

## Structure-based assignment of the biochemical function of a hypothetical protein: A test case of structural genomics

THOMAS I. ZAREMBINSKI\*<sup>†‡</sup>, LI-WEI HUNG\*<sup>†‡</sup>, HANS-JOACHIM MUELLER-DIECKMANN\*<sup>†‡</sup>, KYEONG-KYU KIM\*<sup>†§</sup>, HISAO YOKOTA\*, ROSALIND KIM\*, AND SUNG-HOU KIM\*<sup>†¶</sup>

\*Physical Biosciences Division of Lawrence Berkeley National Laboratory and <sup>†</sup>Department of Chemistry, University of California, Berkeley, CA 94720

Contributed by Sung-Hou Kim, October 28, 1998

**ABSTRACT** Many small bacterial, archaeobacterial, and eukaryotic genomes have been sequenced, and the larger eukaryotic genomes are predicted to be completely sequenced within the next decade. In all genomes sequenced to date, a large portion of these organisms' predicted protein coding regions encode polypeptides of unknown biochemical, biophysical, and/or cellular functions. Three-dimensional structures of these proteins may suggest biochemical or biophysical functions. Here we report the crystal structure of one such protein, MJ0577, from a hyperthermophile, *Methanococcus jannaschii*, at 1.7-Å resolution. The structure contains a bound ATP, suggesting MJ0577 is an ATPase or an ATP-mediated molecular switch, which we confirm by biochemical experiments. Furthermore, the structure reveals different ATP binding motifs that are shared among many homologous hypothetical proteins in this family. This result indicates that structure-based assignment of molecular function is a viable approach for the large-scale biochemical assignment of proteins and for discovering new motifs, a basic premise of structural genomics.

As of October 1998, 16 microbial genomes had been completely sequenced (Web site: [www.tigr.org](http://www.tigr.org)). These genomes are from all three branches of life: four from the Archaea, one from Eukarya, and the rest from Bacteria. To predict a function for each of their predicted protein coding regions or ORFs, the amino acid sequence of the ORF is compared against all functionally assigned sequences in protein sequence databases. If there is significant sequence or motif identity between the ORF and a functionally assigned sequence, then it is assumed that the two sequences share the same function. Unfortunately, up to 62% of the ORFs from these genomes share little or no sequence identity with any assigned sequence and hence are of unknown function (1–15). A major challenge, therefore, is to find ways to reliably and rapidly predict or determine the molecular (biochemical and biophysical) functions as well as cellular functions of these proteins.

One approach for assigning the molecular function of a protein with unknown function is first to determine the three-dimensional structure of the protein by either x-ray crystallography or NMR. The structure, instead of the amino acid sequence, then is compared against those of the protein structure database (Protein Data Bank). If there are one or more significant structural homologs, the hypothetical protein is predicted to have molecular properties similar to the homologs. The predictions then can be tested experimentally. The molecular function then can provide a basis for searching for the cellular function of the protein. This method, structural genomics (16, 17), is far more sensitive than primary sequence comparisons because proteins having insignificant sequence

similarity often adopt similar tertiary structures with similar or related molecular functions. With the increasing advances in computer hardware and software associated with structure determination, this approach will become more viable.

Herein we present an example showing the feasibility of the structural genomics approach by determining the crystal structure of the protein encoded by MJ0577, an ORF of unknown function from the recently sequenced hyperthermophile, *Methanococcus jannaschii* (4). We found that the crystal structure of the gene product, MJ0577, has a bound ATP, immediately suggesting its biochemical function to be either an ATPase or an ATP-binding molecular switch. Preliminary biochemical experiments show that MJ0577 is likely to be the latter, because the protein by itself is not an ATPase, but it hydrolyzes ATP in the presence of *M. jannaschii* crude cell extract, a property analogous to GTP hydrolysis by Ras in the presence of GTPase-activating protein.

