X-ray crystal structure of a hypothetical Sua5 protein from Sulfolobus tokodaii strain 7

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · FEBRUARY 2007

Impact Factor: 2.63 · DOI: 10.1002/prot.21794 · Source: PubMed

8 AUTHORS, INCLUDING:

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

Taisuke Wakamatsu
Kochi University
16 PUBLICATIONS **133** CITATIONS
SEE PROFILE

Yoshitaka Bessho
RIKEN
74 PUBLICATIONS **1,364** CITATIONS
SEE PROFILE

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

WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

X-ray crystal structure of a hypothetical Sua5 protein from *Sulfolobus tokodaii* strain 7

Yoshihiro Agari,[1] Shinya Sato,[1] Taisuke Wakamatsu,[1,2] Yoshitaka Bessho,[1,3] Akio Ebihara,[1] Shigeyuki Yokoyama,[1,3,4] Seiki Kuramitsu,[1,2] and Akeo Shinkai[1]*

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

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

[3] Genomic Sciences Center, Yokohama Institute, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

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

Key words: archaea; ATP hydrolysis; DNA-binding protein; pfam01300; ribosome; RNA-binding protein; Rossmann fold; ST1526; thermophile; translation factor.

INTRODUCTION

The Sua5-yciO-yrdC domain proteins (pfam01300, pfam database[1]) are widely distributed in prokaryotes and eukaryotes (http://pfam.janelia.org/cgi-bin/getdesc?name = Sua5_yciO_yrdC) [Fig. 1(A)]. One of the proteins in this family, *Escherichia coli* YrdC, preferentially binds to double-stranded RNA and DNA.[6] It has been predicted to be a rRNA maturation factor.[7,8] The crystal structure of *E. coli* YrdC has been determined and a possible nucleic acid-binding site has been identified.[6]

The Sua5 protein consists of an N-terminal YrdC domain and a C-terminal Sua5 domain.[6] The *sua5* gene was first identified in *Saccharomyces cerevisiae* as a suppressor of a translation initiation defect in the leader region of the iso-1-cytochrome *c* (*cyc1*) gene, which is caused by the generation of an aberrant ATG start codon upstream of the gene as a consequence of a single base-pair substitution.[9,10] The function and three-dimensional structure of the Sua5 protein remain to be elucidated. In the present study, we determined the crystal structure of the hypothetical Sua5 (ST1526) protein from thermoacidophilic archaeon *Sulfolobus tokodaii* strain 7,[11] which exhibits 49.7% similarity to *S. cerevisiae* Sua5 [Fig. 1(A)].

MATERIALS AND METHODS

Cloning, expression, and purification

The open reading frame of *S. tokodaii* Sua5 (gi:15921819) was cloned into the pET-21a(+) expression vector (*Nde*I-*Bam*HI site) (Novagen), with no purification tag. Selenomethionine-substituted proteins were produced in the *E. coli* methionine auxotroph BL21-CodonPlus (DE3)-RIL-X strain (Stratagene). The cell lysate was heated at 90°C for 10 min. Then the soluble fraction was applied to a Resource ISO column (GE Healthcare Biosciences) equilibrated with 50 m*M* sodium phosphate buffer (pH 7.0) containing 1.5*M* $(NH_4)_2SO_4$, which was eluted with a linear gradient of 1.5–0*M* $(NH_4)_2SO_4$. The fractions containing the Sua5 protein were collected and applied to a Resource Q column

Grant sponsors: RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural, Functional Analyses, Ministry of Education, Culture, Sports, Science and Technology of Japan.
*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 12 July 2007; Revised 17 August 2007; Accepted 24 August 2007
Published online 14 November 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21794

