Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the hypothetical protein AQ_1354 from *Aquifex aeolicus*

Vaheh Oganesyan,[a] Didier Busso,[b] Jeroen Brandsen,[b] Shengfeng Chen,[b] Jaru Jancarik,[b] Rosalind Kim[a] and Sung-Hou Kim[a,b]*

[a]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA, and [b]Department of Chemistry, University of California, Berkeley, California 94720, USA

Correspondence e-mail: shkim@cchem.berkeley.edu

Received 7 April 2003
Accepted 19 May 2003

PDB Reference: AQ_1354, 1oz9, r1oz9sf.

In memory of Jeroen Brandsen.

The crystal structure of a hypothetical protein AQ_1354 (gi 2983779) from the hyperthermophilic bacteria *Aquifex aeolicus* has been determined using X-ray crystallography. As found in many structural genomics studies, this protein is not associated with any known function based on its amino-acid sequence. PSI-BLAST analysis against a non-redundant sequence database gave 68 similar sequences referred to as 'conserved hypothetical proteins' from the uncharacterized protein family UPF0054 (accession No. PF02310). Crystallographic analysis revealed that the overall fold of this protein consists of one central α-helix surrounded by a four-stranded β-sheet and four other α-helices. Structure-based homology analysis with DALI revealed that the structure has a moderate to good resemblance to metal-dependent proteinases such as collagenases and gelatinases, thus suggesting its possible molecular function. However, experimental tests for collagenase and gelatinase-type function show no detectable activity under standard assay conditions. Therefore, we suggest either that the members of the UPF0054 family have a similar fold but different biochemical functions to those of collagenases and gelatinases or that they have a similar function but perform it under different conditions.

1. Introduction

Large-scale genome-sequencing projects reveal that, on average, the functions of more than 50% of the predicted proteins cannot be inferred based on their amino-acid sequences alone. For those with inferred functions, experimental assays may or may not support the functional prediction. A structural approach has sometimes proven to be a valid way of deducing the molecular functions of hypothetical proteins (Zarembinski *et al.*, 1998; Hwang *et al.*, 1999; Teplova *et al.*, 2000; Schulze-Gahmen *et al.*, 2003). Based on such observations, the Protein Structure Inititative of the National Institutes of Health attempts to obtain the structures of representatives from all protein families and to associate one or more molecular function with each protein-sequence family in order to obtain a global view of the protein universe in terms of their structure and function (Hou *et al.*, 2003). Here, we present an example where two different protein-sequence families have a similar fold but may not have similar molecular functions.

2. Materials and methods

2.1. Target

Pfam (http://pfam.wustl.edu) is a large collection of protein-sequence families based on multiple sequence alignments and hidden Markov models. Pfam version 8.0 (February 2003)

contains alignments and models for 5193 protein families, based on the Swiss-Prot 40.31 and SP-TrEMBL 22.0 protein-sequence databases. The Pfam database contains 68 members of the protein family UPF0054 (accession No. PF02130) to which AQ_1354 belongs. These are small evolutionarily related proteins of 17–21 kDa which contain at their C-terminus a conserved region with three histidines, also called the 'UPF0054 signature'. None of the members have substantial sequence homology to proteins of known structure and/or function. The sequence alignment around the 'UPF0054 signature' for several members of this family is shown in Fig. 1.

2.2. Cloning, expression and purification

The sequence encoding AQ_1354 was amplified by the polymerase chain reaction (PCR) from genomic DNA of the hyperthermophilic bacteria *Aquifex aeolicus* (Zhang *et al.*, 2001; Huber *et al.*, 1992) using a forward primer harbouring an *Nde*I site and a reverse primer harbouring a Stop codon and a *Bam*HI site. The resulting PCR product was inserted using the ligation activity of topoisomerase into a TOPO vector (Invitrogen, Carlsbad, California, USA). At this stage, the correctness of the PCR product was confirmed by DNA sequencing. A DNA insert encoding AQ_1354 was prepared by digestion using *Nde*I and *Bam*HI (New England Biolabs, Beverly, MA, USA). After gel purification using the Strata-Prep DNA gel-extraction kit (Stratagene, La Jolla, CA, USA), the DNA insert was ligated into the pHG expression vector digested by the corresponding enzymes, resulting in the pHG-1108B vector. The expression vector encoding for AQ_1354 was transformed into a methionine auxotroph, *Escherichia coli* strain B834(DE3)-pSJS1244 (Leahy *et al.*, 1992; Kim *et al.*, 1998). The expressed protein harboured a six-histidine tag, a GST tag and an m tobacco etch virus (mTEV) cleavage sequence at its N-terminus.

