

# Novel dimer structure of a membrane-bound protease with a catalytic Ser–Lys dyad and its linkage to stomatin

Hideshi Yokoyama,<sup>a</sup> Shiho Hamamatsu,<sup>a</sup> Satoshi Fujii<sup>a</sup> and Ikuo Matsui<sup>b\*</sup>

<sup>a</sup>School of Pharmaceutical Sciences, University of Shizuoka, 52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan, and <sup>b</sup>Biological Information Research Center, National Institute of Advanced Industrial Science and Technology (AIST), 1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan. E-mail: ik-matsui@aist.go.jp

Received 30 July 2007

Accepted 27 December 2007

Membrane-bound proteases are involved in various regulatory functions. A previous report indicates that the N-terminal region of PH1510 (1510-N) from the hyperthermophilic archaeon *Pyrococcus horikoshii* is a serine protease with a catalytic Ser–Lys dyad (Ser97 and Lys138), and specifically cleaves the C-terminal hydrophobic region of the p-stomatin PH1511. According to the crystal structure of the wild-type 1510-N in dimeric form, the active site around Ser97 is in a hydrophobic environment suitable for the hydrophobic substrates. This article reports the crystal structure of the K138A mutant of 1510-N at 2.3 Å resolution. The determined structure contains one molecule per asymmetric unit, but 1510-N is active in dimeric form. Two possible sets of dimer were found from the symmetry-related molecules. One dimer is almost the same as the wild-type 1510-N. Another dimer is probably in an inactive form. The L2 loop, which is disordered in the wild-type structure, is significantly kinked at around A-138 in the K138A mutant. Thus Lys138 probably has an important role on the conformation of L2.

© 2008 International Union of Crystallography  
Printed in Singapore – all rights reserved

**Keywords:** membrane-bound protease; stomatin; ClpP; dimers; *Pyrococcus horikoshii*.

## 1. Introduction

Membrane-bound proteases play several important roles in protein quality control and regulation. Mitochondrial membranes possess quality control systems that remove non-assembled polypeptides and prevent their possibly deleterious accumulation in the membrane (Arnold & Langer, 2002). Several transcription factors are synthesized as membrane proteins and activated upon the specific proteolytic cleavage of solvent-exposed fragments that subsequently enter the nucleus (Brown *et al.*, 2000). Elucidation of these membrane proteases will lead to a better understanding of crucial biological processes and diseases.

Stomatin is one of the major integral membrane proteins of human erythrocytes, but its function is not fully understood. In humans, the absence of the protein is associated with a form of hemolytic anemia known as hereditary stomatocytosis (Stewart *et al.*, 1993). Stomatin-like proteins are found in almost all species of eukaryotes, eubacteria and archaea (Tavernarakis *et al.*, 1999). The prokaryotic members of the stomatin family are called p-stomatin. Interestingly, *p-stomatin* genes and *stopp* (stomatin operon partner protein) genes probably form an operon in at least 19 species of prokaryotes and archaea (Green *et al.*, 2004). In the genomic sequences of the hyperthermophilic archaeon *Pyrococcus horikoshii*, PH1510 and PH1511 correspond to STOPP and p-stomatin, respectively. The N-terminal region of PH1510 (residues 16–236, 1510-N) is a thermostable serine protease with a catalytic Ser–Lys dyad (Ser97 and Lys138), and it

specifically cleaves the C-terminal hydrophobic region of the p-stomatin PH1511 (Yokoyama & Matsui, 2005). The cleavage is involved in a certain regulatory system, but little is known about its physiological role. In order to understand the functions of 1510-N and PH1511, the crystal structure of 1510-N was determined at 3.0 Å resolution in dimeric form (Yokoyama *et al.*, 2006). The 1510-N protease preferentially cleaves hydrophobic substrates. According to the determined crystal structure, each active site around Ser97 of 1510-N is in a hydrophobic environment suitable for hydrophobic substrates. The monomer of 1510-N shows a structural similarity to one monomer of *Escherichia coli* ClpP (Wang *et al.*, 1997). ClpP is a protease with a catalytic Ser–His–Asp triad and degrades misfolded proteins in order to maintain quality control of proteins. Despite the structural similarity between the monomers of 1510-N and ClpP, their oligomeric forms are different. The difference in oligomeric form and catalytic residues between 1510-N and ClpP would lead to a functional difference; 1510-N is likely to have a regulatory function, whereas ClpP is involved in protein quality control.

