PROTEIN STRUCTURE REPORT

Crystal structure of *Homo sapiens* PTD012 reveals a zinc-containing hydrolase fold

BABU A. MANJASETTY,[1,2,7] KONRAD BÜSSOW,[4,5] MARTIN FIEBER-ERDMANN,[1,3] YVETTE ROSKE,[2,4] JOHAN GOBOM,[5] CHRISTOPH SCHEICH,[4,5] FRANK GÖTZ,[2,4] FRANK H. NIESEN,[4,6] AND UDO HEINEMANN[2,3]

[1]Protein Structure Factory, Berlin 12489, Germany
[2]Max-Delbrück-Centrum für Molekulare Medizin, Berlin 13092, Germany
[3]Institut für Chemie und Biochemie, Berlin 14195, Germany
[4]Protein Structure Factory, Berlin 14059, Germany
[5]Max-Planck-Institut für Molekulare Genetik, Berlin 14195, Germany
[6]Institut für Medizinische Physik und Biophysik, Charité Universitätsmedizin Berlin, Berlin 10098, Germany

Abstract

The human protein PTD012 is the longer product of an alternatively spliced gene and was described to be localized in the nucleus. The X-ray structure analysis at 1.7 Å resolution of PTD012 through SAD phasing reveals a monomeric protein and a novel fold. The shorter splice form was also studied and appears to be unfolded and non-functional. The structure of PTD012 displays an αββα four-layer topology. A metal ion residing between the central β-sheets is partially coordinated by three histidine residues. X-ray absorption near-edge structure (XANES) analysis identifies the PTD012-bound ion as Zn^{2+}. Tetrahedral coordination of the ion is completed by the carboxylate oxygen atom of an acetate molecule taken up from the crystallization buffer. The binding of Zn^{2+} to PTD012 is reminiscent of zinc-containing enzymes such as carboxypeptidase, carbonic anhydrase, and β-lactamase. Biochemical assays failed to demonstrate any of these enzyme activities in PTD012. However, PTD012 exhibits ester hydrolase activity on the substrate *p*-nitrophenyl acetate.

Keywords: PTD012 family; alternative splicing; splice variant; structural genomics; Zn-binding site; ester hydrolase

Supplemental material: see www.proteinscience.org

The protein PTD012, encoded by *Homo sapiens* gene *PTD012*, was chosen as a target for the Protein Structure Factory (Heinemann et al. 2003) for three reasons: (1) No homologous structure was known when the work started, indicating that the protein might belong to an unknown structural class; (2) PTD012 represents a sizeable protein family; (3) The *PTD012* transcript undergoes alternative splicing leading to six different products in humans.

Homologous sequences to PTD012 have been found in *Mus musculus* (SwissProt Q91V76, 315 residues, 87% sequence identity), *Brachydanio rerio* (Q6PHI5, 318, 62%), *Drosophila melanogaster* (Q9W0J9, 322, 47%; Q8IRI0, 361, 44%; Q8SXQ8, 322, 44%), *Anopheles gambiae* str. PEST (Q7QI91, 360, 41%), and *Caenorhabditis elegans* (Q18131, 192, 36%). Figure 1A shows an alignment of these proteins with PTD012 (splice variant Q9H0W9).

