsin, with the active-site residues located in a cleft on the exterior of the protein (Fig. 2), and indeed exhibits a high degree of structural homology to other members of the serine hydrolase family, superposing well with bovine and other bacterial trypsins, with an r.m.s.d. of only 1.42 Å to bovine trypsin. Interatomic distances between catalytic residues were measured and compared, demonstrating that the catalytic residues of SET and other serine proteases exist in a well defined arrangement in the unliganded form. Disulfide bonds (Cys27–Cys43 and Cys175–Cys199) adjacent to the active site ensure the precise positioning needed to ensure efficient catalysis. Curiously, the Cys150–Cys164 disulfide bond exhibits two conformations; although the cysteine side chains exist in two different rotomeric states separated by almost 90°, the α-carbon positions remain nearly identical. This was also observed in the 0.75 Å resolution structure of bovine trypsin (PDB entry 4i8h; Liebschner *et al.*, 2013).

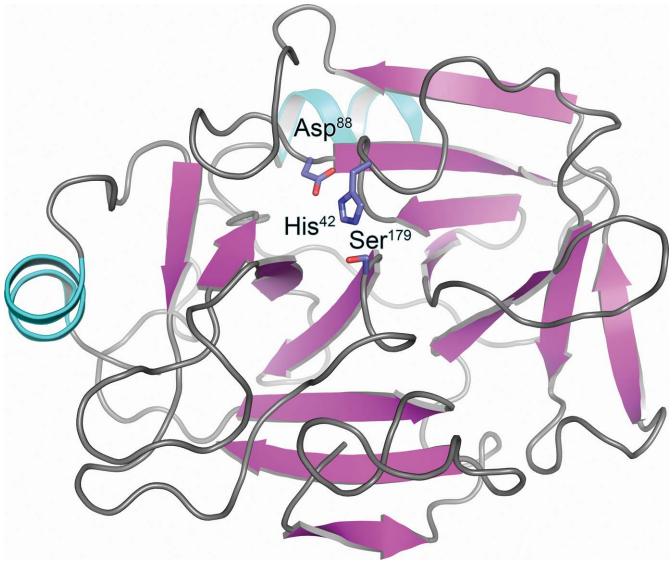

Figure 2
SET exhibits the prototypical trypsin fold. The structure contains two α-helices (cyan) and 14 β-sheets (magenta). 13 β-sheets are present in the high-resolution bovine trypsin structure (PDB entry 4i8h), and an additional β-sheet (B6) was identified during structural analysis of SET. Active-site residues are depicted as blue sticks. H atoms have been removed for clarity.

Few differences in the overall structure of SET compared with the other serine proteases analysed are observed: the autolysis loop observed in bovine trypsin is present, but lacks a lysine or arginine residue and is therefore insensitive to proteolysis. Bovine trypsin contains 13 β-sheets; an additional β-sheet (B6) was added to our structure during refinement. Although many other trypsins are activated by Ca^{2+} binding (Green & Neurath, 1953), SET does not contain the conserved acidic residues forming a Ca^{2+}-binding/activation site and indeed there was no evidence of a Ca^{2+} ion in the model of SET, despite the inclusion of $CaCl_2$ in the crystallization conditions. Although included in the crystallization conditions, Ca^{2+} ions were also not observed in previous SET structures (Yamane *et al.*, 1991, 1995).

3.2. Comparison to other unliganded serine hydrolase structures

In the initial reports of the crystal structure of SET, its structure was compared with the available structures of bovine trypsin and *S. griseus* trypsin, where a high degree of fold similarity was demonstrated (Yamane *et al.*, 1991, 1995). Here, we compare the structure of SET with several high-resolution serine protease structures as well as the structures of two other enzymes possessing a trypsin fold and classical catalytic triad, as the resolution for comparison has increased almost a full angstrom over the previously described SET structures. The structures chosen for comparison are high-resolution serine hydrolase structures crystallized at neutral pH that are not complexed with substrate or inhibitor. The catalytic residues of a bacterial α-lytic protease (PDB entry 1ssx; Fuhrmann *et al.*, 2004), porcine elastase (PDB entry 1qnj; Würtele *et al.*, 2000) and *Fusarium oxysporum* trypsin (PDB entry 1xvo; Schmidt & Lamzin, 2005) were measured and compared with those of SET through structural superposition and interatomic distance and angle calculations. Two additional structures of bovine trypsin were also included in the analysis (PDB entries 3unr and 4i8h; Fisher *et al.*, 2012; Liebschner *et al.*, 2013) although they were complexed with the inhibitor benzamidine, as no unmodified high-resolution structure of the bovine