M9 medium supplied with selenomethionine and appropriate antibiotics was used for protein expression. The culture was grown at 310 K until OD_{600nm} reached 0.5 and was transferred to 303 K before induction with 0.5 m*M* IPTG. After overnight growth, the cells were harvested and resuspended in 80 ml of 50 m*M* HEPES pH 7.0 buffer. The cells were disrupted by microfluidization (Microfluidics, Newton, MA, USA) and centrifuged for 30 min at 35 000*g* and 277 K. The supernatant was supplemented with 100 m*M* NaCl, 5 m*M* imidazole and 10% glycerol before incubation in a batch method with 3 ml of Talon resin (Clontech) equilibrated with buffer *A* (50 m*M* HEPES pH 7.0, 300 m*M* NaCl) containing 10 m*M* imidazole. After 1 h incubation at 277 K, the resin was poured into a column and washed extensively with buffer *A* containing 5 m*M* imidazole. The protein was eluted with one column volume of buffer *A* containing 150 m*M* imidazole and three column volumes of buffer *A* containing 300 m*M* imidazole. All fractions were pooled and diluted twice with buffer *A* containing 300 m*M* imidazole. The protein was dialyzed overnight at room temperature against 2 l buffer *B* (20 m*M* Tris–HCl pH 7.5, 1 m*M* DTT, 1 m*M* EDTA) containing 10 m*M* NaCl in the presence of 175 µg of mTEV (1 µg of mTEV cleaves 1 nmol of recombinant protein). After centrifugation, the supernatant was applied onto a HiTrap Q (1 ml) column equilibrated with buffer *B* containing 10 m*M* NaCl. The cleaved recombinant protein was in the flowthrough. The SDS–PAGE showed one band around 17 kDa and dynamic light scattering (DynaPro 99, Proterion Corporation, Piscataway, NJ, USA) showed a monodisperse peak with a radius of ~2 nm, which corresponds to a molecular weight of ~20 kDa. The sample was concentrated using a Centriprep 3K device (Millipore Corp., Bedford, Massachusetts, USA) to reduce the volume to 500 µl and was further concentrated using an 'Ultrafree' 5K unit (Millipore Corp., Bedford, Massachusetts, USA).

2.3. Crystallization and data collection

The protein was concentrated to 8 mg ml^{-1} in 20 m*M* Tris–HCl buffer pH 7.0, 10 m*M* NaCl, 1 m*M* DTT, 1 m*M* EDTA. Screening for crystallization conditions was performed using the sparse-matrix method (Jancarik & Kim, 1991) implemented in various screens from Hampton Research (Hampton Research, Laguna Niguel, California, USA). The crystal used for data collection was grown in one week in a hanging drop using the vapor-diffusion method. The reservoir contained 200 m*M* sodium acetate, 100 m*M* Tris–HCl pH 8.5 and 30% PEG 4K (directly from condition #22, Crystal Screen).