It is not known how 1510-N recognizes and degrades the p-stomatin PH1511. A structural analysis of 1510-N in complex with its substrate would be needed, but no structures have been determined. This article reports a crystal structure of the K138A mutant of 1510-N at a resolution 2.3 Å higher than that of the wild-type structure reported previously (Yokoyama *et al.*, 2006). Although the structure does not contain a substrate peptide, a novel dimer structure was observed.

## 2. Materials and methods

### 2.1. Crystallization and diffraction data collection

The K138A mutant of 1510-N was prepared as described previously (Yokoyama & Matsui, 2005; Yokoyama *et al.*, 2006). For the structure determination of 1510-N in complex with its substrates, crystallization attempts were carried out using 1510-N K138A mixed with a 16-aa synthetic peptide containing the sequence of PH1511 degraded by 1510-N protease (its amino acid sequence is DKSNVIVLMLPMEMLK). The protein solution contains 5 mg ml<sup>-1</sup> of K138A and 5 mg ml<sup>-1</sup> of peptide in 50 mM Tris-HCl (pH 8.5) containing 0.1 M NaCl, 0.05% (w/v) *N,N*-dimethyldodecylamine *N*-oxide (LDAO, Fluka) and 10% (v/v) dimethyl sulfoxide. Crystallization trials were carried out using the hanging-drop vapor-diffusion method by mixing equal volumes of the protein and reservoir solutions. Crystals were grown at 293 K using a reservoir solution containing 18% (w/v) PEG 4000, 0.1 M MgCl<sub>2</sub> and 0.1 M sodium acetate (pH 4.5). Cubic crystals grew to an approximate size of 0.2 mm on a side.

A crystal was flash-frozen in a 100 K stream of nitrogen gas. The cryoprotectant solution used was 20% (w/v) PEG 4000, 25% (w/v) sucrose, 0.1 M MgCl<sub>2</sub> and 0.1 M sodium acetate (pH 4.5), with Tris-HCl (pH 7.5) added to a final concentration of 0.1 M. The crystal diffracted synchrotron X-rays to 2.3 Å resolution. The asymmetric unit contains one 1510-N K138A molecule, showing a high value for the crystal volume per unit molecular mass of 3.6 Å<sup>3</sup> Da<sup>-1</sup> and a solvent content of 0.66. X-ray diffraction data were collected at beamline BL6A of the Photon Factory in KEK (Tsukuba, Japan) using a Quantum-4R CCD detector, and were processed and scaled using *HKL2000* (Otwinowski & Minor, 1997).

### 2.2. Structure determination

The structure was determined by the molecular replacement method using the program *CNS* (Brünger *et al.*, 1998). The crystal structure of chain *B* of 1510-N [Protein Data Bank (PDB) accession code 2deo] was used as a search model (Yokoyama *et al.*, 2006). After rotation and translation searches, the model was subjected to rigid-body refinement with 20–3.0 Å data using the program *REFMAC* (Murshudov *et al.*, 1997) in the *CCP4* suite (Collaborative Computational Project, Number 4, 1994), and then the model gave an *R* factor of 0.385. The model was subjected to several cycles of crystallographic refinement using *REFMAC*, followed by manual model fitting using the program *COOT* (Emsley & Cowtan, 2004). Data collection and refinement statistics are summarized in Table 1. The buried surface area was calculated using *CNS* (Brünger *et al.*, 1998) with a probe radius of 1.4 Å. Structural geometries were analyzed using *PROCHECK* (Laskowski *et al.*, 1993). Figures were produced using *PyMOL* (<http://www.pymol.org>), *Molscript* (Kraulis, 1991) and *Raster3D* (Merritt & Bacon, 1997). The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank with the accession code 3bpp.

