

Natasha N. Smith and  
D. Travis Gallagher\*

National Institute of Standards and Technology,  
Biochemical Science Division, Gaithersburg,  
MD 20899-8312, USA

Correspondence e-mail:  
gallaghe@umbi.umd.edu

Received 7 July 2008  
Accepted 5 September 2008

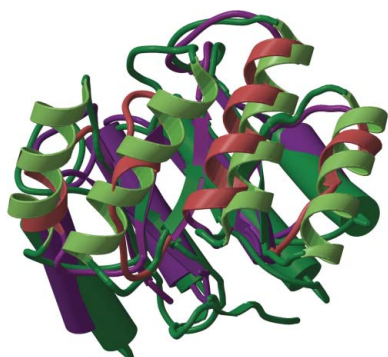
**PDB Reference:** archaeal dehydroquinase,  
2ox1, r2ox1sf.

## Structure and lability of archaeal dehydroquinase

Multiple sequence alignments of type I 3-dehydroquinase dehydratases (DQs; EC 4.2.1.10) show that archaeal DQs have shorter helical regions than bacterial orthologs of known structure. To investigate this feature and its relation to thermostability, the structure of the *Archaeoglobus fulgidus* (Af) DQ dimer was determined at 2.33 Å resolution and its denaturation temperature was measured *in vitro* by circular dichroism (CD) and differential scanning calorimetry (DSC). This structure, a  $P2_12_12_1$  crystal form with two 45 kDa dimers in the asymmetric unit, is the first structural representative of an archaeal DQ. Denaturation occurs at  $343 \pm 3$  K at both low and high ionic strength and at 349 K in the presence of the substrate analog tartrate. Since the growth optimum of the organism is 356 K, this implies that the protein maintains its folded state through the participation of additional factors *in vivo*. The  $(\beta\alpha)_8$  fold is compared with those of two previously determined type I DQ structures, both bacterial (*Salmonella* and *Staphylococcus*), which had sequence identities of ~30% with AfDQ. Although the overall folds are the same, there are many differences in secondary structure and ionic features; the archaeal protein has over twice as many salt links per residue. The archaeal DQ is smaller than its bacterial counterparts and lower in regular secondary structure, with its eight helices being an average of one turn shorter. In particular, two of the eight normally helical regions (the exterior of the barrel) are mostly nonhelical in AfDQ, each having only a single turn of  $3_{10}$ -helix flanked by  $\beta$ -strand and coil. These 'protohelices' are unique among evolutionarily close members of the  $(\beta\alpha)_8$ -fold superfamily. Structural features that may contribute to stability, in particular ionic factors, are examined and the implications of having a  $T_m$  below the organism's growth temperature are considered.