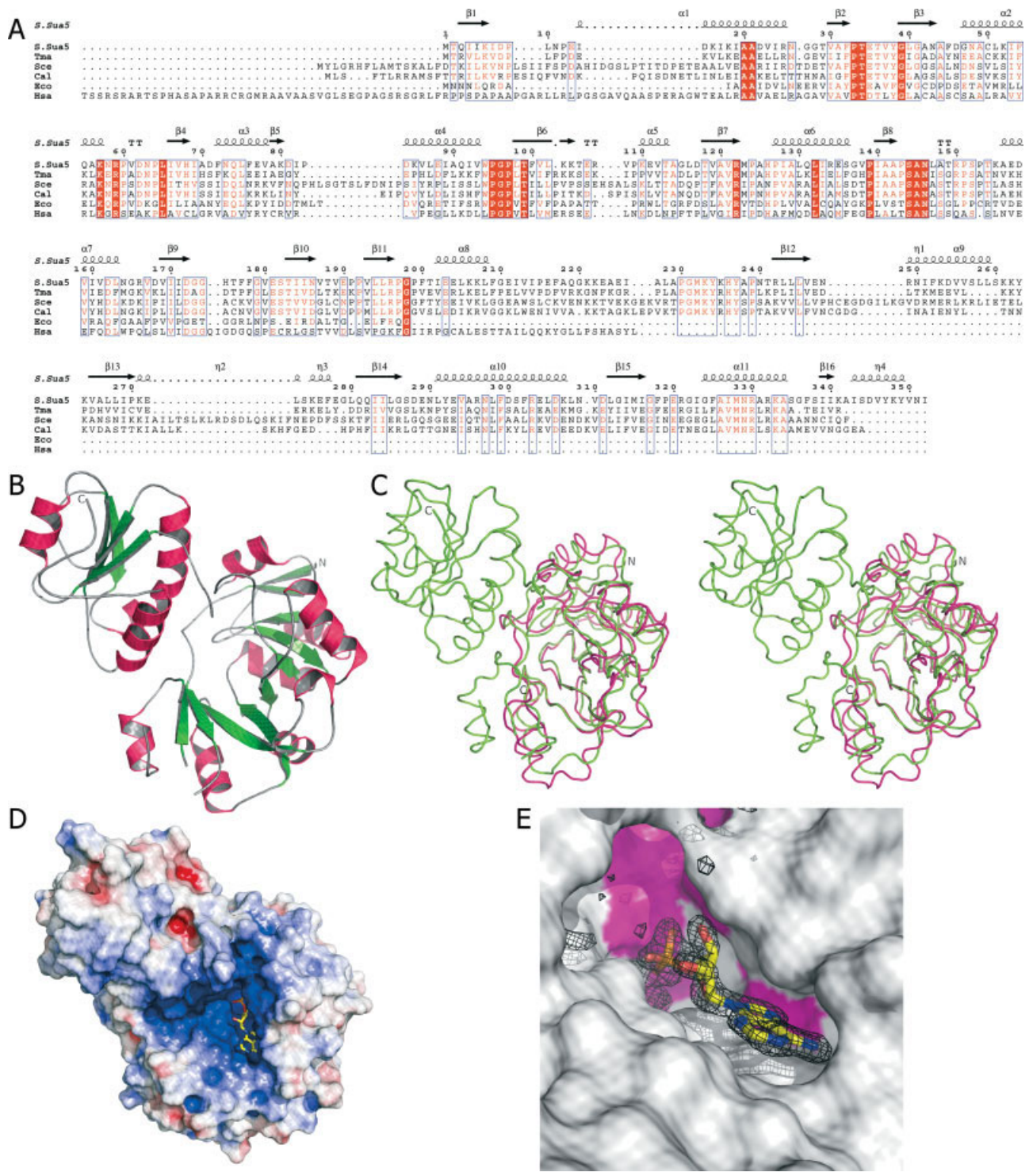

Figure 1

*(**A**) Sequence alignment of Sulfolobus Sua5 (S.Sua5) with Sua5/YrdC/YciO family proteins. Strictly conserved and similar residues are boxed in red and represented by red letters, respectively. Tma, Thermotoga maritima MSB8 (gi:15643615); Sce, Saccharomyces cerevisiae S288C (gi:6321269); Cal, Candida albicans SC5314 (gi:68471205); Eco, E. coli K12 (gi:16131163); and Hsa, Homo sapiens (gi:114325403). The sequences were aligned using Clustal W.[2] The secondary structure was predicted with DSSP,[3] and the figure was generated with ESPript 2.2.[4] (**B**) Ribbon diagram of Sulfolobus Sua5. The α-helices and β-strands are colored magenta and green, respectively. (**C**) Superpositioning of the main chain structures of Sulfolobus Sua5 (green) and E. coli YrdC (magenta). (**D**) Molecular surface representation of Sulfolobus Sua5. Red and blue surfaces represent negative and positive electrostatic potentials (−10 k_BT, +10 k_BT), respectively. AMP is shown as a stick model, in which yellow, red, blue, and orange represent carbon, oxygen, nitrogen, and phosphorus atoms, respectively. (**E**) Molecular surface colored as to sequence conservation, and difference electron density around the large concave. Strictly conserved residues, which are boxed in red in A, are colored magenta. The $F_o - F_c$ omit map (gray mesh) was obtained without the ligand molecule at 1.8 Å resolution, with contouring at 4.0 σ. AMP is represented as in D. The electrostatic potentials were calculated using the Adaptive Poisson-Boltzmann Solver (APBS)[5] with PyMol APBS tools. B–E were generated with program PyMol (http://www.pymol.org).*

(GE Healthcare Biosciences) equilibrated with 20 m*M* Tris-HCl (pH 8.0), which was eluted with a linear gradient of 0–0.5*M* NaCl. The fractions containing the Sua5 protein were collected and applied to a hydroxyapatite CHT10-I column (Bio-Rad) equilibrated with 10 m*M* sodium phosphate buffer (pH 7.0), which was eluted with a linear gradient of 10–500 m*M* sodium phosphate buffer (pH 7.0). The fractions containing the Sua5 protein were collected