Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallographic evidence for noncoplanar catalytic aspartic acids in plasmepsin II resides in the Protein Data Bank

Arthur H. Robbins, Ben M. Dunn, Mavis Agbandje-McKenna and Robert McKenna*

Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, FL 32610, USA

Correspondence e-mail: rmckenna@ufl.edu

Received 17 November 2008
Accepted 8 December 2008

PDB Reference: plasmepsin II, 3f9q, r3f9qsf.

The carboxylate atoms of the two catalytic aspartic acid residues in aspartic proteases are nearly coplanar and in the uncomplexed form share an in-plane nucleophilic water molecule that is central to the mechanism of these enzymes. This note reports that while reviewing the electron-density maps derived from the deposited data for uncomplexed plasmepsin II from *Plasmodium falciparum* [Asojo *et al.* (2003), *J. Mol. Biol.* **327**, 173–181; PDB code 1lf4], it was discovered that the aspartic acid residues in this structure should in fact be distinctly non-coplanar. The crystallographic model from the deposited coordinates has been re-refined against the 1.9 Å resolution published diffraction data to an $R_{\rm cryst}$ of 21.2% and an $R_{\rm free}$ of 22.2%. The catalytic water molecule is present, but the plane of the carboxylate group of Asp214 is rotated by 66° from its original position.

Aspartic proteases are ubiquitous enzymes in all living organisms and contain two aspartic acid residues in their active sites, forming the so-called catalytic dyad. All aspartic proteases are characterized by two conserved DT(S)G sequences and fold into primarily β structures. These enzymes are involved in digestion, regulation of blood pressure and the activation of their own proenzymes and as such have been identified as therapeutic targets (Eder *et al.*, 2007).

One of the simplest of the aspartic proteases, HIV-1 protease, consists of two identical monomers, each of which contributes one of the aspartic acid residues responsible for catalysis. Two 'flaps', one from each monomer, overlap at the tips and close down upon an inhibitor (or presumably substrate) to form a tightly bound complex. Hundreds of examples of these complexes have been deposited in the Protein Data Bank (PDB; Berman *et al.*, 2000), dating back to those of Wlodawer and coworkers (Miller *et al.*, 1989). In aspartic proteases from higher organisms, a single polypeptide folds into two domains: the N-terminal domain contains the single flap region and the N-terminal and C-terminal domains each furnish one of the catalytic aspartic acids.

After an extensive review of the deposited aspartic protease crystal structures in the PDB, it is of note that the carboxylate atoms of the catalytic dyad are located in nearly the same plane as the putative nucleophilic water molecule. Most aspartic protease inhibitors are designed to replace the scissile peptide of a pseudo-substrate with various tetrahedral intermediate mimics, which often displace the nucleophilic water with a hydroxyl O atom.

In fact, to our knowledge no crystal structure of an active aspartic protease has yet displayed noncoplanar carboxylate atoms, either when uncomplexed or in complexes with natural or synthetic inhibitors. Indeed, in the proposed catalytic mechanism, the interactions between the aspartates, the surrounding amino acids and the water molecule are thought to require the planar configuration (Northrop, 2001).

The prediction of an ultrashort low-barrier hydrogen bond (LBHB) between the inner O atoms of the catalytic dyad was reported about a decade ago (Piana & Carloni, 1999; Cleland, 2000; Northrop, 2001). The prediction of the LBHB was followed by a series of papers from Cooper and coworkers, who used high-resolution crystal structures (Coates *et al.*, 2002, PDB codes 1gvt,

Table 1
Refinement statistics of the uncomplexed PM2 structure.

These values were obtained from the deposited PDB data.

