X-ray crystal structure of a CRISPR-associated RAMP superfamily protein, Cmr5, from Thermus thermophilus HB8

6 AUTHORS, INCLUDING:

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

STRUCTURE NOTE

X-ray crystal structure of a CRISPR-associated RAMP superfamily protein, Cmr5, from *Thermus thermophilus* HB8

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

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

[2] RIKEN Systems and Structural Biology Center, Yokohama Institute, Tsurumi, Yokohama 230-0045, Japan

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

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

INTRODUCTION

Clustered regularly interspaced short palindromic repeats (CRISPRs) are DNA direct repeats present in many prokaryotic genomes.[1–4] They are composed of 24–47 bp direct repeats exhibiting weak dyad symmetry, and separated by 26–72 bp nonrepetitive sequences. Interestingly, some of the spacer sequences are found in foreign replicons such as phages, plasmids, or chromosomes.[5–7] CRISPRs are flanked by CRISPR-associated (*cas*) genes, most of which encode uncharacterized proteins. The CRISPR systems (CRISPRs and Cas proteins) are classified into several subtypes, and each subtype contains several different subtype-specific *cas* genes.[8] The amino acid sequences of several Cas family members are similar to those of nucleases, helicases, RNA- and DNA-binding proteins, or transcription factors. On the basis of their properties, CRISPR systems have been hypothesized to be DNA repair ones[9] or prokaryotic host defense ones against invading foreign replicons.[5–7,10] It was experimentally shown that CRISPR systems are involved in resistance against phages.[11–14] Recently, it was shown that one of the Cas proteins, Cas2, comprises a novel family of endoribonucleases, with a double-split β-α-β fold, and cleaves single-stranded RNAs preferentially within U-rich regions.[15] Furthermore, a complex of Cas proteins comprising repeat-associated mysterious protein (RAMP) superfamily proteins, Cse1, Cse2, Cse3, Cse4, and Cse5e,

cleaves a CRISPR RNA precursor in each repeat and retains the cleavage products containing the virus-derived sequence.[14] Cse3, which is essential for this cleavage, adopts a ferredoxin-like fold, and its overall structure is similar to those of many RNA-binding proteins.[10,16] Cse2 adopts a novel fold with large continuous basic patches on one side of its surface.[17]

An extremely thermophilic bacterium, *Thermus thermophilus* HB8,[18] has 'Ecoli subtype'- and 'Mtube subtype'-like CRISPR systems[8] on megaplasmid pTT27, and expression of the *cas* genes is positively regulated by cyclic AMP receptor protein (CRP).[19] This strain has another operon, TTHB160-161-162-163-164-165, encoding RAMP superfamily proteins on pTT27, which is not present in the vicinity of CRISPRs unlike other *cas* genes (NCBI accession no. NC_006462). Transcription of this operon was significantly decreased in a *crp*-deficient strain although the CRP-dependent promoter was not

found upstream of the operon.[19] In the present study, we determined the crystal structure of one of the RAMP superfamily proteins, *T. thermophilus* Cmr5 (TTHB164), comprising a novel homotrimer.

MATERIALS AND METHODS

Cloning, expression, and purification

The open reading frame of *T. thermophilus* Cmr5 was cloned into the pET-11a expression vector (*Nde*I-*Bam*HI site) (Novagen). A selenomethionine-substituted protein was produced in the *E. coli* methionine auxotroph B834(DE3) strain (Novagen). The cell lysate was heated at 70°C for 10 min. Then the soluble fraction was applied to a SuperQ Toyopearl column (Tosoh Corp.) that had been preequilibrated with 20 mM Tris-HCl (pH 8.0), and the flowthrough fraction was collected. This fraction was applied to a Resource S column (GE Healthcare Biosciences) that had been preequilibrated with 20 mM MES buffer (pH 6.0), and the flowthrough fraction was collected. This fraction was applied to a HiTrap heparin column (GE Healthcare Biosciences) that had been preequilibrated with 20 mM MES buffer (pH 6.0) and was eluted with a linear gradient of 0–2M NaCl. The target fractions were collected and applied to a HiLoad 16/60 Superdex 75pg column (GE Healthcare Biosciences) that had been preequilibrated with 20 mM MES buffer (pH 6.0) containing 0.5M NaCl. The purified protein was concentrated to 9.3 mg/mL using a Centriprep concentrator (3 kDa molecular-weight cutoff, Millipore), and dithiothreitol was added to the sample to a final concentration of 1 mM. The molecular mass of the purified protein determined on gel filtration column chromatography on a HiLoad 16/60 Superdex 75pg column (GE Healthcare Biosciences) was 36 kDa, and that determined on light scattering photometry (DynaPro-801 detector, Protein Solutions) was 40 kDa, suggesting that it exists as a trimer in solution.