### MATERIALS AND METHODS

**Expression, Purification, and Crystallization of the Wild-Type Protein and the Selenomethionine Derivative.** The MJ0577 coding region of *M. jannaschii* was prepared by PCR, cloned into the pET-23a vector, and transformed into the *Escherichia coli* host BL21 (DE3) SJS1244 (18) for isopropyl  $\beta$ -D-thiogalactoside-dependent protein expression. The yield of protein was approximately 2.5 mg pure protein/liter of culture. Because the protein is heat stable, the expressed protein was purified initially by an 80°C incubation for 30 min and centrifugation followed by anion exchange chromatography over a DEAE Sepharose FF column (Pharmacia). The crystallization condition was screened by the sparse matrix method (19) using the Hampton Research Crystal Screen (Laguna Niguel, CA) and by footprint screening (20). The crystals were grown by the vapor-diffusion method from a solution containing 2.5 mg/ml of protein, 0.5 mM DTT, 25 mM Tris (pH 8.0), 8% polyethylene glycol (PEG) 4000, 50 mM Imidazole-malate (pH 7.0), and 5 mM MnCl<sub>2</sub> in a drop equilibrated against 16% PEG 4000, 100 mM Imidazole-malate (pH 7.0), and 10 mM MnCl<sub>2</sub>. The selenomethionine mutant proteins were made in a methionine auxotroph, and the crystals were grown under similar conditions but required streak seeding with native crystals for optimally sized crystals. The crystals are in space group P2<sub>1</sub>2<sub>1</sub>2, with unit-cell dimensions  $a = 95.53$  Å,  $b = 96.08$  Å, and  $c = 37.5$  Å at 100 K. The crystals were flash-frozen in the above mother liquor plus 25% glycerol (21) before exposure to x-ray.

Data deposition: The structure reported in this paper has been deposited in the Protein Data Bank, Biology Department, Brookhaven National Laboratory, Upton, NY 11973 (PDB ID code 1mjh).

<sup>‡</sup>T.I.Z., L.-W.H. and H.-J.M.-D. contributed equally to this work.

<sup>§</sup>Present address: Plant Molecular Biology and Biotechnology Research Center, Gyeongsang National University, Chinju 660–701, Korea.

<sup>¶</sup>To whom reprint requests should be addressed. e-mail: [shkim@lbl.gov](mailto:shkim@lbl.gov).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9515189-5\$2.00/0 PNAS is available online at [www.pnas.org](http://www.pnas.org).

Table 1. Statistics for data collection of MJ577, resolution 1.8 Å

Wavelength, Å	Number of reflections		Completeness, %		$R_{\text{sym}}$ , %	
	Measured	Unique	All	Last shell	All	Last shell
$I_1 = 1.00$ (remote)	287,882	32,040	97.6	99.6	5.8	34.8
$I_2 = 0.9806$ (edge)	288,979	31,979	97.5	99.5	7.7	41.0
$I_3 = 0.9799$ (peak)	290,945	31,918	97.6	99.5	6.7	32.3
$I_4 = 0.9683$ (remote)	296,375	32,169	97.5	98.9	7.6	62.6

X-ray diffraction data sets were collected at four wavelengths at the Macromolecular Crystallography Facility at Advanced Light Source of the E. O. Lawrence Berkeley National Laboratory, and the data were processed with DENZO (22) and reduced with SCALEPACK (22). We used the SOLVE program package (23) ([www.solve.lanl.gov](http://www.solve.lanl.gov)) to obtain electron density maps. Four selenium sites were found in one asymmetric unit. Because MJ0577 contains four methionines in its primary sequence, and two molecules exist per asymmetric unit, only two of the four methionines per monomer were ordered. Forty-five percent of the unit cell volume was estimated to be solvent.

The initial multiwavelength anomalous diffraction-derived phases to 1.8-Å Bragg spacings subsequently were improved through solvent flattening, histogram matching, and multiple cycles of model building and refinement. The final model was refined against 1.7-Å resolution data taken at a wavelength away from the Se absorption edge by using 1-Å x-ray at 100 K. Positional refinement, simulated annealing refinement, and temperature refinement were performed with the programs CNS (24), REFMAC (25), and ARP (26). The electron densities are weak for two loops (residues 49–65 of the first monomer and 1049–1064 of the second); hence, we did not construct a model for these regions. Noncrystallographic symmetry (NCS) restraints were applied and gradually released over the course of the refinement. In the final refinement, no NCS constraints or restraints were used. During refinement, the free  $R$  factor was monitored by using 10% of the total reflections as a test data set. After the free  $R$  factor was reduced below 30%, 284 water molecules were identified from well-defined electron densities in the  $F_o - F_c$  map by using ARP (26). The structure has been refined at 1.7-Å resolution with a crystallographic  $R$  value of 21% and a free  $R$  value of 25%. All backbone dihedral angles (except glycines and prolines) fall within allowed regions in a Ramachandran plot. The atomic coordinates and

structure factors have been deposited into the Brookhaven Protein Data Bank.