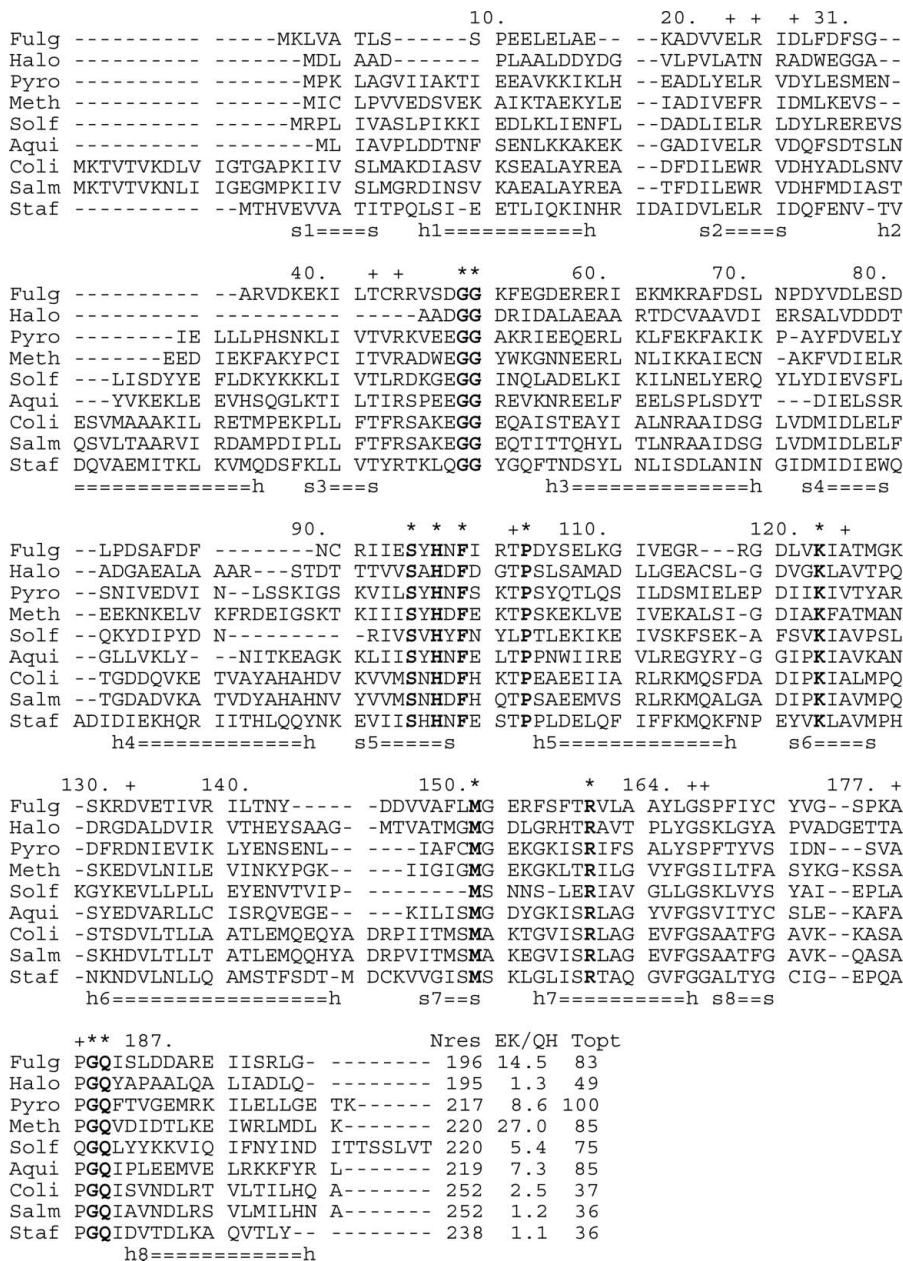
### 1. Introduction

Just as the central dogma of molecular biology has required several elaborations and enhancements during its 50 y history, a secondary paradigm of modern biology, that each gene encodes one macromolecule with one structure that performs one function, is currently in transition because of new findings. In particular, observations of natively unfolded proteins have forced a review of what is a structure and how we define and count functions. Many proteins that are intrinsically unstructured or of transient or ligand-dependent structure are now known; these are primarily non-enzyme proteins from mesophilic organisms (Dyson & Wright, 2005). It has been estimated that ~30% of eukaryotic proteins are at least partially unstructured *in vivo* (Fink, 2005). The corresponding estimate for archaeal proteins, based on sequence, is only 2%, but this number may underaccount for extreme growth conditions, and both chaperones and small-molecule solutes are known to play important roles in maintaining archaeal proteins in a functional state (Scandurra *et al.*, 2000). Here, we describe an archaeal enzyme that is unstructured at neutral pH *in vitro* at the organism's optimal growth temperature of 356 K (Klenk *et al.*, 1997; Rohlin *et al.*, 2005), implying that the enzyme is 'trans-stabilized', *i.e.* that its function requires more than one structure or additional factor(s) *in vivo*. The crystal structure is presented and compared with those of two bacterial mesophilic homologs.



In the aromatic biosynthesis pathway in bacteria, plants and archaea, dehydroquinase (DQ; 3-dehydroquinase dehydratase; aroD; EC 4.2.1.10) performs the first step after the initial cyclization by removing water to convert 3-dehydroquinone to 3-dehydroshikimate. DQ enzymes have been found to occur in two distinct types with different folds and different mechanisms that accomplish the same reaction (Gourley *et al.*, 1999). The type I enzymes are found in archaea, plants and some bacteria, while type II enzymes are found in bacteria and fungi. The type I gene shows a high conservation among

phyla (~25% sequence identity between archaeal and bacterial sequences), although in plants it is found in a bifunctional enzyme that performs two sequential metabolic steps (Singh & Christendat, 2006). The crystal structures of DQ from the bacteria *Salmonella typhimurium* (SalmDQ; PDB codes 1qfe and 1l9w; Gourley *et al.*, 1999; Lee *et al.*, 2002) and *Staphylococcus aureus* (PDB code 1sfl; Nichols *et al.*, 2004) show that type I DQ belongs to the ( $\beta\alpha$ )<sub>8</sub>-barrel fold family; the subunits of about 200–250 residues function as dimers (Reilly *et al.*, 1994). In the *Archaeoglobus fulgidus* genomic sequence



**Figure 1** Multiple alignment of archaeal and bacterial DQ sequences, along with (E + K)/(Q + H) ratios (EK/QH) and the organisms' optimal growth temperatures (Topt). Five archaeal sequences are at the top, followed by four bacterial sequences. The organism names are abbreviated as follows: Fulg, *Archaeoglobus fulgidus* (Stetter *et al.*, 1987); Halo, *Halobacterium* sp. NRC-1 (Robinson *et al.*, 2005); Pyro, *Pyrococcus furiosus* (Fiala & Stetter, 1986); Meth, *Methanococcus jannaschii* (Klipcan *et al.*, 2006); Solf, *Sulfolobus solfataricus* (Klipcan *et al.*, 2006); Aqoi, *Aquifex aeolicus* (Klipcan *et al.*, 2006); Coli, *Escherichia coli* (Klipcan *et al.*, 2006); Salm, *Salmonella typhimurium* (Gray & Fedorka-Cray, 2002); Staf, *Staphylococcus aureus* (Bergdoll, 1989). Note that the archaeon *Halobacterium* is not a thermophile, while the bacterium *Aquifex* is. The three known structures are Fulg (at the top), Salm and Staf (the last two). Residue numbers along the top are for the *A. fulgidus* sequence. Residues that are completely conserved among the nine sequences are shown in bold and are marked with an asterisk above the alignment, while those that are conserved with one or two exceptions are marked with a plus sign. The secondary structure below the alignment, giving numbers for the eight strands and helices of the barrel, corresponds to the *Salmonella* enzyme and corresponds approximately with the others. For the three known structures (Fulg, Salm and Staf), the alignment was structure-based, while for the other sequences it was calculated by *ClustalW* (Thompson *et al.*, 1994) based on sequence homology.

**Table 1**

Sample information.