## 3. Results and discussion

### 3.1. Structure determination

The crystal structure of the K138A mutant of 1510-N was solved at 2.3 Å resolution using the molecular replacement method. Although diffraction data were collected using the cocrystal of 1510-N K138A and the substrate peptide, no electron densities corresponding to the peptide were observed. The refined model contains residues 20–235 and 63 water molecules. Clear and continuous electron densities are

**Table 1**

Data collection and refinement statistics.

Data collection	
Space group	I422
Cell dimensions (Å)	<i>a</i> = 106.3, <i>c</i> = 129.0
Wavelength (Å)	1.0000
Resolution range (Å)	20–2.3 (2.38–2.30)†
<i>R</i> <sub>merge</sub> ( <i>I</i> )‡	0.049 (0.330)
Average <i>I</i> /σ <i>I</i>	71.8 (11.9)
No. unique reflections	16,792
Redundancy	14.1
Completeness (%)	99.9 (100)
Refinement	
Resolution range (Å)	20–2.3
No. reflections used	15063
Completeness (%)	99.8
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> ¶	0.237/0.310
Number of atoms	
Protein	1663
Water	63
Average <i>B</i> factors (Å <sup>2</sup> )	
Protein	47.6
Water	45.3
Main-chain torsion angles (%) in	
Most favored regions	87.3
Additionally allowed regions	11.1
Generously allowed regions	1.1
Disallowed regions	0.5
r.m.s. deviations	
Bond lengths (Å)	0.015
Bond angles (°)	1.507

† Values in parentheses are for the highest-resolution shell. ‡  $R_{\text{merge}}(I) = \sum |I - \langle I \rangle| / \sum I$ , where *I* is the observed diffraction intensity. §  $R = \sum |F_o - F_c| / \sum F_o$ , where *F*<sub>o</sub> and *F*<sub>c</sub> are observed and calculated structure amplitudes, respectively. ¶ *R*<sub>free</sub> is an *R* value for 10% of the reflections chosen randomly and omitted from the refinement.

**Table 2**

The structural difference between the wild type and the K138A mutant.

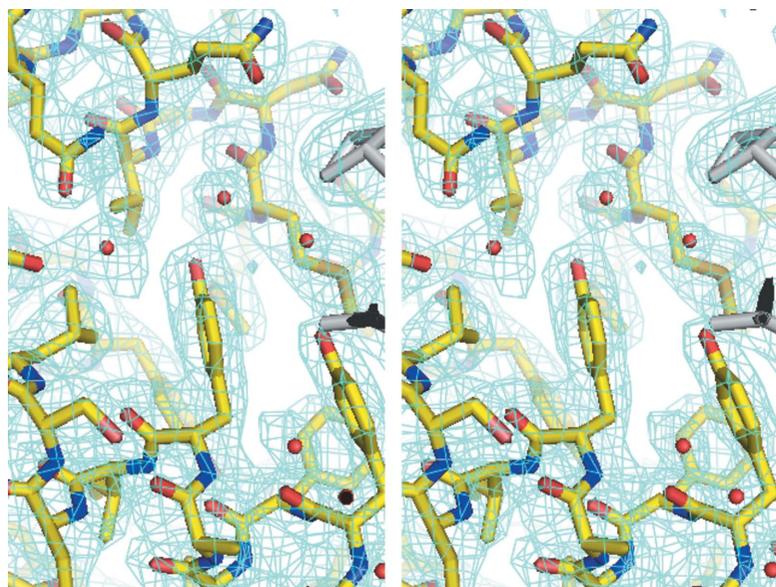
	Wild type	K138A
PDB accession code	2deo	3bpp
Space group	C2	I422
Molecules per asymmetric unit	2	1
Resolution (Å)	3.0	2.3
BSA† of type-1 dimer (Å <sup>2</sup> )	1217	1180 (1427)‡
BSA of type-2 dimer (Å <sup>2</sup> )	917	1498 (1870)

† Buried surface area of the dimer calculated with water molecules omitted. ‡ Values in parentheses are BSAs calculated with water molecules included.