**Preparation of the Crude Extract of *M. jannaschii*.** Frozen *M. jannaschii* cells (1.2 g) were resuspended with 4.65 ml of 50

Table 2. Native data and refinement statistics

Native data at 1.7 Å	
Wavelength	1.00 Å
Reflections (unique)	292,156 (38,706)
Completeness (last shell)	98.0% (85.4%)
$R_{\text{sym}}$ (last shell)	4.9% (24.9%)
No. of non-hydrogen atoms	2,696
No. of substrate atoms	2 ATP, 2 $\text{Mn}^{2+}$
No. of water molecules	318
Refinement information	
$R_{\text{work}}$	20.6%
$R_{\text{free}}$	24.9%
Rmsd	
Bond length	0.015 Å
Angle	0.030 Å
Average B factor	
Protein	31.0 Å <sup>2</sup>
Water	39.1 Å <sup>2</sup>
Ramachandran plot	
Most favored	94.7%
Additional	5.0%
Disallowed	0.0%

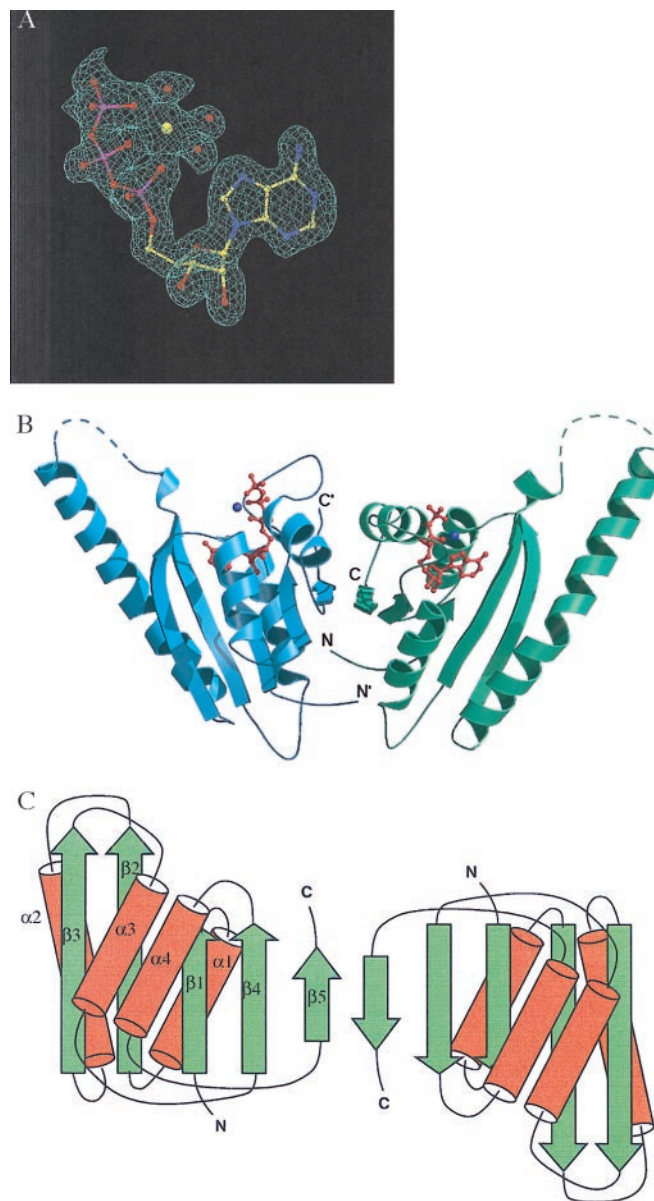


FIG. 1. (A) A sample of the electron density map for ATP in the ATP-binding pocket of MJ0577. The multiwavelength anomalous diffraction-phased electron density map at 1.8-Å resolution is contoured at 1 sigma, with the current model displayed for comparison. A Mn ion and three waters bound to ATP are also shown. (B) Ribbon diagram of the crystal structure of the MJ0577 dimer drawn by using MOLSCRIPT (29) and RASTER3D (30). The secondary structure assignment is based on DSSP (31). Dashed lines represent loops not modeled in the structure. The N and C termini of each monomer are labeled. (C) Schematic structure, where  $\beta$ -strands and  $\alpha$ -helices are represented by arrows and cylinders, respectively.

