

## Relationship between thermal stability and 3-D structure in a homology model of 3-isopropylmalate dehydrogenase from *Escherichia coli*

Csaba Magyar, András Szilágyi and Péter Závodszy<sup>1</sup>

Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Pf. 7, H-1518 Budapest, Hungary

<sup>1</sup>To whom correspondence should be addressed

To reveal the structural basis of the increased thermal stability of 3-isopropylmalate dehydrogenase (IPMDH) from *Thermus thermophilus*, an extreme thermophile, the homology-based structural model of one mesophilic (*Escherichia coli*) counterpart, was constructed. Both IPMDHs are homodimeric proteins. We built a model of one subunit using the 3-D structures of the *Th.thermophilus* IPMDH and the homologous *E.coli* isocitrate dehydrogenase. Energy minimization and molecular dynamics simulated annealing were performed on the dimer, including a surrounding solvation shell. No serious errors were detected in the refined model using the 3-D profile method. The resulting structure was scrutinized and compared with the structure of the *Th.thermophilus* IPMDH. Significant differences were found in the non-specific interactions including the hydrophobic effect. The model predicts a higher number of ion pairs in the *Th.thermophilus* than in the *E.coli* enzyme. An increase was observed in the stabilities of  $\alpha$ -helical regions in the thermophilic protein. The preliminary X-ray coordinates of the *E.coli* IPMDH were received after the completion of this work, allowing an assessment of the model in terms of the X-ray structure. The comparison proved that most of the structural features underlying the stability differences between the two enzymes were predicted correctly.

**Keywords:** homology modeling/3-isopropylmalate dehydrogenase/protein structure prediction/thermophiles/thermostability

### Introduction

A large body of experimental data from related proteins with various thermal stabilities provides convincing evidence that these enzymes are similar in terms of their basic topology, activity and mechanism of action (Amelunxen and Murdock, 1978; Jaenicke, 1981, 1991; Daniel, 1986). The conformational stability of globular proteins can be defined by the free energy difference between the folded and unfolded states. Adaptation to different thermal optima can be accomplished by slight changes in  $\Delta G_{N \rightarrow D}$  (Jaenicke and Závodszy, 1990; Jaenicke, 1991). This relatively small change in  $\Delta G$  can be realized in a large number of ways, because protein stability is a reflection of the balance between an almost infinite number of stabilizing and destabilizing forces (Pace, 1990). It is clear that the relationship between conformational stability and protein structure is complex. A systematic comparison of the structures and functional properties of enzyme pairs of different thermal stabilities is the appropriate way to approach this problem.

3-Isopropylmalate dehydrogenase (IPMDH; EC 1.1.1.85) is a bifunctional enzyme involved in the leucine biosynthesis pathway. It catalyzes the dehydrogenation and decarboxylation of the 3-isopropylmalate substrate, yielding 2-ketoisocaproate and carbon dioxide, with a concomitant reduction of the cofactor NAD<sup>+</sup>. IPMDH and isocitrate dehydrogenase (EC 1.1.1.42) belong to a unique class of metal-dependent decarboxylating dehydrogenases, sharing a common protein fold and having a similar catalytic mechanism, although they differ in their specificities for substrates and cofactors (Hurley *et al.*, 1991; Imada *et al.*, 1991).

Many IPMDH genes from different organisms have been cloned and expressed to characterize the biochemical and biophysical properties of the enzyme. In this study, special attention was devoted to the IPMDH from *Thermus thermophilus*, an extreme thermophile, because this enzyme is much more thermostable than its counterparts isolated from mesophilic sources, despite the high sequence homologies. The X-ray structure of the *Th.thermophilus* IPMDH (TtIPMDH) has been solved in the absence and presence of cofactor (Imada *et al.*, 1991; Hurley and Dean, 1994). This enzyme is a functional dimer composed of two identical subunits each with 345 amino acid residues. The polypeptide chain of a subunit is folded into two domains with similar folding topologies, based on parallel  $\alpha/\beta$  motifs. Site-directed mutagenesis studies were carried out on this enzyme to determine the residues involved in catalysis (Miyazaki and Oshima, 1993) and coenzyme specificity (Miyazaki and Oshima, 1994; Yaoi *et al.*, 1995). Zhang and Koshland (1995) modeled the binding of the substrate and the coenzyme to IPMDH. Kirino *et al.* (1994) mutated two hydrophobic residues at the subunit-subunit interface of TtIPMDH to the corresponding residues found in IPMDH from *Escherichia coli* (L246E, V249M), and examined the thermostability of the mutant. The results demonstrate the contribution of hydrophobic interactions at the subunit-subunit interface to the thermostability of TtIPMDH. The structure of this mutant was later solved by X-ray crystallography (Moriyama *et al.*, 1995). Chimeric IPMDHs from TtIPMDH and IPMDH from the mesophile *Bacillus subtilis* were created and analyzed (Onodera *et al.*, 1994; Numata *et al.*, 1995); the conclusion from these studies was that amino acid residues contributing to the thermal stability distribute themselves approximately evenly along the polypeptide chain of IPMDH.

