Agari, Y., Yokoyama, S., Kuramitsu, S. & Shinkai, A.X-ray crystal structure of a CRISPR-associated protein, Cse2, from Thermus thermophilus HB8. Proteins73, 1063-1067

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · DECEMBER 2008

Impact Factor: 2.63 · DOI: 10.1002/prot.22224 · Source: PubMed

4 AUTHORS:

Yoshihiro Agari
Illumina K.K. (Japan)
36 PUBLICATIONS **364** CITATIONS
SEE PROFILE

Shigeyuki Yokoyama
RIKEN
744 PUBLICATIONS **20,761** CITATIONS
SEE PROFILE

Seiki Kuramitsu
Osaka University
382 PUBLICATIONS **6,646** CITATIONS
SEE PROFILE

Akeo Shinkai
RIKEN
83 PUBLICATIONS **1,345** CITATIONS
SEE PROFILE

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

X-ray crystal structure of a CRISPR-associated protein, Cse2, from *Thermus thermophilus* HB8

Yoshihiro Agari,[1] Shigeyuki Yokoyama,[1,2,3] Seiki Kuramitsu,[1,4] and Akeo Shinkai[1]*

[1] Photon Science Research Division, RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan

[2] Systems and Structural Biology Center, Yokohama Institute, RIKEN, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

[3] Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[4] Department of Biological Sciences, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan

Key words: Cas protein; CRISPR; Cse2; pfam09485; structural genomics; thermophile; TTHB189.

INTRODUCTION

The clustered regularly interspaced short palindromic repeats (CRISPRs) comprise a family of DNA direct repeats present in many prokaryotic genomes.[1–4] The repeats are composed of 24–47 bp exhibiting weak dyad symmetry, and are separated by 26–72 bp nonrepetitive sequences. The CRISPR-associated (Cas) proteins are encoded in the vicinity of the CRISPRs.[2,3,5,6] The CRISPR systems (CRISPRs and Cas proteins) are classified into several subtypes, and each subtype contains several different subtype specific *cas* genes.[6] Of the 45 Cas protein families, the amino acid sequences of several families are similar to those of nucleases, helicases, RNA- and DNA-binding proteins, or transcription factors.[2,6–9] Based on the properties of Cas proteins, the CRISPR systems have been hypothesized to be DNA repair ones[8] or prokaryotic host defense ones against invading foreign replicons.[9–12] Interestingly, it has experimentally been shown that CRISPR systems are involved in resistance against phages.[13–15] Recently, it was shown that one of the Cas proteins, Cas2, comprises a novel family of endoribonucleases, and cleaves single-stranded RNAs preferentially within U-rich regions.[16] However, the biochemical and structural properties of most other Cas proteins remain to be elucidated.

An extremely thermophilic bacterium, *Thermus thermophilus* HB8,[17] has 'Ecoli subtype'- and 'Mtube subtype'-like CRISPR systems[6] on megaplasmid pTT27, and expression of the *cas* genes is positively regulated by cyclic AMP receptor protein, one of the global transcriptional regulators distributed in many bacteria.[18] It has been shown that the three-dimensional structure of *T. thermophilus* Cse3 (TTHB192), which is one of the nine components of the 'Ecoli subtype'-like CRISPR system, has an RNA recognition motif-like domain.[19] In this study, we determined the crystal structure of the *T. thermophilus* Cse2 (TTHB189) protein, another one of the components of the 'Ecoli subtype'-like CRISPR systems, with a novel fold.

METHODS

Cloning, expression, and purification

The open reading frame of *T. thermophilus* Cse2 was cloned into the pET-11a expression vector (*Nde*I-*Bam*HI sites) (Novagen). A selenomethionine-substituted protein was produced in the *E. coli* methionine auxotroph Rosetta834(DE3) strain, which we obtained by introducing the pRARE plasmid (Novagen) into the B834(DE3) strain (Novagen). The cell lysate was heated at 70°C for 13 min. Then the soluble fraction was applied to a Resource PHE column (GE Healthcare UK Ltd.) equilibrated with 50 m*M* sodium phosphate buffer (pH 7.0) containing 1.5*M*

*Correspondence to: Akeo Shinkai, RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan. E-mail: ashinkai@spring8.or.jp
Received 2 June 2008; Revised 2 July 2008; Accepted 27 July 2008
Published online 17 September 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22224

