Crystal structure of an RtcB homolog protein (PH1602-extein protein) from *Pyrococcus horikoshii* reveals a novel fold

Chiaki Okada, Yuki Maegawa, Min Yao, Isao Tanaka *

Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

email: Isao Tanaka (tanaka@castor.sci.hokudai.ac.jp)

*Correspondence to Isao Tanaka, Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

Grant sponsor: National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Introduction. The hypothetical extein, PH1602-extein (53.5 kDa, 481 residues) from the hyperthermophilic archaebacterium, *Pyrococcus horikoshii* OT3, shows sequence similarity to *Escherichia coli* RtcB (33% identity). RtcB homologs are conserved in all archaebacteria, in most eukaryotes, including metazoa, and in a wide variety of eubacteria.

In *E. coli*, the *rtcB* gene together with *rtcA* and *rtcR* comprise the σ54-dependent operon termed the RNA 3'-terminal phosphate cyclase operon[1]. The functions of the gene products RtcA and RtcR have been characterized previously as RNA 3'-terminal phosphate cyclase and σ^{54}-specific regulator, respectively. On the other hand, although RtcB is highly conserved among many organisms, there have been few studies of RtcB and its homologs.

In a previous study, it was suggested that human RtcB homolog may form a heterotrimer[2] with DDX1 (putative RNA helicase[3]) and CGI-99 (ninein-interaction protein[4]). These three proteins were pulled down by double-stranded DNA fragments. The artificial DNA sequences included an 11-bp inverted repeat, thus enabling formation of a cruciform structure. The manner of interaction between the presumptive trimer and double-stranded DNA has not been clarified.

Sequence analysis indicated that the *ph1602* gene encodes an RtcB homolog with an intervening intein[5]. Thus, translation products of *ph1602* are considered to generate two mature proteins, PH1602-intein and PH1602-extein, due to the protein splicing activity of the intein. We cloned the extein sequence of PH1602, and successfully overexpressed, purified, and crystallized the PH1602-extein. Here, we report the crystal structure of PH1602-extein at 2.15 Å resolution, determined by the single-wavelength anomalous diffraction (SAD) method.

Materials and methods. *Cloning of the gene encoding PH1602-extein.* The *ph1602* gene is predicted to encode two mature proteins, an extein and an intein. The gene encoding RtcB homolog protein, PH1602-extein, is separated into two remote regions. Therefore, four primers were designed to fuse the two exteins *via* PCR to obtain the entire PH1602-extein gene. The sequences of the primers

were as follows: F1 primer, 5'-GTTGGGGATACATATGGTGGTTCCCC-3' (*Nde*I recognition site is underlined); F2 primer, 5'-GTTACGACATTAACTGTGGTGTGAGGTTAATAAGAAC-3'; R1 primer, 5'-CCAATCCCTCCTCTTCACGTCGACCACCATCC-3' (*Sal*I recognition site is underlined); and R2 primer, 5'-CTCACACCACAGTTAATGTCGTAACCAATTCCTC-3'. The DNA regions encoding the N-terminal (291 nucleotides) and C-terminal (1152 nucleotides) exteins were amplified with the primer combinations F1-R2 and F2-R1, respectively. To fuse the two exteins together, portions of each PCR product served as templates in a second PCR using primers F1 and R1. The amplified DNA was 1443 nucleotides in length encoding PH1602-extein and several nucleotides from the primers, which contained an *Nde*I and a *Sal*I site at the N and C termini, respectively. The amplified DNA was digested with the restriction endonucleases *Nde*I and *Sal*I and inserted into the corresponding sites of the expression plasmid, pET-22b(+).

Overexpression, purification, and crystallization. Escherichia coli strain Rosetta(DE3) (Novagen, Madison, WI) was transformed with the vector construct pET-22b(+)/PH1602-extein. The cells were grown at 37°C in 2L of LB medium containing 50 μg/mL ampicillin and 34 μg/mL chloramphenicol. The expression of PH1602-extein was induced by 1 mM IPTG. The cells were harvested by centrifugation at 3,000g for 15 min at 4°C and resuspended in cell

lysis buffer (50 mM Tris-HCl, pH8.0, 1 mM EDTA, 1 mM DTT, 1mM PMSF, 50 mM NaCl).

