Crystal structure of a thermophilic alcohol dehydrogenase substrate complex suggests determinants of substrate specficity and thermostability

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · DECEMBER 1999

Impact Factor: 2.63 · DOI: 10.1002/(SICI)1097-0134(19991201)37:4<619::AID-PROT12>3.0.CO;2-H

PROTEINS: Structure, Function, and Genetics 37:619–627 (1999)

Crystal Structure of a Thermophilic Alcohol Dehydrogenase Substrate Complex Suggests Determinants of Substrate Specificity and Thermostability

Chunmin Li, Joel Heatwole, Sandriyana Soelaiman, and Menachem Shoham*
Department of Biochemistry, School of Medicine, Case Western Reserve University, Cleveland, Ohio

ABSTRACT **The crystal structure of a thermophilic alcohol dehydrogenase (TBAD) from *Thermoanaerobacter brockii* has been determined in a binary complex with sec-butanol as substrate to a resolution of 3.0 Å. Van der Waals interactions of the carbon C1 atom of sec-butanol with atoms in His59, Ala85, Trp110, Asp150, and Leu294 account for the substrate preference of this enzyme for secondary over primary alcohols. A crevice from the surface to the active site provides access for substrates and products. This opening is lined with the hydrophobic residues Ile49, Leu107, Trp110, Tyr267, Leu294 as well as Cys283 and Met285 from another molecule within the tetrameric assembly. This might explain the tolerance of this enzyme toward organic solvents. The zinc ion occupies a position in the active site, which is too remote for direct interaction with the alcohol group. A mechanism is suggested whereby the introduction of NADP would trigger a displacement of the zinc ion to its catalytic site. Features important for the unusually high melting temperature of 98°C are suggested by comparison to the crystal structure of a highly homologous mesophilic alcohol dehydrogenase from *Clostridium beijerinckii* (CBAD). The thermophilic enzyme has a more hydrophilic exterior, a more hydrophobic interior, a smaller surface area, more prolines, alanines, and fewer serines than CBAD. Furthermore, in the thermophilic enzyme the number of all types of intersubunit interactions in these tetrameric enzymes is increased: more salt bridges, hydrogen bonds, and hydrophobic interactions. All these effects combined can account for the higher melting temperature of the thermophilic enzyme. Proteins 1999;37:619–627.** © 1999 Wiley-Liss, Inc.

Key words: enzyme; extremophile; heat-stable; substrate binding; crystallography

INTRODUCTION

The alcohol dehydrogenase (TBAD) from *Thermoanaerobacter brockii*, a bacterium isolated from hot springs in Yellowstone National Park, is a nicotinamide adenine dinucleotide phosphate (NADP)-dependent enzyme that catalyzes the oxidation-reduction reactions of alcohols to their corresponding aldehydes or ketones.[1] The best substrates are secondary alcohols.[1] TBAD has been used in organic synthesis as a biocatalyst for the stereospecific reduction of a broad range of ketones to their corresponding secondary alcohols because of its high tolerance toward organic solvents.[2] TBAD is a tetrameric enzyme of identical 37,652 Da subunits composed of 352 amino acids.[3] In contrast to eukaryotic alcohol dehydrogenases that contain a catalytic Zn^{2+} as well as a structural Zn^{2+}, TBAD has only a single Zn^{2+} ion per subunit, reported to be of catalytic function.[4] The most striking property of TBAD is its thermal stability. The temperature at which half the activity is lost after 1 h of incubation is 93°C, and the melting temperature is 98°C.[5]

The thermal stabilization of proteins is an intriguing scientific problem with important evolutionary and biotechnological implications.[6–8] What mutation(s) should be engineered to increase the heat tolerance of a protein? There may be no unique answer to this question. Structural studies of thermophilic proteins and their mesophilic counterparts have indicated that many small contributions, but no single structural feature can fully account for the increased thermostability.[9–14] A powerful way to stabilize oligomeric proteins is to increase the number of intersubunit interactions. However, in a hyperthermophilic organism all proteins have to be heat resistant, even monomeric ones. Factors known to contribute to thermal stability include more hydrophobic contacts in solvent-inaccessible areas, increased numbers of salt bridges, more hydrogen bonds, more proline residues, and a higher degree of compactness.

The crystal structures of several eukaryotic alcohol dehydrogenases have been determined, including those from cod and horse liver (CLAD and HLAD),[15,16] as well as

Abbreviations: TBAD, alcohol dehydrogenase from *Thermoanaerobacter brockii*; CBAD, alcohol dehydrogenase from *Clostridium beijerinckii*; CLAD, alcohol dehydrogenase from cod liver; HLAD, alcohol dehydrogenase from horse liver; NADP, nicotinamide adenine dinucleotide phosphate; MPD, methyl pentane diol; PEG, polyethylene glycol; MES, 2[N-morpholino]ethanesulfonic acid.

Grant sponsor: Army Research Office; Grant numbers: DAAH04–95–1-0396 and DAAG55–97–1-0214.

Chunmin Li's present address is Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada.

*Correspondence to: Menachem Shoham, Department of Biochemistry, Case Western Reserve University School of Medicine, 10900 Euclid Avenue, Cleveland, OH 44106-4935. E-mail: shoham@biochemistry.cwru.edu

Received 24 May 1999; Accepted 22 July 1999

TABLE I. Statistics of Data Collection and Processing

Resolution (Å)	Reflections	Completeness		Redundancy		$\langle I/\sigma \rangle$		R_{merge}[a]
		All	3.10–2.99 Å	All	3.10–2.99 Å	All	3.10–2.99 Å	
2.99	30323	0.863	0.890	4.4	4.8	12.1	4.6	0.083

[a] $R_{merge} = \Sigma_h \Sigma_i |I_{ih} - \langle I_h \rangle| / \Sigma_h \Sigma_i \langle I_h \rangle$ where $\langle I_h \rangle$ is the mean intensity of the i observations of reflection h.

three human recombinant forms.[17–19] The ligands to the catalytic Zn^{2+} in eukaryotic alcohol dehydrogenases are a histidine residue and two cysteine residues. The fourth ligand is usually a water molecule in the absence of substrate. When the substrate is bound to the enzyme, the alcohol oxygen or carbonyl oxygen atom appears to coordinate directly to the Zn^{2+}, displacing the molecule of water. Studies of the pH dependence of binding of a variety of alcohols to HLAD indicate that the alcohol probably binds in the form of alcoholate anion.[20] The negative charge on the oxygen atom makes the alcoholate a better donor of a hydride ion for the reduction of NAD^+, compared with the undissociated alcohol. The positively charged zinc ion stabilizes the negative charge of the alcoholate in HLAD.