$(NH_4)_2SO_4$, which was eluted with a linear gradient of 1.2–0*M* $(NH_4)_2SO_4$. The fractions containing the Cse2 protein were collected and applied to a Resource S column (GE Healthcare UK Ltd.) equilibrated with 20 m*M* MES buffer (pH 6.0) containing 0.15*M* NaCl, which was eluted with a linear gradient of 0.15–0.4*M* NaCl. The fractions containing the Cse2 protein were collected and applied to a hydroxyapatite CHT10-I column (Bio-Rad) equilibrated with 10 m*M* potassium phosphate buffer (pH 7.0), which was eluted with a linear gradient of 10–500 m*M* potassium phosphate buffer (pH 7.0). The fractions containing the Cse2 protein were collected and applied to a HiLoad 16/60 Superdex 75 column (GE Healthcare UK Ltd.) equilibrated with 20 m*M* Tris-HCl (pH 8.0) containing 0.5*M* NaCl. The purified protein was concentrated to 7.9 mg/mL using a Vivaspin 20 concentrator (5000 molcular-weight cutoff, Sartorius), and dithiothreitol was added to the sample to a final concentration of 1 m*M*. The molecular mass of the purified protein estimated on gel filtration column chromatography was 13.7 kDa, suggesting that it exists as a monomer in solution, although this value is smaller compared with that (19.4 kDa) calculated from the amino acid sequence. The molecular masses of many *T. thermophilus*-derived proteins estimated on gel filtration column chromatography are smaller than the calculated ones (data not shown).

Crystallization, data collection, and structure determination

Crystallization of the Cse2 protein was performed by the hanging drop vapor diffusion method by mixing 1 μL of a protein solution with an equal volume of a reservoir solution comprising 0.1*M* sodium cacodylate (pH 5.9), 30% (v/v) PEG 600, 5% (v/v) PEG 1000, and 10% (v/v) glycerol at 20°C. The crystal in the mother liquor was cryo-cooled in a nitrogen-gas stream. Single-wavelength anomalous dispersion (SAD) data were collected at the RIKEN Structural Genomics Beamline I (BL26B1) at SPring-8 (Hyogo, Japan) utilizing the anomalous scattering from Se atoms. The data set was collected at 1.8 Å resolution using a Jupiter 210 CCD detector (Rigaku MSC). The collected data were processed with the HKL2000 program suite.[20] The positions of two Se atoms out of four possible sites in the asymmetric unit of the crystal were determined with program SOLVE,[21] and then density modification was performed with program RESOLVE.[22] The automatic tracing procedure in program ARP/wARP[23] was utilized to build the initial model. The model refinement, initial picking, and manual verifying of water molecules were carried out using programs CNS and XtalView/Xfit.[24,25] The electron densities corresponding to ions and multiple conformers were not identified. According to PROCHECK in the CCP4 suite,[26] 93.1% of the residues in the final model are in the most favored region of a Ramachandran plot, with no residues in disallowed regions. Data collection statistics and processed data statistics are presented in Table I. The coordinates are available in the Protein Data Bank, under accession code 2ZCA.

Table I
X-ray Data Collection and Refinement Statistics

Data collection	
Wavelength (Å)	0.97897
Resolution (Å)	50–1.8 (1.86–1.80)
Space group	P2(1)
No. of molecules in an asymmetric unit	2
Unit cell parameters (Å, °)	a = 52.08, b = 71.24, c = 53.75, α = γ = 90, β =115.78
No. of measured reflections	178949
No. of unique reflections	32766
Completeness (%)	99.9 (100)
Redundancy	5.5 (5.4)
I/σ(I)	27.8 (3.9)
R_{merge}[a] (%)	7.0 (27.8)
Phasing	
No. of Se atoms used	2
Figure of merit	0.31
Figure of merit after density modification	0.65
Refinement	
Resolution (Å)	50–1.8
R_{work}[b] (%)/R_{free}[c] (%)	20.9/23.6
No. of protein atoms/water atoms	2569/240
r.m.s.d. bond lengths (Å)	0.005
r.m.s.d. bond angles (°)	1.1
Wilson B factor (Å^2)	14.8
Average B factor for protein (Å^2)	16.3
Average B factor for water (Å^2)	26.7
Ramachandran plot (%)	
Most favored	93.1
Allowed	6.9
Disallowed	0.0

Values in parentheses are for the highest-resolution shell.
[a] $R_{merge} = \sum_h\sum_i|I_{h,i} - \langle I_h\rangle|/\sum_h\sum_i I_{h,i}$, where $I_{h,i}$ is the ith measured diffraction intensity of reflection h and $\langle I_h\rangle$ is the mean intensity of reflection h.
[b] R_{work} is the R-factor $= \sum||F_o| - |F_c||/\sum|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.
[c] R_{free} is the R-factor calculated using 10% of the data that were excluded from the refinement.