enzyme at a similar pH without an inhibitor bound is available. Because the original structural papers compared the structure of SET with that of *S. griseus* trypsin (SGT), we also chose to include SGT (PDB entry 1os8; Page *et al.*, 2003) in our analysis.

Superposition of all of these serine hydrolase structures confirms the high degree of structural similarity between all of the serine protease models, including that of SET (Supplementary Table S1). We performed structural analysis using the *RAPIDO* server (Mosca & Schneider, 2008), revealing that the average r.m.s.d. between SET and the compared structures is 2.17 Å. Examination of the alignments of the conserved regions in the structures reveal that while the solvent-exposed loops on the protein surface and flexible backbone regions did not exhibit similar conformations, the catalytic sites and other enzymatically important regions of protein superpose closely. Supplementary Table S1 describes the comparative r.m.s.d. values between the overall structures, demonstrating the structural similarity to other serine proteases, especially to mammalian trypsin, with which it has an r.m.s.d. of 1.42 Å.

Analysis of the geometry of the catalytic residues within each structure demonstrates the conserved hydrogen-bonding network of serine protease active sites. Table 2 and Fig. 3 compare the atomic positions of the catalytic residues of the structures, revealing the hydrogen-bonding interactions that exist between catalytic residues and waters in the unoccupied active site. The catalytic residues display a high degree of similarity in position between all of the structures, implying that the hydrogen-bonding network in the active site is conserved, as are the relative locations of the catalytic residues of these enzymes in the unliganded enzyme. Hydrogen bonding between the O^{γ} atom of Ser179 and the $N^{\varepsilon 2}$ atom of His42 is present in all structures, as expected, as His42 acts as the proton acceptor at the onset of catalysis (Radisky *et al.*, 2006). In addition to this, the $N^{\delta 1}$ atom and backbone N atom of His42 form hydrogen bonds to the $O^{\delta 2}$ and $O^{\delta 1}$ atoms of Asp88, respectively. Additionally, our analysis suggests, based upon distance and geometry, a potential hydrogen bond from the His42 $H^{\varepsilon 1}$ atom (the hydrogen bound to the His42 imidazole ring $C^{\varepsilon 1}$ atom) to the backbone carbonyl O atom of Ser194. This C—H···O hydrogen bond is conserved amongst serine hydrolases, and has been proposed to be an essential component of catalysis by facilitating charge distribution on the catalytic His (Derewenda *et al.*, 1994).

Table 2
Structural comparison of the geometry of the catalytic triad in SET and other high-resolution serine hydrolase structures.

As observed in many other high-resolution serine protease structures, a hydrogen bond exists between the His $\varepsilon 1$ carbon and the carbonyl O atom of a nearby serine (Ser194 in SET). This type of C—H···O bond has important implications in catalysis and has been proposed to assist with the distribution of charge on the imidazole ring of the catalytic histidine (Derewenda *et al.*, 1994). Other measurements comparing the relative locations of the active-site residues also demonstrate the conserved geometry of the catalytic site. Angle and distance measurements were performed using *PyMOL* (Schrödinger, New York, USA).

	SET	Fungal trypsin (1xvo)	Bovine trypsin (4i8h)	Porcine elastase (1qnj)	α-Lytic protease (1ssx)
Distances (Å)					
His42 $N^{\varepsilon 2}$ to Ser179 O^{γ}	2.90	2.94	3.09	3.17	2.96
His42 $N^{\delta 1}$ to Asp88 $O^{\delta 2}$	2.79	2.75	2.72	2.70	2.77
His42 N to Asp88 $O^{\delta 1}$	2.82	2.80	2.82	2.82	2.86
His42 $C^{\varepsilon 1}$ to Ser194 O	3.04	3.04	3.01	3.03	3.03
His42 $H^{\varepsilon 1}$ to Ser194 O	2.52	2.36	2.37	2.37	2.33
Angles (°)					
His42 $C^{\varepsilon 1}$—His42 $H^{\varepsilon 1}$—Ser194 O	115.06	130.19	126.06	127.39	126.12
His42 $H^{\varepsilon 1}$—Ser194 O—Ser194 C	115.85	127.25	124.98	123.90	127.07
His42 $H^{\varepsilon 1}$—Ser194 O—Trp195 N	144.45	157.39	154.47	148.36	148.89