In humans, ~30%–60% of the genes undergo alternative splicing (Lander et al. 2001; Venter et al. 2001;
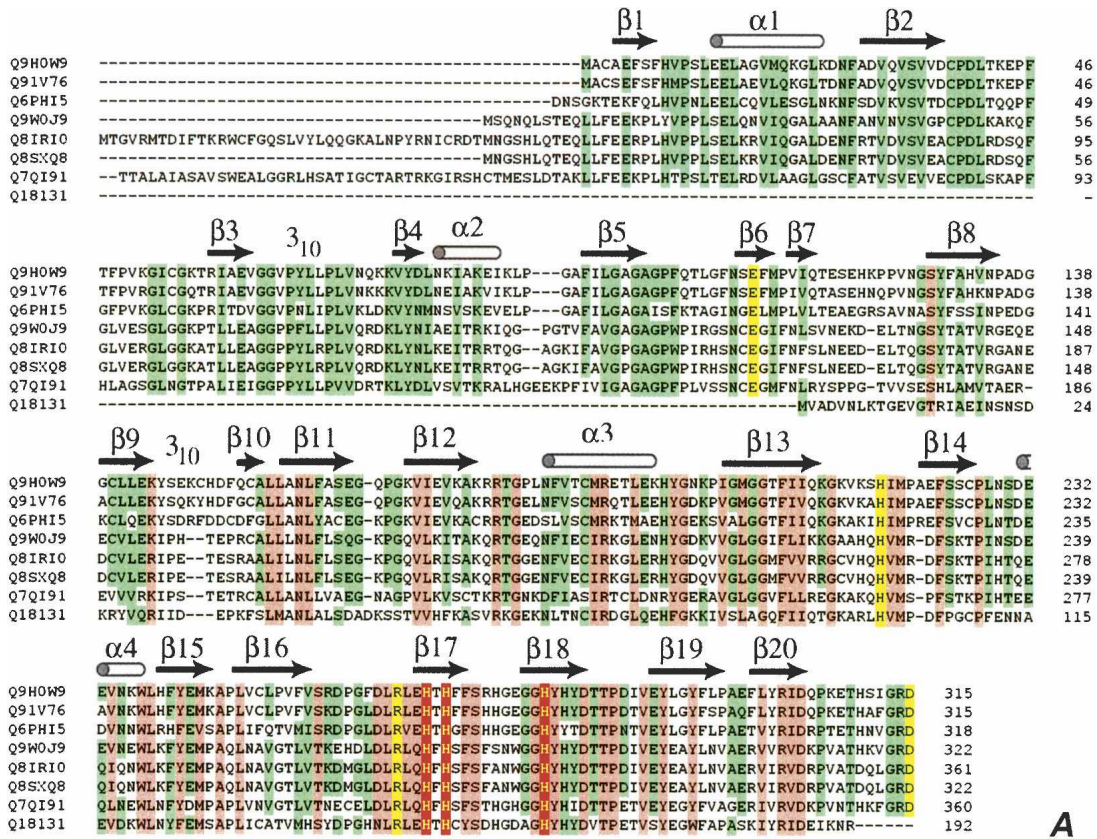

Figure 1. *(A)* Multiple sequence comparison of long splice variant of PTD012 from *H. sapiens* against its homologs. Eight homologs were identified and aligned, SwissProt entries: Q9H0W9 (*Homo sapiens*); Q91V76 (*Mus musculus*); Q6PHI5 (*Brachydanio rerio*); Q9W0J9, Q8IRI0, Q8SXQ8 (*Drosophila melanogaster*); Q7QI91 (*Anopheles gambiae* str. PEST), and Q18131 (*Caenorhabditis elegans*). Fully and highly conserved residues are shaded in khaki and green color respectively. The residues (H266, H268, and H278) coordinated to the Zn^{2+} ion are shaded in red color. The other residues discussed in the text (E110, R263, and D315) are shaded in yellow color. The alignment was generated using the program ClustalW; the β-strands are depicted as arrows and α-helices as cylinders *above* the sequences. *(B)* Ribbon diagram of PTD012 colored by sequence position (progression through the visible spectrum from blue to red starting at the N terminus). The bound acetate buffer molecule is depicted as a ball-and-stick model. The Zn^{2+} ion is shown as pink ball. *(C)* Superposition of β-screw region of PTD012 (red) on the structure of cAMP binding domain of the regulatory subunit of PKA (1GRS, yellow). cAMP bound to the PKA domain and the ion-binding site of PTD012 are also depicted.

Garcia-Blanco et al. 2004) and, in some cases, splice variants are associated with human diseases. The ENSEMBL database (Hubbard et al. 2002) lists a PTD012 splice variant corresponding to a 315-residue protein (SwissProt Q9H0W9) and a second variant lacking 50 amino acids in the central region (Q96EI3). The sub-cellular location of PTD012 was revealed as predominantly nuclear by an automatic phenotyping approach (Conrad et al. 2004); see LIFEdb database (Bannasch et al. 2004). Neither the biochemical function nor the biological role of PTD012 or any of its homologs are known.

Here we report biophysical data (see Supplemental Material) and the crystal structure of the long PTD012 splice form determined by SAD phasing. The protein adopts a zinc-containing hydrolase fold. Zn^{2+} is tetrahedrally coordinated to imidazole nitrogens of three histidine residues and to a carboxylate oxygen of an acetate anion taken up from the crystallization buffer in a way reminiscent of Zn^{2+}-dependent hydrolases. PTD012 exhibits esterase activity toward synthetic substrates.