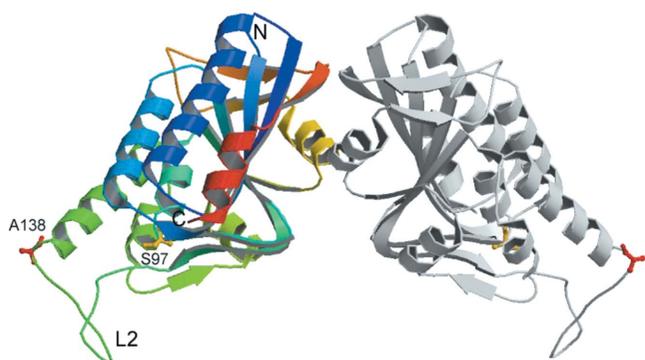
observed for all the residues except for the residues 127–129 with low densities. The 2*F*<sub>o</sub> – *F*<sub>c</sub> electron densities around Tyr146 are shown in Fig. 1. As shown in Table 1, the stereochemistry of the model was evaluated using *PROCHECK*. One residue, Ser97, is in a disallowed region ( $\varphi = 24^\circ$ ,  $\psi = -107^\circ$ ). Ser97 is the catalytic residue and is located as a second residue of a type II' β turn (Hutchinson & Thornton, 1994). ClpP protease shows a similar active-site structure to 1510-N, and the residue (Ser97) corresponding to Ser97 of 1510-N is also in a disallowed region (Wang *et al.*, 1997), as well as that of wild-type 1510-N (Yokoyama *et al.*, 2006).

### 3.2. Overall structure and dimer formation

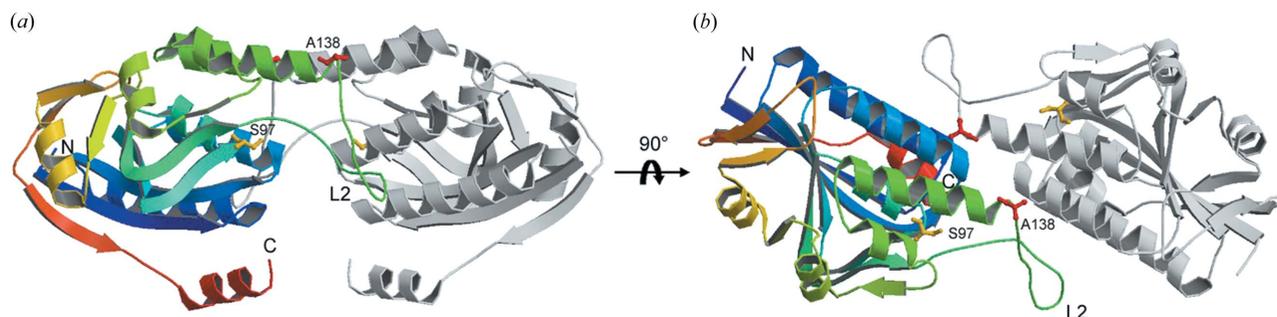
The determined structure of the 1510-N K138A mutant contains one monomer in the asymmetric unit. In contrast, the wild-type 1510-N has two molecules in the asymmetric unit (Yokoyama *et al.*, 2006). In the SDS-PAGE analysis of 1510-N, a major protein band at around 45 kDa (putative dimer) showed proteolytic activity (Yokoyama & Matsui, 2005). From the gel filtration analysis, the expected molecular mass was 40–50 kDa, indicating that 1510-N forms a dimer (data not shown). These results strongly suggest that 1510-N acts as a dimer.



