WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

Crystal structure of a conserved protein of unknown function (MJ1651) from *Methanococcus jannaschii*

Krishnamurthy N. Rao,[1] Stephen K. Burley,[2] and Subramanyam Swaminathan[1]*

[1] Biology Department, Brookhaven National Laboratory, Upton, New York 11973

[2] SGX Pharmaceuticals, Inc., San Diego, California 92121

***Key words:* X-ray structure; MJ1651; archaea; DUF62; Pfam; hexamer.**

INTRODUCTION

Methanococcus jannaschii, a strict anaerobic methane-producing organism, provided the first complete genome sequence reported for an archaeon.[1] This exercise provided an opportunity to compare the three domains of life at the genetic level for the first time. We have cloned the gene MJ1651 (gi: 15669847) from *Methanococcus jannaschii* DSM 2661 and over expressed its 263-residue protein product in *E. coli* to investigate both three-dimensional structure and function. MJ1651 is a member of the DUF62 Pfam family (pfam 01887),[2] which contains both archeal and eubacterial members of no known function. Herein, we report the X-ray crystal structure of MJ1651 determined at 2.5 Å resolution via the single-wavelength anomalous diffraction (SAD) method using selenium-labeled protein. The structure is composed of two domains: an N-terminal αβα sandwich domain and a C-terminal β-strand domain. As expected, the structure of MJ1651 resembles those of other DUF62 family members. Although its structure is most closely related to that of a bacterial fluorinating enzyme, we suggest that the biological function of MJ1651 is different.

MATERIALS AND METHODS

Gene cloning and protein production

The gene for MJ1651 was amplified using polymerase chain reaction from *Methanococcus jannaschii* genomic DNA using a forward (GGGATTTATATGAGGGATGATATTTTAG) and a reverse (CAATCTCAATCTCATCTAAATAATCAACATTC) primer. The amplified gene was gel purified and cloned into TC - psb3 (TC) vector designed to express the protein of interest with a C-terminal hexahistidine affinity tag. Protein expression/purification utilized previously published protocols, which are described in detail in PepcDB (pepcdb.pdb.org).

Crystallization and data collection

Single crystals of MJ1651 were grown via sitting-drop vapor diffusion [2 μL protein solution (6–10 mg/mL) mixed with 2 μL reservoir solution (600 μL) containing 50% MPD, 0.2*M* ammonium dihydrogen phosphate, Tris-HCl, pH 8.0]. Crystals were flash frozen in liquid nitrogen following addition of 15–20% (v/v) glycerol to the mother liquor. Diffraction data were collected under standard cryogenic conditions using beamline X29 at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory. Data were processed using *HKL*2000.[3] Crystals grew in the monoclinic space group C2 (unit-cell parameters: $a = 126.8$ Å, $b = 110.1$ Å, $c = 75.8$ Å, and $\beta = 115.4^\circ$). The calculated Matthews coefficient is 2.52 Å^3/Da, assuming three molecules per asymmetric unit, which corresponds to 51% solvent content by volume. Crystal and data collection statistics are provided in Table I.

*Correspondence to: S. Swaminathan, Biology Department, Brookhaven National Laboratory, Upton, NY 11973. E-mail: swami@bnl.gov

Received 22 May 2007; Accepted 25 May 2007

Published online 1 October 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21646

Table I
Data Collection and Refinement Statistics

Wavelength (Å)	0.9789
Resolution (Å)	40–2.36
Outermost shell (Å)	2.45–2.36
No. of unique reflections	72,852
Redundancy	6.6 (4.0)[a]
R-merge[b] **(%)**	10.5 (46.5)
Overall completeness (%)	99.5 (98.1)
$<I/\sigma(I)>$	14.7
Refinement statistics	
Resolution range (Å)	40–2.50
Outermost shell (Å)	2.61–2.50
No. of reflections	61,146
R-factor[c]	0.237
R-free[d]	0.265
No. of protein atoms	5539
No. of water molecules	31
RMS deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.30
Ramachandran plot statistics	
Residues in (%)	
Core region	85.3
Additionally allowed	3.7
Generously allowed	0.7

[a]Values corresponding to the outermost shell are given within parentheses.
[b]*R*-merge = $\sum_h\sum_i|I_{h,i} - I_h|/\sum_h\sum_i I_{h,i}$ where $<I_h>$ is the mean intensity of symmetry-related reflections, $I_{h,i}$.
[c]*R*-factor = $\sum ||F_o| - |F_c||/\sum|F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.
[d]*R*-free is calculated for the 2% of the data that was withheld from refinement.