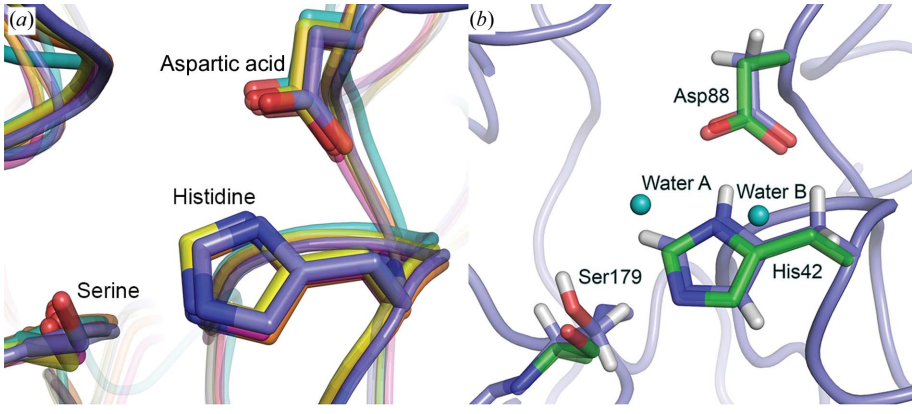

Figure 3
Structural similarity of active-site residues in SET and other serine protease structures. (*a*) Bovine trypsin (PDB entry 3unr, yellow), *F. oxysporum* trypsin (PDB entrey 1xvo, magenta), α-lytic protease (PDB entry 1ssx, cyan) and porcine elastase (PDB entry 1qnj, orange) were superposed on SET (blue). Note that the minor Ser179 conformation is not observed in the other structures. (*b*) High-resolution SET and bovine trypsin structures share similar catalytic residue positions as well as two waters near the active site. The structure of bovine trypsin (PDB entry 4i8h, green) in complex with benzamidine also aligns closely with the catalytic triad (and backbone) of SET (blue). Waters A and B are observed in both structures.

While the locations of the catalytic residues in SET and the compared structures superimpose closely, the observed ordered water molecules in the vicinity of the active sites of these structures show a greater degree of variability in position, whilst remaining within hydrogen-bonding distance of the catalytic triad. Waters near the catalytic histidine in SET are observed in 'homologous' positions in other structures. For example, water B (Fig. 4), which is hydrogen-bonded to His42, is also observed in the apo-enzyme and bovine tryspin structures used for comparison; when the catalytic triads are superposed, all 'homologous' water Bs occupy a 2 Å diameter sphere. It has previously been proposed that an ordered water observed near the active site is poised to act as the nucleophile in acyl intermediates of trypsin (Radisky *et al.*, 2006; Singer *et al.*, 1993). In the structure of SET this water is absent, suggesting that this water is labile and is

perhaps only recruited to the active site upon substrate binding or the attainment of a catalytic intermediate.

3.3. Geometry of the active site

Ser179, His42 and Asp88 together form the catalytic triad of SET. Given the marked increase in resolution from the previous SET structures, we were able to observe two conformations for the γ-oxygen of the hydroxyl group of the catalytic Ser179 (Fig. 4). The major conformer, Ser179$_A$, with an occupancy of 67%, is poised to interact with substrate; any subsequent analysis of the catalytic triad utilized this conformation. The *A* conformer of Ser179 is modelled with a hydrogen attached to the γ-oxygen of this catalytic serine residue, aimed towards the centre of the active site and hydrogen-bonded to the N$^{\varepsilon 2}$ atom of His42, as would be expected based on proximity. Ser179$_B$, with an occupancy of 33%, has its γ-oxygen located approximately 1.3 Å from that of Ser179$_A$; Ser179$_B$ is located within hydrogen-bonding distance of water C (water 736), a well ordered solvent molecule located within the active site that is not modelled in any other unliganded serine protease structure. These results suggest that the hydroxyl group of the catalytic serine has some freedom to rotate as a result of rotation of the C$^{\alpha}$—C$^{\beta}$ bond of this residue. In the 0.75 Å resolution structure of bovine trypsin bound to benzamidine (PDB entry 4i8h), only a single conformation of the catalytic serine (Ser195) is observed, suggesting that perhaps the binding of inhibitor within the S1 pocket may stabilize the serine, although it cannot be excluded that this flexibility is a feature of SET alone rather than of serine hydrolases in the absence of substrate.

