Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# The structure of endothiapepsin complexed with the *gem*-diol inhibitor PD-135,040 at 1.37 Å

**L. Coates,* P. T. Erskine, S. Mall, P. A. Williams, R. S. Gill, S. P. Wood and J. B. Cooper**

School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, England

Correspondence e-mail: leightonc@bigfoot.com

Received 10 December 2002
Accepted 18 March 2003

**PDB Reference:** endothiapepsin–PD-135,040 complex, 1od1, r1od1sf.

The crystal structure of endothiapepsin complexed with the *gem*-diol inhibitor PD-135,040 has been anisotropically refined to a resolution of 1.37 Å. The structure of this inhibitor complex is in agreement with previous structures of endothiapepsin *gem*-diol inhibitor complexes that have been used to develop proposed catalytic mechanisms. However, the increase in resolution over previous structures confirms the presence of a number of short hydrogen bonds within the active site that are likely to play an important role in the catalytic mechanism. The presence of low-barrier hydrogen bonds was indicated in a previous one-dimensional H NMR spectrum.

## 1. Introduction

Endothiapepsin derived from the fungus *Endothia parasitica* is a member of the aspartic proteinase class of enzymes. These enzymes are widely distributed and are found in fungi, plants and vertebrates as well as being found in the HIV retrovirus, where the proteinase is essential for maturation of the virus particle (Cooper, 2002). Aspartic proteinase inhibitors have a proven therapeutic record in the treatment of AIDS. Aspartic proteinases also play major roles in hypertension, amyloid disease and malaria and have been implicated in tumorigenesis. Inhibitors of them are therefore much sought after as potential therapeutic agents.

Aspartic proteinases are comprised of two structurally similar domains; each domain contributes an aspartic acid residue to form a catalytic dyad that acts to cleave the substrate peptide bond (Blundell *et al.*, 1990). A defining feature of this class of enzymes is that they are inhibited by the microbial peptide pepstatin, which contains the unusual amino acid statine (Bailey *et al.*, 1993). The fold of endothiapepsin is mainly $\beta$-sheet with small areas of $\alpha$-helix on the outside of the protein; this fold is typical of all aspartic proteinases (Blundell *et al.*, 1990).

Like most aspartic proteinases, endothiapepsin has an optimal acidic pH (4.5) and cleaves protein substrates with a similar specificity to that of porcine pepsin A; it prefers hydrophobic residues at the $P_1$ and $P_1'$ in the cleavage site.

Two catalytic Asp-Thr-Gly sequences are conserved in these enzymes (Pearl & Blundell, 1984). These residues form the active site, in which the carboxyl groups of the two catalytic aspartates are held coplanar by a network of hydrogen bonds involving the surrounding main chain and conserved amino-acid side-chain groups. A solvent molecule bound tightly to both carboxyls by hydrogen bonds is found in all native aspartic proteinase crystal structures. This bound water molecule is within hydrogen-bonding distance of all four carboxyl O atoms and has been implicated in catalysis. Suguna

*et al.* (1987) have suggested that it becomes partly displaced on binding of the substrate and is polarized by one of the aspartate carboxyls. The water then nucleophilically attacks the scissile-bond carbonyl group. These proposals stem from the failure of numerous chemical studies to trap a covalently bound substrate, indicating that the reaction involves a non-covalently bound intermediate (Hofmann *et al.*, 1984).