RESULTS AND DISCUSSION

T. thermophilus Cse2 is composed of 169 amino acid residues. A BLAST search indicated that close homologs are a conserved hypothetical protein (UBAL2_82410431) from *Leptospirillum* sp. Group II UBA (1e-25), Cse2 (Ppro_2341) from *Pelobacter propionicus* DSM 2379 (2e-23), Cse2 (Pmen_3759) from *Pseudomonas mendocina* ymp (3e-19), a hypothetical protein (Paer2_01000229) from *Pseudomonas aeruginosa* 2192 (7e-18), Cse2 (RoseRS_0649) from *Roseiflexus* sp. RS-1 (1e-12), Cse2 (Sfum_2826) from *Syntrophobacter fumaroxidans* MPOB (5e-12), and a hypothetical protein (RCIX1149) from uncultured methanogenic archaeon RC-I (2e-9) [Fig. 1(A)].

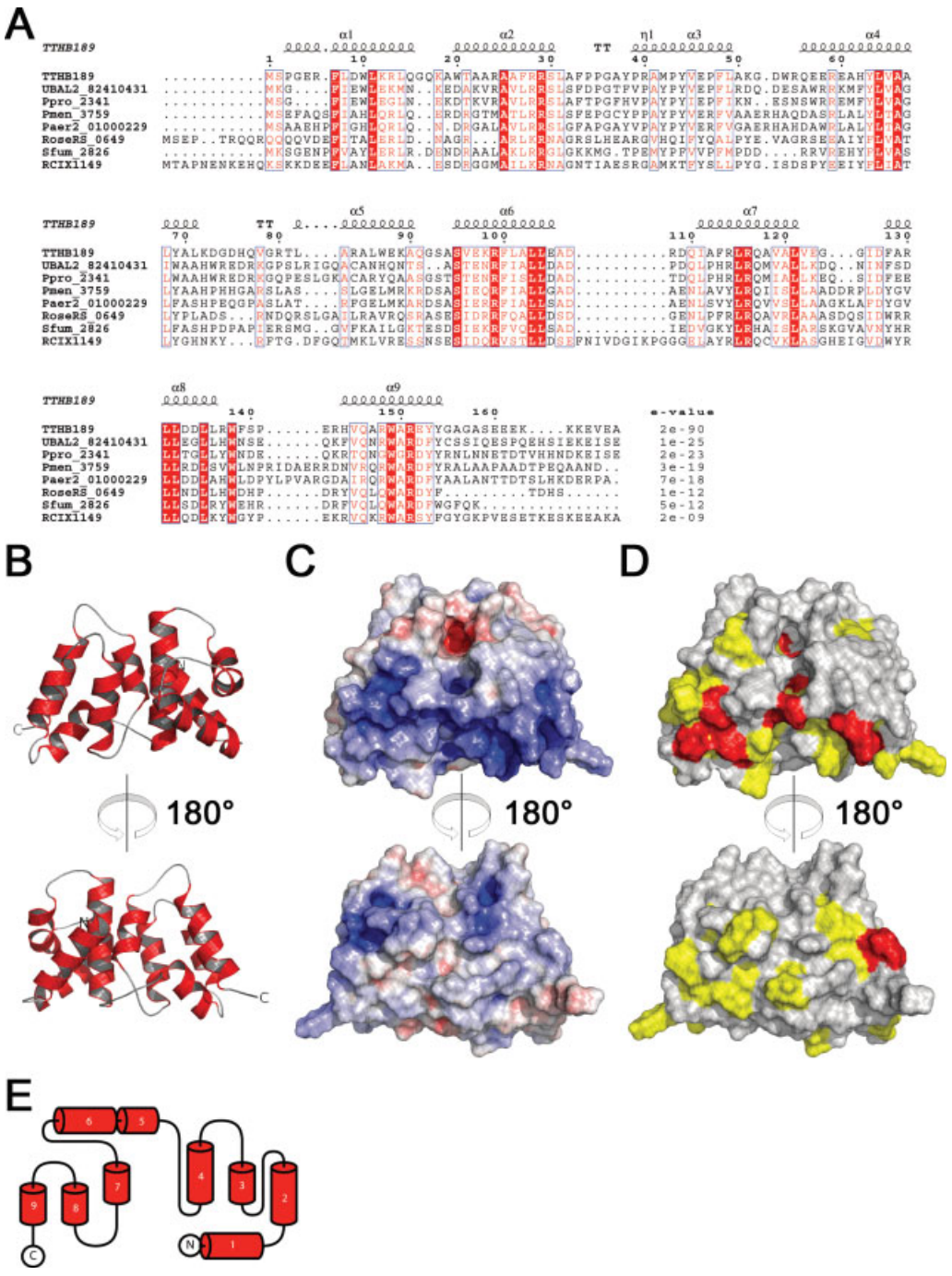

Figure 1

(**A**) Sequence alignment of T. thermophilus Cse2 with representative homologous proteins; UBAL2_82410431, conserved hypothetical protein from *Leptospirillum* sp. Group II UBA; Ppro_2341, Cse2 from *Pelobacter propionicus* DSM 2379; Pmen_3759, Cse2 from *Pseudomonas mendocina* ymp; Paer2_01000229, hypothetical protein from *Pseudomonas aeruginosa* 2192; RoseRS_0649, Cse2 from *Roseiflexus* sp. RS-1; Sfum_2826, Cse2 from *Syntrophobacter fumaroxidans* MPOB; and RCIX1149, hypothetical protein from uncultured methanogenic archaeon RC-I. Strictly conserved and similar residues are boxed in red and represented by red letters, respectively. The sequences were aligned using ClustalW2.[27] The secondary structure was predicted with DSSP,[28] and the figure was generated with ESpript 2.2.[29] (**B**) Ribbon diagram of *T. thermophilus* Cse2 (chain B). The α-helices are colored red. (**C**) Molecular surface representation of *T. thermophilus* Cse2 (chain B). Red and blue surfaces represent negative and positive electrostatic potentials (-10 k_BT and $+10$ k_BT), respectively. The electrostatic potentials were calculated using the Adaptive Poisson-Boltzmann Solver (APBS)[30] with PyMol APBS tools. (**D**) Molecular surface representation of *T. thermophilus* Cse2 (chain B). Strictly conserved and similar residues in the eight Cse2 family proteins (A) are colored red and yellow, respectively. (**E**) A topology diagram of *T. thermophilus* Cse2 (chain B). The α-helices are represented by red cylinders. B–D were generated with program PyMol (http://pymol.sourceforge.net/).