Investigations of the structural background of the thermostability of TtIPMDH are limited by the fact that no 3-D structure of a mesophilic counterpart is known; thus a comparative structural analysis is not possible. To tackle this problem, we built a homology model of the structure of a mesophilic IPMDH from *E.coli* (EcIPMDH), and then analyzed the structures of TtIPMDH and the EcIPMDH model comparatively, in terms of the stabilizing interactions. We were particularly interested in the electrostatic interactions and the hydrophobic effects, because in earlier comparative studies on mesophilic and thermophilic glyceraldehyde-3-phosphate

dehydrogenase we had observed the contribution of these factors to the difference in thermal stability (Szilágyi and Závodszy, 1995). After having completed our modeling studies, we learnt that the structure of EcIPMDH has been solved by X-ray crystallography by Gerlind Wallon in Gregory A.Petsko's laboratory at Brandeis University, MA (Wallon, G., Kryer, G., Lovett, S., Oshima, T., Ringe, D. and Petsko, G.A., 1996, submitted to *Structure*). These authors provided us with the coordinates prior to publication, and consequently we have been able to evaluate the reliability of our model and check our conclusions, demonstrating both the potential and the limitations of homology modeling. We declare here that the modeling procedure was blind, and no corrections were introduced into the model after being given the X-ray coordinates.

## Materials and methods

### Construction of the model

We used the standard homology modeling procedure described by Greer (1990). The molecular graphics program InsightII and its Homology module (Biosym Inc.) were used on a Silicon Graphics Personal IRIS 4D35 workstation.

The structures of three reference proteins were used for the modeling: TtIPMDH [Imada *et al.*, 1991; Protein Data Bank (PDB) entry 1IPD], isocitrate dehydrogenase from *E.coli* (Hurley *et al.*, 1991; PDB entry 1ICD) and a chimeric protein consisting of the N-terminal part of *B.subtilis* IPMDH and the C-terminal part of *Th.thermophilus* IPMDH, denoted 4M6T (Onodera *et al.*, 1994). The atomic coordinates of these proteins were retrieved from the PDB (Bernstein *et al.*, 1977), except for the chimeric enzyme, whose coordinates were obtained from the authors.

### Finding the structurally conserved regions (SCRs) and the structurally variable regions (SVRs)

To identify the SCRs of the reference proteins, we used the fragments defined in the 1IPD entry of the FSSP database (Holm *et al.*, 1992), created by the SUPPOS algorithm (Vriend and Sander, 1991), as a starting point. Two reference proteins, TtIPMDH and the *E.coli* isocitrate dehydrogenase, were superimposed by a least-squares superposition of the C $\alpha$ s in these conserved fragments. The SCRs were derived by adjusting the length of each fragment so that the r.m.s. deviation between the two structures was minimized and each SCR contained only one secondary structural element. Surface loops were classified as SVRs.

The sequence of EcIPMDH (Kirino *et al.*, 1994) was aligned to the other sequences by the Needleman–Wunsch algorithm (Needleman and Wunsch, 1970), using the Dayhoff PAM250 matrix (Barker and Dayhoff, 1972).

### Construction of a rough model

In building the model of the EcIPMDH subunit, the SCRs were constructed first. The main-chain atom positions in each SCR were copied from the corresponding region of the TtIPMDH. The SVRs were taken from the reference protein with the highest sequence similarity in the corresponding region. In the case of two loop regions, Glu274–Gly278 and Ala129–Gly131, no suitable fragment could be found among the reference proteins. In the first case, we applied a loop search procedure (Jones and Thirup, 1986) to a subset of the PDB which contained all available structures of at least 2.0 Å resolution; a suitable fragment (Gly78–Asn82) was found in the protein rhizopuspep-

sin from *Rhizopus chinensis* (Suguna *et al.*, 1987; PDB entry 2APR). In the case of the Ala129–Gly131 loop, a loop search procedure gave no satisfactory results, so we applied the random tweak method (Shenkin *et al.*, 1987) to find a suitable conformation. During model building, we made the necessary side-chain replacements and adjusted the conformation of each side chain using the  $\chi$  torsion angles of the corresponding residues in the reference proteins, where possible.