The X-ray diffraction data were collected from a single flash-frozen (100 K) SeMet-containing crystal to 1.89 Å resolution at a single wavelength corresponding to the selenium absorbance peak (λ = 0.9792 Å) at the Macromolecular Crystallography facility beamline 5.0.2 of the Advanced Light Source at Lawrence Berkeley National Laboratory (ALS at LBNL, Berkeley, CA, USA). A Quantum 4 charge-coupled device detector from Area

```
Y053_SYNY3      TGDRRIQELNRKFRHQDKP..TDVLAFAALEGDFP....LIESEETEEPLYLGDIV
YQFG_BACSU      VSNDDIHQINKEYRGKDAP..TDVISFALEEEGEG....EIEIVGAEMPPVLGNII
YB60_HELPY      VSDETIREINKDLRGCDYA..TDVLSFPLEAI.P..............HTPLGSVV
YN67_MYCLE      VLLDTAAMANLHMRWMDLPGPTDVMSFPMDEFEPG....GRPDAAEPGPSMLGDIV
YBEY_HAEIN      VDEAESHELNLTYRGKDRP..TNVLSFPFECPDE............VELPLLGDLV
Y741_RICPR      TNTAEILTLNQQFRSIEKA..TNVLSFPNNELNWHNLYSKLEFLYYSDYMHLGDIA
  AQ_1354       TDDQEIRELNKTYRKKDKP..TDVLSFPMGEEFGG............YKILGDVV

Y053_SYNY3      ISLERADHQARERGHSAKLEVVWLTAHGLLHLLGWDHPDEASLTTM.....LSEQ
YQFG_BACSU      ISADRTREQAEEYNHSFKRELGFLAVHGFLHLLGYDHMTKEEEEEM.....FTKQ
YB60_HELPY      INAPLAQTNALKLGHSLENEIALLFIHGVLHLLGYDHEKDKGEQ.......RQKE
YN67_MYCLE      LCPEFAAQQAAAEGHSLGHELALLTIHGVLHLLGYDHGEPDEEKEMFALQGRLLE
YBEY_HAEIN      ICRQVVEREASEQEKPLMAHWAHMVVHGSLHLLGYDHIEDDEAEEM.....ESLE
Y741_RICPR      FCYEVIYNEACEQQKTFENHFIHMLIHGILHLIGFDHQNDTDANIM.....ESLE
  AQ_1354       ISQDTAERQARELGHSLEEEVKRLIVHGIVHLLGYDHEKGGEEEKKFRELENYVL
```

Figure 1
Amino-acid sequence alignment of the C-terminal region of several members of the UPF0054 family. Identical residues are shown in red. The aligned sequences and their length were chosen by the program *PFAM*8.0 (http://pfam.wustl.edu/) as representatives. AQ_1354 is added to show the degree of similarity. The alignment was generated using the program *ClustalW* (Altschul *et al.*, 1997).

Detector Systems Co. (Poway, CA, USA) was used. The crystal-to-detector distance was set to 200 mm. 246 images were collected (120 and 126 images for the direct and inverse beams, respectively) in several sweeps with a 1° oscillation range. The X-ray diffraction data were processed and scaled using *DENZO* and *SCALEPACK* from the *HKL* program suite (Otwinowski & Minor, 1996). The crystal belongs to the primitive tetragonal space group $P4_32_12$, with unit-cell parameters $a = b = 58.02$, $c = 110.26$ Å.

With one molecule in the asymmetric unit, the volume fraction of the unit cell occupied by atoms is 50.8%. A least-squares straight line approximates the *B* factor as around 30 Å^2. After being flash-frozen, the mosaicity of the crystal was 0.5°. The total number of reflections used for anomalous scaling was 262 069. The data statistics are summarized in Table 1.

2.4. Structure determination and refinement

The crystal structure of AQ_1354 was solved using the single-wavelength anomalous diffraction (SAD) method. The protein contained only one SeMet residue per 150 residues (excluding the N-terminus).The *SOLVE*/*RESOLVE* program suite (Terwilliger, 2002) was used to locate that single Se atom and improve the phases by using statistical density modification. The *SOLVE* program gave consistent results at different resolution cutoffs when space group $P4_32_12$ was used. The best figure of merit of 0.294 was achieved at a resolution of 2.5 Å. The *RESOLVE* program was used to extend phases to 2.2 Å and to apply density modification. The resulting electron density allowed tracing of 140 residues including their side chains using the program *O* (Jones *et al.*, 1991). Refinement was carried out using *REFMAC*5 from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) at full resolution to 1.89 Å. Water molecules initially were placed by hand and in the final stages the water-picking procedure in *CCP*4 was used with a σ-cutoff of 2.5. The refinement converged after several alternating cycles of manual model building and refinement. Eight residues from the N-terminus and two residues from the C-terminus are not visible in the electron density. The phasing and refinement statistics are summarized in Table 2.