	1lf4	3f9q
Reflections used (no. used for R_{free})	28457 (1443)	28457 (1443)
R_{cryst}† (%)	21.7	21.2
R_{free}‡ (%)	22.8	22.2
Model statistics		
No. of protein atoms	2570	2569
No. of water atoms	290	213
Mean *B* factors (Å^2)		
Protein atoms	33.1	31.9
Water atoms	45.1	37.5
R.m.s. deviations		
Bond distances (Å)	0.005	0.006
Angles (°)	1.4	1.4
Ramachandran plot (%)		
Favored	87.9	86.9
Allowed	11.4	13.1
Generously allowed	0.4	0.0
Disallowed	0.4	0.0

† $R_{cryst} = (\sum |F_o| - |F_c| / \sum |F_o|) \times 100$. ‡ R_{free} is calculated in the same way as R_{cryst}, except that it uses 5% of reflection data that were omitted from refinement.

1gvu, 1gvv, 1gvw and 1gvx; Erskine *et al.*, 2003, PDB codes 1oew and 1oex; Coates *et al.*, 2003, PDB code 1od1; Coates *et al.*, 2008, PDB code 2jjj), proton nuclear magnetic resonance (Coates *et al.*, 2002) and neutron diffraction (Coates *et al.*, 2001, PDB code 1gkt; Coates *et al.*, 2008, PDB code 2vs2) to determine the proton positions on the catalytic aspartates in endothiapepsin. In all cases, the proton resides on the outer carboxylate O atom of the second catalytic aspartate in the amino-acid sequence and this proton is involved in a hydrogen bond to a hydroxyl O atom of many inhibitors. All of the cited X-ray crystal structures, in addition to the structure of *Irpex lacteus* aspartic protease (Fujimoto *et al.*, 2004; PDB code 1wkr), were determined at resolutions higher than 1.4 Å and all reinforce the concept of coplanar catalytic residues in endothiapepsin.

Hence, during a routine survey of PDB-deposited structures it was surprising to find crystallographic evidence that the crystal structure of the M205S mutant of uncomplexed plasmepsin II (PM2) from *Plasmodium falciparum* (Asojo *et al.*, 2003; PDB code 1lf4) should have been refined with noncoplanar carboxylates in its catalytic dyad. The Met-to-Ser mutation creates an enzyme that lacks an autocatalytic cleavage site and yet had no effect on the catalytic potential of the active site. It is revealing to note that the original authors also neglected to make the methionine-to-serine change in the 1lf4 deposited model.

Atomic coordinates and $2|F_o - F_c|$ and $|F_o - F_c|$ difference electron-density map calculated with model phases were obtained from the EDS website (Kleywegt *et al.*, 2004) and were displayed using *Coot* (Emsley & Cowtan, 2004). The environment near Asp214 contained pairs of $|F_o - F_c|$ peaks contoured at $\pm 3.5\sigma$, one pair positive and one pair negative (Fig. 1*a*). Hence, the side chain of Asp214 was modeled into the positive density, the M205S mutation was built and a limited number of other amino acids and solvent molecules were refitted into the electron density. The model then was re-refined against the 1.9 Å resolution deposited X-ray data using *CNS* (Brünger *et al.*, 1998; Brunger, 2007) with Engh and Huber geometric restraints (Engh & Huber, 1991). The R_{cryst} and R_{free} values did not change appreciably based upon these minor adjustments of the molecular model (Table 1). Least-squares planes and root-mean-square deviations (r.m.s.d.s) were calculated with the *MOLEMAN*2 software (Kleywegt, 1999). Only the four side-chain atoms of each aspartate residue and the nucleophilic water molecule (nine atoms) were used to calculate the planes.