### Refinement

The rough model was refined by forcefield-based methods, using the Discover molecular dynamics software (Biosym Inc.) with the consistent valence forcefield (CVFF). We used the double cut-off method, as implemented in the Discover program; the inner and outer cut-off distances were 12 and 21 Å, respectively. Energy minimizations were carried out by performing 250 steps by the steepest descents algorithm, followed by a minimization using the conjugate gradients algorithm until the maximum derivative became  $<0.5$  kcal/Å/mol.

The distorted peptide bonds at the boundaries of SCRs and SVRs were corrected by energy minimization using omega-forcing with a force constant of 150 kcal/rad/mol; all atoms were fixed during this minimization except two residues on both sides of each splice site.

In all cases where a side chain had been replaced by a longer one during model building, we corrected the conformation of the new side chain by performing a systematic search using the Ponder and Richards rotamer library (Ponder and Richards, 1987), accepting the conformations with the lowest energy values.

Before further refinement, we assembled the dimeric enzyme form from two monomers using the observed symmetry of the TtIPMDH dimer. A 7 Å-thick water shell was generated to solvate the dimer. During subsequent minimizations and molecular dynamics calculations, the oxygen atoms of waters in the inner 4 Å-thick shell were tethered with a force constant of 50 kcal/Å/mol, while oxygens in the outer shell were fixed; this prevented water molecules from 'boiling off'. To speed up calculations, we used the symmetry of the dimer by restricting calculations to one subunit and a 10 Å-thick contact region from the other subunit; the rest of the second subunit was held fixed. The main-chain atoms in the SCRs were tethered by a force constant of 120 kcal/Å/mol.

After minimizing the dimer, we performed a molecular dynamics simulated annealing procedure on the whole system. The parameters for the simulated annealing protocol were determined as described previously (Szilágyi and Závodszy, 1995). The simulation was started at 1000 K and the temperature was decreased by 5% in every step, until 300 K was reached. At each temperature, the system was simulated for 420 fs (180 fs using direct velocity scaling and 240 fs with coupling to a heat bath; Berendsen *et al.*, 1984). The total simulation time was 10.5 ps.

A final energy minimization was performed using a 0.2 kcal/Å/mol convergence criterion for the maximum derivative.

### Checking the quality of the model

The quality of the model was first assessed by comparing the number of overlapping atoms and the value of forcefield energy with those quantities obtained from the X-ray structure of the TtIPMDH. For a more sensitive assessment, we used the 3-D profile method (Bowie *et al.*, 1991), employing 'profile window plots' as described by Lüthy *et al.* (1992).

### Analysis of protein stability

Ion pairs were defined using a search procedure, as described previously (Szilágyi and Závodszy, 1995). First, we took all

residue pairs with opposite charges; these were potential ion pairs. Then, all the combinations of rotameric conformations from the Ponder and Richards rotamer library (Ponder and Richards, 1987) were set for both side chains in each residue pair. Combinations with overlapping atoms were rejected. Residue pairs were defined as ion pairs if their oppositely charged atoms were closer than 6.0 Å in at least 15% of the remaining, energetically allowable rotamer combinations. This way of defining ion pairs proved to be useful in eliminating possible artefacts resulting from forcefield-based structure refinement or from the modification of surface side-chain conformations by the crystal-line environment in a crystallographically determined structure (Szilágyi and Závodszy, 1995). Charge clusters were defined as networks of ion pairs, i.e. groups of charged side chains in electrostatically favorable arrangements. A charge cluster should not be localized to one spot in the protein; it can contain chains of adjacent ion pairs that may span a considerable distance in space. We classified charge clusters by the number of their constituting ion pairs.

The hydrogen bonds were defined using the HB2 module of the WHAT IF program (Vriend, 1990). This program finds the best hydrogen bond network in the protein, scoring the hydrogen bonds using a special forcefield developed for WHAT IF using a database of accurate small molecule structures as parameter source. This method is more sophisticated than using a simple 'yes or no' distance (and angle) criterion.