Results and Discussion

Overall structure and fold identification

Two independent molecules of human PTD012 are present in the asymmetric unit. The two molecules are essentially identical with an RMSD of C^α positions for residues 3 to 315 of 0.27 Å. The solvent accessible surface buried between the two monomers is 633 $Å^2$, corresponding to 5.2% of the total surface area (12,120 $Å^2$ per molecule). Since the fraction of surface area buried in dimers tends to be significantly larger (Jones and Thornton 1995), we conclude that PTD012 exists as a monomer in solution.

Each PTD012 molecule is a single polypeptide chain, composed of 315 residues and roughly globular in shape with a diameter of ~45 Å. The overall structure is mainly composed of 20 β-strands and 4 α-helices which are numbered as shown in Figure 1B. The secondary structure assignment was made using PROMOTIF (Hutchinson and Thornton 1996). The PTD012 structure is organized into a four-layer αββα topology. Structure-based searches using DALI (Holm and Sander 1993) and VAST (Gibrat et al. 1996) did not point to any homologous structures in the Protein Data Bank. Distant structural homology appears to exist with PDB entry 1XV2, a *Staphylococcus aureus* protein sharing sequence similarity with α-acetolactate decarboxylase (A. Murzin, pers. comm.). Even though it is a four-layer sandwich, the topology of the PTD012 monomer does not fall under any existing classification of four-layer structures (Carfi et al. 1995; Guddat et al. 1999; Knöfel and Sträter 1999; Kumaran et al. 2003), since the order of β-strands, the connectivity, and crossovers of the sheets are highly different. The shorter splice variant, which lacks the central 50 residues, is expressed strongly in *Escherichia coli* but is completely insoluble. The PTD012 structure suggests that a protein lacking these 50 residues cannot fold and is likely non-functional.

Nine antiparallel β-strands in the order β2, β12, β19, β13, β16, β17, β18, β14, and β15 form a β-screw architecture which is rolling around helix α3, and the rest of the 11 mixed β-strands form a β-layer, which is partially twisted adjacent to helix α2. The interface of the β-screw and β-layer divides the molecule into two halves. However, the connecting loops between them have filled up the empty gaps. In addition, the helices α1, α4, and two 3_{10} helices are flanking the surface of the molecule. A further database search was carried out with VAST (Gibrat et al. 1996) to identify structural neighbors of PTD012 sub-structures. For the β-layer region (aa 51–165), the search did not yield any structural neighbors. However, a search using the β-screw region of PTD012 (aa 171–296) revealed structural similarity to the cyclic adenosine 3′,5′-monophosphate (cAMP) binding domain of the regulatory subunit of protein kinase A (1GRS, 53 matched residues out of 125, RMSD = 2.7 Å, Z-score 7.8, 13% sequence identity) (Su et al. 1995; Wu et al. 2004). The superposition of the matched regions is shown in Figure 1C. In addition to the conservation of much of the secondary structure, we note that the cAMP-binding site of the PKA regulatory subunit superimposes onto the Zn^{2+}-binding site of PTD012 (see below).