Structure determination

The selenium substructure was determined using SOLVE,[4] permitting location of 10 of a possible 15 heavy atom sites. Automated model building with RESOLVE yielded structural information for ~50% of the three polypeptide chains comprising the asymmetric unit. An entire polypeptide chain, barring a few segments, was assembled from the output of RESOLVE and the balance of the model was manually built using O.[5] The complete asymmetric unit containing all three protomers (A, B, and C) was then generated using the requisite non-crystallographic symmetry (NCS) operators. Refinement was performed using CNS.[6] The following termini and loop segments were not visible in the experimental electron density map and were not included in structure refinement (A: 1–8, 87–90, 132–139, 190–199, and 265–273; B: 1–10, 87–90, 132–140, 195–196, and 265–273; C: 1–10, 87–90, 132–139, 192–194, and 265–273). The final refined model of MJ1651 consists of three protomers and 31 solvent atoms. Refinement statistics are provided in Table I.

Chlorinase assay

The assay was carried out in a total reaction volume of 200 μL. Fifteen micrograms of protein was incubated with 0.8 m*M* SAM (*S*-adenosyl-L-methionine), 10 m*M*

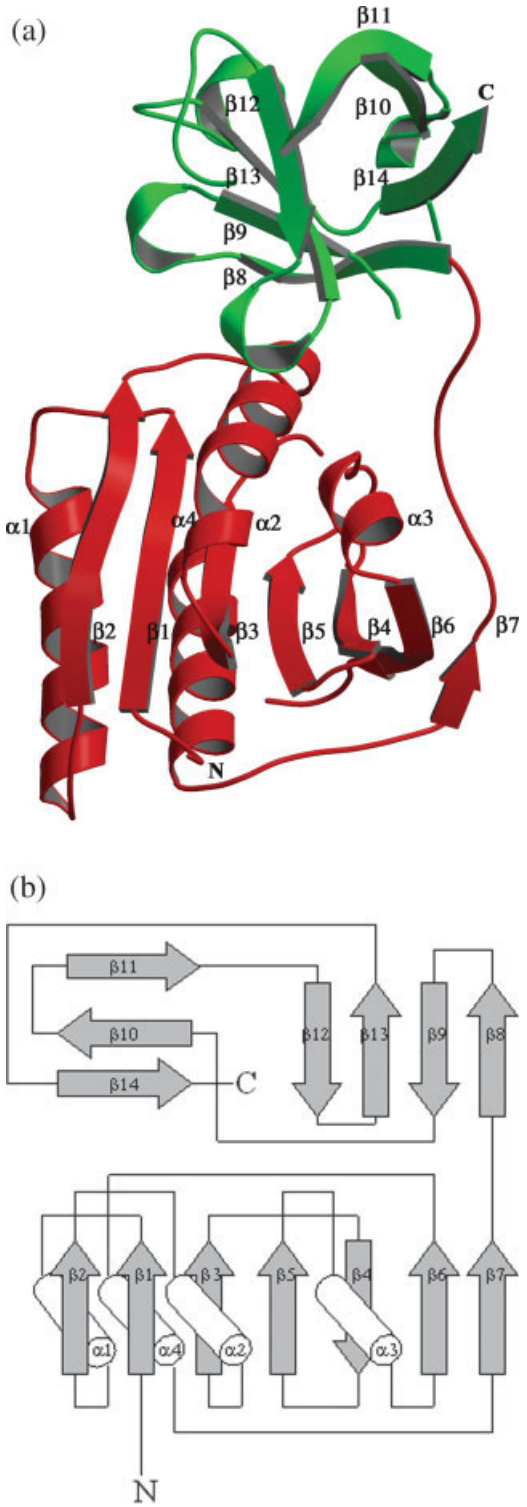

Figure 1
*(**a**) Ribbons representation of the structure of MJ1651 with the N-terminal domain in red and the C-terminal domain in green. The figure was prepared using Molscript.[9] (**b**)Topology diagram depicting the secondary structure of MJ1651. β-strands are shown as arrows and α-helices as cylinders.*