The cells were disrupted using a French Press. The homogenate was clarified by centrifugation at 40,000g for 30 min at 4°C. The supernatant of the cell extract was incubated for 30 min at 70°C and then centrifuged at 40,000g for 30 min at 4°C. Thereafter, ammonium sulfate was added to the supernatant at a final concentration of approximately 3.5 M and gently stirred for 30 min at 4°C, followed by centrifugation at 40,000g for 30 min at 4°C. The collected protein precipitate was dissolved in buffer A (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM DTT, 50 mM NaCl), and the protein solution was dialyzed against buffer A overnight at 4°C. After dialysis, the protein solution was filtrated with a 0.22 μm filter and applied to a HiTrap SP-XL column (Amersham Biosciences Corp., Arlington Heights, IL), which had been equilibrated with buffer A (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM DTT, 50 mM NaCl). After washing with buffer A, the bound protein was eluted with a linear gradient of 0.05–0.5 M NaCl in 100 mL of buffer. The fractions containing PH1602-extein were dialyzed against buffer A for 2 hours at 4°C and applied to a HiTrap Heparin column (Amersham Biosciences Corp.), which had been equilibrated with buffer A. After washing with buffer A, the bound protein was eluted with a linear gradient of 0.05–0.5 M NaCl in 100 mL of buffer. The fractions containing PH1602-extein were pooled and dialyzed against buffer B (10 mM Tris-HCL,

pH8.0, 200 mM NaCl) and concentrated by ultrafiltration using Apollo (Orbital Biosciences Topsfield, MA) to a final concentration of 10 mg/mL. The purity of the protein was analyzed by MALDI-TOF using a Mass Voyager DE-Pro (Applied Biosystems, Foster City, CA).

The crystallization conditions for PH1602-extein consisted of 0.1 M HEPES, pH7.5, 1.6 M ammonium sulfate, 100 mM NaCl. Crystals were grown at 20°C up to a size of 0.05 mm × 0.07 mm × 0.5 mm within a few weeks.

Crystallization of selenomethionine substituted PH1602-extein. Selenomethionyl recombinant PH1602-extein was prepared from methionine auxotroph *E. coli* cells Rosetta(DE3) (Novagen) transformed with pET-22b(+)/PH1602-extein plasmid. Se-Met PH1602-extein crystals (Se-Met crystals) were grown under the same conditions as native crystals.

Data collection. X-ray diffraction data of native and Se-Met crystals were collected to 2.15 and 2.5 Å resolution, respectively, at the beamline BL41XU in SPring-8 (Hyogo, Japan). The native and Se-Met crystals belonged to space group $P2_12_12_1$ with similar cell parameters [Table I]. The asymmetric unit contained two PH1602-extein molecules. Native data were collected at a wavelength of 0.9000 Å under cryogenic conditions (100 K) after soaking the crystal in cryoprotectant solution supplemented with 30% (w/v) sucrose. The data were integrated and scaled using the MOSFLM[6] and SCALA[7] programs. Single-wavelength anomalous diffraction (SAD) data were collected at a

wavelength of 0.9796 Å from a Se-Met crystal under the same conditions as native data. The data were integrated and scaled using HKL2000[8].

Structure determination and refinement. Selenium sites were determined and the initial phases were calculated by SAD method using SOLVE[9]. After improving phases, approximately 50% of the overall structure of PH1602-extein was automatically built by RESOLVE[10]. The remaining 50% of the protein was traced manually over the initial map using O[11].

The initial model was refined using native data by positional and B-factor refinement with the CNS program[12]. During refinement, 10% of all reflection data were set aside for calculation of the free R-factor. The final model consisted of 960 residues, 871 water molecules, 9 sulfate ions, and 2 sucrose molecules with an R of 16.8% and R_{free} of 20.4% at 2.15 Å. The model quality was checked using PROCHECK[13]. The atomic coordinates of PH1602-extein have been deposited in the Protein Data Bank with access code 1UC2.

Results and discussion. The crystal structure of the RtcB homolog, PH1602-extein, was determined at 2.15 Å resolution [Table I]. The asymmetric unit contains two PH1602-extein molecules with root mean square deviation of 0.877 Å. Although these two molecules are related by local two-fold axis, the results of gel filtration chromatography showed that PH1602-extein is a monomer in solution (data not shown). These results indicate that

intermolecular interface between two molecules in the asymmetric unit is only formed in crystal state.