3. Results and discussion

3.1. Overall structure

AQ_1354 is a single-domain molecule of 150 residues with an overall topologic arrangement of alternating β-strands and α-helices. There is one four-stranded mixed β-sheet with strand order 1243 and five α-helices. The first three strands (residues 8–17, 40–48 and 68–78) of the β-sheet are parallel and each strand is followed by an α-helix (22–37, 49–61 and 91–102). The fourth helix consists of residues 105–122 and is positioned in the center of the molecule and is surrounded by all the other secondary-structure elements. The fifth helix (residues 129–148) completes the circle around the

Table 1
Statistics of the X-ray diffraction data.

Values in parentheses are for the outermost resolution shell.

Wavelength (Å)	0.9792 (peak)
Resolution (Å)	41.2–1.89 (1.97–1.89)
Redundancy	9.3
Unique reflections (with Bijvoet pairs)	28054
Completeness (%)	96.3 (85.8)
$I/\sigma(I)$	24.5 (2.25)
$R_{\rm sym}$† (%)	5.5 (46)

† $R_{\rm sym} = \sum_h \sum_i |I_{h,i} - \langle I \rangle| / \sum_h \sum_i |I_{h,i}|$.

Table 2
Phasing and refinement statistics of the SAD X-ray diffraction data.

Phasing statistics	
Wavelength (Å)	0.9792 (peak)
Maximum resolution (Å) for phasing	2.5
Maximum resolution (Å) for density modification	2.2
Total reflections	269062
Unique reflections (with Bijvoet pairs)	28054
Se atoms	1
FOM	0.294
FOM after density modification	0.457
Refinement statistics	
Resolution (Å)	1.89
Unique reflections (Bijvoet pairs merged)	15,227
R factor	0.201
$R_{\rm free}$†	0.234
No. of atoms	1202
Protein	1151
Water	51
R.m.s. deviation	
Bond distance (Å)	0.013
Angle distance (°)	1.524
Residues in most favored region	122 (96.1%)
Residues in additional allowed regions	5 (3.9%)
Residues in other regions‡	0

† $R_{\rm free}$ calculated as *R* factor but on 5% of data set aside from refinement. ‡ Number of end residues = 2, number of glycines = 10, number of prolines = 2; a total of 141 residues.

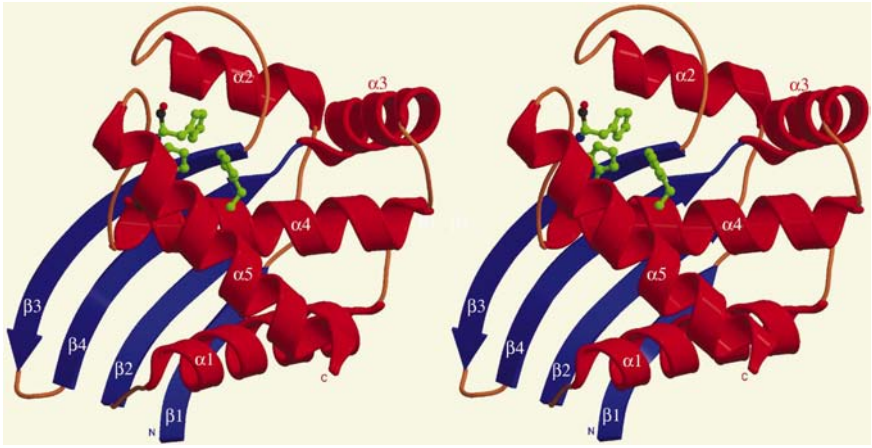

Figure 2
Overall fold of AQ_1354. Histidine residues conserved throughout the UPF0054 family and some metalloproteinases are shown in green. Illustration prepared with *MOLSCRIPT* (Kraulis, 1991) and *Raster3D* (Merritt & Murphy, 1994).

central fourth helix. The overall fold of AQ_1354 is shown in Fig. 2. The electron density around the hypothetical active site is shown in Fig. 3.