Crystallization, data collection, and structure determination

Crystallization of the Cmr5 protein was performed by the sitting drop vapor diffusion method by mixing 0.6 μL of a protein solution with an equal volume of a reservoir solution comprising 0.1M sodium acetate (pH 4.6), 1.7M ammonium sulfate, and 30 mM dithiothreitol at 20°C. Single-wavelength anomalous dispersion data were collected with the RIKEN Structural Genomics Beamline II (BL26B2) at SPring-8 (Hyogo, Japan) using the anomalous scattering from Se atoms. A data set was collected at 2.05 Å resolution using a Mar Mosaic225 CCD detector (Rayonix, LLC). The collected data were processed with the HKL2000 program suite.[20] The positions of 10 of 12 possible Se atoms in the asymmetric unit of the crystal were determined with program SOLVE,[21] and

Table I
X-Ray Data Collection and Refinement Statistics

Data collection	
Wavelength (Å)	0.9790
Resolution (Å)	50–2.05 (2.12–2.05)
Space group	I222
No. of molecules in an asymmetric unit	6
Unit cell parameters (Å, °)	a = 102.122
	b = 114.801
	c = 122.845
	$\alpha = \beta = \gamma = 90$
No. of measured reflections	473,635
No. of unique reflections	45,600
Completeness (%)	99.9 (100)
Redundancy	10.4 (10.6)
I/σ(I)	51.1 (9.4)
R_{merge}[a](%)	6.4 (31.4)
Phasing	
No. of Se atoms used	10
Figure of merit	0.42
Figure of merit after density modification	0.64
Refinement	
Resolution (Å)	43.62–2.10
R_{work}[b](%)/R_{free}[c](%)	23.5/24.9
No. of protein atoms/water atoms	5,299/217
r.m.s.d. bond lengths (Å)	0.006
r.m.s.d. bond angles (°)	1.0
Wilson B factor (Å^2)	21.2
Average B factor for protein (Å^2)	30.8
Average B factor for water (Å^2)	34.6
Ramachandran plot (%)	
Most favored	99.0
Allowed	1.0
Disallowed	0.0

Values in parentheses are for the highest-resolution shell.
[a]$R_{\text{merge}} = \Sigma_h \Sigma_i |I_{h,i} - <I_h>|/\Sigma_h \Sigma_i I_{h,i}$, where $I_{h,i}$ is the ith measured diffraction intensity of reflection h and $<I_h>$ 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.

then density modification was performed with program RESOLVE.[22] The automatic-tracing procedure in program ARP/wARP[23] was used to build the initial model. The model refinement, initial picking, and manual verifying of water molecules were performed using programs CNS and XtalView/Xfit.[24,25] Seven electron densities that probably correspond to sulfate ions were detected in the asymmetric unit. Electron densities corresponding to multiple conformers could not be detected. According to PROCHECK in the CCP4 suite,[26] 99.0% 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 2ZOP.