Crystallization	
Crystallization method	Vapor diffusion, hanging drop
Temperature (K)	278
Crystallization solutions	
Macromolecule	12 g l <sup>-1</sup> protein, 10 mM sodium cacodylate pH 6.8
Precipitant	Same as reservoir
Reservoir	0.2 M ammonium acetate, 0.1 M sodium acetate, 30% PEG 4000 pH 4.6
Cryo treatment	
Final cryoprotection solution	7:3 mixture of reservoir and glycerol
Soaking time (s)	5
Cooling	Immersion in liquid N <sub>2</sub>
Crystal data	
Crystal colour	Colorless
Crystal shape	Rectangular bars
Crystal dimensions (µm)	50 × 100 × 400
Matthews coefficient (Å <sup>3</sup> Da <sup>-1</sup> )	2.91
Solvent content (%)	57.7
Unit-cell data	
Space group	<i>P</i> <sub>2</sub> <sub>1</sub> <sub>2</sub> <sub>1</sub>
Unit-cell parameters (Å, °)	<i>a</i> = 91.36, <i>b</i> = 96.06, <i>c</i> = 117.87. <i>α</i> = <i>β</i> = <i>γ</i> = 90
No. of molecules (dimers) in unit cell ( <i>Z</i> )	8

(Klenk *et al.*, 1997), the DQ (AfDQ) gene is labeled AF0228 and is encoded by Genbank gene AAB91004. The protein has 196 residues, with a molecular weight of 22.2 kDa, and its calculated pI is 5.0.

Extensive efforts to understand the structural basis of protein thermostability have been made by comparison of thermophilic proteins with mesophilic homologs. Several hundred such mesophile-thermophile structure comparisons are now available, including several that focus on (*β*α)<sub>8</sub> enzymes (Lo Leggio *et al.*, 1999; Lorentzen *et al.*, 2003; Hocker *et al.*, 2001). Several structural features that correlate with increased stability have been identified, such as ion pairs, the presence of proline at helix N-termini, the presence of charges near helix termini, insertions/deletions with steric effects, internal packing, surface-polarity and subunit-interface properties. Alignment of the AfDQ sequence with bacterial homologs (Fig. 1) shows that many of the helical regions, especially helices 2 and 4, appear to be absent or significantly shorter in the archaeal protein, prompting the question of how the common fold accommodates the smaller helices.

In order to advance understanding of protein structure and stability, we describe here the crystal structure of this enzyme from *A. fulgidus*, the first reported DQ structure from an archaeal organism, and report its melting temperature under various conditions. In addition to the goal of understanding thermostability, studies on DQ are motivated by potential synthetic applications of the aromatic biosynthetic pathway. Furthermore, since the pathway is absent in animals, DQ and other aromatic pathway enzymes are potential targets for the development of antibiotics and benign herbicides.

## 2. Materials and methods

The native form of the AfDQ gene was expressed using a pET3a vector transformed into BL21 (DE3) Codon Plus cells. The cell lysate was heated at 343 K for 10 min to precipitate *Escherichia coli* proteins and the remaining supernatant was bound to a column of HiTrap-Q anion-exchange media (GE Life Sciences). Pure AfDQ was eluted by a NaCl gradient at 20% NaCl. Molecular-biology reagents were purchased from Sigma-Aldrich. The protein was exchanged into 10 mM cacodylate buffer pH 6.8, concentrated to 12 mg ml<sup>-1</sup> (0.5 mM protein monomer) and stored at 200 K. The

**Table 2**

Data-collection and structure-solution statistics.

Values in parentheses are for the outer shell.	
Temperature (K)	105
Diffraction source	Rotating anode
Diffraction protocol	Single wavelength
Monochromator	Mirror coating
Wavelength (Å)	1.54
Detector	Rigaku R-Axis IV image plate
No. of frames	360
Software	<i>CrystalClear</i>
Reflection data: native	
Resolution range (Å)	20.0–2.33 (2.41–2.33)
No. of unique reflections	43264 (4135)
Completeness (%)	96.2 (93.3)
Redundancy	3.36 (3.27)
<i>I</i> / <i>σ</i> ( <i>I</i> )	17.7 (4.0)
<i>R</i> <sub>merge</sub>	0.04 (0.293)
Reflection data: derivative	
Resolution range (Å)	20.0–2.91 (3.04–2.91)
No. of unique reflections	23607 (2914)
Completeness (%)	99.7 (99.8)
Redundancy	6.86 (6.47)
<i>I</i> / <i>σ</i> ( <i>I</i> )	19.2 (6.5)
<i>R</i> <sub>merge</sub>	0.06 (0.271)
Phasing method	SIRAS
No. of derivatives	1
Phasing resolution range (Å)	8.0–2.9
Figure of merit (all data)	0.59
Solution software	<i>SOLVE/RESOLVE</i>

protein concentration was measured using the Micro BCA kit with bovine serum albumin as the standard (Pierce, Rockford, Illinois, USA). Size-exclusion chromatography was used to assess the oligomeric state by comparing the elution time with those of molecular-weight standards. The purified native DQ enzyme eluted as if its size were 52 kDa, which is about 2.3 times the monomer size. This result is consistent with a dimer with a large axial ratio, similar to the dimer observed for bacterial homologs.

Melting temperatures were measured both by DSC (differential scanning calorimetry) and by CD (circular dichroism). DSC measurements used a model VP calorimeter from Microcal Inc. with a scan rate of 15 K h<sup>-1</sup>. The protein was at 1 mg ml<sup>-1</sup> in 10 mM sodium cacodylate pH 6.8 or in this buffer plus 1 M NaCl. In a series of scans, the sample and reference chamber (0.5 ml) were first filled with buffer and then scanned from 288 to 368 K, cooled to 288 K and re-scanned. The sample chamber was then emptied, filled with protein solution by syringe and scanned again. After the completion of a set of protein scans, a second buffer *versus* buffer scan was taken as the baseline and subtracted from each of the protein scans. The resulting net protein *versus* buffer scans were converted to heat capacity *versus* temperature for analysis. CD measurements were performed with a Jasco spectropolarimeter, model J-720, using water-jacketed quartz cells with path lengths of 1 cm or 0.1 mm depending on the protein concentration, which was 50 µg ml<sup>-1</sup> for the tartrate-free scans and 3 mg ml<sup>-1</sup> for the scan with tartrate. CD spectra were measured in 100 mM potassium phosphate buffer, in 100 mM potassium phosphate buffer plus 1 M KCl and in 100 mM potassium phosphate buffer plus the substrate analog sodium tartrate (Singh & Christendat, 2006) at 0.1 M. Temperature control was provided by a Neslab RTE-110 circulating water bath interfaced with an MTP-6 temperature programmer. Ellipticities at 222 nm were continuously monitored at a scanning rate of 1 K min<sup>-1</sup> over the temperature range 298–373 K.