and applied to a HiLoad 16/60 Superdex 200 column (GE Healthcare Biosciences) equilibrated with 20 m*M* Tris-HCl (pH 8.0) containing 150 m*M* NaCl. The purified protein was concentrated to 9.6 mg/mL using a Vivaspin 20 concentrator (10,000 molcular-weight cutoff, Sartorius), and dithiothreitol was added to the sample to a final concentration of 1 m*M*. The N-terminal amino-acid sequence of the purified protein was TQIIK, which means that the N-terminal methionine had been deleted. The molecular mass of the purified protein estimated on light scattering photometry was 38.5 kDa, suggesting that it exists as a monomer in solution.

Crystallization, data collection, and structure determination

Crystallization of the Sua5 protein was performed by the sitting 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 acetate (pH 4.6) and 4*M* sodium nitrate (Salt RX, no.46, Hampton Research) at 20°C. The crystal in the mother liquor containing 20% glycerol was cryo-cooled in a nitrogen-gas stream. Multiple-wavelength anomalous dispersion (MAD) data were collected at the RIKEN Structural Genomics Beamline II (BL26B2) at SPring-8 (Hyogo, Japan) utilizing the anomalous scattering from Se atoms. The data sets were collected at 1.8 Å resolutions using a Jupiter 210 CCD detector (Rigaku MSC). The collected data were processed with the HKL2000 program suite.[12] The atomic positions of the Se atoms in the unit cell were determined with program SOLVE,[13] and then density modification was performed with program RESOLVE.[14] The automatic-tracing procedure in program ARP/wARP[15] 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.[16,17] In order to compensate for excess electron density of the N-terminal domain, AMP was presumptively used as a ligand [Fig. 1(E)]. According to PROCHECK in the CCP4 suite,[18] 91.9% of the residues in the final model are in the most favored region of a Ramachandran plot, with no residues in the disallowed regions. The data collection statistics and processed data statistics are presented in Table I. The coordinates are available in the Protein Data Bank, under accession code 2EQA.