The sequence of TBAD is remotely related to those of eukaryotic alcohol dehydrogenases with a degree of sequence identity around 30%. In contrast, TBAD is highly homologous to a mesophilic prokaryotic alcohol dehydrogenase from *Clostridium beijerinckii* (CBAD) with a sequence identity of 75%.[21] CBAD and TBAD differ in their thermal stability and in their resistance to denaturing agents. The extent of resistance to denaturation in TBAD depends on the presence of the cofactor.[21] The temperature at which half the activity is lost within 1 h is 67°C and 93°C for CBAD and TBAD, respectively.[21] The increased thermostability in TBAD must be localized to those 25% of the residues that are different in the two enzymes.

The crystal structure of TBAD in a binary cofactor complex with NADP has been determined at 2.5 Å resolution while this study was in progress.[22] We report here the crystal structure of TBAD in a binary substrate complex with sec-butanol to a resolution of 3.0 Å. Catalysis does not take place in these crystals because of the absence of cofactor.

MATERIALS AND METHODS

Crystallization, X-Ray Data Collection and Processing

A recombinant form of TBAD, expressed in *Escherichia coli*, was obtained from the laboratory of Dr. Yigal Burstein at the Weizmann Institute of Science. TBAD crystals were grown by the vapor diffusion method against a well solution containing 10% PEG-8000, 5.8% (v/v) methyl pentane diol (MPD), 2.5% (v/v) sec-butanol, 50 mM magnesium acetate, 0.1 M 2-[*N*-morpholino]ethanesulfonic acid (MES) buffer, pH 5.8. The crystals belong to space group $P2_12_12_1$, with unit-cell dimensions $a = 80.45$ Å, $b = 123.08$ Å, $c = 168.03$ Å. There are four molecules of TBAD per asymmetric unit. The volume-to-mass ratio is 2.86 $Å^3Da^{-1}$, which is close to the average for protein crystals.[23] The solvent content is 57%.

Before data collection, the crystals were dialyzed overnight against a cryoprotectant solution containing 10% PEG-8000, 20% (v/v) MPD, 2.5% (v/v) sec-butanol, 50 mM magnesium acetate, 0.1 M MES buffer, pH 5.8. X-ray diffraction data were collected at CHESS beamline A1 by using synchrotron radiation ($\lambda = 0.908$ Å) at 100 K, with a crystal-to-detector distance of 253 mm. The crystals diffracted to 2.99 Å. X-ray diffraction data were processed by using the program DENZO and reduced with SCALEPACK.[24] Table I shows the statistics of data collection and processing.

Determination of Zinc Content

Several TBAD crystals were dissolved in distilled water, and the solution was analyzed for metal ion content by inductively coupled plasma mass spectroscopy by using a VG Elemental Plasmaquad^{2+} inductively coupled plasma mass spectrometer (VG Elemental, Beverly, MA), equipped with a CETAT technologies U-6000 ultrasonic nebulizer (CETAC Technologies, Omaha, NE).

Structure Determination and Refinement

The structure of TBAD was determined by molecular replacement with the program AmoRe[25] within the CCP4 program package,[26] using reflections with $F < \sigma_F$ in the resolution range from 15 to 4 Å. The crystal structure of the CBAD monomer (PDB code 1KEV) devoid of NADP and water molecules was chosen as a starting model.[22] (The coordinates for the TBAD-NADP binary complex were not available at the time this structure was determined.) No outstanding rotation function solutions were found. Translation-function searches were subsequently performed at the same resolution range for the top 35 rotation solutions. Four of the solutions, ranked 1, 3, 13, and 26 in the rotation search, were found to be outstanding in the translation search with correlation coefficients at least 38% higher than the rest of solutions. The correctness of these four solutions was validated by the fact that they pack well to form the tetramer, and they conform to the subunit relationships revealed by a self-rotation search (data not shown). The tetramer has local 222 symmetry. The model of the tetramer generated from the four top solutions yielded an R-factor of 36.1% and a correlation coefficient of 63.6% at 15–4 Å resolution range.

The molecular-replacement model was subsequently modified by mutating nonconserved residues to Ala and setting all B factors to 20 $Å^2$. Refinement was performed by using programs X-PLOR, versions 3.1 and 3.8[27] as well as CNS.[28] All data to 2.99 Å were used in the refinement. Rigid-body refinement was performed to improve the

positional and orientational parameters of the four subunits. Strict noncrystallographic symmetry was used in the initial refinement cycles. A series of 20-residue omit (F_o–F_c) maps were calculated and used to retrace the whole model using the TBAD amino acid sequence. The Zn^{2+} ion and the substrate were located from a residual (F_o–F_c) map after all protein residues were accounted for. Manual rebuilding between refinement cycles was performed with the computer graphics software O.[29] The model was subjected to several cycles of simulated annealing by "heating" the system to 3,500 K and slowly "cooling" it to 300 K in time steps of 0.00025 ps. This reduced the R value to 26.3%. Positional parameters and temperature factors were then refined in alternating cycles. A ($2F_o$–F_c) Fourier map was computed at every stage of the refinement, and the model was adjusted manually to fit this electron-density map until the R factor and R_{free} factor were reduced to 22.3% and 27.7%, respectively. At this stage, water molecules were added. If the temperature factor of a solvent molecule exceeded 70 Å^2, or a solvent molecule moved out of density, it was deleted from the model. At this stage, all four monomers in the asymmetric unit were refined with restraints to the local symmetry. Effective force constants of 20 and 10 Kcal mol^{-1} Å^{-2} were used for main chain atoms and side chain atoms, respectively.

Eight of the solvent molecules exhibited temperature factors of 2 Å^2, which is the lower limit set by the program. This was taken as an indication that these atoms should have more electrons than oxygen has. Four of these solvent molecules were assigned as Mg^{2+} ions and the other four as Cl^- ions. The four ion pairs are related to each other by the local 222 symmetry. Each pair of Mg^{2+} and Cl^- ions forms a bridge between two subunits at the interface via an intervening water molecule. The Mg^{2+} ion is bound to a carboxylate oxygen of Asp307 and to the water molecule, which in turn binds to the Cl^- ion. The other ligands to the chloride ion include the guanido group of Arg 193 and the ϵ-amino group of Lys 234. The ligands to the Mg^{2+} and Cl^- ions belong to different subunits. This explains the observation that the addition of Mg^{2+} ions to the crystallization mixture improves the quality of the crystals.

The Ramachandran plot shows that 83.2% of the residues are within most favored regions with none in disallowed regions, as calculated with the program PROCHECK.[30] The statistics of the crystallographic analysis and the final model are listed in Table II. The coordinates have been deposited at the Protein Data Bank (code 1BXZ).