Cation-binding site

Difference Fourier maps revealed strong electron density at the β-screw/β-layer interface within coordination distance (~2 Å) to atoms N^δ of H278, N^ϵ of H266 and N^ϵ of H268, suggesting that a metal atom is bound to the protein at an HxHxxxxxxxxH motif (Fig. 2A). The metal ion bound to PTD012 was determined to be zinc. In a fluorescence spectrum scan (hν = 9670 eV) the X-ray emission (fluorescence) line Kα of Zn and the elastic scattered line were clearly resolved (Fig. 2B). Furthermore, the X-ray absorption near-edge structure (XANES) could be recorded by taking the counts solely in the region where the Zn Kα line is located. Moreover, the excitation photon energy was scanned through the XANES energy range of the Zn K-edge (Fig. 2C). In the structure, the tetrahedral Zn coordination is completed with density likely to represent an acetate molecule. Acetate may have been recruited from the crystallization buffer in which it was present at a 200-mM concentration. Occupation of the fourth coordination site by a molecule is very suggestive of an enzymatic role for the metal ion, as Zn^{2+} is similarly
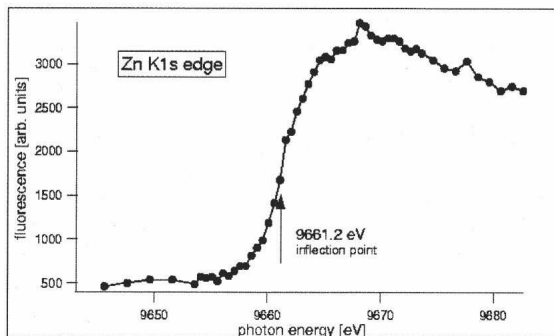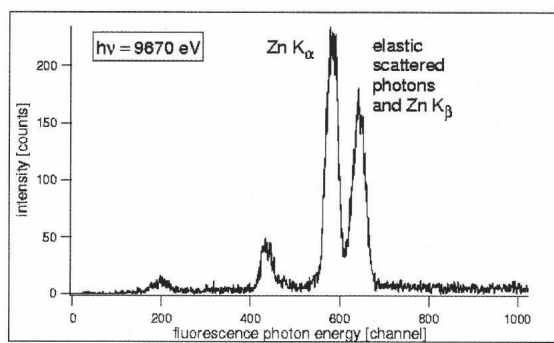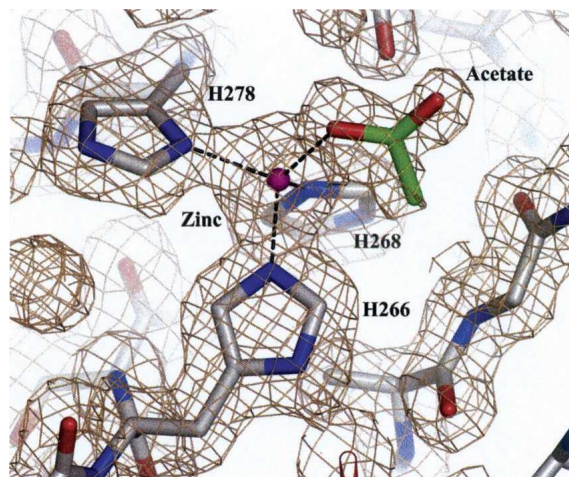
Figure 2. Ion-binding site of PTD012. *(A)* Putative active site architecture of PTD012. Zn-mediated binding mode of the acetate molecule. The $2F_o$-F_c electron density is shown at 1.2σ contour level in ochre. *(B)* Fluorescence spectrum at a photon energy of 9670 eV, taken by an Amptek-detector XR-100CR. *(C)* X-ray absorption near-edge structure (XANES) at the Zn Kα line.

coordinated in many zinc-containing enzymes (Lipscomb and Sträter 1996; Hayashi 2004).

Comparison with Zn-containing enzymes

The structure of PTD012 exhibits striking similarities with the structure of α-carbonic anhydrase (α-CA) (Tripp et al. 2001), even though no sequence homology

is found between them. First, the dominating secondary structure is a 10-stranded, twisted β-sheet. Second, an HxH motif is utilized in zinc coordination. In addition to this conservation, the carboxylate group of residue E110 resides 3.8 Å away from the Zn^{2+} ion (5 Å away in α-CA). Furthermore, an acetate molecule (structurally analogous to bicarbonate) is bound in the crystal structure of PTD012, resembling one of the reaction products within the pathway of reversible hydration of carbon dioxide to carbonic acid (or to bicarbonate), which is catalyzed by α-CA (Liljas and Laurberg 2000; Tripp et al. 2001). In view of this, an assay for carbonic anhydrase activity was performed (Supplemental Material). In this assay, PTD012 did not show carbonic anhydrase activity, suggesting that PTD012 does not belong to the α-CA family.