**Figure 1**  
Stereoview of the  $2F_o - F_c$  electron densities contoured at  $1.0\sigma$ . Each residue around Tyr146 (center) is shown as a stick model (C atoms, yellow; N atoms, blue; O atoms, red; S atoms, orange). Water molecules are shown as red spheres. Grey sticks show the residues of a symmetry-related molecule.



**Figure 2**  
Ribbon representation of 1510-N K138A. A monomer is colored in a rainbow ramp from purple at the N-terminus to red at the C-terminus. The catalytic residue Ser97 is drawn as yellow sticks and Ala138 is drawn as red sticks. L2 corresponds to the disordered loop in the 1510-N wild-type structure (Yokoyama *et al.*, 2006). 'N' and 'C' denote the N and C termini, respectively. One symmetry-related molecule is also shown in grey. A crystallographic twofold axis relating two molecules runs vertically in the figure plane. The dimer (type 1) was observed in the 1510-N wild-type structure (Yokoyama *et al.*, 2006).



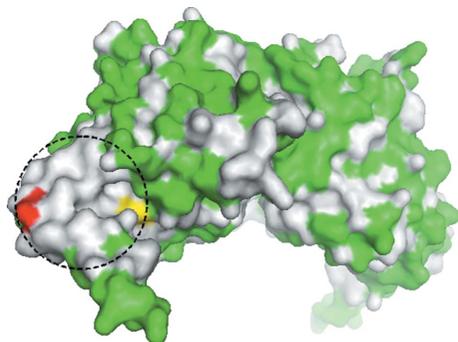
**Figure 3**  
The other dimer form (type 2). A monomer is colored as in Fig. 2. The symmetry-related molecule is shown in grey. (a) Side view of the dimer. A crystallographic twofold axis relating two molecules runs vertically in the figure plane. (b) Top view of the dimer.

The crystal of 1510-N K138A belongs to space group *I*422, and one molecule makes contact with five symmetry-related molecules. The buried surface area (BSA) between one monomer and each symmetry-related molecule was calculated, and two large BSA values above  $1000 \text{ \AA}^2$  were found (Table 2). Thus two possible sets of dimer were observed (type 1 in Fig. 2 and type 2 in Fig. 3), and both dimers are related by a crystallographic twofold axis. As shown in Table 2, the BSA for the type-1 dimer is  $1180 \text{ \AA}^2$ , and the dimer (Fig. 2) exhibits almost the same structure as the wild-type dimer (Yokoyama *et al.*, 2006). For the type-1 dimer, the r.m.s. difference of C $\alpha$  atoms of the K138A and wild-type dimers is  $1.7 \text{ \AA}$ . The r.m.s. difference of the K138A monomer and the wild-type chain B is  $1.5 \text{ \AA}$ . The relatively large difference is derived from large deviation of the regions around Arg66 (which corresponds to the L1 loop in the wild-type structure) and Arg121 (located at the base of the L2 loop). On the other hand, the BSA for the type-2 dimer of the K138A mutant is  $1498 \text{ \AA}^2$ . In the structure, the L2 loop mainly constitutes the dimer interface, as shown in Fig. 3 (among 20 residues forming the dimer interface, nine residues belong to the L2 loop). In the K138A mutant, the novel type-2 dimer ( $1498 \text{ \AA}^2$ ) shows a buried surface area larger than the type-1 dimer ( $1180 \text{ \AA}^2$ ). In wild-type 1510-N, the corresponding type-2 dimers are formed using symmetry-related molecules, but the r.m.s. difference of C $\alpha$  atoms of the wild-type and K138A dimers is large ( $4.1 \text{ \AA}$ ). In wild-type 1510-N, the residues 122–139 corresponding to the L2 loop are disordered and are not involved in the formation of the dimer interface. Thus the BSA of the type-2 dimer ( $917 \text{ \AA}^2$ ) is narrower than that of type 1 ( $1217 \text{ \AA}^2$ ) in the wild type.

