Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

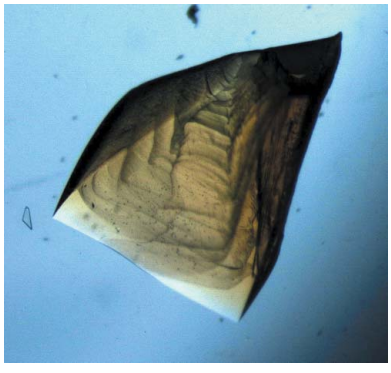
Preliminary neutron and ultrahigh-resolution X-ray diffraction studies of the aspartic proteinase endothiapepsin cocrystallized with a *gem*-diol inhibitor

Han-Fang Tuan,[a] Peter Erskine,[b] Paul Langan,[c,d] Jon Cooper[b] and Leighton Coates[a,d]*

[a]Spallation Neutron Source, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA, [b]Laboratory for Protein Crystallography, Centre for Amyloidosis and Acute Phase Proteins, UCL Department of Medicine (Hampstead Campus), Rowland Hill Street, London NW3 2PF, England, [c]Bioscience Division, Mailstop M888, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, and [d]Department of Chemistry, University of Toledo, Toledo, OH 53606, USA

Correspondence e-mail: coatesl@sns.gov

Received 3 October 2007
Accepted 20 November 2007



Endothiapepsin has been cocrystallized with the *gem*-diol inhibitor PD-135,040 in a low solvent-content (39%) unit cell, which is unprecedented for this enzyme–inhibitor complex and enables ultrahigh-resolution (1.0 Å) X-ray diffraction data to be collected. This atomic resolution X-ray data set will be used to deduce the protonation states of the catalytic aspartate residues. A room-temperature neutron data set has also been collected for joint refinement with a room-temperature X-ray data set in order to locate the H/D atoms at the active site.

1. Introduction

Endothiapepsin from *Endothia parasitica* is a member of the aspartic proteinase class of enzymes that are widely distributed among fungi, plants, vertebrates and viruses. A defining feature of this class is inhibition by the microbial peptide pepstatin, which contains the unusual amino acid statine (Bailey *et al.*, 1993). In the HIV retrovirus, the proteinase (PR) is essential for maturation of the virus particle (Cooper, 2002) and PR 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 are much sought after as potential therapeutic agents. Aspartic proteinases comprise two structurally similar domains or subunits, each of which contributes an aspartic acid residue to form a catalytic dyad that acts to cleave the substrate peptide bond (Blundell *et al.*, 1990). The fold of endothiapepsin is mainly β-sheet with small areas of α-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. Two catalytic Asp-Thr-Gly sequences are conserved in these enzymes (Pearl & Blundell, 1984), forming the active site, in which the carboxyl groups of the 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 water 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 substrate binding and is polarized by one of the aspartate carboxyls. The water then nucleophilically attacks the scissile-bond carbonyl group. The failure of numerous chemical studies to trap a covalently bound substrate supports the notion of a reaction mechanism that involves a noncovalently bound transition-state intermediate (Hofmann *et al.*, 1984).

Short oligopeptide inhibitors of aspartic proteinases bind in β-strand conformations in the active-site cleft between the two domains, which is around 30 Å in length. The best synthetic inhibitors are those that mimic one or both of the hydroxyls 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 hydroxyl

group alone. Three X-ray structures of endothiapepsin complexed with this class of inhibitor have been refined to atomic resolution (Coates *et al.*, 2002) and reveal a number of very short potential hydrogen bonds 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 suggested that it binds to the enzyme in a hydrated *gem*-diol form [$-C(OH)_2-CH_2-$] (Rich *et al.*, 1982). Current mechanistic proposals have been based on the X-ray structures of *gem*-diol transition-state analogues of the difluoroketone class [$-C(OH)_2-CF_2-$] (Veerapandian *et al.*, 1992; James *et al.*, 1992; Coates *et al.*, 2003). These structures have been refined at resolutions between 1.37 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 the charges on the active-site aspartates.

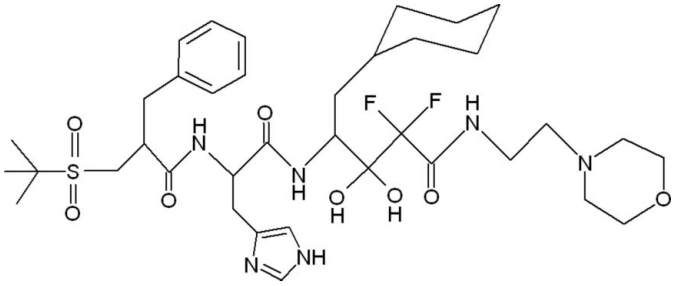


Figure 1
The chemical structure of PD-135,040 in the hydrated *gem*-diol form.