Table I
X-ray Data Collection and Refinement Statistics

	Remote	Peak	Edge
Data collection			
Wavelength (Å)	0.9000	0.9792	0.9797
Resolution (Å)		50–1.8 (1.86–1.80)	
Space group		C222	
No. of molecules in an asymmetric unit		1	
Unit cell parameters (Å, °)		$a = 85.39$, $b = 120.84$, $c = 67.55$ $\alpha = \beta = \gamma = 90$	
No. of measured reflections	228,949	226,736	229,625
No. of unique reflections	32,797	32,701	32,831
Completeness (%)	100 (100)	99.9 (100)	99.9 (100)
Redundancy	7.0 (7.3)	6.9 (7.2)	7.0 (7.2)
I/σ(I)	50.3 (10.9)	52.5 (12.1)	50.4 (9.9)
R_{merge}[a](%)	4.5 (13.9)	5.1 (13.9)	4.8 (15.2)
Refinement			
Resolution (Å)		50–1.8	
R_{work}[b](%)/R_{free}[c](%)		20.8/23.3	
No. of protein atoms/water atoms		2563/180	
RMSD bond lengths (Å)		0.005	
RMSD bond angles (°)		1.3	
Ramachandran plot (%)			
Most favored		91.5	
Allowed		8.5	
Disallowed		0.0	

Values in parentheses are for the highest-resolution shell.
[a] $R_{merge} = \Sigma_h\Sigma_i|I_{h,i} - \langle I_h \rangle| / \Sigma_h\Sigma_i I_{h,i}$, 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 $= \Sigma||F_o|-|F_c||/\Sigma|F_o|$, 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

The overall structure of *Sulfolobus* Sua5 is shown in Figure 1(B). Note that there are two disordered regions, residues 196 and 215–233, which are not included in the model. Further, the side chains of Glu190 and Tyr234 could not be observed. The N-terminal domain of *Sulfolobus* Sua5 (residues 2–214) consists of eight α-helices and eleven β-strands. The overall fold of the N-terminal domain is similar to that of *E. coli* YrdC[6] that exhibits 42.9% similarity in amino acid sequence with the N-terminal domain of *Sulfolobus* Sua5, the *Z*-score being 21.3 and the r.m.s.d. value being 2.4 Å [Fig. 1(C)].[19] A large concave surface exhibiting a positive electrostatic potential, which is similar to that in YrdC, where nucleic acids are predicted to interact,[6] was found in *Sulfolobus* Sua5 [Fig. 1(D)]. Interestingly, excess electron density that might be due to an *E. coli*-derived nucleotide was observed on this concave surface of *Sulfolobus* Sua5 [Fig. 1(D,E)].

The C-terminal Sua5 domain (residues 234–352) of *Sulfolobus* Sua5 consists of three α-helices, α9, residues 249–259; α10, residues 291–307; and α11, residues 324–

DOI 10.1002/prot

335, and five β-strands, β12, residues 242–246; β13, residues 264–270; β14, residues 281–286; β15, residues 322–326; and β16, residues 338–340, which adopt a Rossmann fold. The C-terminal domain structure was compared with the previously determined structures in the PDB database, using the DALI server.[19] The closest structure was that of *Methanocaldococcus jannaschii* HypB, a GTP-binding protein that regulates metal binding[20] (PDB code: 2HF9 chain *A*), the *Z*-score being 7.7 and the r.m.s.d. value being 3.2 Å, despite the low sequence identity (IDE = 14%). The structural homologs that showed high *Z*-scores were hypothetical protein TM0796 from *Thermotoga maritima* (PDB code: 2G0T chain *A*, *Z* = 7.4, r.m.s.d. = 2.8 Å, IDE = 11%), the GTP-binding domain of Fth from *Thermus aquaticus*[21] (PDB code: 1NG1, *Z* = 6.4, r.m.s.d. = 3.1 Å, IDE = 10%), and *E. coli* cytidine triphosphate synthetase[22] (PDB code: 1S1M chain *A*, *Z* = 5.6, r.m.s.d. = 2.6 Å, IDE = 8%). Thus, the C-terminal Sua5 domain might be involved in nucleotide binding.