The arrangement of residues R263 and E110 near the cation-binding site of PTD012 is very similar to the arrangement of R482 and E626 near the active site of human carboxypeptidase A2 (CPA2) (Reverter et al. 2000). In CPA2, the residue E626 (E110 of PTD012) acts as a general acidic catalyst to form a metal-bound hydroxide species (zinc hydroxide). The zinc hydroxide then acts as the nucleophile to attack the peptide carbonyl function that is polarized with the assistance of R482 (R263 in PTD012). In addition to the conservation of these two residues in PTD012, the carboxylate group of the PTD012 C terminus resides very close to the cation-binding site. One of the C-terminal carboxylate oxygens of E315 interacts with the imidazole N^ε of the Zn^{2+}-coordinating residue H278 within a distance of 2.8 Å, and the other C-terminal oxygen of E315 forms a hydrogen bond with the guanidinum group of R263 (over 3.3 Å). Interestingly, the amino acid sequence at the C terminus, a GRD motif, is well conserved in all PTD012 homologs (Fig. 1A). Moreover, the mode of binding of the acetate carboxylate group to the zinc ion is reminiscent of the binding of the leech carboxypeptidase inhibitor (LCI) C terminus to the human CPA2 active site (Reverter et al. 2000). In view of the conservation of residues R263 and E110 near the cation-binding site, the binding of the fourth ligand (acetate), and the interaction between the C termini and the active site, a possible carboxypeptidase activity was tested by incubating a number of peptides with the protein, followed by mass spectrometric analysis of the reaction mixtures (see Supplemental Material). No proteolytic activity was observed on any of the standard peptides, ruling out the possibility that PTD012 might belong to the carboxypeptidase family of enzymes.

PTD012 and β-lactamase show similar four-layered αββα sandwich structures and Zn^{2+}-binding sites at the interface of two sheets, although the overall structural topology between the two proteins is rather poor.

β-lactamase utilizes Zn^{2+} to hydrolyze a series of β-lactam antibiotics, such as penicillin and cephalosporin derivatives (Massova and Mobashery 1999). A hydrolytic activity of PTD012 on ampicillin as substrate was not detectable. However, when the ester hydrolase activity of PTD012 was assayed with the substrate *p*-nitrophenyl acetate (pNPAc), this compound was hydrolyzed to nitrophenolate and acetate. The crystal structure showing acetate bound to the Zn^{2+}-binding site of PTD012 may thus represent an enzyme–product complex. The enzymatic turnover number for PTD012 was 0.042 sec^{-1}. In comparison, for bovine carbonic anhydrase II, which also catalyzes the ester hydrolysis of pNPAc, a turnover number of 1 sec^{-1} was determined. Therefore, PTD012 is an ester hydrolase and might belong to the superfamily of metallo-β-lactamase fold proteins.

Conclusion

The long splice variant of human PTD012 is a monomeric protein with a novel overall fold. The shorter transcript does not seem to encode a folded protein. PTD012 contains a Zn^{2+} ion coordinated to three histidines and an acetate anion at a site that likely supports the enzymatic activity of the protein as an ester hydrolase. The anion binding to the putative active site may prove useful in studies of the properties of the metal center, particularly the search for the natural substrate and various specific inhibitors to derive the enzymatic mechanism and to understand the underlying biological function.

Materials and Methods

Cloning and expression

A cDNA fragment corresponding to the open reading frame of one of the two long splice variants of PTD012 (GenBank accession no. CAB66540) was amplified by PCR from the cDNA clone DKFZp564H1122 of the German cDNA consortium (Wiemann et al. 2001) obtained from the German Resource Center (RZPD, http://www.rzpd.de). The PCR product was cloned into the vector pQTEV (GenBank accession no. AY243506) using the restriction enzymes BamHI and NotI (resulting in addition of a His_6 affinity tag and a cleavage site for TEV protease to the N terminus of the expressed protein). The resulting plasmid was introduced into *E. coli* SCS1 cells carrying pRARE (Novagen). The resulting clone was used for expression of unlabeled protein (PSF clone ID 110535, RZPD ID PSFEp250B014). For expression of selenomethionine-labeled protein, the plasmid of clone 110535 was used to transform B834 *E. coli* cells (Novagen) carrying the helper plasmid pSE111. pSE111 provides resistance to 15 μg/mL kanamycin and carries the *lacI^Q* repressor and the *argU* gene for a rare tRNA (PSF clone ID 111491).

Fermentation and cultivation of E. coli clones

Clone 110,535 was grown at 37°C to an OD_{600} of 8.5 in SB medium (12 g/L Bacto-tryptone, 24 g/L yeast extract, 0.4% (v/v) glycerol, 17 mM KH_2PO_4, 72 mM K_2HPO_4), supplemented with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol in shaker flasks. Gene expression was induced by addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After 3.5 h the cells were harvested by centrifugation (6000*g*, 15 min, 4°C) and stored at –80°C. Clone 111491 was expressed using an in-house protocol mainly based on the stepwise cultivation method of Budisa et al. (1995) (see Supplemental Material).