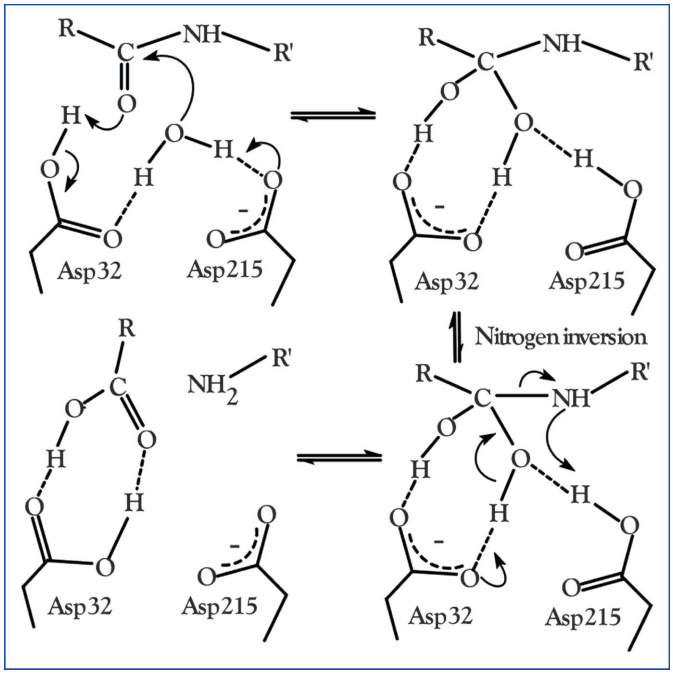


Figure 2
The catalytic mechanism proposed by Veerapandian *et al.* (1992), which is based on the X-ray structure of a difluoroketone (*gem*-diol) inhibitor bound to endothiapepsin. A water molecule tightly bound to the aspartates in the native enzyme is proposed to nucleophilically attack the scissile-bond carbonyl. The resulting tetrahedral intermediate (top right) 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.

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: the inner O atom of Asp32 or the outer O atom of Asp215. A proton in either of the other two positions, *i.e.* the outer O atom of Asp32 or the inner O atom of Asp215, would not have good enough geometry to form hydrogen bonds to the inhibitor hydroxyl. A neutron diffraction study of endothiapepsin complexed with the hydroxyethylene inhibitor H261 (Coates *et al.*, 2001) showed that the outer O atom of Asp215 was protonated with the inhibitor bound, suggesting that Asp32 is negatively charged in the transition-state complex. The inhibitor H261 mimics one of the two hydroxyls thought to exist in the transition state. Until recently, it has not been possible to grow large protein crystals of endothiapepsin with any *gem*-diol inhibitor that are suitable for neutron diffraction.

Here, we report the preparation of D_2O-exchanged crystals of endothiapepsin complexed with a *gem*-diol inhibitor that are suitable for neutron data collection. We also report preliminary X-ray and neutron data collection at room temperature and X-ray data collection to ultrahigh resolution at low temperature. These studies provide the prospect of locating the catalytic H atoms in a complex that will give an unprecedented view of the tetrahedral intermediate.

2. Materials and methods

2.1. Protein crystallization

The endothiapepsin–PD-135,040 cocrystals were grown at room temperature (298 K) by the batch method using 2 ml of a 2.0 mg ml^{-1} protein solution at pH 4.6 in 0.1 *M* sodium acetate buffer. A tenfold excess of the *gem*-diol inhibitor PD-135,040 was added to the crystallization solution. Ammonium sulfate was added to 55% saturation. A large (1.3 × 1.3 × 1.6 mm; Fig. 3) crystal and numerous smaller crystals (0.5 × 0.3 × 0.1 mm) appeared over a period of three years. The reasons for this extended crystallization period are not clear.

The larger protein crystal was mounted within a 3 mm diameter fused quartz capillary for room-temperature neutron data collection. Hydrogen–deuterium exchange was conducted by vapor diffusion within the capillary using a deuterated mother-liquor solution.

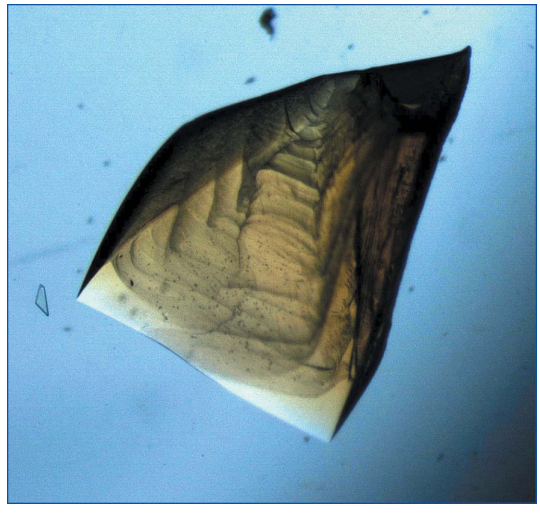


Figure 3
A large crystal of endothiapepsin cocrystallized with PD-135,040. The volume of this crystal is approximately 2.7 mm^3.