Crystal growth was observed in several trial conditions following screening with both Hampton and Emerald reagent sets. The largest crystal form was optimized to the conditions given in Table 1. These crystals belonged to space group *P*<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> and grew as slightly oblique



**Table 3**

Structure refinement and model validation.

Values in parentheses are for the outer shell.

Refinement software	REFMAC5.2.0005
$\sigma$ cutoff	—
Resolution range (Å)	10.00–2.33 (2.387–2.330)
No. of reflections used in refinement	40343 (2708)
No. of reflections above $\sigma$ cutoff in final cycle	40343 (2708)
Final overall <i>R</i> factor	0.208 (0.249)
Overall average <i>B</i> factor (Å <sup>2</sup> )	30.2
No. of protein atoms	6244
No. of solvent atoms	409
Total No. of atoms	6653
Noncrystallographic symmetry restraints	None
Final <i>R</i> <sub>work</sub>	0.204 (0.249)
No. of reflections for <i>R</i> <sub>free</sub>	2168 (137)
Final <i>R</i> <sub>free</sub>	0.277 (0.312)
Ramachandran data, non-Gly residues in (%)	
Most favored regions	90.3
Additional allowed regions	9.7
Missing/partial residues	None
Omitted residues	None

bars that diffracted to about 2.4 Å resolution. A phasing derivative was formed by soaking a crystal (of dimensions 50 × 120 × 300 μm) in 5 mM HgCl<sub>2</sub> for 5 min. The protomer contains three cysteines and Hg binding to the crystals was confirmed using MALDI–TOF mass spectrometry, which showed an Hg-induced peak at 200 mass units above the native peak at 22.2 kDa. Native and derivative diffraction data were collected using a rotating-anode generator equipped with R-AXIS IV image plates. Just before diffraction, crystals were dipped into a 30% glycerol solution for 5 s and cooled to 100 K.

Structure determination took place by the SIRAS method using the program *SOLVE/RESOLVE* (Terwilliger & Berendzen, 1999) as detailed in Table 2. Initial phase calculations, which were limited by the diffraction of the derivative, were performed at 2.9 Å resolution. Manual model building into a *RESOLVE* map at 2.33 Å resolution using the program *XFIT* (McRee, 1999) led to a 96% complete chain trace. Since many of the side chains were still truncated at this point (e.g. modeled as Ala when known to be Lys), the atom fraction of 4134 out of 4898 is a better measure of the model's completeness. Before any refinement, this model had an *R* value of 0.45 over the resolution range 10–4 Å. Structure refinement against data in the 10–2.33 Å shell utilized the program *REFMAC* v.5.2 (Murshudov *et al.*, 1997). Noncrystallographic symmetry was not restrained during the refinement in order to preserve any differences between the subunits. Continued model building, including the placement of waters in suitable difference density peaks, refinement and adjustment of atom positions, were carried out in several iterations. The refined model includes 247 water molecules and four glycerols, which were found to occupy the active site. Refinement statistics are given in Table 3.

### 3. Results and discussion

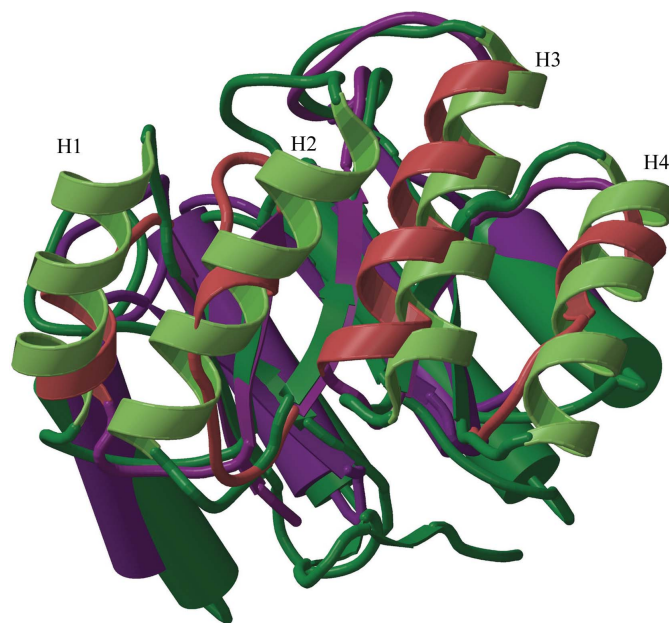
#### 3.1. Structure

The final structural model contains two dimers in the asymmetric unit, including all 196 residues in each of the four chains. Because the four chains overlay closely (the r.m.s.d. value is below 1.2 Å for all C $\alpha$  atoms in all pairwise comparisons) and the features described below are common to all four protomers, the following description and comparisons are based on subunit *A*. As expected from homology, AfDQ is a dimer of all-parallel ( $\beta\alpha$ )<sub>8</sub> barrels. The central  $\beta_8$  core of the barrel overlays both the archaeal and bacterial structures closely (r.m.s.d. = 1.4 Å for 48 core C $\alpha$  positions between AfDQ and SalmDQ). Most of the following archaeal *versus* bacterial structural

comparisons utilize SalmDQ (DQ from *Salmonella typhimurium*; PDB code 119w) for reasons of economy; in general, comparison with StafDQ (DQ from *Staphylococcus aureus*; PDB code 1sfl) could equally suffice as the two bacterial structures are relatively similar. Fig. 1 contains the aligned sequences, while Table 4 reports the overall  $\alpha$  and  $\beta$  content. Although the lengths and positions of the eight  $\beta$ -strands are conserved, their sequences contain only 14 identities in the 48 positions, which is slightly below the overall fraction conserved between AfDQ and SalmDQ. Sequence identities are distributed nearly equally among the strands, loops and helices, but the helices and their connecting loops show extensive structural variation as described below.