### 3.3. Possible dimer interface

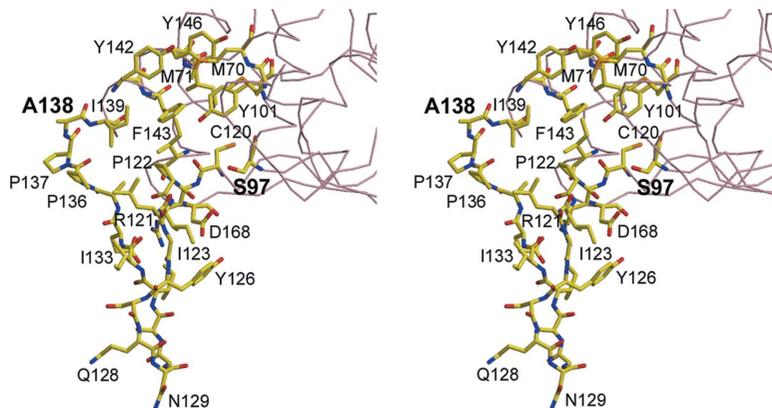
In the K138A structure, the molecular surface representation indicates that the active site around S97 and A138 (mutated from Lys) is in a hydrophobic environment (Fig. 4). In particular, the aromatic residues Tyr101, Tyr142, Phe143 and Tyr146 form a cluster along with Met70 and Met71 (Fig. 5). The hydrophobic region is probably the binding site of hydrophobic substrates, and 1510-N preferentially degrades hydrophobic substrates in the region.

The type-1 dimer exposes the hydrophobic active site to the solvent or the membrane (Fig. 4). Meanwhile, the type-2 dimer uses the hydrophobic active site to form the dimeric interface (Fig. 3). Then the type-2 dimer is blocked off to the solvent or the membrane, and therefore it would be inactive. The 1510-N protease specifically degrades the C-terminal hydrophobic region of PH1511, 1511-C



**Figure 4**

Molecular surface of the type-1 dimer of 1510-N K138A viewed from almost the same direction as Fig. 2. Polar residues (Asp, Glu, His, Lys, Arg, Gly, Ser, Thr, Cys, Asn and Gln) are shown in green, and hydrophobic and aromatic residues are in white. Ser97 is shown in yellow and Ala138 is in red. The dashed circle indicates the hydrophobic region around the active site.



**Figure 5**

Stereo representation of 1510-N K138A viewed from almost the same direction as Fig. 2. Ser97, Ala138, the L2 loop and neighboring residues are shown in stick models of the same color as Fig. 2, and  $\alpha$  traces are shown in pink.

(189–266), in which the residue number is presented in the parentheses (Yokoyama & Matsui, 2005). Although 1510-N K138A does not degrade the substrate 1511-C (189–266), trace amounts of putative acyl-enzyme intermediates are detected in the SDS–PAGE analysis (unpublished data). Therefore, 1510-N K138A would bind to the substrate in solution. Owing to the crystal packing constraint that all monomers forming the type-1 dimer also form the type-2 dimer, the 16-aa synthetic peptide might not enter the active site in the K138A crystal. A Lon protease from *Methanococcus jannaschii* forms a dimer in the crystal, and the active sites of the two monomers are both completely buried, preventing exposure of catalytic residues to the solvent. Therefore, the authors suggest that the observed dimeric structure is an artifact (Im *et al.*, 2004). In the 1510-N K138A structure, the same discussion may be possible for the type-2 dimer, but the type-2 dimer may exist as an inactive form. The type-1 dimer is probably in an active form for proteolysis because the structures in both the wild type and the K138A mutant are almost the same.