TABLE II. Refinement Statistics

Resolution (Å)	50–2.99
Reflections (work/free)	25745/2575
R[a]/Rfree[b]	0.214/0.265
No. of protein atoms	2638
No. of water molecules	248
No. of ions	12
rmsd bond lengths (Å)	0.013
rmsd bond angles (°)	1.69
Average B factor (Å^2)	41.9

[a]R = Σ|Fo − Fc|/ΣFo Fo, observed structure factor; Fc, calculated structure factor. [b]R_{free}, defined as in R for a randomly selected set of 10% of the reflections.

RESULTS AND DISCUSSION

Structure Description

There are four molecules of this enzyme-substrate complex in the asymmetric unit. During the course of the refinement, these four molecules were restrained to each other by the local 222 symmetry. The rms deviations of corresponding backbone atoms from the mean for a pair of subunits ranges from 0.16 to 0.21 Å. Because of this high degree of similarity, the following description applies to all four molecules.

As in all other alcohol dehydrogenases of known structure, the monomer is made up of two domains, the catalytic domain and the cofactor-binding domain (Fig. 1). The catalytic domain contains the N- and C-terminal segments, consisting of residues 1–149 and 297–352. There are two β-sheets and five α-helices in the catalytic domain. Strands 1 and 2 form a small antiparallel β-meander on the surface of this domain, whereas strands 3, 4, 5, 12, and 13 form a mixed β-sheet. The loop between strands 3 and 4 is very long and contains a helical stretch α1. After β5 there are four α-helical segments in a row. The last of these, α5, links the catalytic domain to the nucleotide-binding domain.

The nucleotide-binding domain consists of the contiguous stretch of residues 157–292. It forms a typical Rossmann fold with six parallel β-strands, β6–β11. This β-sheet is flanked by α-helical segments, α6 and α7 on one side of the sheet, and α8 and α9 on the other side of the sheet. After the last strand in this sheet, β11, there is a long loop that meanders back into the catalytic domain to lead into helix α10, strand β12, helix α11 and finally strand β13. The two domains are separated by a deep active site cleft that is accessible to the solvent.

The Active Site

The location of sec-butanol and the zinc ion in the active site cleft was determined from a residual (F_o–F_c) difference Fourier map (Fig. 2). The binding site for the substrate sec-butanol consists of the hydrophilic ligands His59 and Asp150 as well as the hydrophobic residues Ala85, Trp110, Leu294, and Met285 from another subunit in the tetrameric assembly. The hydroxyl group of the substrate is hydrogen bonded to Nϵ2 of His59 and one of the carboxylate oxygen atoms of Asp150 with distances of approximately 3.8 and 3.4 Å, respectively. These two residues have been shown by alanine mutagenesis to be essential for catalysis.[4] The C1 atom of sec-butanol is located within van der Waals distance of atoms in the following five residues: His59, Ala85, Trp110, Asp150, and Leu 294 (Fig. 3). In the absence of the C1 atom these hydrophobic interactions would be lost, the substrate would presumably not bind as well, and the K_M value would be higher. This structural feature can account for the observation

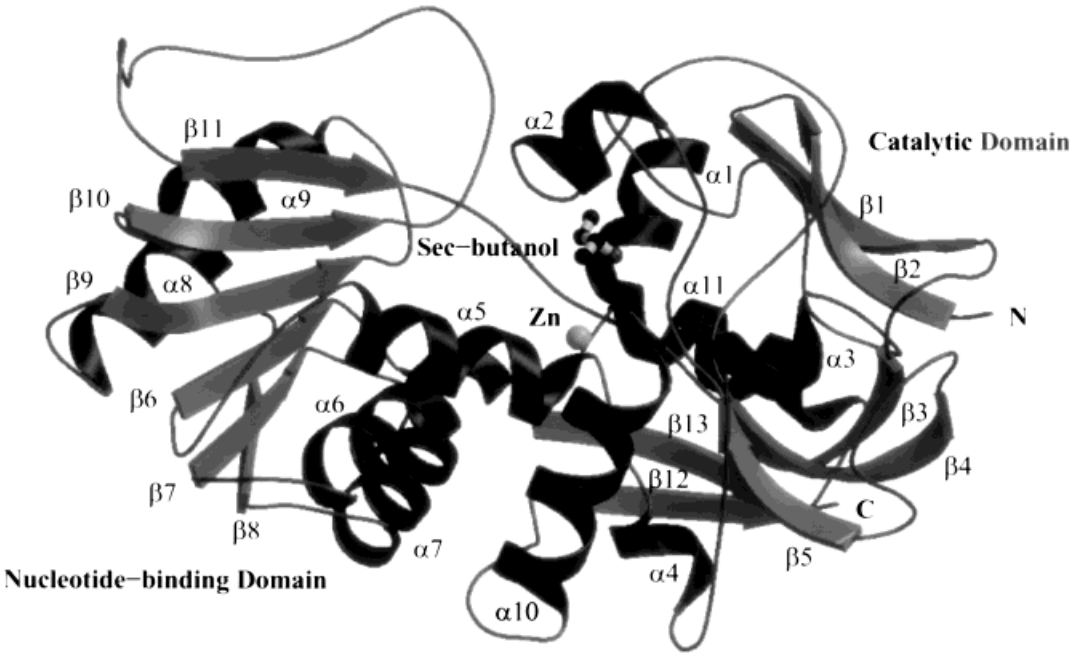

Fig. 1. Ribbon drawing of the TBAD monomer. Helices and strands are shown in black and gray, respectively. The amino and carboxy termini are labeled N and C, respectively. The enzyme is made up of two domains, the nucleotide-binding domain and the catalytic domain. The two domains are separated by a deep active site cleft that is accessible to the solvent. The substrate sec-butanol is depicted in ball-and-stick. The Zn^{2+} ion is shown as a sphere at the bottom of the cleft in the catalytic domain. This figure was created with programs MOLSCRIPT[46] and RASTER3D.[47]

that secondary alcohols are much better substrates for TBAD than primary alcohols.[1]

There is a crevice from the active site to the surface of the protein for substrates and products to diffuse in and out. This opening is lined with the hydrophobic residues Ile49, Leu107, Trp110, Tyr267 as well as Cys283 and Met285 from another polypeptide chain in the tetramer. The preponderance of hydrophobic residues in this access channel could explain the tolerance of TBAD toward organic solvents.

The enzyme contains zinc ions, as established by inductively coupled plasma mass spectroscopy (see Materials and Methods). A sample of dissolved crystals contained 11.2 μ g/L or 0.175 μ M Zn^{2+}. The protein concentration of this solution was 5.6×10^3 μg/L or 0.149 μ M as determined by absorption at 280 nm. Hence, the measured stoichiometry is 1.17 zinc ions per monomer. It follows that the amount of zinc is sufficient to occupy a site in 100% of the monomers. The refined isotropic temperature factor for the four zinc ions ranges from 26.15 to 28.00 Å^2 The location of the zinc ion was unequivocally determined from the highest peak (7σ) in a residual (F_o–F_c) difference Fourier map (Fig. 2). Residues surrounding the zinc ion include Cys37, Thr38, Met337, and Met151. His59 and Asp150 are not in direct contact with the zinc ion. Surprisingly, the zinc ion is not in direct proximity to the substrate, but it is located 7 Å away from the hydroxyl oxygen atom of the substrate.