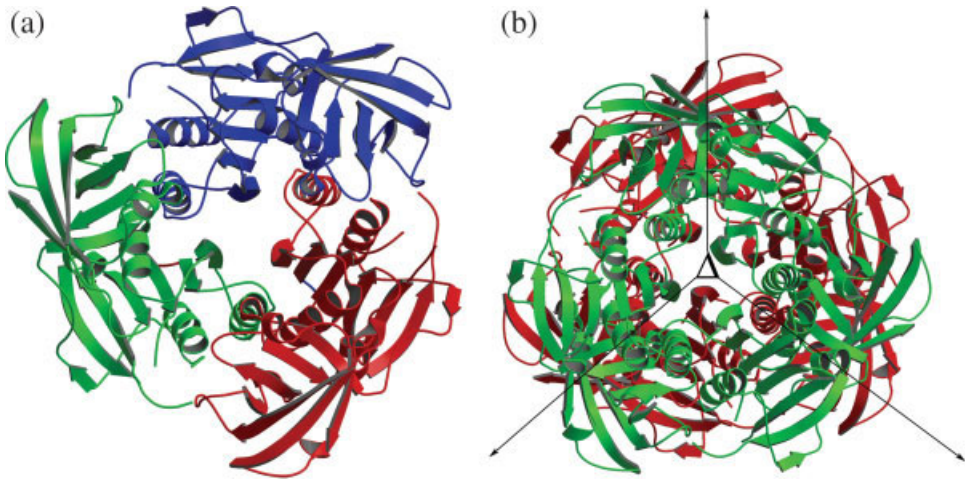

Figure 2

*(**a**) Ribbon diagram of MJ1651 trimer viewed down the non-crystallographic threefold axis. Each monomer is shown in different color. (**b**) Ribbon diagram showing the MJ1651 hexamer assembly with the two trimers colored red and green.*

NaCl, 10 μg of L-amino acid oxidase, and 50 m*M* Tris-HCl (pH 7.9) at 37°C for 3 h. Samples were boiled at 95°C for 3 min to stop the reaction. The reaction product 5′-ClDA (5′-chloro-5′-deoxyadenosine) was monitored by liquid chromatography mass spectrometry (LC-MS) using synthetic 5′-ClDA as a standard.[7,8] SAM, 5′-ClDA, and L-amino acid oxidase were purchased from Sigma-Aldrich Protein Production.

RESULTS AND DISCUSSION

Structure of MJ1651

MJ1651 adopts the α/β fold [Fig. 1(a)]. The molecule has approximate dimensions of 55 × 35 × 35 Å^3, and consists of two domains: the N-terminal core domain and the small C-terminal domain. The N-terminal domain (residues 9–168) is composed of a seven-stranded mixed β-sheet (β1–β7) sandwiched by two α-helices on each side (α1 and α4; α2 and α3). The C-terminal domain (residues 174–264) is composed of two β-sheets. The four (β8, β9, β12, β13) and three (β10, β11, β14) antiparallel stranded β-sheets are arranged perpendicular to one another forming a small β barrel. A cartoon representation of the secondary structure is given in Figure 1(b). A 10-residue random coil segment connects the N-terminal and C-terminal domains.

Biological unit

The crystallographic asymmetric unit contains three protomers arranged as a trimer about a threefold NCS axis [Fig. 2(a)]. The three N-terminal domains are tightly clustered around the threefold axis, making intimate contact with each other. The C-terminal domains make no intermolecular contacts with each other. Instead, they interact with the N-terminal domains of other monomers making up the trimer. Within the monoclinic crystal form we obtained, MJ1651 forms a hexamer via crystallographic symmetry with the two asymmetric units forming a dimer of trimers. The two trimers associate front-to-front [Fig. 2(b)] via interactions mediated by three β11 strand from each trimer. After such an association, three six-stranded extended β-sheets formed firmly holds the hexamer assembly. Strands β10, β11, and β14 of each monomer from both the trimers associate to form such three six-stranded continuous β-sheets. We believe that this larger assembly represents the biologically active form of MJ1651, because analytical gel filtration documented that the protein occurs as a hexamer in solution.