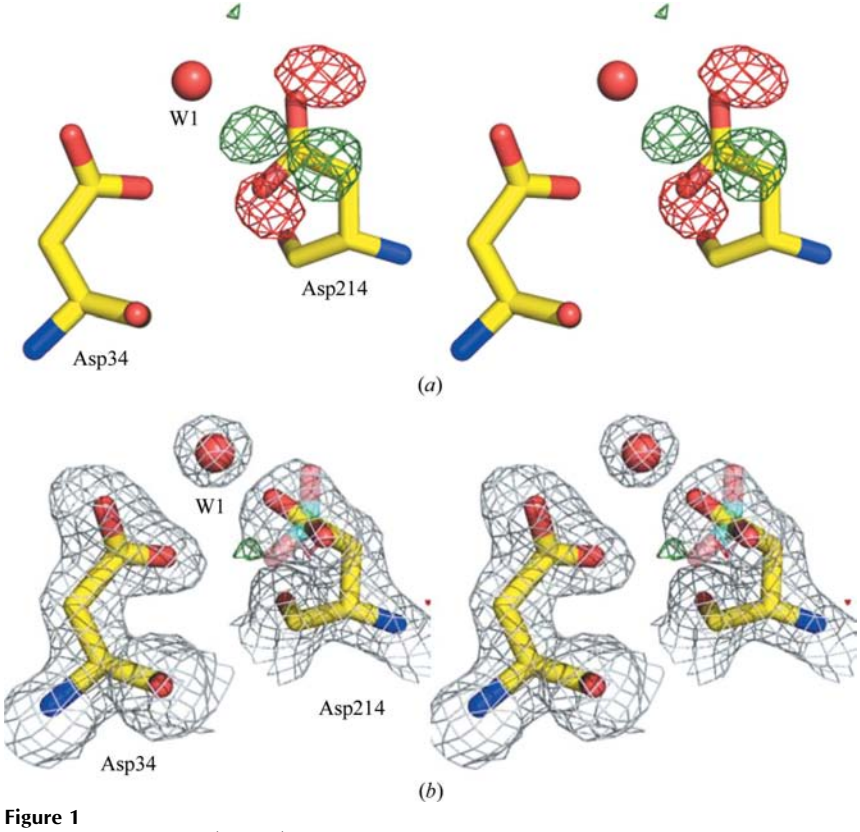

Figure 1
Catalytic dyad. (*a*) Stereo $|F_o - F_c|$ electron-density map calculated from the crystal structure of the M205S mutant of uncomplexed plasmepsin II from *P. falciparum* (Asojo *et al.*, 2003; PDB code 1lf4), contoured at $\pm 3.5\sigma$. Note the positive (green) and negative (red) pairs of electron-density difference peaks for Asp214. (*b*) The re-refined structure shows the $|2F_o - F_c|$ electron-density map (gray) contoured at 1.5σ and the recalculated $|F_o - F_c|$ difference electron-density map contoured at $\pm 3.5\sigma$ (as above). The deposited 1lf4 coordinates of Asp214 (transparent cyan) are superposed onto the re-refined structure. Note the approximately 66° rotation of the Asp214 carboxylate group with respect to that in the deposited 1lf4 model. This figure was generated using *PyMOL* (DeLano, 2002).

Upon inspection of the new difference electron-density maps calculated from the resulting coordinates, it was clear that the significant $|F_o - F_c|$ electron-density difference peaks at the Asp214 side chain had been removed (Fig. 1*b*) and that there was now an approximately 66° rotation of the Asp214 carboxylate group with respect to that in the deposited 1lf4 model. The nucleophilic water molecule (O29 in 1lf4 and this work), often referred to as W1, maintains hydrogen bonds to both OD1 and OD2 of Asp214 and the inner O atom OD2 of Asp34.

In aspartic proteases, a number of hydrogen bonds are seen from the surrounding amino acids to the catalytic aspartate OD1 and OD2 atoms. These interactions help to define the coplanarity of the catalytic aspartic acids. Using the atomic resolution crystal structure of endothiapepsin (Erskine *et al.*, 2003; PDB code 1oew), it can be seen that Asp218 OD1, the 'outer' O atom, forms hydrogen bonds to the catalytic water molecule W1 and Thr221 OG1. In the present work, rota-

tion of the carboxylate group causes the corresponding Asp214 OD1 to form hydrogen bonds to W1 and a second water molecule, O17. The OD2 in the 1oew structure forms hydrogen bonds to W1, the main-chain amide N atom of Gly220 and a weak hydrogen bond to Gly37 N. In the re-refined structure, OD2 also forms a hydrogen bond to W1, but the bond to Gly216 N is now a weak interaction with poor geometry; however, it is augmented by a third hydrogen bond to Thr217 OG1.