After tracing and refinement of the structure was finished, a DALI search (Holm & Sander, 1993) was initiated to detect similarity in fold to proteins of known spatial structure. The search was performed against a database consisting of 3241 polypeptide chains as of November, 2002 (non-redundant at less than 30% of amino-acid sequence identity). A highest score of 6.6 was obtained which corresponded to human fibroblast collagenase (PDB code 1hfc; Pfam accession No. PF00413; extracellular metalloproteinases, such as collagenase and stromelysin, which degrade the extracellular matrix, are known as matrixins, M10 subfamily; Spurlino *et al.*, 1994). The next two highest scores of 6.0 and 5.6 corresponded to deuterolysin (PDB code 1eb6; accession No. PF02102; deuterolysin metalloproteinase family, M35 subfamily; McAuley *et al.*, 2001) and type IV collagenase (PDB code 1ck7, full length; accession No. PF00413; gelatinase, M10 subfamily; Morgunova *et al.*, 1999), respectively. All three proteins found by DALI are classified as matrix metalloproteinases. Detailed analysis with each of the three matrix metalloproteinases has shown that the highest structural similarity is with the protein with the highest score in the DALI search. The amino-acid sequence alignment between AQ_1354 and those three proteins is shown in Fig. 4. Sequence identities for these fold homologues range from 9.33% with 1eb6 to 16.0% with 1hfc.

The superposition of those four structures revealed that only the region around the 'UPF0054 signature' motif H*xxx*H*xx*G*xx*H is well conserved (with one exception, where the last His is a Gln in 1eb6). As expected, the highest sequence identity with human fibroblast collagenase also gave the highest structural similarity with that enzyme. When the three-dimensional structures of AQ_1354 and human fibroblast collagenase were superimposed, the best fit occurred in the region of α4, consisting of the conserved residues. It should be noted here that despite the absence of Zn^{2+} ion in the AQ_1354, which in our opinion was a result of 0.1 m*M* EDTA and 0.1 m*M* DTT being used in protein purification and crystallization, the side chains for those histidine residues are very well superimposed with those of the zinc-containing structures of collagenases and gelatinases. Besides the central helix, all the other secondary-structure elements either superimposed much more poorly or do not have counterparts in the other structures at all. The superimposition of AQ_1354 and human fibroblast collagenase is shown in Fig. 5.