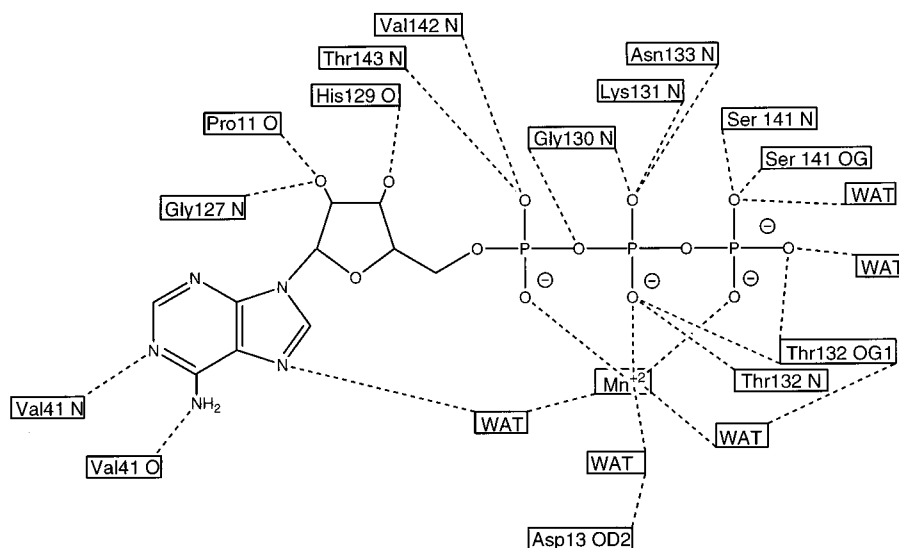


FIG. 2. Schematic drawing showing all of the hydrogen bonds (dashed lines) involving ATP and coordination bonds (dotted lines) involving the Mn<sup>2+</sup> ion. The protein residues and atoms involved in each hydrogen bond are shown in the boxes.

mM Tris (pH 8.0), 5 mM DTT, 1 mM phenylmethylsulfonyl fluoride, and 0.1 mg/ml of lysozyme. The solution was left on ice for 30 min followed by brief sonication. The lysate then was centrifuged to remove insoluble cell debris, and the supernatant was frozen immediately in liquid nitrogen and stored at -80°C.

**ATP Hydrolysis Assays.** Purified MJ0577 (1.3 mg) was incubated alone or in the presence of crude *M. jannaschii* cell extract (corresponding to about 0.18 g) in 50 mM Tris at pH 8.0, 1 mM DTT at a final volume of 500 µl for 1 hr at either 4°C or 80°C. MJ0577 then was repurified by anion exchange chromatography [Pharmacia Q column (Hi trap 1 ml) by using 40 mM Tris (7.5) and a NaCl gradient from 0 to 1 M)]. In experiments where *M. jannaschii* extract was mixed with pure MJ0577 protein, an additional gel filtration step was done with Pharmacia PD-10 columns to remove any small molecules from the *M. jannaschii* extract that cochromatographed with

MJ0577. The protein then was concentrated by using nanosep 3K filters (Pall Filtron) and extracted with an equal volume of phenol/chloroform/isoamyl alcohol (27). The aqueous phase containing the nucleotide(s) then was subjected to anion exchange chromatography, essentially as above, and the resulting peaks were visualized at 254 nm.

### RESULTS AND DISCUSSION

The crystal structure of MJ0577 was determined to 1.8-Å resolution by using its selenomethionyl derivative and the multiwavelength anomalous diffraction method (28). Crystallographic statistics for x-ray diffraction data at four different wavelengths as well as model refinement are given in Tables 1 and 2. A sample of the electron density map is shown in Fig. 14. Four methionines exist in the protein sequence, only two of which were detected in the final structure as selenomethi-