### 3.4. Role of the L2 loop

In the crystal structure of the wild-type 1510-N, the residues 122–139 corresponding to the L2 loop are disordered (Yokoyama *et al.*, 2006). The structure of the K138A mutant contains the L2 loop in spite of the high average *B* factors of 86.2 Å<sup>2</sup> (the average *B* factors of all protein atoms are 47.6 Å<sup>2</sup>). Ala138 is located at the base of the loop. Interestingly, the L2 loop is significantly kinked at around Ala138 (Figs. 2 and 5). ClpP protease has a long handle region corresponding to the L2 loop, but because of the kink around Ala138 the structure of the L2 loop is quite different from the handle region of ClpP. Lys138 of 1510-N is located at an N-terminal end of the following  $\alpha$ -helix. Considering the  $\alpha$ -helical dipole moment, the side chain of Lys138 with positive charge would destabilize the neighboring structure. Lys138 of 1510-N would probably have an important role on the conformation of L2. Cys120 is also located at the base of the L2 loop and is located in the vicinity of Ser97 (Fig. 5). The side chain of Cys120 points to the protein interior. In wild-type 1510-N, however, the side chain of Cys120 is relatively solvent exposed. Cys120 would probably have an effect on the conformation of L2. Three proline residues (Pro122, Pro136 and Pro137) located at the base of the L2 loop would also contribute to the kinked structure of L2.

The 1510-N protease specifically cleaves the C-terminal hydrophobic region of the p-stomatatin, PH1511 (Yokoyama & Matsui,

2005). Although the function of stomatin is not clear, stomatin is a lipid raft protein and could function as an oligomeric scaffolding protein or as an active signalling component involved in vesicle trafficking (Umlauf *et al.*, 2004). In the case of p-stomatatin PH1511 being degraded, 1510-N would attack the C-terminal hydrophobic region of PH1511. If both the type-1 and the type-2 dimers are present and can change mutually at physiological conditions, the interconversion of the dimers may regulate the protease activity (active type 1 and inactive type 2). Lys138 or Cys120 would be involved in the possible conformational change of L2. In order to elucidate how 1510-N degrades p-stomatatin PH1511, structure determination of 1510-N and its ligand would be needed.

We thank the Photon Factory staff for assistance in data collection. We also thank E. Matsui and Y. Urushibata (AIST, Tsukuba, Japan) for helpful discussions and E. Yamamoto (AIST, Tsukuba, Japan) for technical assistance.

### References

- Arnold, I. & Langer, T. (2002). *Biochim. Biophys. Acta*, **1592**, 89–96.
- Brown, M. S., Ye, J., Rawson, R. B. & Goldstein, J. L. (2000). *Cell*, **100**, 391–398.
- Brünger, A. T. *et al.* (1998). *Acta Cryst.* **D54**, 905–921.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Green, J. B., Fricke, B., Chetty, M. C., von Düring, M., Preston, G. F. & Stewart, G. W. (2004). *Blood Cells Mol. Dis.* **32**, 411–422.
- Hutchinson, E. G. & Thornton, J. M. (1994). *Protein Sci.* **3**, 2207–2216.
- Im, Y. J., Na, Y., Kang, G. B., Rho, S.-H., Kim, M. -K., Lee, J. H., Chung, C. H. & Eom, S. H. (2004). *J. Biol. Chem.* **279**, 53451–53457.
- Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Stewart, G. W., Argent, A. C. & Dash, B. C. J. (1993). *Biochim. Biophys. Acta*, **1225**, 15–25.
- Tavernarakis, N., Driscoll, M. & Kyriakides, N. C. (1999). *Trends Biochem. Sci.* **24**, 425–427.
- Umlauf, E., Csaszar, E., Moertelmaier, M., Schuetz, G. J., Parton, R. G. & Prohaska, R. (2004). *J. Biol. Chem.* **279**, 23699–23709.
- Wang, J., Hartling, J. A. & Flanagan, J. M. (1997). *Cell*, **91**, 447–456.
- Yokoyama, H. & Matsui, I. (2005). *J. Biol. Chem.* **280**, 6588–6594.
- Yokoyama, H., Matsui, E., Akiba, T., Harata, K. & Matsui, I. (2006). *J. Mol. Biol.* **358**, 1152–1164.