Short oligopeptide inhibitors of aspartic proteinases bind in $\beta$-strand conformations in the active-site cleft, which is around 30 Å in length and is situated between the two domains of the protein. The best synthetic inhibitors are those that mimic one or both of the hydroxyls found in the putative transition state. One hydroxyl occupies the same position as the water molecule in the native enzyme and binds *via* hydrogen bonds to both the catalytic aspartates (Bailey *et al.*, 1993). Most of the transition-state analogues, such as statine-based inhibitors, mimic this group alone. Three X-ray structures of endothiapepsin complexed with this class of inhibitor have recently been refined to atomic resolution (Coates *et al.*, 2002) and a number of very short hydrogen bonds were identified within the active site. In contrast, fluoroketone analogue inhibitors ($-CO-CF_2-$) such as PD-135,040 (Fig. 1) mimic both hydroxyls in the putative intermediate, as they readily hydrate to the *gem*-diol form [$-C(OH)_2-CF_2-$]. Initial NMR studies using a ketone analogue of the scissile peptide bond (Rich *et al.*, 1982) suggested that it binds to the enzyme in a hydrated *gem*-diol form [$-C(OH)_2-CH_2$]. Current mechanistic proposals have been based on the X-ray structures of *gem*-diol transition-state analogues [difluoroketone; $-C(OH)_2-CF_2-$] (Veerapandian *et al.*, 1992; James *et al.*, 1992). These structures have been refined at resolutions between 2.0 and 2.3 Å. The mechanism based on the structure of a fluoroketone as proposed by Veerapandian *et al.* (1992) is shown in Fig. 2. Early catalytic mechanisms differed in the assignment of charges on the active-site aspartates.

Owing to the low pH optimum of the enzyme (Fruton, 1976) and the proximity of the aspartates within the active site, it is likely that only one of the catalytic carboxyls will be charged at the optimal pH (4.5). The structure of the active site with a single hydroxyl transition-state analogue inhibitor bound allows only two possible positions for a proton on the catalytic aspartates. This could reside either on the inner oxygen of Asp32 or the outer oxygen of Asp215. A proton in either of the other two positions, *i.e.* the outer oxygen of Asp32 or the inner oxygen of Asp215, would be too distant to form hydrogen bonds to the inhibitor hydroxyl. A recent neutron diffraction study of endothiapepsin complexed with the statine-based inhibitor H261 (Coates *et al.*, 2001) showed that the outer oxygen of Asp215 was protonated with the inhibitor bound, suggesting that Asp32 is negatively charged in the transition-state complex.

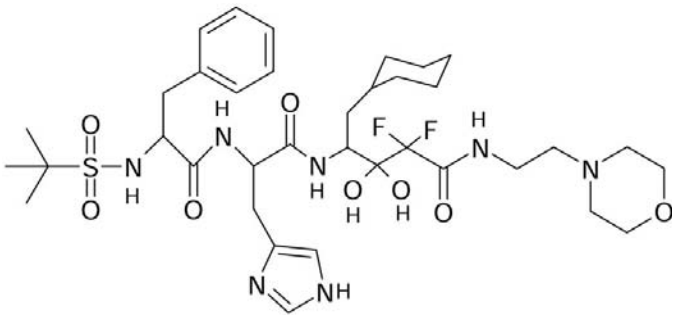

**Figure 1**
The chemical structure of PD-135,040 (shown in the hydrated *gem*-diol form).

## 2. Experimental methods

### 2.1. Protein crystallization

The endothiapepsin–PD-135,040 co-crystals were grown by the hanging-drop vapour-diffusion method in which 5 µl of 2.0 mg ml$^{-1}$ protein solution at pH 4.6 in 0.1 *M* sodium acetate buffer was placed on siliconized cover slips and then mixed with 5 µl of well solution. Each cover slip was then sealed with high-vacuum grease over a well containing 0.7 ml of 100% saturated ammonium solution and 0.7 ml of 0.1 *M* sodium acetate buffer solution at pH 4.7. Small crystals (0.5 × 0.3 × 0.1 mm) appeared after three weeks.

### 2.2. Data collection and refinement

Diffraction data were collected at the ESRF, Grenoble (beamline ID-29). A high-resolution pass of 180° of data was