Multiple sequence alignment based on these three structures (Fig. 1) facilitates the comparison of structural elements among nine DQ enzymes: five archaeal and four bacterial. One obvious difference between the thermophilic sequences and the *Salmonella* and *E. coli* homologs is the 15-residue N-terminal extension in the mesophiles (which forms a  $\beta$ -hairpin that lies over one end of the barrel). Since the extension is present in SalmDQ (and in the close homolog from *E. coli*) but not in StafDQ, it is not an archaeal *versus* bacterial difference; it appears to be an adaptation acquired by some mesophiles in order to limit access to the active site inside the barrel. In both StafDQ and AfDQ, both ends of the barrel are open. An additional feature common to all three known DQ structures, one archaeal and two bacterial, is that they share a peculiarity that is absent in other ( $\beta\alpha$ )<sub>8</sub>-fold families: the adjacent barrel strands 7 and 8 have no backbone-to-backbone hydrogen bonds, making the barrel core discontinuous.

The most striking differences between the bacterial and archaeal structures are found in the first five of the eight helical regions of the barrel, which are those that are not involved in the dimer interface. While in both bacteria all eight segments on the barrel exterior are  $\alpha$ -helical, with their numbers of turns given by (4, 4, 4, 4, 4, 5, 4, 4), in AfDQ the helical regions are generally shorter and less helical, with numbers of turns given by (3, 1, 4, 1, 3, 4, 4, 3). Of these, the second



**Figure 2** Superposition of AfDQ (red and purple) and bacterial (*Salmonella*, green) structures, showing the key differences in helical secondary structure for helices H1–H4. Note that H1, H2 and H4 have fewer helical turns in the AfDQ structure. The AfDQ H4 structure is also shown in Fig. 3.

**Table 4**

Comparison of structural composition and intrasubunit ion pairs (IPs; distance cutoff 3.3 Å) in archaeal and bacterial DQs based on the A chains of 2ox1, 119w and 1sfl.

Protein	DEKR residues (%)	$\alpha/\beta$ content (%)	IPs	IPs per residue	IPs per DEKR residue
AfDQ (2ox1)	33	27/22	16	0.082	0.246
Salm (119w)	24	43/19	6	0.024	0.099
Staf (1sfl)	25	41/20	6	0.025	0.101

and fourth helices contain no  $\alpha$ -helix; instead, they are short segments with mostly  $\beta$  conformation and with one turn of  $3_{10}$ -helix each. Fig. 2 shows an overall superposition of the AfDQ and SalmDQ structures emphasizing these differences. Fig. 3 focuses on helix 4 in AfDQ, showing that this region is only marginally helical.

The active site is generally similar in terms of both its key residues and their geometry. Of the seven residues within 3.2 Å of the bound product in the SalmDQ product-complex structure 119w, five are well conserved in the alignment of Fig. 1 and are expected to play similar roles in AfDQ. These five are Glu23, Arg45, His98, Lys122 and Arg159 in AfDQ, corresponding to Glu46, Arg82, His143, Lys170 and Arg213 in SalmDQ. The greatest differences between AfDQ and SalmDQ in the active site are the two nonconserved product-binding residues, Thr6 and Lys178 in AfDQ (corresponding to Ser21 and Ser232 in SalmDQ), and two other residues that form van der Waals interactions with the SalmDQ product. These are Phe225, which stacks against the product in 119w and is replaced by Tyr171 in AfDQ, and Met203 (adjacent to the product in 119w), which is replaced by Phe149 in AfDQ. An analysis of the structural mechanisms of DQ catalysis has been made by Gourley *et al.* (1999).

The dimer interface involving helices 6, 7 and 8 is similar in its gross area and composition, but differs in detail and in the distribution of aliphatic and aromatic components. The main difference is that the greater ionic content of the archaeal protein provides it with several additional charges on the periphery of the interface and some of these form ion pairs across the interface. As a result, the dimer

**Table 5**

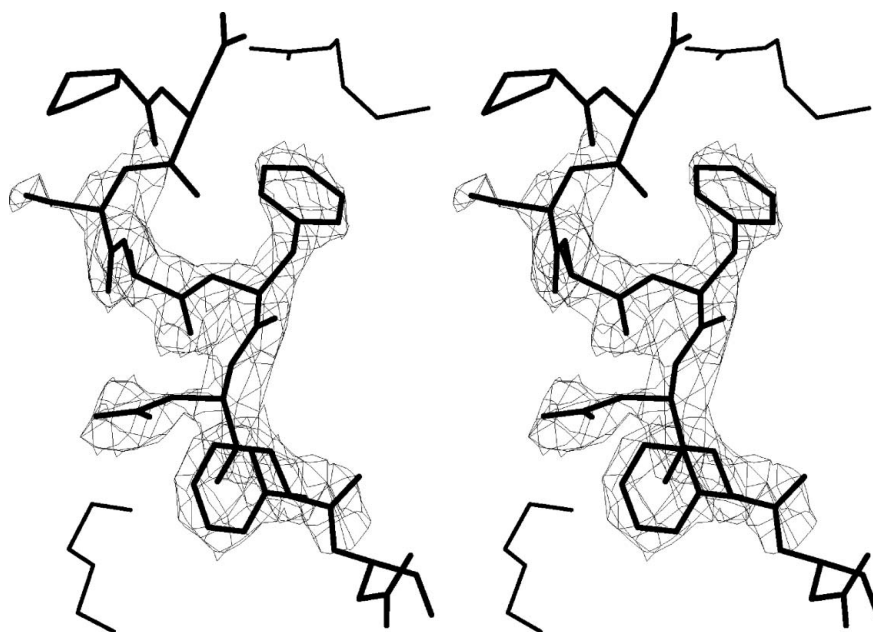
Melting temperature of purified AfDQ under various conditions.