Protein purification

Cell pastes of clones 110535 (unlabeled) and 111491 (SeMet-labeled) were resuspended in extraction buffer (20 mM Tris-HCl at pH 8.0, 300 mM NaCl, 0.5 mM EDTA, 1 mM PMSF) containing lysozyme and benzonase, and either 4 mM or 10 mM 2-mercaptoethanol, respectively. The cells were disrupted by ultrasonic treatment in a chilling water bath, using a Bandelin Sonoplus sonifier. They were centrifuged 20 or 45 min at 55,000*g* and filtered using 0.45 μm syringe filters. The pH of the filtrate was adjusted to pH 7.4 and then applied to a 10-mL TALON column, pre-equilibrated with buffer (20 mM Tris-HCl at pH 7.4, 50 mM NaCl, 10 mM imidazole). After thorough washing, proteins were eluted with 200 mM imidazole in 20 mM Tris-HCl, 50 mM NaCl (pH 7.4). After removal of the His_6 tag by overnight incubation with TEV protease at 4°C, protein solutions were subjected to POROS 20 S cation-exchange chromatography on a column pre-equilibrated with 20 mM Tris-HCl buffer, 15 mM NaCl (pH 7.4). Upon application of a linear NaCl gradient the PTD012 proteins were eluted separately from the TEV protease. The proteins were further purified on a POROS HQ anion-exchange column under the same buffer conditions, concentrated using Amicon Ultra-15 PLTK centrifugal filter units (10 kDa cut-off), and applied to a Superdex 75 column equilibrated with buffer (15 mM Tris-HCl at pH 7.4, 50 mM NaCl, 0.1 mM EDTA, 0.02% NaN_3, and 2 or 5 mM DTT). For crystallization, the proteins were concentrated to 32 mg/mL each.

Esterase assay

Esterase activities were measured according to Rozema and Gellman (1996) using *p*-nitrophenyl acetate (pNPAc) as substrate. Samples containing 20 μg of enzyme were pipetted into 20 mM Tris-HCl (pH 7.7) to give a final volume of 400 μL. After addition of 45 μL of 52 mM pNPAc in dry acetonitrile and mixing, the formation of the hydrolysis product, *p*-nitrophenolate (ε_{400nm} = 19,300 $M^{-1}cm^{-1}$), was monitored at room temperature by measuring the linear increase in absorbance at 400 nm between 30 and 60 sec (background hydrolysis was subtracted).

Crystallization and Data Collection

SeMet-derivatized crystals were grown from a protein solution at 32 mg/mL containing 20% PEG 4000, 0.2 M ammonium acetate, 0.1 M bis-Tris (pH 5.3) at 20°C. The crystals belong to space group P2_1 with cell parameters a = 44.5 Å, b = 52.9 Å, c = 123.5 Å, and β = 90.8°. Native crystals were grown from a

Table 1. *Crystallographic data*

	Native	SeMet
Data sets		
Resolution (Å)	1.7	2.0
Wavelength (Å)	0.9184	0.9797
Total reflections	228,515	268,487
Unique reflections	61,064	74,935
Completeness (%)	95.6 (95.5)	97.8 (92.2)
Average $I/\sigma(I)$	22.2 (6.8)	30.5 (4.6)
R(merge) (%)	5.5 (21.5)	6.7 (20.4)
Phasing		
Resolution (Å)		2.0
Se atoms per asym. unit		12
Figure of merit (SOLVE)		0.35
Figure of merit (RESOLVE)		0.64
Refinement		
Resolution range (Å)	20–1.7	
Total reflections used in refinement	57,927	
No. of reflections in R(free) set	4161	
R factor[a]	0.190	
R(free)[b]	0.222	
RMSDs		
Bond distances (Å)	0.011	
Bond angles (°)	1.310	
No. of atoms per asym. unit		
Protein (A and B)	4860	
Waters	310	
Zinc	2	
Acetate	2	