RESULTS AND DISCUSSION

The *T. thermophilus* Cmr5 monomer is composed of 114 amino acid residues whose molecular mass is calculated to be 13 kDa. A BLAST search revealed that homol-
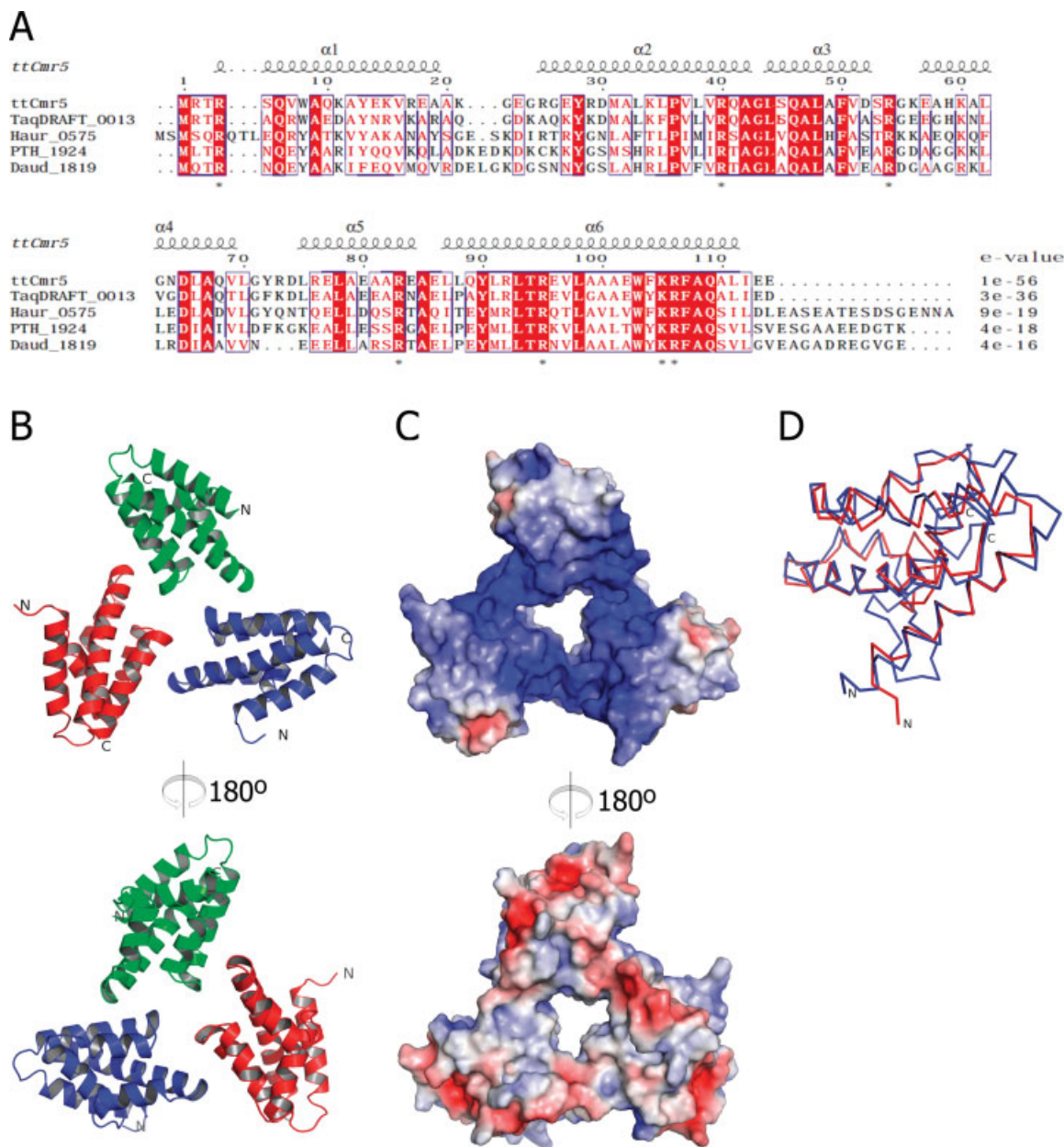
Figure 1

(**A**) Sequence alignment of *T. thermophilus* Cmr5 (ttCmr5) with representative homologous proteins; TaqDRAFT_0013 from *T. aquaticus* Y51MC23; Haur_0575, Cmr5 family protein from *H. aurantiacus* ATCC 23779; PTH_1924, hypothetical protein from *P. thermopropionicum* SI; and Daud_1819, Cmr5 family protein from *C. desulforudis audaxviator* MP104C. Strictly conserved and similar residues are boxed in red and represented by red letters, respectively. Basic residues are indicated by asterisks. The sequences were aligned using ClustalW2.[28] The secondary structure was generated with ESpript 2.2.[29] (**B**) Ribbon diagram of the *T. thermophilus* Cmr5 trimer; chains A (red), B (blue), and C (green). Sulfate ions identified in the structure are not shown. (**C**) Molecular surface representation of the *T. thermophilus* Cmr5 trimer (chains ABC). Red and blue surfaces represent negative and positive electrostatic potentials ($-5\ k_BT$, $+5\ k_BT$), respectively. The electrostatic potentials were calculated using the Adaptive Poisson-Boltzmann Solver (APBS)[30] with PyMol APBS tools. (**D**) Superpositioning of the main chain structures of *T. thermophilus* Cmr5 (red) and AF1862 from *A. fulgidus* DSM 4304 (PDB code: 2OEB) (blue). *B–D* were generated with program PyMol which is available at: http://pymol.sourceforge.net/.