His42 is assumed to be neutral based on the pK_a of this residue (6.9; Sakiyama & Kawata, 1983) and the pH of the crystallization conditions (7.9), so no hydrogen is modelled on the N$^{\varepsilon 2}$ atom. No residual $F_o - F_c$ peak was observed for this hydrogen when it was omitted from refinement, further suggesting the unprotonated species. The O$^{\delta 2}$ atom of the carboxyl group of Asp88 is located within hydrogen-bonding distance of His42 and is geometrically positioned such that it can hydrogen bond to the δ1 nitrogen of the catalytic histidine. Ser194 also plays a role in the geometry of the active site, as the backbone carbonyl O atom of this residue has been shown to form an O$\cdots$H—C hydrogen bond to the H$^{\varepsilon 1}$ atom of His42 and has been proposed to facilitate serine protease catalysis by assisting in distribution of charge on the imidazole ring (Derewenda *et al.*, 1994).

Two additional ordered water molecules (numbered 659 and 550; here referred to as waters A and B) are observed within hydrogen-bonding distance of the histidine in the catalytic triad, and one ordered water molecule (numbered 736; here referred to as water C) is located within hydrogen-bonding distance of Ser179$_B$. Waters near the active site may play roles in catalysis by acting as a nucleophile in the deacylation reaction and in stabilization of the residues near the oxyanion hole, as described in the structure of *F. oxysporum* trypsin (PDB entry 1xvo; Schmidt & Lamzin, 2005); a water molecule observed in *F. oxysporum* trypsin (PDB entry 1pq8) located in a similar position was proposed to be the water that attacks the acyl-enzyme (Schmidt *et al.*, 2003). In addition to waters A, B and C, two additional waters (waters 586 and 738) are located in the vicinity of the active site of SET and are more distant from the catalytic residues, but remain within hydrogen-bonding distance of each other and waters A and B, forming a continuous hydrogen-bonded network of solvent. An area of electron density near the catalytic serine in the active site is left unmodeled. A sulfate molecule, hydrogen-bonded to nearby residues, has been observed in this location in several other serine protease structures which were crystallized in ammonium sulfate, perhaps stabilizing the active site (PDB entries 3unr, 1xvo, 1ssx, 1qnj and 4i8h); however, in our maps this region was too disordered to fit a sulfate unambiguously and anyhow the residual $F_o - F_c$ density is more extensive than a sulfate. The active site of SET is illustrated in detail in Fig. 4, and measurements corresponding to the geometry of the catalytic triad and its potential hydrogen-bonding network are presented in Table 2. These figures demonstrate the locations of the catalytic residues in relation to each other and to ordered solvent molecules, illustrating the extent of the hydrogen-bonding network in the vicinity of the substrate-binding and active sites.