The three-dimensional crystal structure of *T. thermophilus* Cse2 was determined at a resolution of 1.8 Å, with crystallographic R_{work} and R_{free} factors of 20.9% and 23.6%, respectively (Table I). The asymmetric unit of the crystal contained two monomers of Cse2 (designated as chains A and B), which are similar, as shown by the r.m.s.d. value of 0.45 Å for corresponding main chain atoms. There are disordered regions, i.e., residues 157–169 in chain A and residues 160–169 in chain B, which are not included in the model. The overall structure of Cse2 (chain B) is shown in Figure 1(B). The structure consists of nine α-helices: α1, residues 3–15; α2, residues 20–30; α3, residues 39–49; α4, residues 56–71; α5, residues 82–90; α6, residues 95–105; α7, residues 111–121; α8, residues 128–136; and α9, residues 144–154. Three main hydrophobic cores are present in the structure. One is composed of A83, L86, and A90 in α5; V96, F100, L103, and L104 in α6; W138 between α8 and α9; and V145, W149, and Y153 in α9. Another one is composed of A61, L64, V65, and L68 in α4; L82 in α5; L115 and V119 in α7; I126 in the loop between α7 and α8; and F128, L131, and L135 in α8. The third one is composed of F7, W10, L11 in α1; W20, A23, and F27 in α2; V45, F48, and L49 in α3; and H62, Y63, A66, and A70 in α4. The *T. thermophilus* Cse2 structure was compared with previously determined structures in the PDB database, using the Secondary structure matching (SSM) server.[31] As a result, only a few proteins exhibiting matches with lower similarity were found, i.e, the closest structure was that of TetR-family regulator (SCO0857) from *Streptomyces coelicolor*, the r.m.s.d. value being 3.19 Å, and the Q-, P-, and Z-scores being 0.12, 0.0, and 1.6, which means that the matches are insignificant. This indicates that *T. thermophilus* Cse2 adopts a novel fold [Fig. 1(B,E)]. A conserved concavity is present on the surface [Fig. 1(C,D)]. *T. thermophilus* Cse2 has a high theoretical isoelectric point of ~9.6, and it has large continuous basic patches on one side of its surface, which are highly conserved in the seven closest homologs to the Cse2 [Fig. 1(C,D)]. This structural feature is also observed in *T. thermophilus* Cse3, which is another of the nine components of the 'Ecoli subtype'-like CRISPR system, although its overall structure differs from that of the Cse2.[19] The *cas* gene products including *T. thermophilus* Cse3 have been suggested to be involved in DNA/RNA metabolism.[2,6–9,15,19] The Cse2 might interact with nucleic acids at its basic patches.