Comparison with homologous structures

A DALI[10] search revealed a number of structures similar to that of MJ1651, including PH0463 from *Pyrococcus horikoshii* OT3 [PDB ID: 1WU8, *Z* score = 23.8, 36% sequence identity, and root-mean-square deviation (rmsd) = 2.1 Å for 212 structurally equivalent α-carbon atomic pairs], a conserved protein of unknown function (TTHA0338) from *Thermus thermophilus* HB8 (2CW5, *Z* = 24.2, sequence identity = 26%, and rmsd = 2.0 Å for 212 pairs), and a fluorinase from *Streptomyces cattleya* (1RQR/2C2W, *Z* = 21.0, sequence identity = 23%, and rmsd = 2.8 Å for 217 pairs). Each of these homologs, including the fluorinase (1RQR), belongs to the same Pfam family, DUF62 Pfam.

DOI 10.1002/prot

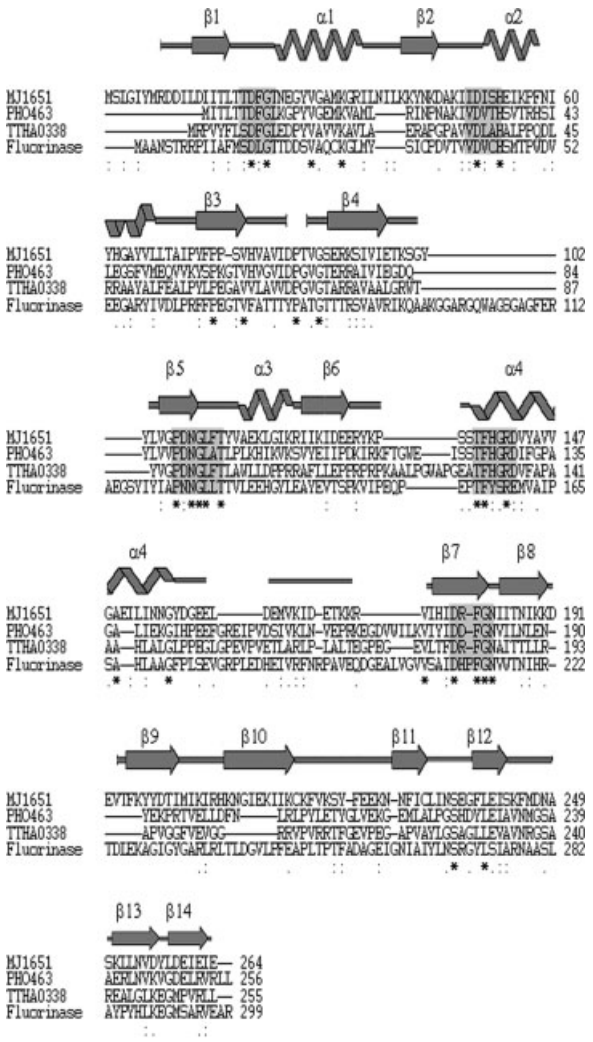

Figure 3

Multiple sequence alignment of MJ1651 with other available DUF62 family structures in PDB. Residues belonging to conserved motifs are highlighted in gray. The alignment was constructed using CLUSTAL-W[12] *with the bottom line representing the degree of homology – "*" identical, ":" conserved, and "." semiconserved.*

Although the fold of the N-terminal domain of all the DUF62 family structures is the same there are some differences in the structure of C-terminal domain. The four-stranded β-sheet (β8, β9, β12, β13) is common to all the structures. The other β-sheet is made up of five strands in both 1WU8 and fluorinase, four strands in 2CW5, and three strands in MJ1651 [β10, β11, β14; Fig. 1(b)]. Despite this difference, the strand β11 is conserved in both MJ1651 and fluorinase and they could form the same hexamer assembly.