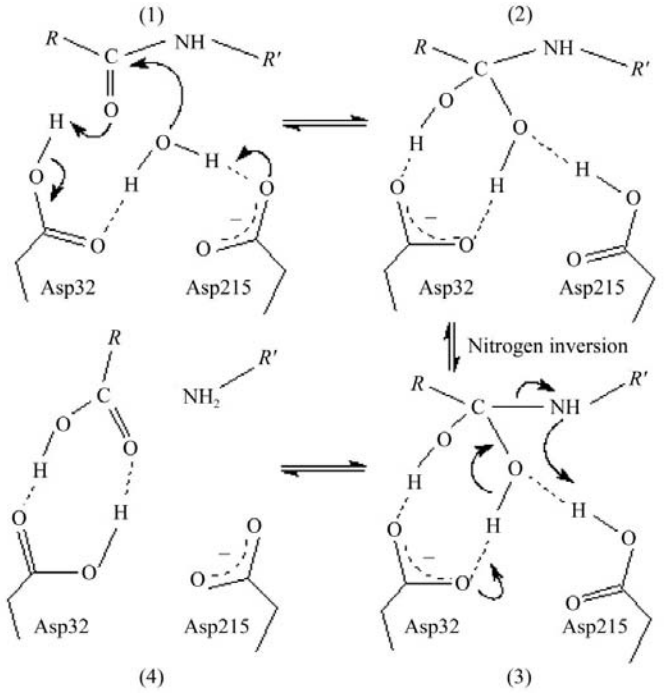


**Figure 2**
The catalytic mechanism proposed by Veerapandian *et al.* (1992), which is based on the X-ray structure of a difluoroketone (*gem*-diol) inhibitor bound to endothiapepsin. A water molecule tightly bound to the aspartates in the native enzyme is proposed to nucleophilically attack the scissile-bond carbonyl. The resulting tetrahedral intermediate (2) is stabilized by hydrogen bonds to the negatively charged carboxyl of Asp32. Fission of the scissile C—N bond is accompanied by transfer of a proton to the leaving amino group either from Asp215 (with nitrogen inversion) or from bulk solvent. Dashed lines indicate hydrogen bonds.

collected with an oscillation angle of 1° with a crystal-to-detector distance of 180 mm and an exposure time of 1 s. A low-resolution pass of 200° of data was also collected with an oscillation angle of 2° and a crystal-to-detector distance of 200 mm. For the low-resolution pass the beam was attenuated using the in-line attenuation system and the exposure time for this pass was 6 s owing to the attenuation of the beam. All data were processed using *MOSFLM* (Leslie, 1992) and scaled with *SCALA* (Collaborative Computational Project, Number 4, 1994). A set of reflections representing 5% of the unique reflections were chosen at random for inclusion in an $R_{\rm free}$ set (Brünger, 1992). An initial rigid-body refinement was performed with *SHELX* (Sheldrick, 1998*a*,*b*), after which a number of amino-acid side chains were rebuilt using *XTALVIEW* (McRee, 1999) with $\sigma_A$-weighted maps (Read, 1986). In further refinement rounds the ligand (PD-135,040) was modelled into electron density, water molecules were then added to the structure and finally riding H atoms were added. The introduction of anisotropic displacement parameters into the model using *SHELX* reduced the *R* factor and $R_{\rm free}$ by 3.20 and 3.16%, respectively. The final refinement and structure quality statistics are given in Table 1. The Ramachandran plot of the final structure shows that 94.6% of the residues are in the most favoured region, with the other 5.4% in the additionally allowed region. Following refinement, the least-squares matrix was inverted with all restraints and shift dampeners removed; owing to computer limitations the ADPs were omitted from this calculation.

**Table 1**
Data-processing and structure-refinement statistics.

Values for the outer shell are given in parentheses.

| | |
|---|---|
| Unit-cell parameters (Å, °) | $a$ = 53.20, $b$ = 73.25, $c$ = 46.00, $\beta$ = 110.10 |
| Space group | $P2_1$ |
| No. of unique reflections | 68317 |
| Resolution range (Å) | 10–1.37 (1.44–1.37) |
| Multiplicity | 6.80 (4.00) |
| $I/\sigma(I)$ | 3.20 (1.70) |
| $R_{\rm merge}$ (%) | 11.30 (36.5) |
| Data completeness (%) | 98.3 (95.5) |
| $R_{\rm factor}$ (%) | 11.53 |
| $R_{\rm free}$ (%) | 15.34 |
| No. of water molecules | 414 |
| Average $B_{\rm iso}$, protein atoms | 14.98 |
| Anisotropic data-to-parameter ratio | 2.48 |