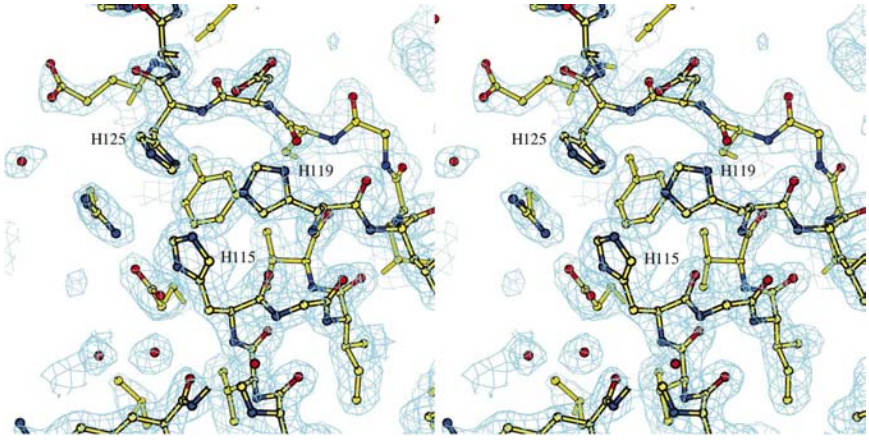

Figure 3
The electron-density map around conserved histidine residues. The contour level corresponds to 1.3σ. Illustration prepared using *BOBSCRIPT* (Esnouf, 1997, 1999).

```
1hfc    ---------------- ------VLTEGNPRW EQTHLTYRIENYTPD LPRADVDHAIEKAFQ  39
1ck7    DQNTIETMRKPRCGN PDVANYNFFPRKPKW DKNQITYRIIGYTPD LDPETVDDAFARAFQ 120
AQ_1354 ---------------- -------MSSTKRQK NRVLVKLKKRKVRKD KIEKWAELALS---A  35
1eb6    ---------------- ----TEVTDCKGDAE SSLTTALSNAAKLAN QAAEAAESGDESKFE  41

1hfc    LWSNVTPLTFTKVSE GQADIMISFVRGDHR --DNSPFDGPGGNLA HAFQPGPGIGGDAHF  97
1ck7    VWSDVTPLRFSRIHD GEADIMINFGRWEHG --DGYPFDGKDGLLA HAFAPGTGVGGDSHF 178
AQ_1354 LGLNNVELSVYITDD QEIRELNKTYRKKD- --K--PTDVLSFPMG EEPG-GYKILGDVVI  89
1eb6    EYFKTTDQQTRTTVA ERLRAVAKEAGSTSG GSTTYHCNDPYGYCE PNVLAYTLPSKNEIA 101

1hfc    DEDERWTN------- ---------------- --NFREYN------- ---------------- 111
1ck7    DDDELWTLGEGQVVR VKYGNADGEYCKFPF LFNGKEYNSCTDTGR SDGFLWCSTTYNFEK 238
AQ_1354 SQDTAERQ------- ---------------- ---ARELG------- ---------------- 102
1eb6    NCDIYYSELPP---- ---------------- --LAQKCH------- ---------------- 118

1hfc    ---------LHRVAA HELGHSLGLSHSTDI GALMYPSYTFSGDVQ LAQDDIDGIQAIYGR 162
1ck7    GFCPDQGYSLFLVAA HEFGHAMGLEHSQDP GALMAPIYTYTKNFR LSQDDIKGIQELYGA 418
AQ_1354 ---HSLEEEVKRLIV HGIVHLLGYDHE--K GGEEEKKFRELENYV LSKLSKAL------- 150
1eb6    ------AQDQATTTL HEFTHAPGVYQP--- GT-EDLGYGYDAATQ LSAQDALNNADSYAL 168
```

Figure 4
Amino-acid sequence alignment of AQ_1354 and threefold homologues picked by the DALI engine. Identical residues are shown with green shading. Residues in yellow are responsible for binding the 'structural' Zn ion, which is 12 Å away from the active site; in magenta is one of the key residues in the active site of matrix metalloproteinases (in 1ck7 the glutamic acid was substituted with alanine in order to prevent autolysis).

An important feature of matrix metalloproteinases is the glutamic acid residue within the motif. The HE*xx*H sequence is found in three out of five matrixin clans. According to the amino-acid sequence and architecture, AQ_1354 is closer to those three clans, but neither it nor any other member of UPF0054 contains the glutamic acid residue within that motif that is essential for catalytic activity (Hangauer *et al.*, 1984).

Despite the differences between AQ_1354 and matrix metalloproteinases, we attempted a functional assay using the EnzChek Gelatinase/Collagenase Assay Kit from Molecular Probes Inc. (Eugene, Oregon, USA). It failed to detect collagenase or gelatinase activity in AQ_1354 under standard assay conditions.

In attempts to find functional information about AQ_1354, several other functional assays were performed. Among these were a collagenase assay with furyl-acryloyl-Leu-

Figure 5
Stereoview of superimposition of the structures of AQ_1354 (gold) and human fibroblast collagenase (green). When all the identical residues from sequence alignment are superimposed, the r.m.s. displacement is 9.73 Å (over those residues only); when only residues from a conserved motif are used the r.m.s. displacement is 0.49 Å (over main-chain atoms for non-identical residues and over all atoms for identical ones within the motif; shown in figure). The r.m.s. displacement calculated by DALI is 3.4 Å (over 106 residues of 141 in the coordinate file). The structures are superimposed using the *LSQKAB* program from the *CCP*4 suite. The illustration was prepared using *MOLSCRIPT* (Kraulis, 1991).