ogous proteins that are most closely related to this protein are TT_P0119 from *T. thermophilus* HB27 (5e-56), there only being one amino acid substitution (D65H), TaqDRAFT_0013 from *Thermus aquaticus* Y51MC23 (3e-36), Haur_0575 from *Herpetosiphon aurantiacus* ATCC 23779 (9e-19), PTH_1924 from *Pelotomaculum thermopropionicum* SI (4e-18), and Daud_1819 from *Candidatus desulforudis audaxviator* MP104C (4e-16) [Fig. 1(A)].

The three-dimensional crystal structure of *T. thermophilus* Cmr5 was determined at a resolution of 2.1 Å, with crystallographic R_{work} and R_{free} factors of 23.5 and 24.9%, respectively (Table I). The asymmetric unit of the crystal comprised two trimers of Cmr5 (trimer 1, chains ABC [Fig. 1(B)]; trimer 2, chains DEF), which are related through a noncrystallographic pseudo-symmetric two-fold axis. The structures of the six monomers are similar, as shown by the r.m.s.d. values of 0.30–0.45 Å for the corresponding Cα atoms. The Cα r.m.s.d. values for the two trimers are 0.75–1.22 Å. The two trimers exhibit the highest correlation when they are related through a pseudo-two-fold axis. There are disordered regions, that is, residue 114 in chain A, residues 113 and 114 in chain B, residues 1–3, 54, 113, and 114 in chain C, residues 1–5, 55, and 114 in chain D, residues 1, 56, and 114 in chain E, and residues 1, 55, and 114 in chain F, which are not included in the model. The Cmr5 monomer consists of six α-helices (chain A: α1, residues 4–19; α2, residues 25–42; α3, residues 44–52; α4, residues 57–69; α5, residues 75–84; and α6, residues 87–111) [Fig. 1(A,B)]. Hydrophobic cores are formed at the trimer interfaces, which are composed of L38, A42, A47, and F50 of one chain, and L87, L88, and L91 of the other chain. The interaction surface areas at the interfaces of chains AB and AC, calculated with the AreaI-Mol program,[26] are 224 and 280 Å^2, which account for 3.4 and 4.2% of the total surface area of chain A. The *T. thermophilus* Cmr5 structure was compared with previously determined structures in the PDB database, using the secondary-structure matching server.[27] The closest structure was that of hypothetical protein AF1862, a Cmr5 family protein from *Archaeoglobus fulgidus* DSM 4304 (PDB code: 2OEB), the root mean square deviation (r.m.s.d.) being 1.59, and the Q-, P-, and Z-scores being 0.53, 3.4, and 5.3 (sequence identity = 28%) [Fig. 1(D)]; however, the AF1862 protein is a monomer unlike *T. thermophilus* Cmr5. *T. thermophilus* Cmr5 has a high theoretical isoelectric point of approximately 9.8, and has large continuous basic patches on one side of its surface, and acidic patches on the other side [Fig. 1(C)]. The amino acid residues comprising the basic and acidic patches are conserved in the four closest homologs to *T. thermophilus* Cmr5 [Fig. 1(A)]. Furthermore, this structural feature is conserved in AF1862, although its basic patches are smaller than those of *T. thermophilus* Cmr5. RAMP has been predicted to be a RNA-binding protein in a putative host defense system,[10] and it was experimentally shown that a complex of RAMP superfamily proteins cleaves RNA.[14] It might be that Cmr5 interacts with RNA through its basic patches.

ACKNOWLEDGMENTS

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. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 2008;321:960–964.
15. 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.
16. 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.
17. Agari Y, Yokoyama S, Kuramitsu S, Shinkai A. X-ray crystal structure of a CRISPR-associated protein, Cse2, from *Thermus thermophilus* HB8. Proteins 2008;73:1063–1067.

18. 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.

19. 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.

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. 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.

28. 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.

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.