The overall structure of PH1602-extein forms a single rigid domain. The domain showed a novel fold and unique topology [Fig. 1(a), 1(b)]. No structure exhibiting significant similarity to PH1602-extein was found by a DALI structure similarity search[14]. There are three beta-sheets in PH1602-extein. The beta-sheet the including N-terminal strand consists of β1-β2-β3-β4-β10-β9, the tiny beta-sheet consists of β11-β14, and branched beta-sheet including the C-terminal strand is in the order β8-β5-β7-β6 (-β16-β13-β12) -β17-β15. These beta-sheets are surrounded by 15 alpha-helices, one of which (α10) is unique in that it is 50 Å in length (35 residues). The electrostatic potential on the accessible molecular surface of PH1602-extein revealed a characteristic positively charged cleft, with a hydrophilic pocket in the middle of the cleft [Fig. 1(c)]. The length of the basic cleft is about 35 Å, and its width and depth are 9 Å. The hydrophilic pocket is located at the center of the molecule, and its diameter is over 10 Å. Although the human RtcB homolog, DDX1, and CGI-99 may form a heterotrimer and the presumptive trimer binds to double-stranded DNA[2], the positively charged cleft of the RtcB homolog is too narrow to bind to double-stranded DNA.

Sequence comparison indicated that the hydrophilic residues creating the pocket are well conserved [Fig. 1(d)]. Therefore, the hydrophilic pocket is

expected to be formed in all RtcB homologs, and to have important biological functions.

Previously, the possible involvement of a metal ion in RtcB function was suggested because of the presence of conserved histidine residues[1]. In the present study, we confirmed that three conserved histidines and a conserved cysteine (His203, His234, His404, and Cys98) are in close proximity to each other at the bottom of the hydrophilic pocket. We searched the Protein Data Bank to find a protein possessing a cysteine and several histidines positioned adjacent to each other in a similar manner as in PH1602-extein. The search results included several zinc metalloenzymes, such as LuxS quorum sensing proteins[15], and some tRNA synthetases[16]. Indeed, zinc ions show a stronger preference toward nitrogen and sulfur-containing ligands, such as His and Cys[17]. The locations of conserved histidines and a cysteine of PH1602-extein suggest that RtcB homologs may be zinc ion binding proteins. However, no metal ion was confirmed in the electron density distribution of PH1602-extein crystals. It is possible that EDTA in the buffer used for protein purification affected metal binding to PH1602-extein.

Acknowledgments. We thank M. Kawamoto and H. Sakai for their help during data collection on the beamline BL41XU of SPring-8, Japan. This work was supported by a research grant from the National Project on Protein Structural

and Functional Analyses from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

1. Genschik P, Drabikowski K, Filipowicz W. Characterization of the *Escherichia coli* RNA 3'-terminal phosphate cyclase and its σ^{54}-regulated operon. Journal of Biological Chemistry 1998;273(39):25516–25526.
2. Drewett V, Molina H, Millar A, Muller S, von Hesler F, Shaw PE. DNA-bound transcription factor complexes analysed by mass-spectrometry: binding of novel proteins to the human c-*fos* SRE and related sequences. Nucleic Acids Research 2001;29(2):479–487.
3. Chen HC, Lin WC, Tsay YG, Lee SC, Chang CJ. An RNA helicase, DDX1, interacting with poly(A) RNA and heterogeneous nuclear ribonucleoprotein K. Journal of Biological Chemistry 2002;277(43):40403–40409.
4. Howng SL, Hsu HC, Cheng TS, Lee YL, Chang LK, Lu PJ, Hong YR. A novel ninein-interaction protein, CGI-99, blocks ninein phosphorylation by GSK3 beta and is highly expressed in brain tumors. Febs Letters 2004;566(1-3):162–168.
5. Perler FB. InBase: The Intein Database. Nucleic Acids Research 2002;30(1):383–384.

6. Leslie AGW. Recent changes to the MOSFLM package for processing film and imaging plate data. Joint CCP4 and ESF-EACMB newsletter on protein crystallography, Vol. 26. Isaacs SB, editor. Warrington: Daresbury Laboratory; 1992. p 44–51
7. Evans PR. Proceedings of the CCP4 Study Weekend on Data Collection & Processing. Warrington: Daresbury Laboratory; 1993. p 114–122
8. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Meth Enzymol 1997;276:307–326.
9. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallographica Section D-Biological Crystallography 1999;55:849–861.
10. Terwilliger TC. Maximum-likelihood density modification. Acta Crystallographica Section D-Biological Crystallography 2000;56:965–972.
11. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved Methods for Building Protein Models in Electron-Density Maps and the Location of Errors in These Models. Acta Crystallographica Section A 1991;47:110–119.
12. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallographica

Section D-Biological Crystallography 1998;54:905–921.

13. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck—a Program to Check the Stereochemical Quality of Protein Structures. Journal of Applied Crystallography 1993;26:283–291.
14. Holm L, Sander C. Protein-Structure Comparison by Alignment of Distance Matrices. Journal of Molecular Biology 1993;233(1):123–138.
15. Lewis HA, Furlong EB, Laubert B, Eroshkina GA, Batiyenko Y, Adams JM, Bergseid MG, Marsh CD, Peat TS, Sanderson WE, Sauder JM, Buchanan SG. A structural genomics approach to the study of quorum sensing: Crystal structures of three LuxS orthologs. Structure 2001;9(6):527–537.
16. Newberry KJ, Hou YM, Perona JJ. Structural origins of amino acid selection without editing by cysteinyl-tRNA synthetase. Embo Journal 2002;21(11):2778–2787.
17. Dudev T, Lin YL, Dudev M, Lim C. First-second shell interactions in metal binding sites in proteins: A PDB survey and DFT/CDM calculations. Journal of the American Chemical Society 2003;125(10):3168–3180.
18. Kraulis PJ. Molscript: a Program to Produce Both Detailed and Schematic Plots of Protein Structures. Journal of Applied Crystallography 1991;24:946–950.
19. Nicholls A, Sharp KA, Honig B. Protein Folding and Association: Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons.

Proteins-Structure Function and Genetics 1991;11(4):281–296.

Table I. Summary of data collection and refinement statistics

Data collection		
	SAD	Native
Wavelength (Å)	0.9796	0.9000
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell parameters (Å)	a=81.5	a=81.3
	b=138.1	b=137.6
	c=149.1	c=148.5
Resolution (Å)[a]	50-2.50	39.39-2.15
	(2.59-2.50)	(2.27-2.15)
Number of Unique reflections[a]	58915 (5761)	90861 (13018)
Completeness (%)[a]	99.9(99.7)	99.7(99.7)
Average redundancy[a]	7.2(6.4)	5.3(4.5)
Average I/σ(I)[a]	13.7(2.8)	8.5(2.4)
R_{sym} (%)[a, b]	8.5(39.0)	7.7(30.5)
Refinement statistics		
Resolution range (Å)		10-2.15
Number of reflections		90104
Completeness (%)		99.4
Total number of non-hydrogen atoms		
Protein		7520
Others		962
R (%)[c]		16.8
R_{free} (%)[d]		20.4
r.m.s. deviation from standard values		
Bonds (Å)		0.0120
Bond angles (deg.)		1.598
Average B-factor (Å^2)		28.71
Ramachandran Plot		
Residues in most favored regions (%)		91.5
Residues in additional allowed regions (%)		8.1
Residues in generously allowed regions (%)		0.1
Residues in disallowed regions (%)		0.2