In the absence of a structure for the ternary (substrate-NADP-enzyme) complex, one can gain some insight into the structure of such a ternary complex before catalysis by

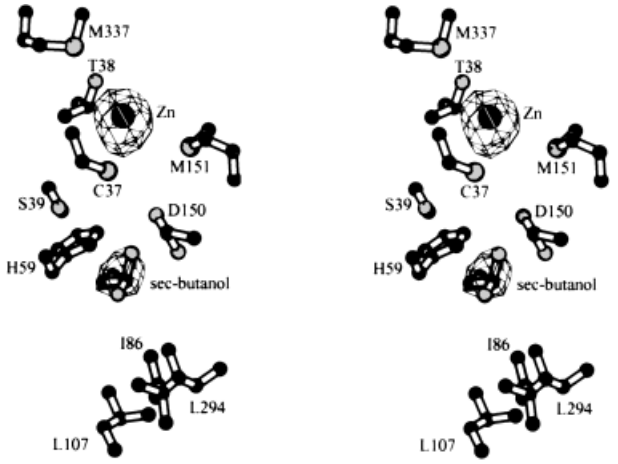

Fig. 2. Ball-and-stick representation of residues in the vicinity of the active site and residual (F_o–F_c) difference Fourier map contoured at a level of 2.8 standard deviations above the mean. F_c contains contributions from all protein atoms but no zinc or sec-butanol. The two peaks shown here are the two most prominent peaks in this map with peak heights of 7.0 and 4.3 σ, respectively. The higher peak has been interpreted as the zinc ion because of its higher electron density and its proximity to Cys37. The other peak has been assigned to sec-butanol because of its proximity to the catalytic residues His59 and Asp150 and a hydrophobic region of residues Ile86, Leu107, and Leu294. The zinc ion is not in direct proximity to the substrate but is located 7 Å away from the hydroxyl oxygen atom of the substrate. Residues surrounding the zinc ion include Cys37, Thr38, Met151, and Met337. The active site is similar in all four subunits of the asymmetric unit. Subunit A is shown in this figure. The program BOBSCRIPT was used to make this figure.[46,48]

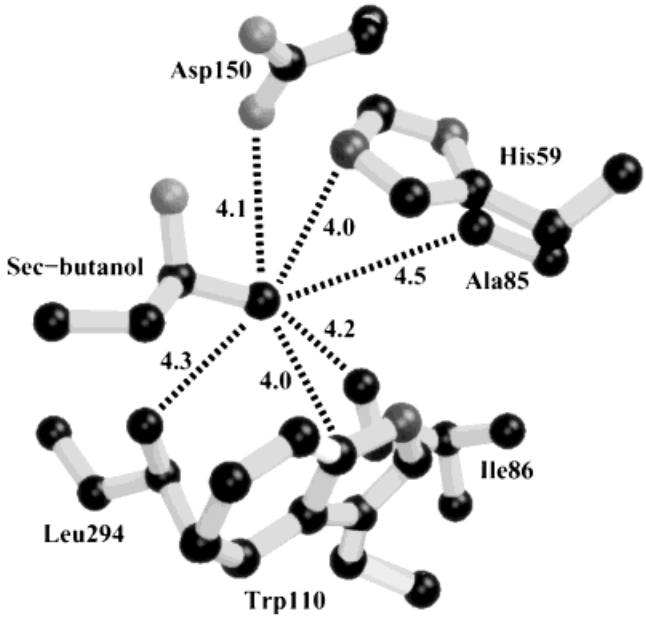

Fig. 3. Van der Waals distances in Å from atom C1 of sec-butanol in the active site of TBAD. These hydrophobic interactions could explain why secondary alcohols are better substrates than primary alcohols.[1]

superimposing the respective crystal structures of the two binary complexes, one with NADP[22] and the other with sec-butanol (this study). The two structures superimpose very well with rms deviations of 0.49 Å for all backbone atoms and 0.94 Å for all atoms, respectively. The only discrepancy in the active site region is the location of the zinc ion (Fig. 4). For the zinc ion to be involved in catalysis, a conformational change would be required, presumably on NADP binding, which would bring the zinc ion closer to the substrate. Indeed, substantial conformational changes between the apo and holo forms of liver alcohol dehydrogenases have been observed.[31] In the superposition, the hydroxyl group of the substrate is 2.94 Å away from the zinc ion position in the NADP complex and 2.35 Å apart from the atom C4 of the nicotinamide ring. Thus, this zinc ion is within contact distance of both the substrate and the cofactor. The role of the zinc ion seems to be to facilitate the hydrogen transfer from the alcohol substrate to the cofactor, as is the case for the eukaryotic alcohol dehydrogenases.[20]

Structural Features That Account for the Thermostability

The degree of sequence identity between TBAD and CBAD is 75%. Residues that account for the increased thermostability in TBAD must therefore be located in the 25% of the residues that are different between the two enzymes (Fig. 5). The sequence alignment is unambiguous because there are no gaps. There are a total of 87 residues that are different between the two enzymes. As shown in Table III, the most significant differences in amino acid composition between the two enzymes are for alanine, proline, and serine.

Proline residues

The number of proline residues increases from 13 in the mesophilic enzyme to 21 in the thermophilic enzyme (Table III). Proline reduces the main-chain flexibility of the backbone, thereby decreasing the entropy and increasing the free energy of the unfolded state.[32] Consequently, replacing any amino acid residue with a proline should destabilize the unfolded state, thereby stabilizing the folded state. However, replacements to proline do not occur at random positions but at preferential sites. Prolines 177, 222, and 316 are found at the N-terminal position of helices. A residue in this position is partly outside the helix and partly inside, i.e., the proline ring is not part of the helix but the carbonyl oxygen forms a α-helical hydrogen bond with the amide nitrogen four residues upstream. Single proline site-directed mutagenesis of CBAD revealed the following changes in the melting temperature: an increase of 0.5°C for the alanine-to-proline substitution at position 177, 10.8°C the leucine-to-proline substitution at position 316, and a decrease of 6.5°C for the histidine-to-proline substitution at position 222.[5] Position 222 is on the surface of the protein. Replacing a histidine for a proline at this position leads to a loss of solvation energy, which would account for the observed destabilization. Pro316 is the first residue in a four-residue long stretch of 3_{10} helical conformation, linking an α-helix to a β-strand. Replacing a bulky leucine for the smaller proline may relieve some overcrowding in this region containing Met147, Met152, and Val309. Therefore, one may conclude that the introduction of a proline at the beginning of a helix does not stabilize the protein unless it is a special position. Proline residues rarely occur in helix interiors where they would cause a kink in the helix.[33,34] Side-chain hydrogen bonding to the backbone in a helix is rarely to the amide nitrogen of the first residue but frequently to the second or third residue. The lack of a hydrogen atom on the amide nitrogen of a proline residue in position 1 of a helix seems therefore not to be a hindrance. Pro24 is located at position 2 in a β-turn (Fig. 6), a position that is frequently populated by prolines. The serine-to-proline mutant of CBAD at this position has a 3.9°C higher melting temperature.[5] The introduction of a kink at this position is facilitated by the fixed Φ angle in proline.