A PSI-BLAST[11] search for sequence homologs of MJ1651 identified a number of homologs from both eubacteria and archaea, with several conserved motifs (Fig. 3), including TDFGT (motif 1, residues 19–24 forming β1−α1 loop), IDISH (motif 2, residues 49–53 in the β2-strand), PDNGLFT (motif 3, residues 107–113 of helix α3), TFHGRD (motif 4, residues 137–142 forming α4−β7 loop), and DRFGN (motif 5, residues 179–183 forming β8−β9 loop). Motif 1 of molecule A interacts with motif 5 of molecule C. The IDISH conserved motif 2 residues from each monomer are symmetrically arrayed around the threefold NCS axis near the center of the trimer, where they appear to engage in stabilizing interactions. Residues of motif 3 interact with the loop that connects the N-terminal to C-terminal domain. Motif 4, which is not visible in our structure, is located close to where the N-terminal domain of molecule A contacts the C-terminal domain of molecule C in known structural homologs. At the trimer interface, the interaction is mainly between the residues belonging to α1 helix of molecule A and the residues of β1 strands, β2 strands, and α2 helix of molecule C. In addition, residues belonging to loop β1−α1 of molecule A interact with residues of loop β8−β9 of molecule C (molecule B interacts with molecule A in a similar way the molecule A interacts with molecule C). In the sequence alignment illustrated in Figure 3, an extended loop of about 15 residues in the N-terminal domain occurs only in the fluorinase. From our structural analysis it appears that the conserved motifs are important for the structural integrity and quaternary structure of the DUF62 family proteins.

Quaternary structure

Both MJ1651 and the flourinase form front-to-front hexamers via crystallographic symmetry with the two asymmetric units forming a dimer of trimers. Interaction between the two trimers occurs exclusively via C-terminal domain for both MJ1651 and the fluorinase. The three β11 strands from each trimer associate to form a six-stranded continuous β-sheet in MJ1651 and a 10-stranded β-sheet in fluorinase. It is interesting to see here that though MJ1651 and fluorinase differ in the number of strands in this β-sheet (β10, β11, β14 in MJ1651), they form the same hexamer assembly. It is remarkable that the secondary structural elements required for both the trimer and hexamer structural association are preserved in MJ1651 and fluorinase structures. In contrast, 1WU8 forms a back-to-back hexamer via crystallographic symmetry, via interactions between N-terminal domains.

Chlorinase activity assay and putative active site identification

X-ray crystal structures of the fluorinase bound to substrate and product has been reported.[7] This enzyme converts inorganic fluorine to organic fluoride by catalyzing C—F bond formation. Fluorinase also acts as a chlorinase.[8] Among the DUF62 homolog proteins, the fluori-