Statistics for the highest resolution shell are given in parentheses.
[a] $R = \Sigma \ \Sigma_h ||F_o(h)| - k|F_c(h)|| / \Sigma_h |F_o(h)|$.
[b] Free R factor was calculated using a randomly selected subset of reflections.

protein solution at 31.8 mg/mL containing 25% PEG 4000, 0.2 M ammonium acetate, 12% glycerol, 0.1 M Na acetate (pH 4.6), at 20°C. Also in P2$_1$, their cell parameters are a = 44.6 Å, b = 52.9 Å, c = 124.8 Å, β = 88.9°. For both protein forms, the solvent content was calculated to 41% (V_M = 2.1 Å^3/Da) (Matthews 1968) assuming two molecules with a total molecular mass of $2 \times 35,117$ Da in the asymmetric unit. Single-wavelength anomalous diffraction data at the Se edge were collected at 100 K to 2.0 Å resolution on the SeMet-derivatized crystals and additionally high-resolution 1.7 Å native data were collected on the MAR345 imaging plate detector at beamline PSF-ID14.2, BESSY, Berlin. The data sets were processed and scaled using the program HKL2000 (Otwinowski and Minor 1997). The data collection statistics are as shown in Table 1.

Phasing, model building, and refinement

The 12 selenium sites corresponding to 2 PTD012 molecules in the asymmetric unit were identified, refined, and phases were calculated using SOLVE (Terwilliger and Berendzen 1999). Solvent flattening and automatic model building using RESOLVE (Terwilliger 2001) allowed to trace 564 out of 630 residues. For phase extension to 1.7 Å resolution, the partial model was placed into the native unit cell by molecular replacement using MOLREP (Vagin and Teplyakov 2000) of the

CCP4 suite (Collaborative Computational Project Number 4 1994). The rest of the residues were built manually with O (Jones et al. 1991), and REMAC5 (Murshudov et al. 1997) was employed to refine the model. Solvent molecules were located using the ARP/wARP suite (Morris et al. 2003). TLS parameters (Winn et al. 2001) were determined and refined for two molecules in the asymmetric unit. The final model consists of 626 residues, two zinc ions, two acetate, and 310 water molecules. The final refinement statistics are shown in Table 1. The stereochemical parameters of the protein model were evaluated with PROCHECK (Laskowski et al. 1993). The atomic coordinates and structure factors are available from the Protein Data Bank under accession code 1XCR. Figures were drawn with MOLSCRIPT (Kraulis 1991) and PYMOL (DeLano 2003).

References

Bannasch, D., Mehrle, A., Glatting, K.H., Pepperkok, R., Poustka, A., and Wiemann, S. 2004. LIFEdb: A database for functional genomics experiments integrating information from external sources, and serving as a sample tracking system. *Nucleic Acids Res.* **32:** D505–D508.

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. 1995. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli. Eur. J. Biochem.* **230:** 788–796.

Carfi, A., Pares, S., Duee, E., Galleni, M., Duez, C., Frere, J.M., and Dideberg, O. 1995. The 3-D structure of a zinc metallo-β-lactamase from *Bacillus cereus* reveals a new type of protein fold. *EMBO J.* **14:** 4914–4921.

Collaborative Computational Project Number 4. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **50:** 760–763.

Conrad, C., Erfle, H., Warnat, P., Daigle, N., Lorch, T., Ellenberg, J., Pepperkok, R., and Eils, R. 2004. Automatic identification of subcellular phenotypes on human cell arrays. *Genome Res.* **14:** 1130–1136.

DeLano, W.L. 2003. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA, http://www.pymol.org.

Garcia-Blanco, M.A., Baraniak, A.P., and Lasda, E.L. 2004. Alternative splicing in disease and therapy. *Nat. Biotechnol.* **22:** 535–546.

Gibrat, J.F., Madej, T., and Bryant, S.H. 1996. Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.* **6:** 377–385.

Guddat, L.W., McAlpine, A.S., Hume, D., Hamilton, S., de Jersey, J., and Martin, J.L. 1999. Crystal structure of mammalian purple acid phosphatase. *Structure.* **7:** 757–767.

Hayashi, T. 2004. Zinc-containing enzymes and their models. In *Encyclopedia of supramolecular chemistry.* (eds. J.L. Atwood and J.W. Steed), pp. 1631–1638, Marcel Dekker, New York.