Method	Protein conditions	$T_m$ (K)
CD	0.1 M potassium phosphate pH 7.2	345
CD	+ 1 M KCl	345
CD	+ 0.1 M tartrate	349
DSC	10 mM sodium cacodylate	341
DSC	+ 1 M NaCl	340

interface in AfDQ has six ion pairs, while in SalmDQ it has two. The buried area of the AfDQ dimer interface (using a 1.4 Å probe) is 1094 Å<sup>2</sup>, which is 11.3% of the protomer surface. The corresponding values for the bacterial dimers are 1101 Å<sup>2</sup> (10.3% of the protomer surface) for SalmDQ and 972 Å<sup>2</sup> (8.7% of the protomer surface) for StafDQ.

### 3.2. Thermal lability

The melting behavior of purified AfDQ under five different solution conditions was analyzed by CD and by DSC. The resulting values of  $T_m$ , all of which were in the range of 340–349 K, are given in Table 5. These findings may be compared with the  $T_m$  of 330 K measured for *E. coli* DQ (Kleanthous *et al.*, 1991). In all cases, the observed  $T_m$  of AfDQ was well below the growth optimum of the organism (356 K). The CD measurement was carried out under three different conditions to test the effect of 1 M salt and to test the effect of an active-site ligand. The salt was slightly destabilizing, but the substrate analog tartrate (Singh & Christendat, 2006) stabilized the protein by about 4 K. While in this case the tartrate-stabilized protein was still denatured below *in vivo* growth temperatures, the substrate (dehydroquinone) is likely to have a stronger stabilization effect, consistent with previous findings for the *E. coli* DQ (Reilly *et al.*, 1994). Fig. 4 shows the CD traces with and without 1 M KCl. The CD results in 1 M KCl show a bimodal curve, indicating a secondary melting transition that may arise from disruption of the dimer (Fig. 4). The AfDQ  $T_m$  was measured by DSC in 10 mM buffer at two ionic



**Figure 3**

Stereo diagram (convergent) showing the marginally helical segment in AfDQ that corresponds to helix 4 in the bacterial structures. The view direction is from outside the barrel. Electron density ( $2F_o - F_c$ ) for residues 85–89 contoured at  $1.2\sigma$  is also shown. Some parts of the adjacent helices 3 and 5 are visible as thinner lines to the sides.

strengths: with and without 1 M NaCl. Each of the resulting scans showed a single peak indicating cooperative unfolding. The relative insensitivity of  $T_m$  to ionic strength may be the result of a close balance between a stabilizing effect arising from shielding of unfavorable repulsions between like charges and a destabilizing effect of shielding between opposite charges, which may include disruption of ion pairs.

### 3.3. Residue distributions

Previous genomic studies have shown that proteins from thermophiles have a higher proportion of charged residues than their mesophilic counterparts (30% versus 24%; Deckert *et al.*, 1998). This tendency is maintained in the DQ family (Table 4), with the five thermophile sequences in Fig. 1 having an average of  $30.1 \pm 3.0\%$  DEKR (Asp, Glu, Lys and Arg) residues, while the four mesophilic counterparts average  $21.9 \pm 1.4\%$ . Furthermore, consistent with the genomic analysis, the distribution of polar uncharged surface residues is globally similar among the DQs of known structure, but with a tendency for uncharged residues to be replaced by charged residues in the thermophile. The ratio  $(E + K)/(Q + H)$  has been found to correlate with thermostability (Farias & Bonato, 2003); in AfDQ this ratio is 14.5. Fig. 1 includes the optimal growth temperatures of the nine organisms for which the DQ sequences are compared, along with the values of  $(E + K)/(Q + H)$  based on the sequence. The thermophiles all have higher values (ranging from 5.7 to 27.0) of this ratio than the mesophiles (range 1.1–2.5). (Note that the archaeon *Halo bacterium* is not a thermophile, but the non-archaeon *Aquifex* is.)

The 65 DEKR residues in AfDQ form 16 intrasubunit ion pairs (IPs; 3.3 Å distance cutoff). There are only six in each of the larger

bacterial subunit structures (Table 4). The AfDQ structure has over twice as many IPs per residue compared with its bacterial homologs; this increase in IP correlates closely with the increase in charged-residue (DEKR) content. Among the IPs are two that are conserved among all three structures, involving Arg25 and Arg45 in AfDQ. These two are close to the product in the *Salmonella* structure 119w and appear to be involved in substrate binding. Protein compactness (calculated as the reciprocal of the radius of gyration, normalized by the cube root of the reciprocal mass) was found not to vary significantly between AfDQ and SalmDQ. Similarly, neither the locations of prolines nor the placement of charges with respect to the helices show a significant difference between the thermophilic and mesophilic structures.