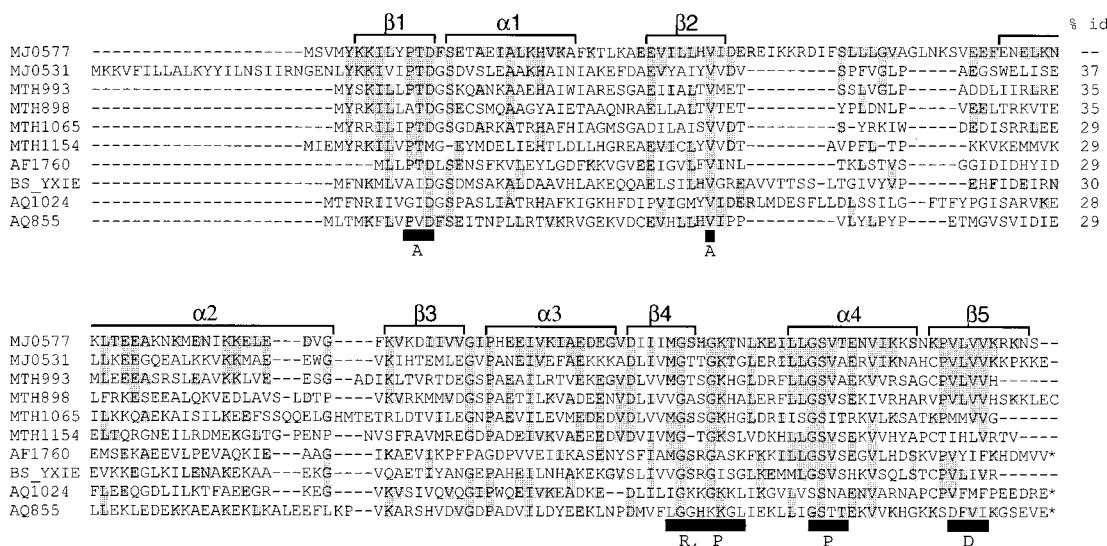


FIG. 3. Multiple alignment of conserved regions of the MJ0577 superfamily showing four sequence motifs. It was constructed by using PSI-BLAST (38) followed by CLUSTALW (39). The first column shows the protein identifier with the letters designating the organism code and the subsequent numbers designating the gene. Organism codes: MJ, *M. jannaschii*; MTH, *Methanobacterium thermoautotrophicum*; AF, *Archaeoglobus fulgidus*; BS, *Bacillus subtilis*; AQ, *Aquifex aeolicus*. Conserved residues are shaded. The black boxes below the sequence indicate conserved residues contacting ATP (A: adenine; R: ribose; P: phosphate; D: dimer interface).  $\alpha$ -helices and  $\beta$ -strands are represented by brackets above the sequences. \* at the C terminus designate those sequences with tandem homologs of MJ0577. Percent identities between the sequences are at far right.

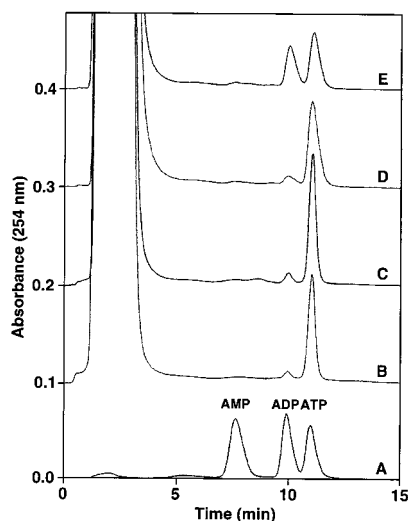


FIG. 4. Measurement of ATP hydrolysis. Anion exchange chromatograms from four different MJ0577 reactions processed as described in *Materials and Methods* are shown. The samples are: A, a reference sample containing 10 nmol each of AMP, ADP, and ATP; B, MJ0577 alone (4° C); C, MJ0577 alone (80° C); D, MJ0577 plus *M. jannaschii* extract (4° C); E, MJ0577 plus *M. jannaschii* extract (80° C). Migration positions of AMP, ADP, and ATP are indicated above sample A. The initial large peak is the result of residual phenol/chloroform/isoamyl alcohol during the nucleotide extraction step (see *Materials and Methods*).

onine-substituted in the derivative. The current refined model at 1.7 Å ( $R$  factor 21.0%,  $R_{\text{free}}$  25%) includes 287 aa per dimer, but loop residues 49–65 on the first monomer and loop residues 49–64 on the second monomer are disordered. The model also includes two manganese, two ATPs, and 284 water molecules per dimer.

The MJ0577 monomer structure is an open-twisted five-stranded parallel  $\beta$ -sheet with two helices on each side of the sheet (Fig. 1B). The topological structure is shown in Fig. 1C. The structure has a nucleotide-binding pocket surrounded by motifs with limited similarities to those commonly found among ATP-binding proteins (32, 33), but the sequential arrangement of the motifs and the spacings between the motifs are very different from other ATP-binding proteins. Thus, this structure represents a different family of ATP-binding molecules.