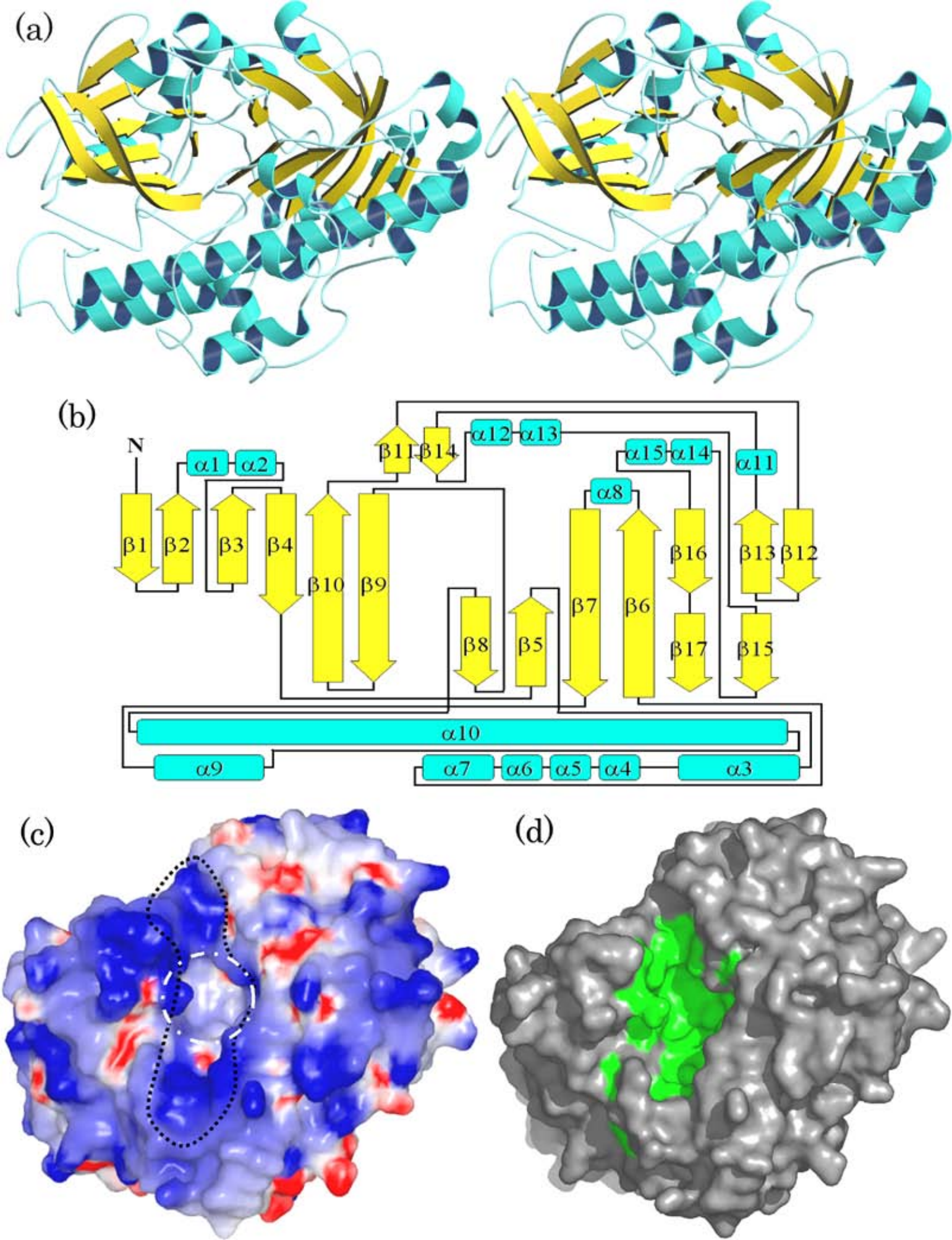
[a]Values within parentheses are for the highest resolution shell. [b]$R_{sym} = \Sigma_h\Sigma_i | I_{hi} - \langle I_h \rangle | / \Sigma_h\Sigma_i | I_{hi} |$, I_{hi} is the scaled intensity of the ith measurement of reflection h, and $\langle I_h \rangle$ is the mean intensity for the reflection of h.

[c]$R = \Sigma_h | | F_o | - | F_c | | / \Sigma_h | F_o |$, where F_o and F_c are the observed and calculated structure factors, respectively.

[d]R_{free} was calculated for R, using only an unrefined subset of reflections data (10%).

(Figure caption)

Fig. 1. (**a**) Stereo view of monomeric PH1602-extein. Alpha-helices are represented in cyan, and three beta-sheets in yellow. The figure was prepared with MOLSCRIPT[18]. (**b**) Schematic representation of the PH1602-extein topology, in the orientation of (a). The colors of helices and sheets are in accordance with (a). (**c**) Electrostatic surface potential of the PH1602-extein (blue, positive potential; red, negative potential). The surface shows a positively charged cleft (black dashed line). A hydrophilic pocket exists in the center of the cleft (white circle). The figure was prepared with GRASP[19]. (**d**) (same orientation as in (c)). The surface consisting of conserved residues is colored green. The hydrophilic pocket is formed by conserved residues. The residues have been completely conserved in homologs from *Pyrococcus horikoshii*, *Aeropyrum pernix*, *Sulfolobus solfataricus*, *Drosophila melanogaster*, *Homo sapiens*, *Caenorhabditis elegans*, *Nostoc sp*, *Escherichia coli*, and *Streptomyces avermitilis*. The figure was prepared with PYMOL (DeLano Scientific).


(a)
(b)
N
α1
α2
β11
β14
α12
α13
α15
α14
α11
α8
β1
β2
β3
β4
β10
β9
β16
β13
β12
β8
β5
β7
β6
β17
β15
α10
α9
α7
α6
α5
α4
α3
(c)
(d)