### 3.4. Loops

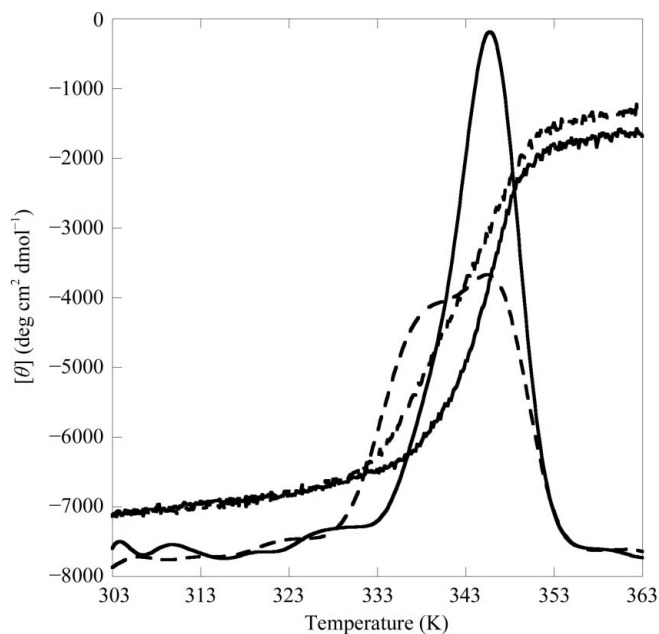
Formally, there are 15 connections between the 16 sequential  $\beta$  and  $\alpha$  elements in the  $(\beta\alpha)_8$  fold. While the two bacterial structures are very similar, about half of the loops have a different conformation in AfDQ compared with the bacterial DQs. The differences in loop conformations are usually associated with insertions in the bacterial sequences, but in some cases appear to be the consequence of sequence differences alone. They affect both ends of the barrel about equally. Loop s2–h2 (connecting strand 2 to helix 2) has a 14-residue insertion in the bacterial sequences that extends h2 and, together with a six-residue s1–h1 insertion, adds bulk to this end of the barrel. Additional six-residue insertions in the h4–s5 and h6–s7 loops extend these helices and add bulk to the other end of the barrel.

Fig. 1 shows that the archaeal sequences have shorter loops and in some cases shorter secondary-structural elements. Structure-based alignment shows that in going from the archaeal to the bacterial proteins both termini are extended and there are six internal insertions. The lengths of the eight augmentations (termini and insertions) are 15, six, two, 14, six, three, six and two residues for SalmDQ with respect to AfDQ. The sizes and positions of the eight augmentations are given by the positive values between secondary-structure elements s1, h1, s2 *etc.* in the string (+15, s1, +6, h1, +2, s2, +14, h2s3h3s4h4, +6, s5h5, +3, s6h6, +6, s7h7s8h8, +2). Most (five of the eight) of the augmentations are at the C-terminal end of a helix. Most of the inserted residues are in loops, but the helices are also lengthened.

One possible explanation for the larger loops generally observed in mesophilic proteins is that they could provide a method of tuning stability to provide for turnover *via* denaturation and/or as protease-sensitive sites for degradation. Alternatively, they could exist as an evolutionarily neutral background produced by random sequence extensions, with a low enough stability and metabolic cost that they are maintained. In this latter scenario, the functionless loop extensions enable a low-cost search over evolutionary time for favorable new structural components by providing raw material for occasional development of new specific interactions and functions. Consistent with either hypothesis, hyperthermophilic proteins require shorter loops, as the loops are expected to have increased stability cost at higher temperature.

## 4. Conclusions

The structure of an archaeal DQ enzyme enables analysis of the structure–stability relationship and comparison with bacterial homologs. Structure comparisons are limited to the three known DQ structures, AfDQ and two bacterial mesophiles, while comparisons of aligned sequences can include both archaeal mesophiles and bacterial



**Figure 4**  
Plot of circular dichroism during thermal denaturation of AfDQ. The mean residue ellipticity at 222 nm is plotted *versus* temperature in the range 303–363 K. The temperature profiles were recorded using a 1 cm cylindrical cuvette with protein concentrations of  $50 \mu\text{g ml}^{-1}$  in 0.1 M potassium phosphate pH 7.2 (solid line) and 0.1 M potassium phosphate pH 7.2 plus 1 M KCl (dashed line). The ellipticity results are expressed as mean residue ellipticity  $[\theta]$ . Also shown are the corresponding derivative curves (no KCl, solid line; 1 M KCl, dashed line). The peaks of these curves correspond to the inferred  $T_m$  values, which are 345 K for both conditions. The curves with KCl show bimodal melting behavior with a lower temperature shoulder, indicating a partial loss of structure before the main transition.



thermophiles as in Fig. 1. While the general shortening of loops in thermophiles is as observed in many such comparisons, the shortening and disruption of helices is unusual. It suggests an evolutionary pressure to minimize the size of the protein, even at some cost in the stability associated with main-chain hydrogen bonding. Destabilization owing to the lower fraction of helix in AfDQ (Table 4) appears to be compensated by increases in IPs, by shorter loops before and after the shortened helices and by hydrophobic placement whereby the usual helical periodicity in hydrophobicity is altered to a more  $\beta$ -like periodicity. For example, the segment Phe87-Asp-Phe-Asn, which belongs to the h4 region but is not helical (Figs. 2 and 3), places the aromatic residues inward against other apolar residues and the polars outward towards solvent. Nonregular structural elements such as these are much more common in AfDQ compared with the bacterial DQs (Table 4) and appear to be stabilized by such hydrophobic placements, perhaps as a means to locally compensate for the lower incidence of main-chain hydrogen bonding.

As protein-lability measurements become more common, clearer terms and metrics may benefit the description and categorization of labile proteins. It is important to distinguish between lability in isolation, *i.e.* the lability of the purified protein *in vitro*, and the more restrictive (but harder to prove) condition of native lability, *i.e.* lability *in vivo*. Spanning these poles is a range of conditionally labile cases in which the stability and structure of a protein are influenced by various ligands. A large class of proteins is now known for which the structure depends on the ligand. Further variations and subclasses are inevitable, with each structure being somewhat flexible and somewhat influenced by interactions. In the most extreme test of classification schemes, we can imagine a protein whose several subdomains vary independently or cooperatively among random coil, molten globule and multiple distinct well ordered states, depending upon specific and nonspecific interactions. Protein-lability studies in general would be advanced by routine, perhaps automated, measurement of  $T_m$  *in vitro* for proteins produced in proteomics structure projects. The cost of such a measurement by CD is extremely low compared with structure investigation. Furthermore, it would be beneficial to have such data archived in a public database.