In order to compare the coplanarity of the catalytic dyad in the re-refined structure to those in 1lf4 and other uncomplexed aspartic proteases, least-squares planes were calculated for endothiapepsin (Erskine *et al.*, 2003; PDB code 1oew), porcine pepsin (Sielecki *et al.*, 1990; PDB code 4pep) and feline immunodeficiency virus (Laco *et al.*, 1997; PDB code 2fiv). The values for the r.m.s.d.s from the planes for all nine atoms were 0.16 Å for 1oew, 0.18 Å for 1lf4, 0.19 Å for 4pep, 0.52 Å for 3f9q (this work) and 0.63 Å for 2fiv. It should be noted that the FIV protease was an inactive mutant in which both catalytic aspartic acids were replaced by asparagines.

In conclusion, the crystal structure of uncomplexed PM2 from *P. falciparum* has been re-refined and as a result the catalytic aspartic acid carboxyl groups are clearly shown to be noncoplanar. In order to fully understand the cause of this phenomenon, additional uncomplexed plasmepsin crystal structures are needed. It will be interesting to learn whether this noncoplanarity is limited to one or more of the plasmepsins or may yet be found in other uncomplexed aspartic proteases.

A referee of this work succinctly stated that 'refinement should not be driven by the expectations, but rather by the reality of the data'; we fully agree with this reviewer's statement.

References

Asojo, O. A., Gulnik, S. V., Afonina, E., Yu, B., Ellman, J. A., Haque, T. S. & Silva, A. M. (2003). *J. Mol. Biol.* **327**, 173–181.

Berman, H. M., Westbook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Brunger, A. T. (2007). *Nature Protoc.* **2**, 2728–2733.

Cleland, W. W. (2000). *Arch. Biochem. Biophys.* **382**, 1–5.

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. A. A. & Cooper, J. B. (2001). *Biochemistry*, **40**, 13149–13157.

Coates, L., Tuan, H.-F., Tomanicek, S., Kovalevsky, A., Mustyakimov, M., Erskine, P. & Cooper, J. (2008). *J. Am. Chem. Soc.* **130**, 7235–7237.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, Palo Alto, California, USA. http://www.pymol.org.

Eder, J., Hommel, U., Cumin, F., Martoglio, B. & Gerhartz, B. (2007). *Curr. Pharm. Des.* **13**, 271–285.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Erskine, P. T., Coates, L., Mall, S., Gill, R. S., Wood, S., Myles, D. A. A. & Cooper, J. B. (2003). *Protein Sci.* **12**, 1741–1749.

Fujimoto, Z., Fugii, Y., Kaneko, S., Kobayashi, H. & Mizuno, H. (2004). *J. Mol. Biol.* **341**, 1227–1236.

Kleywegt, G. J. (1999). *Acta Cryst.* D**55**, 1878–1884.

Kleywegt, G. J., Harris, M. R., Zou, J., Taylor, T. C., Wählby, A. & Jones, T. A. (2004). *Acta Cryst.* D**60**, 2240–2249.

Laco, G. S., Schalk-Hihi, C., Lubkowski, J., Morris, G., Zdanov, A., Olson, A., Elder, J. H., Wlodawer, A. & Gustchina, A. (1997). *Biochemistry*, **36**, 10696–10708.

Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. & Wlodawer, A. (1989). *Science*, **246**, 1149–1152.

Northrop, D. B. (2001). *Acc. Chem. Res.* **34**, 790–797.

Piana, S. & Carloni, P. (1999). *Proteins*, **39**, 26–36.

Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S. & James, M. N. (1990). *J. Mol. Biol.* **214**, 143–170.