Heinemann, U., Büssow, K., Mueller, U., and Umbach, P. 2003. Facilities and methods for the high-throughput crystal structural analysis of human proteins. *Acc. Chem. Res.* **36:** 157–163.

Holm, L. and Sander, C. 1993. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233:** 123–138.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Ox, T., Cuff, J., Curwen, V., Down, T., et al. 2002. The Ensembl genome database project. *Nucleic Acids Res.* **30:** 38–41.

Hutchinson, E.G. and Thornton, J.M. 1996. PROMOTIF—A program to identify and analyze structural motifs in proteins. *Protein Sci.* **5:** 212–220.

Jones, S. and Thornton, J.M. 1995. Protein–protein interactions: A review of protein dimer structures. *Prog. Biophys. Mol. Biol.* **63:** 31–65.

Jones, T.A., Zhou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47:** 110–119.

Knöfel, T. and Sträter, N. 1999. X-ray structure of the *Escherichia coli* periplasmic 5′-nucleotidase containing a dimetal catalytic site. *Nat. Struct. Biol.* **6:** 448–453.

Kraulis, P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24:** 946–950.

Kumaran, D., Eswaramoorthy, S., Gerchman, S.E., Kycia, H., Studier, F.W., and Swaminathan, S. 2003. Crystal structure of a putative CN hydrolase from yeast. *Proteins* **52:** 283–291.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26:** 283–291.

Liljas, A. and Laurberg, M. 2000. A wheel invented three times. The molecular structures of the three carbonic anhydrases. *EMBO Rep.* **1:** 16–17.

Lipscomb, W.N. and Sträter, N. 1996. Recent advances in zinc enzymology. *Chem. Rev.* **96:** 2375–2434.

Massova, I. and Mobashery, S. 1999. Structural and mechanistic aspects of evolution of β-lactamases and penicillin-binding proteins. *Curr. Pharm. Des.* **5:** 929–937.

Matthews, B.W. 1968. Solvent content of protein crystals. *J. Mol. Biol.* **33:** 491–497.

Morris, R.J., Perrakis, A., and Lamzin, V.S. 2003. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374:** 229–244.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* **53:** 240–255.

Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276:** 307–326.

Reverter, D., Fernandez-Catalan, C., Baumgartner, R., Pfander, R., Huber, R., Bode, W., Vendrell, J., Holak, T.A., and Aviles, F.X. 2000. Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. *Nat. Struct. Biol.* **7:** 322–328.

Rozema, D. and Gellman, S.H. 1996. Artificial chaperone-assisted refolding of carbonic anhydrase B. *J. Biol. Chem.* **271:** 3478–3487.

Su, Y., Dostmann, W.R., Herberg, F.W., Durick, K., Xuong, N.H., Ten Eyck, L., Taylor, S.S., and Varughese, K.I. 1995. Regulatory subunit of protein kinase A: Structure of deletion mutant with cAMP binding domains. *Science* **269:** 807–813.

Terwilliger, T.C. 2001. Maximum-likelihood density modification using pattern recognition of structural motifs. *Acta Crystallogr. D. Biol. Crystallogr.* **57:** 1755–1762.

Terwilliger, T.C. and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **55:** 849–861.

Tripp, B.C., Smith, K., and Ferry, J.G. 2001. Carbonic anhydrase: New insights for an ancient enzyme. *J. Biol. Chem.* **276:** 48615–48618.

Vagin, A. and Teplyakov, A. 2000. An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D. Biol. Crystallogr.* **56:** 1622–1624.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* **291:** 1304–1351.

Wiemann, S., Weil, B., Wellenreuther, R., Gassenhuber, J., Glassl, S., Ansorge, W., Böcher, M., Blöcker, H., Bauersachs, S., Blum, H., et al. 2001. Toward a catalog of human genes and proteins: Sequencing and analysis of 500 novel complete protein coding human cDNAs. *Genome Res.* **11:** 422–435.

Winn, M.D., Isupov, M.N., and Murshudov, G.N. 2001. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D. Biol. Crystallogr.* **57:** 122–133.

Wu, J., Brown, S., Xuong, N.H., and Taylor, S.S. 2004. RIα subunit of PKA: A cAMP-free structure reveals a hydrophobic capping mechanism for docking cAMP into site B. *Structure* **12:** 1057–1065.