Alanine and serine residues

As shown in Table III, TBAD has eight more alanine residues than CBAD. Three of the replacements to alanine occur in α-helical regions. An alanine increases the stability of a helix because of the unrestricted rotational freedom of the methyl side chain.[35,36] Introduction of alanine residues at other positions involves replacement of polar side chains such as serine and arginine in buried regions (Ser35, Ser168, Ser246, Ser250, Ser263, and Arg238), as well as replacement of a valine on the surface of the protein at position 32 (Fig. 5). These replacements also provide entropic stabilization. Some serine residues in the core of the protein are replaced with the more apolar threonine residues without loss of hydrogen bond capacity (Ser92, Ser154, and Ser169). Two exposed serine residues

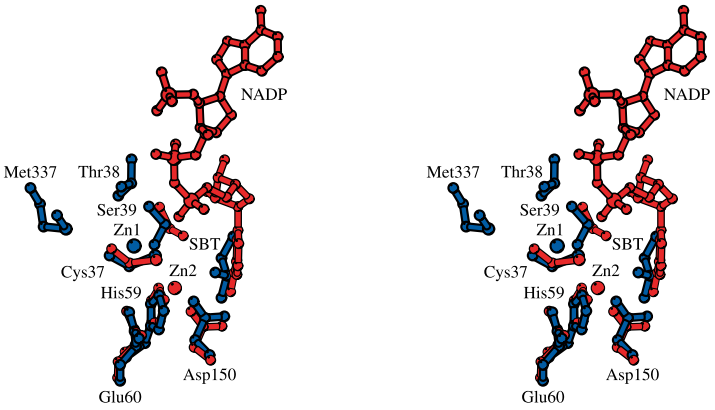

Fig. 4. Superposition of the active sites of two binary complex structures of TBAD: the NADP-complex (PDB code 1YKF; red)[22] and the substrate complex with sec-butanol (this study; blue). The only significant difference is in the location of the zinc ion. Positions Zn1 and Zn2 are occupied in the substrate and NDAP complexes, respectively. For catalysis to take place the zinc ion has to be in position Zn2.

```
CBAD      1 MKGFAMLGINKLGWIEKERPVAGSYDAIVRPLAVSPCTSDIHTVFEGALGDRKNMILGHE
TBAD      1        S G V      K AP PF         A             I E H
                   e e b      e ee ee         b             e e e

CBAD     61 AVGEVVEVGSEVKDFKPGDRVIVPCTTPDWRSLEVQAGFQQHSNGMLAGWKFSNFKDGVF
TBAD     61                        V  AI     TS   R YH   G          V
                                   b  bb     be   e ee   e          e

CBAD    121 GEYFHVNDADMNLAILPKDMPLENAVMITDMMTSGFHGAELADIQMGSSVVVIGIGAVGL
TBAD    121   F           H   EI   A    P    T          EL AT A L   P
              e           e   eb   b    b    b          eb bb b b   b

CBAD    181 MGIAGAKLRGAGRIIGVGSRPICVEAAKFYGATDILNYKNGHIVDQVMKLTNGEGVDRVI
TBAD    181  AV            A     V  D   Y      V   D P ES I N  E K   AA
             bb            b     e  e   e      b   e e ee b e  e e   bb

CBAD    241 MAGGGSETLSQAVSMVKPGGIISNINYHGSGDALLIPRVEWGCGMAHKTIKGGLCPGGRL
TBAD    241 I   NADIMAT  KI     T A V  F E EV PV  L
            b   ebebbbb  eb     b b b  b e ee eb  e

CBAD    301 RAEMLRDMVVYNRVDLSKLVTHVYHGFDHIEEALLLMKDKPKDLIKAVVIL-
TBAD    301  M R I L F K   P       FR   N  K FM           P    A
             b e b b e e   b       ee   e  e be           b    e
```

Fig. 5. Sequence alignment of TBAD with CBAD. The rows marked "TBAD" contain only changes from the CBAD sequence. Residues marked "b" are buried (accessibility < 25 Å^2), residues marked "e" are exposed in the tetramer (accessibility > 25 Å^2).

TABLE III. Differences in the Amino Acid Composition[†]

	TBAD	CBAD	TBAD-CBAD
Ala	35 (.099)	27 (.077)	**8 (0.022)**
Arg	14 (.040)	14 (.040)	0
Asn	10 (.028)	13 (.037)	−3 (−0.009)
Asp	20 (.057)	21 (0.060)	−1 (−0.003)
Cys	4 (0.011)	5 (0.014)	−1 (−0.003)
Gln	3 (0.009)	6 (0.017)	−3 (−0.008)
Glu	21 (0.060)	18 (0.051)	3 (0.009)
Gly	43 (0.122)	45 (0.128)	−2 (−0.006)
His	10 (0.028)	11 (0.03.1)	−1 (−0.003)
Ile	26 (0.074)	26 (0.074)	0
Leu	23 (0.065)	28 (0.080)	−5 (−0.015)
Lys	24 (0.068)	21 (0.060)	3 (0.008)
Met	15 (0.043)	18 (0.051)	−3 (−0.008)
Phe	14 (0.040)	11 (0.031)	3 (0.009)
Pro	21 (0.060)	13 (0.037)	**8 (0.023)**
Ser	9 (0.026)	17 (0.048)	**−8 (−0.022)**
Thr	13 (0.037)	11 (0.031)	2 (0.008)
Trp	4 (0.011)	4 (0.011)	0
Tyr	6 (0.017)	7 (0.020)	−1 (−0.003)
Val	37 (0.105)	35 (0.100)	2 (0.005)

[†]The amino acid composition is based on SWISSPROT entries P14941 and P25984 for TBAD and CBAD, respectively. Values in parentheses represent mole fractions. Differences >2% are shown in bold.

at positions 254 and 270 are replaced with the more hydrophilic lysine and glutamic acid, respectively.