The ATP is anchored by octahedral coordination with a divalent cation, three water molecules, and several protein contacts (Fig. 2). This cation is believed to be manganese because it was used as a crystallization additive, and ATP phosphates usually bind either magnesium or manganese. Furthermore, this density occurs at a distance appropriate for a Mn–O bond (about 2.3 Å). Specific hydrogen bondings involving ATP are shown in Fig. 2.

Interestingly, ATP was never added during purification or crystallization. Hence, MJ0577 must have scavenged the ATP from its *E. coli* host during overexpression and neither released it during purification nor hydrolyzed it. These data suggest that, assuming MJ0577 hydrolyzes ATP *in vivo*, MJ0577 requires an additional factor(s) present in *M. jannaschii* but not in *E. coli* for hydrolysis.

MJ0577 homodimerizes in the crystal (Fig. 1B and C) via antiparallel hydrogen bonding of the highly conserved residues 153–158 in the fifth  $\beta$ -strand on each subunit. An accessible surface area of 1,056 Å<sup>2</sup> (15%) from each monomer is buried at the dimer interface. Interestingly, several MJ0577 homologs contain tandem MJ0577 homologues in the same polypeptide chain, suggesting that oligomerization is important for the function of this family of proteins (see \* in Fig. 3).

To predict a possible molecular (biochemical or biophysical) function of MJ0577, we compared the crystal structure with all representatives of the protein folds in the Protein Data Bank with the program DALI (34). The comparison revealed that this structure has fold similarities to several proteins despite the lack of significant sequence similarities: the  $\alpha$  and  $\beta$  subunits of the human electron transfer flavoprotein (35) [Protein Data Bank (PDB) ID codes: 1efv-B and 1efv-A; Z scores: 10.4 and 8.2; 252 and 312 residues, 17% and 13% sequence identity, respectively], DNA photolyase (36) (PDB ID code: 1qnf; Z score: 9.5; 475 residues, 12% sequence identity), and the tyrosyl-tRNA synthetase (37) (PDB ID code: 2ts1; Z score: 7.1; 317 residues, 11% sequence identity), among others. Despite the significant structural similarities in parts, the similarities did not provide functionally useful information, because each of these proteins is substantially larger than MJ0577.

Biochemical experiments showed that MJ0577 has no appreciable ATPase activity by itself. However, when *M. jannaschii* cell extract was added to the reaction mixture, 50% of ATP was hydrolyzed to ADP in 1 hr at 80°C (Fig. 4). This result indicates that MJ0577 requires one or more soluble components to stimulate ATP hydrolysis, perhaps an ATPase-activating protein(s) (40) analogous to the GTP-activating protein for Ras protein, Ras GAP (41). Furthermore, these factor(s) are probably specific to *M. jannaschii* because ATP hydrolysis did not occur within *E. coli* during the overexpression or purification of MJ0577. Although the biochemical function of the protein is a factor-dependent ATPase, the factors need to be purified and identified to understand the cellular function of this protein as, perhaps, a molecular switch for a cellular process.

In conclusion, we have presented an example of the structure-based assignment of the biochemical function of a hypothetical protein: the protein's structure revealed a bound ATP, which immediately suggested a small number of possible biochemical functions of the protein. Biochemical assays allowed us to identify the biochemical function of this "hypothetical" protein, which also revealed a different ATP-binding protein family.

We thank Dr. David King (Department of Molecular and Cell Biology, University of California, Berkeley) for mass spectrometry experiments, Dr. Thomas Earnest and Dr. Gerry McDermott (Advanced Light Source, Lawrence Berkeley National Laboratory) for help in multiwavelength anomalous diffraction data collection and processing, and Dr. Douglas Clark (Department of Chemical Engineering, University of California, Berkeley) for the *M. jannaschii* cells. We also thank Xinlin Du and Dr. Ed Berry for technical advice and Drs. Chao Zhang, Adam Arkin, Steve Holbrook, and Jeroen Brandsen for critical reading of the manuscript. The work was supported by a grant from the Office of Health and Environmental Research, Office of Energy Research, U.S. Department of Energy to R.K. and S.H.K. (DE-AC03-76SF00098). H.-J. M.-D. was supported by the Deutsche Forschungsgemeinschaft in part.