## 3. Results and discussion

The overall structure is largely similar to that of other endothiapepsin–*gem*-diol complexes (Bailey & Cooper, 1994; Veerapandian *et al.*, 1992). However, the low atomic positional e.s.d. values in the active site (0.03 Å) allow the identification of a number of short hydrogen bonds between the catalytic aspartates and the *gem*-diol inhibitor as shown in Fig. 3. The one-dimensional H NMR spectra of this inhibitor bound to the active site of endothiapepsin (Coates *et al.*, 2002) shows the presence of a peak at 16.2 p.p.m., which is well outside the normal range of protein peaks (1–11 p.p.m.). Using the data of McDermott & Ridenour (1996), this suggests hydrogen-bond length(s) of around 2.5 to 2.6 Å, which is in agreement with the bond lengths shown in Fig. 3. The electron density at the active site is well defined for both the inhibitor and the catalytic aspartates (Fig. 4); however, there is no electron density for the hydrogen positions as might be expected at this resolution.

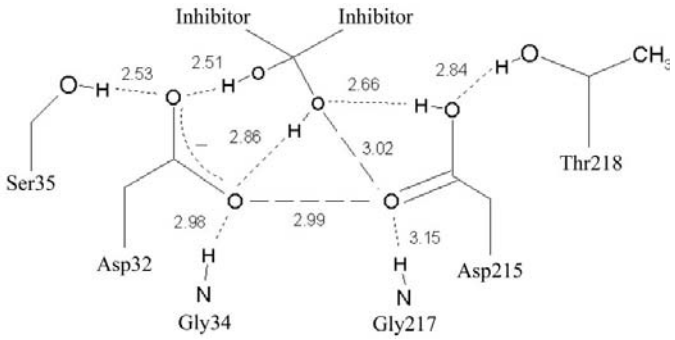


**Figure 3**
The active site of endothiapepsin bound to the transition-state analogue inhibitor PD-135,040. The atomic positional e.s.d. values for the atoms shown are all around 0.03 Å. All lengths are given in ångstroms with dotted lines indicating hydrogen bonds. For hydrogen bonds the donor–acceptor atom distances are shown.

In earlier atomic resolution X-ray studies of endothiapepsin complexed with statine-based inhibitors, a short hydrogen bond was found between the inner oxygen of Asp32 and the inhibitory hydroxyl (Coates *et al.*, 2002). In contrast, in the PD-135,040 structure a different short hydrogen bond is formed between the outer oxygen of Asp32 and the outermost of the two inhibitory hydroxyls. In the *gem*-diol structure the hydrogen bond between the inner oxygen of Asp32 and the inner hydroxyl is stretched to around 2.86 Å, which is still acceptable for a normal hydrogen bond. The structure also shows that Asp32 is involved in four hydrogen bonds, one with Ser35 (2.53 Å), one with the amide of Gly34 (2.98 Å) and two others with the hydroxyl groups on the inhibitor (2.51, 2.86 Å).

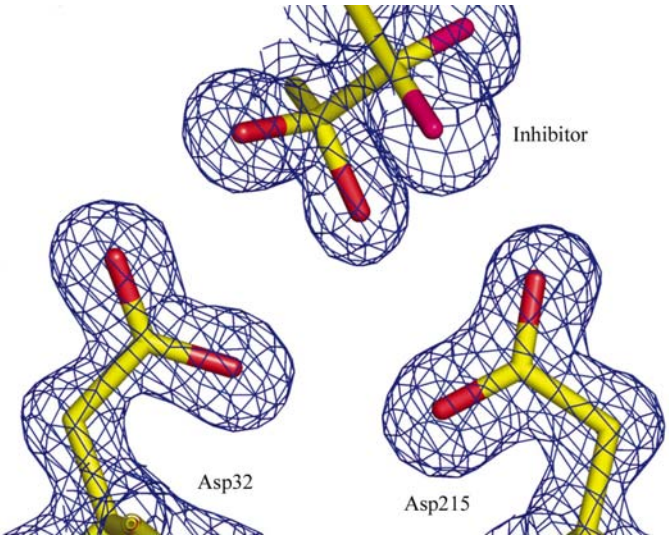


**Figure 4**
The $2mF_o - DF_c$ electron-density map contoured at 1.5 r.m.s. around the active site. The figure was created using *PyMol* (http://www.pymol.org).