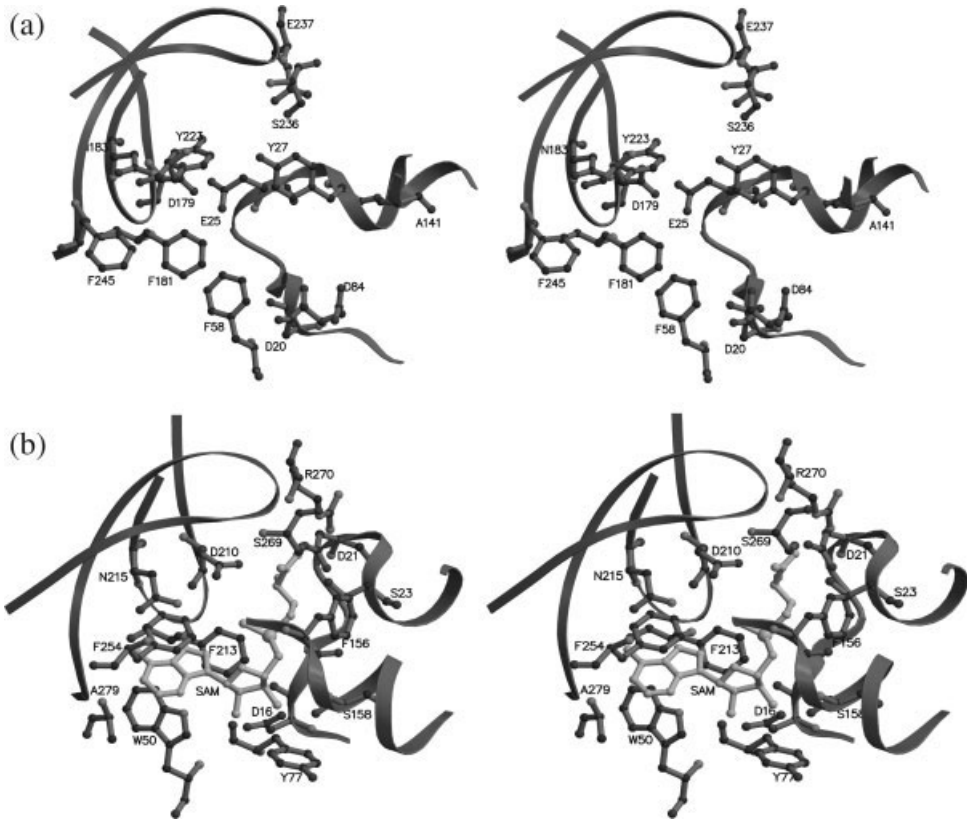

Figure 4

*Stereodrawing of (**a**) putative active site of MJ1651 and (**b**) active site of fluorinase in complex with the substrate, SAM (S-adenosyl-L-methionine). The residues, both conserved and variant, seen near the active site are shown in ball and stick model. The conserved secondary structure elements close to the active site are shown in ribbons representation.*

nase (chlorinase) from S*treptomyces cattleya* is the only enzyme with defined biological activity. Despite very low sequence identity (15%), the structures of MJ1651 and fluorinase are quite similar (rmsd = 2.8 Å). An assay of halogenase (chlorinase) activity with MJ1651 used for crystallization failed to show significant chlorinase activity.

The X-ray crystal structure of 1WU8, a protein of unknown function, showed an adenosine molecule bound in position analogous to that occupied by the adenine moiety of SAM in the active site of the fluorinase.[7] Comparison of this region of the structures 1WU8, the fluorinase, and MJ1651 revealed considerable similarity, which suggests that it represents the active site of DUF62 family members. The fluorinase and MJ1651 share a number of conserved residues in this putative active site, including Asp20 (Asp16 in fluorinase), Phe136 (Phe156), Asp179 (Asp210), Phe181 (Phe213), Asn183 (Asn215), and Ser236 (Ser269). Other residues lining the putative active site are Glu25 (Asp21 in fluorinase), Tyr27 (Ser23), Asp84 (Tyr77), Ala141 (Ser158), Glu237 (Arg270), and Phe245 (Ala279) [Fig. 4(a,b)]. Residues from Asp179 onwards are contributed by the second promoter of the asymmetric unit. In the fluorinase, Ser158 participates in catalysis, binding to the halogen and O_4 of SAM. We suggest that these differences in the putative active site reflect functional differences between the fluorinase and MJ1651.

CONCLUSIONS

We described the determination of the structure of MJ1651 from *Methanococcus jannaschii*. Structure and sequence comparison with homologous proteins helped us to identify common structural features that explain trimeric/hexameric self-association of DUF62 family members. We also identified the putative active site of MJ1651, and documented that the protein is not a chlorinase.

The structure has been deposited to the Protein Data Bank (ID: 2F4N).

ACKNOWLEDGMENTS

Research was supported by a U54 award from the National Institute of General Medical Sciences to the NYSGXRC (GM074945; PI: Stephen K. Burley) under

DOI 10.1002/prot

DOE Prime Contract No. DEAC02-98CH10886 with Brookhaven National Laboratory. We thank Dr. Howard Robinson for providing data collection facilities at the National Synchrotron Light Source. We also thank Dr. Chang-Jun Liu for his help in LC-MS analyses.

REFERENCES

1. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NS, Venter JC. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 1996;273:1058–1073.
2. 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.
3. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
4. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 2002;58(Part 10, No. 2):1772–1779.
5. Jones TA, Zou J, Cowtan S, Kjeldgaard M. Improved methods in building protein models in electron density map and the location of errors in these models. Acta Crystallogr A 1991;47:110–119.
6. Brunger AT, Adams PD, Clore GM, Delano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Somonsom T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
7. Dong C, Huang F, Deng H, Schaffrath C, Spencer JB, O'Hagan D, Naismith JH. Crystal structure and mechanism of a bacterial fluorinating enzyme. Nature 2004;427:561–565.
8. Deng H, Cobb SL, McEwan AR, McGlinchey RP, Naismith JH, O'Hagan D, Robinson DA, Spencer JB. The fluorinase from *Streptomyces cattleya* is also a chlorinase. Angew Chem Int Ed Engl 2006;45:759–762.
9. Kraulis P. MOLSCRIPT: a program to produce both detailed and schematic plots of proteins. J Appl Crystallogr 1991;24:946–950.
10. Holm L SC. A network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.
11. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
12. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the clustal series of programs. Nucleic Acids Res 2003;31:3497–3500.