Surface characteristics and subunit interfaces

TBAD is more compact than its mesophilic counterpart. The accessible surface area in TBAD is 7.7% smaller than in CBAD (Table IV). Tighter packing has also been observed in the case of a thermophilic 3-isopropyl malate dehydrogenase.[9] Moreover, the area buried on oligomerization is significantly larger in TBAD; there are more extensive contacts between the subunits in the thermophilic enzyme. These intermolecular contacts include salt bridges, hydrogen bonds, and hydrophobic interactions.

Salt bridges, hydrogen bonds, and van der Waals contacts

The number of van der Waals contacts, hydrogen bonds, and salt bridges are larger in TBAD than in CBAD, particularly those at subunit interfaces (Table V). Although the TBAD monomer has fewer hydrogen bonds than the CBAD monomer because of the loss of buried polar residues, this loss is more than offset by the creation of extra salt bridges and van der Waals contacts both

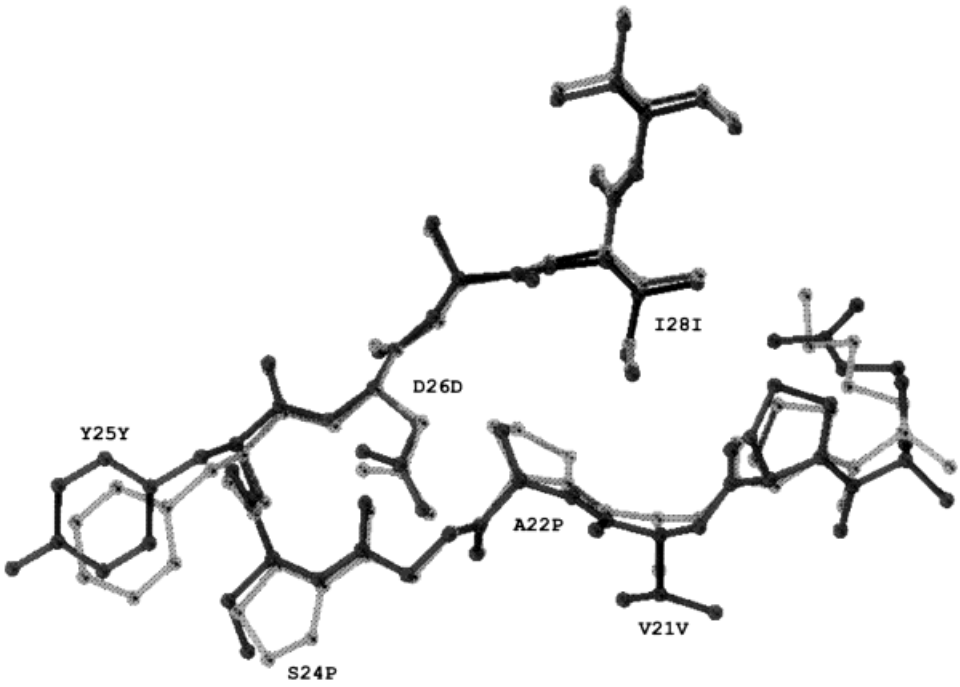

Fig. 6. Ball-and-stick representation of a region containing three proline residues in the thermophilic enzyme at positions 20, 22, and 24. TBAD and CBAD[22] are shown in dark and light gray, respectively. Subunits A of both TBAD and CBAD are used for the superposition. Pro20 is conserved in both enzymes, whereas residues 22 and 24 in CBAD are alanine and serine, respectively. Introduction of a proline in position 22 increases the number of van der Waals contacts with nearby Ile28. Residue 24 is located at position two of a β-turn, a position that is frequently populated by prolines. There is little or no entropy loss during folding for a proline at this position.

TABLE IV. Surface Area of the Tetramer

	TBAD	CBAD[22]	Difference (%)
Total accessible surface area (Å^2)	44996	48775	−7.7
Area buried in tetramer interfaces (Å^2)	14684	14397	2.0
Fraction of buried area	0.326	0.295	3.1
Percentage of exposed hydrophilic residues	71.4	69.2	2.2
Percentage of exposed hydrophobic residues	28.6	30.8	−2.2
Percentage of buried hydrophilic residues	19.4	24.5	−5.1
Percentage of buried hydrophobic residues	80.6	75.5	5.1

TABLE V. Interactions in CBAD and TBAD†

	Intermolecular contacts			Intramolecular contacts		
	H-bonds	Salt bridges	van der Waals	H-bonds	Salt bridges	van der Waals
CBAD[22]	167	8	774	538	16	4383
TBAD	218	13	862	476	18	4689
Difference	51	5	88	−62	2	306

†The contacts were calculated with program CONTACT within the CCP4 crystallographic suite of programs.[26] Threshold values: salt bridges 4.0 Å, H-bonds 3.4 Å, and van der Waals contacts 4.0 Å.

within and between the monomers. Furthermore, additional intermolecular hydrogen bonds are present in TBAD. There are 218 intermolecular hydrogen bonds in TBAD versus 167 in CBAD (Table V). Salt bridges are important stabilizing forces in proteins if they occur in regions inaccessible to the solvent.[37–39] The strength of a salt bridge depends on the environment (the local dielectric constant) and the bond distance. There are 31 salt bridges in TBAD versus 24 in CBAD (Table V). Some of the salt bridges in TBAD are involved in extensive networks. The energy gain in a network is larger than the sum of the pairwise interactions between the constituent charges.[40] An example of a sequestered network of six salt bridges—some intramolecular and some intermolecular— is depicted in Figure 7. Only two of these salt bridges exist in CBAD.

Extended networks of ion-pair interactions have also been observed in thermophilic glutamate dehydrogenases.[41] In this family the thermal stability correlates well with the charge density on the protein.[42] The most thermostable member of this family, the *Pyrococcus furiosis* enzyme with a melting temperature of 113°C, has a vast number of salt bridges, including ion-pair networks that involve up to 18 residues.[43] The existence of a strong correlation between the number of ion pairs and enzyme thermostability has also been corroborated by engineering extra salt bridges into the less thermostable glutamate dehydrogenase from *Thermatoga maritima*.[44,45]

Accurate accounting of salt bridges and hydrogen bonds is hampered by the limited resolution of the TBAD crystal structure. Thus, a few genuine interactions may have been missed, and the validity of others may be questioned. However, significant overall differences in the number of