Gly-Pro-Ala, protease assays with benzoyl-Arg-*para*-nitroanilide (BAPNA), Leu-pNA, Pro-pNA, Suc-Phe-pNA and Suc-Ala-Ala-Ala-pNA, several phosphatase assays, a phosphodiesterase/nuclease assay, an esterase/lipase assay, dehydrogenase (amino acids, alcohols and organic acids), oxidase and sulfatase assays. We obtained negative results in all the above-mentioned experiments.

4. Conclusion

The crystal structure of the hypothetical protein AQ_1354 from *A. aeolicus* solved at a resolution of 1.89 Å revealed a certain degree of fold similarity to matrix metalloproteinases. Moreover, they share a conserved zinc-binding motif, which represents the active site of metalloproteinases. However, the collagenase/gelatinase functional assay for AQ_1354 performed under standard conditions did not detect any activity. Three out of six secondary-structure elements are not conserved either in amino-acid sequence or in structure. It very well may be that a similar type of motif is used by nature to perform different functions.

We thank Dr David King of University of California, Berkeley for mass-spectrometric analysis of the protein and the staff of Advanced Light Source beamline 5.0.2 of Lawrence Berkeley National Laboratory for help during diffraction data collection. We are grateful to B. Gold and H. Yokota for cloning, J.-M. Chandonia for a bioinformatics search of the gene and B. Martinez and M. Henriquez for technical help. We are also grateful to Alexander Yakunin and Kate Kuznetsova for performing functional assays. The genomic DNA of *A. aeolicus* was a gift from DIVERSA Corporation (San Diego, USA). This work was supported by Grant GM 62412 to the Berkeley Structural Genomics Center (http://www.strgen.org) from the National Institute of General Medical Sciences, National Institutes of Health.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Hangauer, D. G., Monzingo, A. F. & Matthews, B. W. (1984). *Biochemistry*, **23**, 5730–5741.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Hou, J., Sims, G., Zhang, C. & Kim, S.-H. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 2386–2390.
Huber, R., Wilharm, T., Huber, D., Trincone, A., Koenig, H., Rachel, R., Rockinger, I., Fricke, H. & Stetter, K. O. (1992). *Syst. Appl. Microbiol.* **15**, 340–351.
Hwang, K. Y., Chung, J. H., Kim, S.-H., Han, Y. S. & Cho, Y. (1999). *Nature Struct. Biol.* **6**, 691–696.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kim, R., Sandler, S. J., Goldman, S., Yokota, H., Clark, A. J. & Kim, S.-H. (1998). *Biotechnol. Lett.* **20**, 207–210.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Leahy, D. J., Hendrickson, W. A., Aukhil, I. & Erickson, H. P. (1992). *Science*, **258**, 987–991.
McAuley, K. E., Jia-Xing, Y., Dodson, E. J., Lehmbeck, J., Ostergaard, P. R. & Wilson, K. S. (2001). *Acta Cryst.* D**57**, 1571–1578.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Morgunova, E., Tuuttila, A., Bergmann, U., Isupov, M., Lindqvist, Y., Schneider, G. & Tryggvason, K. (1999). *Science*, **284**, 1667–1670.
Otwinowski, Z. & Minor, W. (1996). *Methods Enzymol.* **276**, 307–326.
Schulze-Gahmen, U., Pelaschier, J., Yokota, H., Kim, R. & Kim, S.-H. (2003). *Proteins*, **50**, 526–530.
Spurlino, J. C., Smallwood, A. M., Carlton, D. D., Banks, T. M., Vavra, K. J., Johnson, J. S., Cook, E. R., Falvo, J., Wahl, R. C. & Pulvino, T. A. (1994). *Proteins*, **19**, 98–109.
Teplova, M., Tereshko, V., Sanishvili, R., Joachimiak, A., Bushueva, T., Anderson, W. F. & Egli, M. (2000). *Protein Sci.* **9**, 2557–2566.
Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1937–1940.
Zarembinski, T. I., Hung, L.-W., Mueller-Dieckmann, H.-J., Kim, K.-K., Yokota, H., Kim, R. & Kim, S.-H. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 15189–15193.
Zhang, X., Meining, W., Fischer, M., Bacher, A. & Ladenstein, R. (2001). *J. Mol. Biol.* **306**, 1099–1114.