- Hodgkin, J., Plasterk, R. H. A. & Waterston, R. H. (1995) *Science* **270**, 410–414.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., *et al.* (1995) *Science* **269**, 496–512.
- Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., *et al.* (1995) *Science* **270**, 397–403.
- Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., *et al.* (1996) *Science* **273**, 1058–1073.
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirose, M., Sugiura, M., Sasamoto, S., *et al.* (1996) *DNA Res.* **3**, 109–136.
- Himmelreich, R., Hilbert, H., Plagens, H., Pirkel, E., Li, B. C. & Herrmann, R. (1996) *Nucleic Acids Res.* **24**, 4420–4449.

7. Goffeau, A., Aert, M. L., Agostini-Carbone, M. L., Ahmed, A., Aigle, M., Alberghina, L., Albermann, K., Albers, M., Aldea, M., Alexandraki, D., *et al.* (1997) *Nature (London)* **387**, Suppl., 5–105.
8. Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., *et al.* (1997) *Nature (London)* **388**, 539–547.
9. Blattner, F. R., Plunkett, G. 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., *et al.* (1997) *Science* **277**, 1453–1474.
10. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., *et al.* (1997) *J. Bacteriol.* **179**, 7135–7155.
11. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., *et al.* (1997) *Nature (London)* **390**, 249–256.
12. Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., *et al.* (1997) *Nature (London)* **390**, 364–370.
13. Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., *et al.* (1997) *Nature (London)* **390**, 580–586.
14. Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., *et al.* (1998) *Nature (London)* **392**, 353–358.
15. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E. 3rd, *et al.* (1998) *Nature (London)* **393**, 537–544.
16. Rost, B. (1998) *Curr. Biol.* **6**, 259–263.
17. Kim, S.-H. (1998) *Nat. Struct. Biol.* **5**, Suppl., 643–645.
18. Kim, R., Sandler, S. J., Goldman, S., Yokota, H., Clark, A. J. & Kim, S.-H. (1998) *Biotechnol. Lett.* **20**, 207–210.
19. Jancarik, J. & Kim, S.-H. (1991) *J. Appl. Crystallogr.* **24**, 409–411.
20. Stura, E. A., Nemerow, G. R. & Wilson, I. A. (1992) *J. Crystallogr. Growth* **122**, 273–285.
21. Garman, E. F. & Schneider, T. R. (1997) *J. Appl. Crystallogr.* **30**, 211–237.
22. Otwinowski, Z. (1993) in *Data Collection and Processing*, eds. Sawyer, L., Isaacs, N. & Bailey, S. (SERC Daresbury Laboratory, Warrington, U.K.), pp. 56–62.
23. Terwilliger, T. C. (1997) *Methods Enzymol.* **276**, 530–537.
24. Brünger, A., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., *et al.* (1998) *Acta Crystallogr. D* **54**, 899–904.
25. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) *Acta Crystallogr. D* **53**, 240–255.
26. Lamzin, V. S. & Wilson, K. S. (1993) *Acta Crystallogr. D* **49**, 129–147.
27. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
28. Hendrickson, W. A. (1991) *Science* **254**, 51–58.
29. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950.
30. Merritt, E. A. & Bacon, D. J. (1997) *Methods Enzymol.* **277**, 505–524.
31. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22**, 2577–2637.
32. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982) *EMBO J.* **1**, 945–951.
33. Traut, T. W. (1994) *Eur. J. Biochem.* **222**, 9–19.
34. Holm, L. & Sander, C. (1993) *J. Mol. Biol.* **233**, 123–128.
35. Roberts, D. L., Frerman, F. E. & Kim, J.-J. P. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 14355–14360.
36. Tamada, T., Kitadokoro, K., Higuchi, Y., Inaka, K., Yasui, A., de Rooter, P. E., Eker, A. P. & Miki, K. (1997) *Nat. Struct. Biol.* **4**, 887–891.
37. Irwin, M. J., Nyborg, J., Reid, B. R. & Blow, D. M. (1976) *J. Mol. Biol.* **105**, 577–586.
38. 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.
39. Higgins, D., Thompson, J. & Gibson, T. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
40. Scheffzek, K., Ahmadian, M. R. & Wittinghofer, A. (1998) *Trends Biochem. Sci.* **23**, 257–262.
41. Trahey, M. & McCormick, F. (1987) *Science* **238**, 542–545.