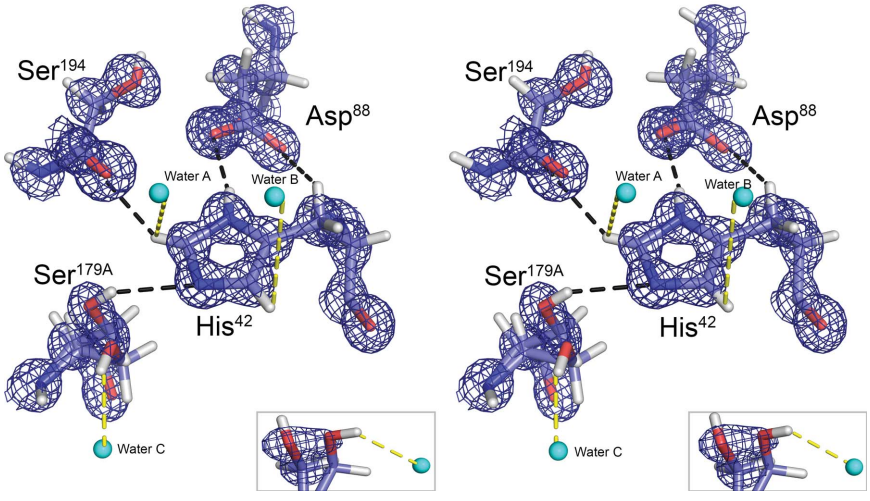

Figure 4
The catalytic triad of SET and nearby waters exhibit a hydrogen-bonding network typical of serine hydrolases. The catalytic triad consisting of Ser179, His42 and Asp88 is shown here (in addition to the neighboring residue Ser194) with $2F_o - F_c$ density (blue) corresponding to the active-site residues (map contoured at 2.0σ). Waters within hydrogen-bonding distance of the active site are shown as cyan spheres. Hydrogen-bonding interactions between active-site residues are shown as black dashed lines and hydrogen-bonding interactions between active-site residues and waters are shown as yellow dashed lines. Inset: close-up of Ser179 and water C. The structure is shown in crossed-eyed stereo.

3.4. S1 substrate-binding site

The S1 substrate-binding pocket utilizes a conserved aspartic acid located at the bottom of the pocket as one of the primary residues for substrate recognition, exploiting an ionic interaction between this residue and the arginyl or lysyl side chain of the substrate (Perona & Craik, 1995). In the absence of substrate, this pocket is occupied by a network of ordered solvent molecules. For comparison, trypsin structures modelled with bound substrate (PDB entry 1pq8) and an inhibitor (PDB entry 4i8k) are shown (Fig. 5). This comparison demonstrates that five ordered solvent molecules occupy the S1 site, forming a connected network of hydrogen-bonded interactions, necessitating their reordering or movement out of the S1 substrate-recognition pocket upon substrate binding.