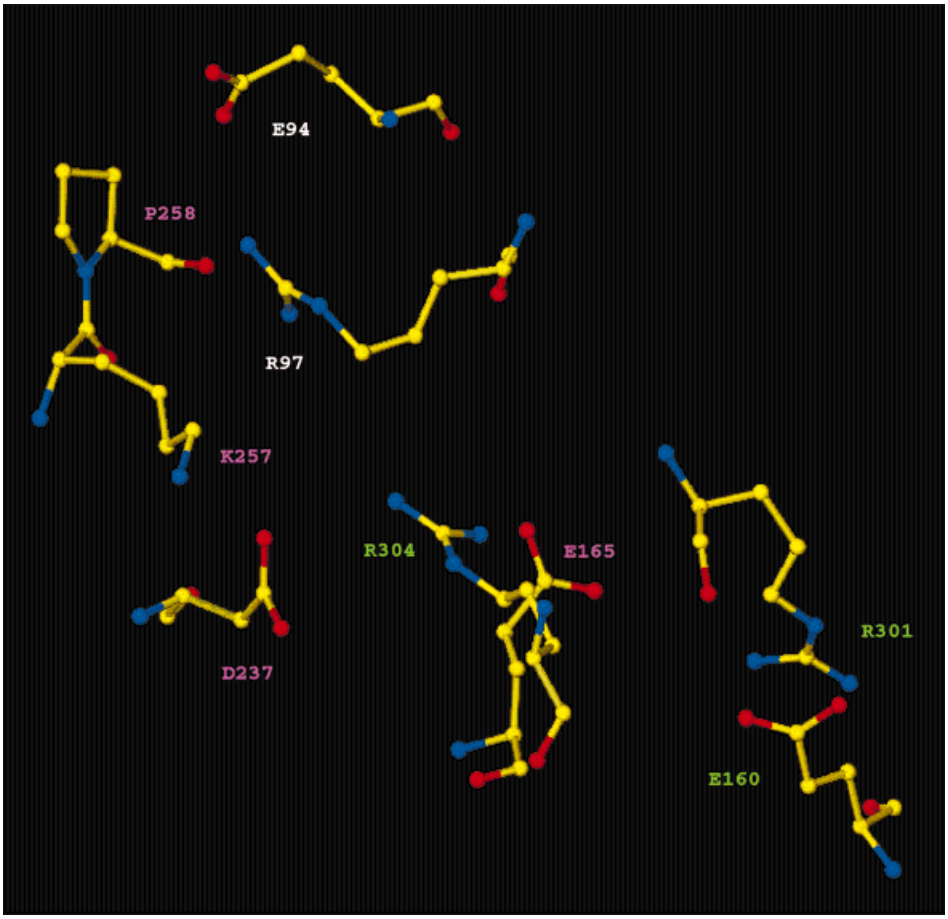

Fig. 7. Ball-and-stick representation of the extensive network of salt bridges involving three subunits in TBAD. Residues are labeled red, green, and white, according to the subunit to which they belong. The guanido group of R97 from the "white" subunit forms an intramolecular salt bridge with the carboxyl group of E94 (distances range from 3.83 to 4.01 Å for four molecules of the tetramer). Across the subunit interface from the guanido group of R97 a carboxylate oxygen of D237 from the "red" subunit is located at an average distance of 5.3 Å (range from 4.83 to 5.70 Å for four molecules of the tetramer), normally considered too long for a salt bridge. However, in a milieu of low-dielectric constant these two opposite charges might still exert some attractive force on each other. The same oxygen atom of residue D237 also interacts with the ϵ-amino group of K257 from the same subunit and with the guanido group of R304 from the "green" subunit at average distances of 3.1 and 3.4 Å, respectively (range from 2.63 to 2.83 Å and from 3.12 to 3.55 Å, respectively, for the tetramer). The other nitrogen of the guanido group of R304 from the "green" subunit in turn interacts with the carboxylate oxygen of E165 from the "red" subunit at an average distance of 3.9 Å (range from 3.68 to 4.05 Å). The other carboxylate oxygen atom of E165 from the "red" subunit is about 7.3 Å (range from 6.76 to 7.66 Å for the tetramer) from the guanido group of R301 of the "green" subunit, too far to be considered a salt bridge. Finally, the guanido group of this R301 interacts with the carboxylate of E160 from the same "green" subunit at an average interatomic distance of 3.0 Å (range from 2.78 to 3.14 Å). There are five salt bridges depicted in this figure—six if you count the long intermolecular bridge between R97 and D237. Of these, only two are present in mesophilic CBAD. A network of salt bridges provides more stabilization than the sum of individual pairwise interactions.

interactions between the thermophilic and the mesophilic enzyme clearly suggest that these differences enhance the thermal stability of TBAD.

ACKNOWLEDGMENTS

The authors thank Yigal Burstein, Oren Bogin, and Moshe Peretz from the Weizmann Institute of Science for providing us with TBAD as well as for continuous encouragement. This work was supported by grant DAAH04–95–1-0396 from the Army Research Office. SS is supported by augmentation award for science and engineering research training DAAG55–97–1-0214 from the Army Research Office. We thank the staff of the Cornell High Energy Synchrotron Radiation Source for assistance during data collection and David Shapiro and Nancy Milam for critically reading this manuscript.

REFERENCES

1. Lamed RJ, Zeikus JG. Novel NADP-linked alcohol—aldehyde/ketone oxidoreductase in thermophilic ethanologenic bacteria. Biochem J 1981;195:183–190.
2. Keinan E, Seth KK, Lamed R, Ghirlando R, Singh SP. Thermostable enzymes in organic Synthesis, 4. TBADH-catalyzed preparation of bifunctional chirons, total synthesis of S-(+)-Z-tetradeca-5-en-13-olide. Biocatalysis 1990;3:57–71.
3. Peretz M, Burstein Y. Amino acid sequence of alcohol dehydrogenase from the thermophilic bacterium *Thermoanaerobacter brockii*. Biochemistry 1989;28:6549–6555.