Table 1
Data-collection statistics.

Values in parentheses are for the outer resolution shell.

	Ultrahigh-resolution X-ray data set	Room-temperature X-ray data set	Neutron diffraction data set
Unit-cell parameters (Å, °)	a = 42.69, b = 74.34, c = 42.53, β = 97.5	a = 43.01, b = 75.71, c = 42.97, β = 97.0	a = 43.03, b = 75.72, c = 42.97, β = 97.0
Space group	$P2_1$	$P2_1$	$P2_1$
No. of unique reflections	135452 (14236)	36612 (4716)	20262 (2475)
Resolution range (Å)	74.33–1.0 (1.05–1.0)	18.24–1.56 (1.65–1.56)	42.72–1.80
Multiplicity	5.8 (2.8)	5.5 (5.1)	2.8 (1.9)
$I/\sigma(I)$	5.3 (10.3)	8.7 (2.4)	2.8 (1.8)
$R_{\rm merge}$† (%)	8.5 (7.3)	4.8 (26.0)	23.3 (36.7)
Data completeness (%)	95.5 (70.6)	94.3 (82.9)	80.3 (67.5)
Mosaicity (°)	0.3	0.2	0.3

† $R_{\rm merge}$ is the conventional merging R factor defined by $\sum_{hkl}\sum_i |I_i(hkl) - \overline{I(hkl)}|/\sum_{hkl}\sum_i I_i(hkl)$.

Interestingly, the protein crystallized in the lower solvent-content (39%) unit cell in comparison to the normal higher solvent-content unit cell (55%). This is unprecedented for this enzyme–inhibitor complex with these well established crystallization conditions and is presumed to be the primary reason for the increase in resolution over previous structures of the same protein–inhibitor complex (Coates *et al.*, 2003). The crystal was deuterated by equilibration with 90% D_2O mother liquor followed by 95% D_2O mother liquor prior to data collection.

2.2. X-ray and neutron data collection

The atomic resolution X-ray diffraction data were collected at the ESRF, Grenoble (beamline ID-29) at 100 K. A high-resolution pass of 180° of data was collected with an oscillation angle of 1°, a crystal-to-detector distance of 110 mm and an exposure time of 5 s (Fig. 4). A medium-resolution pass was also collected with an oscillation angle of 2° and a crystal-to-detector distance of 175 mm. For the low-resolution pass the crystal-to-detector distance was set to 230 mm and 1° oscillations were again used with an exposure time of 1 s. The beam was attenuated as appropriate for each data set using the in-line attenuation system.

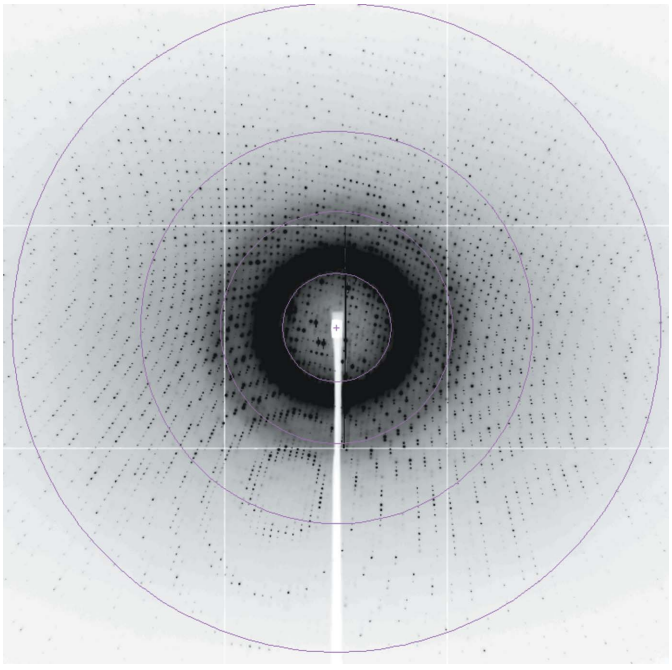


Figure 4
An X-ray diffraction image from the high-resolution data set. This image was collected for a high-resolution pass with an oscillation angle of 1°, a crystal-to-detector distance of 110 mm and an exposure time of 5 s. Several resolution shells are indicated by pink circles at (inner to outer) 4.3, 2.16, 1.43 and 1.08 Å.