This implies that it is negatively charged when a *gem*-diol inhibitor is bound at the active site, which is in agreement with a number of previous neutron and X-ray diffraction studies (Coates *et al.*, 2001, 2002). The outer oxygen of Asp215 is involved in a hydrogen bond (2.84 Å) with the side-chain oxygen of Thr218 and a short hydrogen bond (2.66 Å) to the inner inhibitory hydroxyl, whilst the inner oxygen forms a weak hydrogen bond with the amide group of Gly217 (3.15 Å).

## 4. Conclusions

X-ray crystallography has indicated the presence of a number of short hydrogen bonds within the active site of endothiapepsin complexed with a *gem*-diol inhibitor at high resolution. The lengths of these hydrogen bonds have been confirmed using one-dimensional H NMR. Two of these hydrogen bonds involve the outer oxygen of Asp32, while the inner oxygen does not participate in any short hydrogen bonds. This is a slightly different arrangement compared with that observed when a single hydroxyl inhibitory group is bound to the active site of endothiapepsin (Coates *et al.*, 2002). In this case, the short hydrogen bonds are formed between the inhibitory hydroxyl and the inner oxygen of Asp32. However, the hydrogen-bonding pattern in the active site still indicates that Asp32 is the negatively charged aspartate when a *gem*-diol inhibitor is bound to the active site.

We gratefully acknowledge the BBSRC (UK) for project grant support and the EPSRC for funding a research studentship (to LC). The ESRF (Grenoble, France) is thanked for the provision of X-ray beam time and support during data collection.

## References

Bailey, D. & Cooper, J. B. (1994). *Protein Sci.* **3**, 2129–2143.
Bailey, D., Cooper, J. B., Veerapandian, B., Blundell, T. L., Atrash, B., Jones, D. M. & Szelke, M. (1993). *Biochem. J.* **289**, 363–371.
Blundell, T. L., Jenkins, J. A., Sewell, B. T., Pearl, L. H., Cooper, J. B., Tickle, I. J., Veerapandian, B. & Wood, S. P. (1990). *J. Mol. Biol.* **211**, 919–941.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
Coates, L., Erskine, P. T., Crump, M. P., Wood, S. P. & Cooper, J. B. (2002). *J. Mol. Biol.* **318**, 1405–1415.
Coates, L., Erskine, P. T., Wood, S. P., Myles, D. & Cooper, J. B. (2001). *Biochemistry*, **40**, 13149–13157.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cooper, J. B. (2002). *Curr. Drug Targets*, **3**, 155–174
Fruton, J. S. (1976). *Adv. Enzymol. Relat. Areas Mol. Biol.* **44**, 1–36.
Hofmann, T., Dunn, B. M. & Fink, A. L. (1984). *Biochemistry*, **23**, 5247–5256.
James, M. N. G., Sielecki, A. R., Hayakawa, K. & Gelb, M. H. (1992). *Biochemistry*, **31**, 3872–3886.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **26**.
McDermott, A. & Ridenour, C. F. (1996). *Encyclopaedia of NMR*. New York: John Wiley.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Pearl, L. & Blundell, T. (1984). *FEBS Lett.* **174**, 96–101.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Rich, D. H., Bernatowicz, M. S. & Schmidt, P. G. (1982). *J. Am. Chem. Soc.* **104**, 3535–3536.
Sheldrick, G. M. (1998*a*). *Direct Methods for Solving Macromolecular Structures*, pp. 131–141. Oxford University Press.
Sheldrick, G. M. (1998*b*). *Direct Methods for Solving Macromolecular Structures*, pp. 401–411. Oxford University Press.
Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D. & Davies, D. P. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 7009–7013.
Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosatti, R. L., Dominy, B. W., Damon, D. B. & Hoover, D. J. (1992). *Protein Sci.* **1**, 322–328.