ACKNOWLEDGMENTS

The authors thank Yoko Ukita for the protein purification, Kazuko Agari for the crystallization, and Seiki Baba and Yoshiaki Kitamura for the data collection at SPring-8. The authors also thank Koji Takio and Naoko Takahashi for the gel filtration analysis of the Cse2 protein.

REFERENCES

1. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J Bacteriol 1987;169:5429–5433.
2. Jansen R, van Embden JD, Gaastra W, Schouls LM. Identification of a novel family of sequence repeats among prokaryotes. OMICS 2002;6:23–33.
3. Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol 2007;8:R61.
4. Mojica FJ, Diez-Villasenor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of archaea, bacteria and mitochondria. Mol Microbiol 2000;36:244–246.
5. Godde JS, Bickerton A. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J Mol Evol 2006;62:718–729.
6. Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLo S Comput Biol 2005;1:e60.
7. Brüggemann H, Chen C. Comparative genomics of *Thermus thermophilus*: Plasticity of the megaplasmid and its contribution to a thermophilic lifestyle. J Biotechnol 2006;124:654–661.
8. Makarova KS, Aravind L, Grishin NV, Rogozin IB, Koonin EV. A DNA repair system specific for thermophilic archaea and bacteria predicted by genomic context analysis. Nucleic Acids Res 2002;30:482–496.
9. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct 2006;1:7.
10. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 2005;151:2551–2561.
11. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol 2005;60:174–182.
12. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 2005;151:653–663.
13. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. Science 2007;315:1709–1712.
14. Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol 2008;190:1390–1400.
15. Sorek R, Kunin V, Hugenholtz P. CRISPR–a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol 2008;6:181–186.
16. Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, Kudritska M, Kochinyan S, Wang S, Chruszcz M, Minor W, Koonin EV, Edwards AM, Savchenko A, Yakunin AF. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J Biol Chem 2008;283:20361–20371.
17. Oshima T, Imahori K. Description of *Thermus thermophilus* (Yoshida and Oshima) com. nov., a non-sporulating thermophilic bacterium from a Japanese thermal spa. Int J Syst Bacteriol 1974;24:102–112.
18. Shinkai A, Kira S, Nakagawa N, Kashihara A, Kuramitsu S, Yokoyama S. Transcription activation mediated by a cyclic AMP receptor protein from *Thermus thermophilus* HB8. J Bacteriol 2007;189:3891–3901.

19. Ebihara A, Yao M, Masui R, Tanaka I, Yokoyama S, Kuramitsu S. Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. Protein Sci 2006;15:1494–1499.
20. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
21. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D 1999;55:849–861.
22. Terwilliger TC. Map-likelihood phasing. Acta Crystallogr D 2001;57:1763–1775.
23. Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS. ARP/wARP and molecular replacement. Acta Crystallogr D 2001;57:1445–1450.
24. 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 Crystallogr D 1998;54:905–921.
25. McRee DE. XtalView/Xfit–A versatile program for manipulating atomic coordinates and electron density. J Struct Biol 1999;125:156–165.
26. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D 1994;50:760–763.
27. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–2948.
28. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
29. Gouet P, Robert X, Courcelle E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 2003;31:3320–3323.
30. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 2001;98:10037–10041.
31. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D 2004;60:2256–2268.