In addition to the atomic resolution data set, a room-temperature X-ray data set was collected from one of the smaller crystals with an exposure time of 20 min per frame and a crystal-to-detector distance of 90 mm. In comparison to the atomic resolution X-ray data set, the room-temperature X-ray data set had unit-cell parameters that were similar to those of the room-temperature neutron data set. This will allow joint refinement of an X-ray and neutron structure using a newly adapted version of *CNS*, called *nCNS* (M. Mustyakimov, P. Afonone, P. D. Adams & P. Langan, Los Alamos National Laboratory), developed by the Macromolecular Neutron Crystallography (MNC) consortium (an NIH–NIGMS-funded consortium between LANL and LNBL to develop computational tools for neutron protein crystallography). Joint neutron and X-ray refinement increases the data-to-parameter ratio and allows a more accurate overall structure.

All X-ray data were processed using *MOSFLM* (Leslie, 1992) and scaled with *SCALA* (Collaborative Computational Project, Number 4, 1994). Initial data-processing statistics are shown in Table 1. Initial inspection of the difference electron-density maps confirms the binding of the inhibitor at the active site of the enzyme.

The larger protein crystal was screened for neutron diffraction quality at room temperature using the protein crystallography station (PCS) located at the Los Alamos Neutron Scattering Center (LANSCE; Langan *et al.*, 2004). An exposure time of 24 h produced diffraction extending to beyond 2.0 Å. Neutron data were processed using a modified version of *d*trek* (Langan & Greene, 2004). Initial data-collection statistics for the X-ray and neutron diffraction data sets are given in Table 1.

3. Results and discussion

Endothiapepsin was successfully cocrystallized with the *gem*-diol inhibitor PD-135,040 and the crystals belonged to space group $P2_1$. Two X-ray data sets were collected: one at room temperature (1.56 Å) and another at 100 K (1.0 Å). A room-temperature neutron data set was collected from a larger protein crystal for joint refinement with the room-temperature X-ray data set. The unit-cell parameters, data-collection and processing statistics for the X-ray data sets are given in Table 1. A rearrangement of Cruickshank's diffraction component precision index (DPI) formulae shows that data completeness can give a linear improvement of the DPI (Blow,

2002). The outer resolution shell of the ultrahigh-resolution data set has a data completeness of 70.6%. The primary reason for this is that even with the minimal crystal-to-detector distance, reflections of ∼1 Å are on the edge of the detector (Fig. 4). Hence, the outer-shell data have an $I/\sigma(I)$ as high as 10.3. It is hoped that the missing data in the outer shell can be collected on another beamline in order to improve the DPI.

The earlier study of endothiapepsin complexed with H261 (Coates *et al.*, 2001) was conducted on the reactor-based instrument LADI. In contrast, the current data collection was conducted on a spallation source-based instrument. Spallation-source instrumentation reduces reflections in spatial and harmonic overlaps, giving an improved signal-to-noise ratio for recorded reflections (Langan *et al.*, 2004).

4. Conclusions

High-quality atomic resolution and room-temperature X-ray data sets have been collected from endothiapepsin complexed with the *gem*-diol inhibitor PD-135,040. The atomic resolution data set will allow analysis of the protonation states of the catalytic aspartate residues based on differences in carboxyl bond lengths, while the neutron data set will be used to examine the neutron density at the active site in order to investigate protonation states.

This research was sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory (ORNL), managed by UT-Battelle LLC for the US Department of Energy under Contract No. DE-AC05-00OR22725. The PCS is funded by the Office of Biological and Environmental Research of the Department of Energy. PL was supported by an NIH–NIGMS-funded consortium (1R01GM071939-01) between LANL and LNBL to develop computational tools for neutron protein crystallography and a LANL LDRD grant (20070131ER). We also are grateful to the ESRF (Grenoble, France) for beamtime and travel support.

References

Bailey, D., Cooper, J. B., Veerapandian, B., Blundell, T. L., Atrash, B., Jones, D. M. & Szelke, M. (1993). *Biochem. J.* **289**, 363–371.
Blow, D. M. (2002). *Acta Cryst.* D**58**, 792–797.
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.
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., Mall, S., Williams, P. A., Gill, R. S., Wood, S. P. & Cooper, J. B. (2003). *Acta Cryst.* D**59**, 978–981.
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.
Langan, P. & Greene, G. (2004). *J. Appl. Cryst.* **37**, 253–257.
Langan, P., Greene, G. & Schoenborn, B. P. (2004). *J. Appl. Cryst.* **37**, 24–31.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **26**.
Pearl, L. & Blundell, T. (1984). *FEBS Lett.* **174**, 96–101.
Rich, D. H., Bernatowicz, M. S. & Schmidt, P. G. (1982). *J. Am. Chem. Soc.* **104**, 3535–3536.
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.