The three-dimensional structure of *Sulfolobus* Sua5 showed that both the N- and C-terminal domains were possibly involved in nucleotide binding or metabolism, which is supported by the observation that *Sulfolobus* Sua5 showed ATP hydrolysis activity, AMP being produced, and the apparent rate constants were 0.04 and 0.1 s^{-1} at 37 and 60°C, respectively, in the presence of 1 m*M* ATP, 5 m*M* $MgCl_2$, 50 m*M* HEPES, and 100 m*M* KCl, pH 7.5.

ACKNOWLEDGMENTS

The authors wish to thank Emiko Matsunaga for protein purification, Takeshi Nagira for crystallization, and Seiki Baba and Yoshiaki Kitamura for data collection at SPring-8.

REFERENCES

1. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL. The Pfam protein families database. Nucleic Acids Res 2002;30:276–280.
2. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
3. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
4. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 1999;15:305–308.
5. 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.
6. Teplova M, Tereshko V, Sanishvili R, Joachimiak A, Bushueva T, Anderson WF, Egli M. The structure of the *yrdC* gene product from *Escherichia coli* reveals a new fold and suggests a role in RNA binding. Protein Sci 2000;9:2557–2566.
7. Kaczanowska M, Rydén-Aulin M. Temperature sensitivity caused by mutant release factor 1 is suppressed by mutations that affect 16S rRNA maturation. J Bacteriol 2004;186:3046–3055.
8. Kaczanowska M, Rydén-Aulin M. The YrdC protein—a putative ribosome maturation factor. Biochim Biophys Acta 2005;1727:87–96.
9. Hampsey M, Na JG, Pinto I, Ware DE, Berroteran RW. Extragenic suppressors of a translation initiation defect in the *cyc1* gene of *Saccharomyces cerevisiae*. Biochimie 1991;73:1445–1455.
10. Na JG, Pinto I, Hampsey M. Isolation and characterization of *SUA5*, a novel gene required for normal growth in *Saccharomyces cerevisiae*. Genetics 1992;131:791–801.
11. Kawarabayasi Y, Hino Y, Horikawa H, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A, Kosugi H, Hosoyama A, Fukui S, Nagai Y, Nishijima K, Otsuka R, Nakazawa H, Takamiya M, Kato Y, Yoshizawa T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Masuda S, Yanagii M, Nishimura M, Yamagishi A, Oshima T, Kikuchi H. Complete genome sequence of an aerobic thermoacidophilic crenarchaeon. *Sulfolobus tokodaii* strain7 DNA Res 2001;8:123–140.
12. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
13. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D 1999;55(Part 4):849–861.
14. Terwilliger TC. Map-likelihood phasing. Acta Crystallogr D 2001;57(Part 12):1763–1775.
15. Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS. ARP/wARP and molecular replacement. Acta Crystallogr D 2001;57(Part 10):1445–1450.
16. 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 and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D 1998;54(Part 5):905–921.
17. McRee DE. XtalView/Xfit–a versatile program for manipulating atomic coordinates and electron density. J Struct Biol 1999;125: 156–165.
18. Collaborative Computational Project N. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D 1994;50(Part 5):760–763.
19. Holm L, Sander C. Touring protein fold space with Dali/FSSP. Nucleic Acids Res 1998;26:316–319.
20. Gasper R, Scrima A, Wittinghofer A. Structural insights into HypB, a GTP-binding protein that regulates metal binding. J Biol Chem 2006;281:27492–27502.
21. Freymann DM, Keenan RJ, Stroud RM, Walter P. Functional changes in the structure of the SRP GTPase on binding GDP and Mg^{2+}GDP. Nat Struct Biol 1999;6:793–801.
22. Endrizzi JA, Kim H, Anderson PM, Baldwin EP. Crystal structure of *Escherichia coli* cytidine triphosphate synthetase, a nucleotide-regulated glutamine amidotransferase/ATP-dependent amidoligase fusion protein and homologue of anticancer and antiparasitic drug targets. Biochemistry 2004;43:6447–6463.