4. Bogin O, Peretz M, Burstein Y. *Thermoanaerobacter brockii* alcohol dehydrogenase: characterization of the active site metal and its ligand amino acids. Protein Sci 1997;6:450–458.
5. Bogin O, Peretz M, Hacham Y, et al. Enhanced thermal stability of *Clostridium beijerinckii* alcohol dehydrogenase after strategic substitution of amino acid residues with prolines from the homologous thermophilic *Thermoanaerobacter brockii* alcohol dehydrogenase. Protein Sci 1998;7:1156–1163.
6. Goodenough PW. A review of protein engineering for the food industry. Mol Biotechnol 1995;4:151–166.
7. Lasa I, Berenguer J. Thermophilic enzymes and their biotechnological potential. Microbiologia 1993;9:77–89.
8. Coolbear T, Daniel RM, Morgan HW. The enzymes from extreme thermophiles: bacterial sources, thermostabilities and industrial relevance. Adv Biochem Eng Biotechnol 1992;45:57–98.
9. Wallon G, Kryger G, Lovett S, Oshima T, Ringe D, Petsko GA. Crystal structures of *Escherichia coli* and *Salmonella typhimurium* 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart from *Thermus thermophilus*. J Mol Biol 1997;266:1016–1031.
10. Matthews BW Studies on protein stability with T4 lysozyme. Adv Protein Chem 1995;46:249–278.
11. Rees DC, Adams MW. Hyperthermophiles: taking the heat and loving it. Structure 1995;3:251–254.
12. Spassov VZ, Karshikoff AD, Ladenstein R. The optimization of protein-solvent interactions: thermostability and the role of hydrophobic and electrostatic interactions. Protein Sci 1995;4:1516–1527.
13. Day MW, Hsu BT, Joshua-Tor L, et al. X-ray crystal structures of the oxidized and reduced forms of the rubredoxin from the marine hyperthermophilic archaebacterium *Pyrococcus furiosus*. Protein Sci 1992;1:1494–1507.
14. Russell RJ, Taylor GL. Engineering thermostability: lessons from thermophilic proteins. Curr Opin Biotechnol 1995;6:370–374.
15. Ramaswamy S, Ahmad ME, Danielsson O, Jornvall H, Eklund H. Crystal structure of cod liver class I alcohol dehydrogenase: substrate pocket and structurally variable segments. Protein Sci 1996;5:663–671.
16. Eklund H, Ramaswamy S, Plapp BV, et al. Crystallographic investigations of alcohol dehydrogenases. EXS 1994;71:269–277.
17. Davis G J, Bosron W F, Stone CL, Owusu-Dekyi K, Hurley TD. X-ray structure of human β3β3 alcohol dehydrogenase: the contribution of ionic interactions to coenzyme binding. J Biol Chem 1996;271:17057–17061.
18. Xie P, Parsons SH, Speckhard DC, Bosron WF, Hurley TD. X-ray structure of human class IV σσ alcohol dehydrogenase: structural basis for substrate specificity. J Biol Chem 1997;272:18558–18563.
19. Hurley TD, Bosron WF, Stone CL, Amzel LM. Structures of three human beta alcohol dehydrogenase variants: correlations with their functional differences. J Biol Chem 1994;23:415–429.
20. Eklund H, Plapp BV, Samama JP, Branden CI. Binding of substrate in a ternary complex of horse liver alcohol dehydrogenase. J Biol Chem 1982;257:14349–14358.
21. Korkhin Y, Frolow F, Bogin O, Peretz M, Kalb AJ, Burstein Y. Crystalline alcohol dehydrogenases from the mesophilic bacterium *Clostridium beijerinckii* and the thermophilic bacterium *Thermoanaerobacter brockii*: preparation, characterization and molecular symmetry. Acta Crystallogr D Biol Crystallogr 1996;52:882–886.
22. Korkhin Y, Kalb (Gilboa) AJ, Peretz M, Bogin O, Burstein Y, Frolow F. NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of *Clostridium beijerinckii* and *Thermoanaerobacter brockii*. J Mol Biol 1998;278:967–981.
23. Matthews BW. Solvent content of protein crystals. J Mol Biol 1968;33:491–497.
24. Otwinowski Z. In: Sawyer L, Isaacs N, Bailey S, editors. Proceedings of the CCP4 Study Weekend: Data Collection and Processing. Warrington, UK: SERC Daresbury Laboratory; 1993. P 56–62.
25. Navaza J. AMoRe: an automated package for molecular replacement. Acta Crystallogr A 1994;50:157–163.
26. Collaborative Computational Project, Number 4. Acta Crystallogr 1994;D50:760–763.
27. Brünger AT. X-PLOR, Version 3.1 manual: a system for X-ray crystallography and NMR. New Haven, CT: Yale University Press; 1993.
28. Brünger AT, Adams PD, Clore GM, et al. Crystallography & NMR system: a new software system for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
29. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47:110–119.
30. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
31. Collona-Cesari F, Perahia D, Karplus M, Eklund H, Branden CI, Tapia O. Interdomain motion in liver alcohol dehydrogenase. Structural and energetic analysis of the hinge bending mode. J Biol Chem 1986;261:15273–15280.
32. Matthews BW, Nicholson H, Becktel WJ. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. Proc Natl Acad Sci USA 1987;84:6663–6667.
33. Richardson JS, Richardson DC. Amino acid preferences for specific locations at the ends of alpha helices. Science 1988;240:1648–1652.
34. Aurora R, Rose, GD. Helix capping. Protein Sci 1998;7:21–38.
35. Menendez-Arias L, Argos P. Engineering protein thermal stability: sequence statistics point to residue substitutions in alpha-helices. J Mol Biol 1989;206:397–406.
36. Kelly CA, Nishiyama M, Ohnishi Y, Beppu T, Birktoft JJ Determinants of protein thermostability observed in the 1.9 Å crystal structure of malate dehydrogenase from the thermophilic bacterium *Thermus flavus*. Biochemistry 1993;32:3913–3922.
37. Nakamura H. Roles of electrostatic interaction in proteins. Q Rev Biophys 1996;29:1–90.
38. Shortle D. Mutational studies of protein structures and their stabilities. Q Rev Biophys 1992;25:205–250.
39. Barlow DJ, Thornton JM. Ion-pairs in proteins. J Mol Biol 1983;168:867–885.
40. Horovitz A, Serrano L, Avron B, Bycroft M, Fersht, AR. Strength and co-operativity of contributions of surface salt bridges to protein stability. J Mol Biol 1990;216:1031–1044.
41. Knapp S, de Vos WM, Rice D, Ladenstein R. Crystal structure of glutamate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima* at 3.0 Å resolution. J Mol Biol 1997;267:916–932.
42. Yip KS, Britton KL, Stillman TJ, et al. Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamate dehydrogenase family. Eur J Biochem 1998;255:336–346.
43. Yip KS, Stillman TJ, Britton KL, et al. The structure of *Pyrococcus furiosis* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure 1995;3:1147–1158.
44. Lebbink JH, Knapp S, van der Oost J, Rice D, Ladenstein R, de Vos WM. Engineering activity and stability of *Thermotoga maritima* glutamate dehydrogenase. I. Introduction of a six-residue ion-pair network in the hinge region. J Mol Biol 1998;280:287–296.
45. Lebbink JH, Knapp S, van der Oost J, Rice D, Ladenstein R, de Vos WM. Engineering activity and stability of *Thermotoga maritima* glutamate dehydrogenase. II. Construction of a 16-residue ion-pair network at the subunit interface. J Mol Biol 1999;289:357–369.
46. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24:946–950.
47. Merritt EA, Murphy ME. Raster3D Version 2.0. A program for photorealistic molecular graphics. Acta Crystallogr D Biol Crystallogr 1994;50:869–873.
48. Esnouf R. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. Acta Crystallogr D Biol Crystallogr 1999;55:938–940.