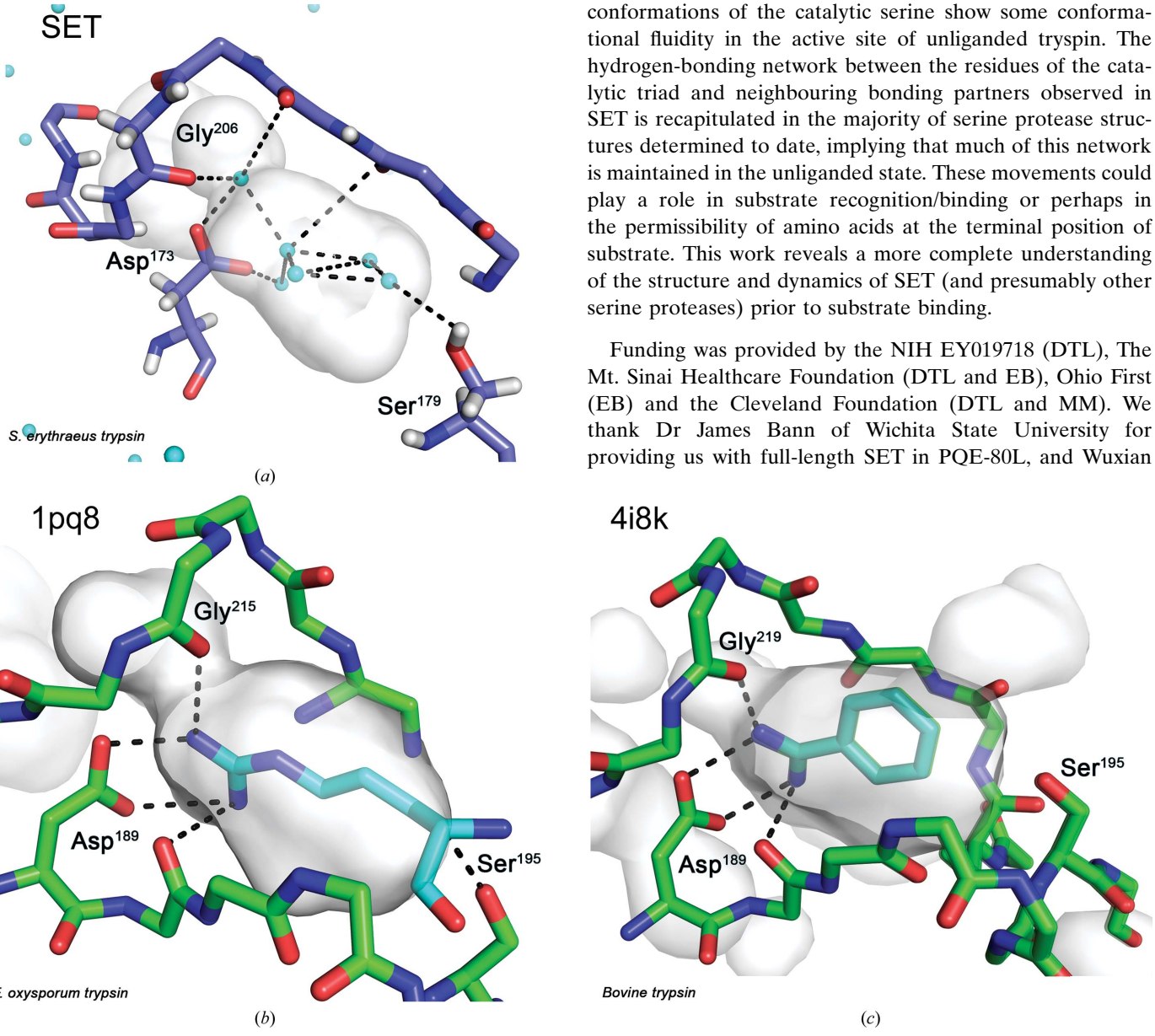

Figure 5
Comparison of the S1 binding pockets in (*a*) SET, (*b*) *F. oxysporum* trypsin and (*c*) bovine trypsin show the hydrogen-bonding interactions in the substrate-recognition site. The S1 pocket of trypsin in the unliganded form, bound to natural substrate (Arg conformation *A* in PDB entry 1pq8) and bound to inhibitor (benzamidine in PDB entry 4i8k) are compared. In the absence of ligand/inhibitor, a network of ordered solvent molecules occupies the pocket.

4. Discussion

SET was crystallized at pH 7.9 in its unliganded state and its structure was determined. Based on evidence from comparative analysis with other serine protease structures, SET can be utilized as a prototypical model for the structural study of serine proteases, and given the advantages with respect to crystallographic study it may be an excellent model with which to pursue high-resolution studies of substrate binding and catalysis. Comparisons with other serine protease structures demonstrate the structural homology of SET, while detailed comparisons of the catalytic triad confirm the conserved position of the catalytic residues in an unoccupied active site, while demonstrating far greater mobility of the catalytic serine than observed in other serine protease structures. The two conformations of the catalytic serine show some conformational fluidity in the active site of unliganded tryspin. The hydrogen-bonding network between the residues of the catalytic triad and neighbouring bonding partners observed in SET is recapitulated in the majority of serine protease structures determined to date, implying that much of this network is maintained in the unliganded state. These movements could play a role in substrate recognition/binding or perhaps in the permissibility of amino acids at the terminal position of substrate. This work reveals a more complete understanding of the structure and dynamics of SET (and presumably other serine proteases) prior to substrate binding.

Funding was provided by the NIH EY019718 (DTL), The Mt. Sinai Healthcare Foundation (DTL and EB), Ohio First (EB) and the Cleveland Foundation (DTL and MM). We thank Dr James Bann of Wichita State University for providing us with full-length SET in PQE-80L, and Wuxian