The type I dehydroquinase from *A. fulgidus* has been found to denature *in vitro* at about 343 K, regardless of salt. Since the organism's growth optimum is 356 K, AfDQ belongs to the set of isolation-labile proteins. A small stabilizing effect was observed in the presence of the substrate analog tartrate, suggesting that AfDQ is ligand-stabilized and may belong to the set of ligand-ordered (*i.e.* disordered until ligand-bound) proteins. Although it is expected that AfDQ is folded when performing catalysis *in vivo*, its structural status during the 'off' phase of its catalytic duty cycle *in vivo* is unknown. It may maintain its fold by retaining its product until another substrate is available, it may adopt an unliganded and partly unfolded (perhaps with the barrel opened) state or else its fold may be stabilized by different ligand(s) such as a chaperonin or by one of the 'compatible solutes' that have been found to stabilize proteins in archaea (Roberts, 2000; Santos & Costa, 2002).

The authors gratefully acknowledge the gift of the cloned gene from Hal Monbouquette and Imke Schroeder and helpful discussions with and assistance from Phil Bryan and Fred Schwarz. Funding for this project was provided by the US National Institute of Standards and Technology. Identification of specific instruments and products in this paper is solely to describe the experimental procedures and does not imply recommendation or endorsement.

## References

- Bergdoll, M. S. (1989). *Foodborne Bacterial Pathogens*, edited by M. P. Doyle, pp. 463–523. New York: Marcel Dekker.
- 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., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J. & Swanson, R. V. (1998). *Nature (London)*, **392**, 353–358.
- Dyson, H. J. & Wright, P. E. (2005). *Nature Rev. Mol. Cell Biol.* **6**, 197–208.
- Farias, S. T. & Bonato, M. C. M. (2003). *Genet. Mol. Res.* **2**, 383–393.
- Fiala, G. & Stetter, K. O. (1986). *Arch. Microbiol.* **145**, 56–60.
- Fink, A. L. (2005). *Curr. Opin. Struct. Biol.* **15**, 35–41.
- Gourley, D. G., Shrive, A. K., Polikarpov, I., Krell, T., Coggins, J. R., Hawkins, A. R., Isaacs, N. W. & Sawyer, L. (1999). *Nature Struct. Biol.* **6**, 521–525.
- Gray, J. T. & Fedorka-Cray, P. J. (2002). *Foodborne Diseases*, edited by D. O. Cliver & H. Riemann, pp. 55–68. New York: Academic Press.
- Hocker, B., Jurgens, C., Wilmanns, M. & Sterner, R. (2001). *Curr. Opin. Biotechnol.* **12**, 376–381.
- Kleanthous, C., Reilly, M., Cooper, A., Kelly, S., Price, N. C. & Coggins, J. R. (1991). *J. Biol. Chem.* **266**, 10893–10898.
- Klenk, H. P. *et al.* (1997). *Nature (London)*, **390**, 364–370.
- Klipcan, L., Safro, I., Temkin, B. & Safro, M. (2006). *FEBS Lett.* **580**, 1672–1676.
- Lee, W.-H., Perles, L. A., Nagem, R. A. P., Shrive, A. K., Hawkins, A., Sawyer, L. & Polikarpov, I. (2002). *Acta Cryst. D* **58**, 798–804.
- Lorentzen, E., Pohl, E., Zwart, P., Stark, A., Russell, R. B., Knura, T., Hensel, R. & Siebers, B. (2003). *J. Biol. Chem.* **278**, 47253–47260.
- Lo Leggio, L., Kalogiannis, S., Bhat, M. K. & Pickersgill, R. W. (1999). *Proteins*, **36**, 295–306.
- McRee, D. (1999). *J. Struct. Biol.* **125**, 156–165.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst. D* **53**, 240–255.
- Nichols, C. E., Lockyer, M., Hawkins, A. R. & Stammers, D. K. (2004). *Proteins*, **56**, 625–628.
- Reilly, A., Morgan, P., Davis, K., Kelly, S. M., Greene, J., Rowe, A. J., Harding, S. E., Price, N. C., Coggins, J. R. & Kleanthous, C. (1994). *J. Biol. Chem.* **269**, 5523–5526.
- Roberts, M. F. (2000). *Frontiers Biosci.* **5**, 796–812.
- Robinson, J. L., Pyzyna, B., Atrasz, R. G., Henderson, C. A., Morrill, K. L., Burd, A. M., DeSoucy, E., Fogleman, R. E. III, Naylor, J. B., Steele, S. M., Elliott, D. R., Leyva, K. J. & Shand, R. F. (2005). *J. Bacteriol.* **187**, 923–929.
- Rohlin, L., Trent, J. D., Salmon, K., Kim, U., Gunsalus, R. P. & Liao, J. C. (2005). *J. Bacteriol.* **187**, 6046–6057.
- Santos, H. & Costa, M. S. D. (2002). *Environ. Microbiol.* **4**, 501–509.
- Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L. & Engel, P. C. (2000). *Frontiers Biosci.* **5**, 787–795.
- Singh, S. A. & Christendat, D. (2006). *Biochemistry*, **45**, 7787–7796.
- Stetter, K. O., Lauerer, G., Thomm, M. & Neuner, A. (1987). *Science*, **236**, 822–824.
- Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst. D* **55**, 849–861.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.