Shi for her assistance with data collection. This work is based upon research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines, which are supported by a grant from the National Institute of General Medical Sciences (P41 GM103403) from the National Institutes of Health. Use of the Advanced Photon Source, an Office of Science User Facility operated for the US Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the US DOE under Contract No. DE-AC02-06CH11357. Initial data for this study were measured on beamline X29 of the National Synchrotron Light Source. Financial support comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy, and from the National Center for Research Resources (P41RR012408) and the National Institute of General Medical Sciences (P41GM103473) of the National Institutes of Health.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Derewenda, Z. S., Derewenda, U. & Kobos, P. M. (1994). *J. Mol. Biol.* **241**, 83–93.
Di Cera, E. (2009). *IUBMB Life*, **61**, 510–515.
Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst.* D**69**, 1215–1222.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. R. (2011). *Acta Cryst.* D**67**, 282–292.
Fisher, S. J., Blakeley, M. P., Cianci, M., McSweeney, S. & Helliwell, J. R. (2012). *Acta Cryst.* D**68**, 800–809.
Fuhrmann, C. N., Kelch, B. A., Ota, N. & Agard, D. A. (2004). *J. Mol. Biol.* **338**, 999–1013.
Green, N. M. & Neurath, H. (1953). *J. Biol. Chem.* **204**, 379–390.
Haddad, K. C., Sudmeier, J. L., Bachovchin, D. A. & Bachovchin, W. W. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 1006–1011.
Hartley, B. S. & Kilby, B. A. (1954). *Biochem. J.* **56**, 288–297.
Hedstrom, L. (2002*a*). *Chem. Rev.* **102**, 4501–4524.
Hedstrom, L. (2002*b*). *Curr. Protoc. Protein Sci.*, Unit 21.10. doi:10.1002/0471140864.ps2110s26.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Kiser, J. Z., Post, M., Wang, B. & Miyagi, M. (2009). *J. Proteome Res.* **8**, 1810–1817.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Liebschner, D., Dauter, M., Brzuszkiewicz, A. & Dauter, Z. (2013). *Acta Cryst.* D**69**, 1447–1462.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Mosca, R. & Schneider, T. R. (2008). *Nucleic Acids Res.* **36**, W42–W46.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Nagamine-Natsuka, Y., Norioka, S. & Sakiyama, F. (1995). *J. Biochem.* **118**, 338–346.
Page, M. J., Wong, S.-L., Hewitt, J., Strynadka, N. C. J. & MacGillivray, R. T. (2003). *Biochemistry*, **42**, 9060–9066.
Perona, J. J. & Craik, C. S. (1995). *Protein Sci.* **4**, 337–360.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Radisky, E. S., Lee, J. M., Lu, C.-J. K. & Koshland, D. E. Jr (2006). *Proc. Natl Acad. Sci. USA*, **103**, 6835–6840.
Read, R. J. & James, M. N. G. (1988). *J. Mol. Biol.* **200**, 523–551.
Sakiyama, F. & Kawata, Y. (1983). *J. Biochem.* **94**, 1661–1669.
Schmidt, A., Jelsch, C., Ostergaard, P., Rypniewski, W. & Lamzin, V. S. (2003). *J. Biol. Chem.* **278**, 43357–43362.
Schmidt, A. & Lamzin, V. S. (2005). *Acta Cryst.* D**61**, 1132–1139.
Schonbaum, G. R., Zerner, B. & Bender, M. L. (1961). *J. Biol. Chem.* **236**, 2930–2935.
Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). *Mol. Syst. Biol.* **7**, 539.
Singer, P. T., Smalås, A., Carty, R. P., Mangel, W. F. & Sweet, R. M. (1993). *Science*, **259**, 669–673.
Topf, M., Várnai, P., Schofield, C. J. & Richards, W. G. (2002). *Proteins*, **47**, 357–369.
Várallyay, E., Pál, G., Patthy, A., Szilágyi, L. & Gráf, L. (1998). *Biochem. Biophys. Res. Commun.* **243**, 56–60.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Würtele, M., Hahn, M., Hilpert, K. & Höhne, W. (2000). *Acta Cryst.* D**56**, 520–523.
Yamane, T., Iwasaki, A., Suzuki, A., Ashida, T. & Kawata, Y. (1995). *J. Biochem.* **118**, 882–894.
Yamane, T., Kobuke, M., Tsutsui, H., Toida, T., Suzuki, A., Ashida, T., Kawata, Y. & Sakiyama, F. (1991). *J. Biochem.* **110**, 945–950.
Yoshida, N., Sasaki, A. & Inoue, H. (1971). *FEBS Lett.* **15**, 129–132.
Yoshida, N., Sasaki, A. & Inouye, K. (1973). *Biochim. Biophys. Acta*, **321**, 615–623.