Buried surface areas were calculated by the method of Shrake and Rupley (1973), using the atomic radii found in Oobatake and Ooi (1993). The denatured state was modeled by setting all dihedral angles to the values specified in Oobatake and Ooi (1993).

To estimate the role of the non-specific interactions in protein stability, we used the method of Oobatake and Ooi (1993). This method allows the calculation of the unfolding enthalpy, entropy and free energy as a function of temperature from the 3-D structure of a protein. The method uses the difference between the accessible surface areas of seven atomic groups in the folded and unfolded states, and parameters derived from experimental data. The method gives a relatively good estimation of the thermodynamic quantities, although it does not take specific interactions into account.

The stability of  $\alpha$ -helices was evaluated by calculating the helical behavior of individual amino acid residues using the program AGADIR (Muñoz and Serrano, 1994, 1995a,b). This program calculates the helical content and the average helicity per residue for peptides from the sequence alone, using the experimental energetic contributions of various interactions that are important for the stabilization of  $\alpha$ -helices, including the intrinsic helical propensities of amino acids, side chain–side chain interactions, main chain–main chain hydrogen bonds, helix macrodipole effects and capping effects. We assumed a pH of 7.0 and a temperature of 300 K when calculating the average helicities per residue. The results of this calculation can be used to estimate the stability of helical regions in proteins in the absence of tertiary interactions.

## Results and discussion

### Model construction and quality of the model

We built a homology model for EcIPMDH by the conventional fragment assembly method, following the guidelines described by Greer (1990) (for details see Materials and methods). Three known 3-D structures were available as reference proteins



Fig. 1. 3-D alignment (spatial superposition) of subunits from two reference proteins (TiIPMDH and *E.coli* isocitrate dehydrogenase). Solid lines, SCRs; dashed lines, SVRs.

```

EcIPMDH:          MSKNYHIAVLPGDGIQPEVMTQALKVLDVA
TiIPMDH:          MKVAVLPGDGIQPEVTEARLKVLRRL
EcICD :          SKVVVPAQGGKITLQNGKLVNPNPIPIYIKGDGIGVDVTPANLKVVDAA

EcIPMDH:  RNRFAHRIIT---SHYDVGGAAIDNH---GQLPPTAVEGCEQADAVLFGS
TiIPMDH:  DEAEGLGLAY---EVFPFGAAIDAF---QEPFPEPTREGVEEAVALLGS
EcICD :  VERAYKGERKISWHETTYGEBEQVYGGDVMLEPARTLDLIREYVAIKGP

EcIPMDH:  VGGPKWEHLPPDQPPERGALLPLRKHFKLFSNLRPAKLYQGLEAFCLPRA
TiIPMDH:  VGGPKWDGLPRKISPETG---LLSLRKSQDLFANLRPAKLVVPLERLSPLE
EcICD :  EITPVGGGIR---EL---RVALRQLRDLTYICLRPVRYVYQGTSPVK---HP

EcIPMDH:  DIAANGFDILCVRELTCGGIYFGQPKRGESGQYKAFDTEVY-----
TiIPMDH:  ELAR---GNDVLIIVRELTCGGIYFGPRGMSAE---ANNTERY-----
EcICD :  ELT---DMVIFRNSEDIYAGISWKADSADA---EKVIKFLREMGVKK

EcIPMDH:  -----HRFEIERIARIAPESARKRRHKVTSIDKANVLQS-----
TiIPMDH:  -----SKFEVERKARVAFESARKRRHKVSVSDKANVLEV---
EcICD :  IRFFEHCGIGIKPCSEECTKRLVRAATRYAANDRADSVTLVHGKHKFT

EcIPMDH:  ---SILWREIVNEIATEYFDV-----ELAHMYIDNA---TM
TiIPMDH:  ---GEFWRKTVVEEYGRGYPDV-----ALEHQYVDAN---AM
EcICD :  EGAFKDWGYQLAREEFGGELIDGGPWLKVKNPNTGKIVIKDIADAFIQ

EcIPMDH:  QLIKDPSQFDVLLCSNLFQDILSDCEAMITGSHGHLPSASLNQQGQPOLYE
TiIPMDH:  HLVRSPARFDVVVTCRIFDGLSDLAQVLPQSLGLLPSASLGR---GTPVPE
EcICD :  QILLRPAEYDVIACMNMGDYISDALAQVGGIGIIPGANIGD---ECALPE

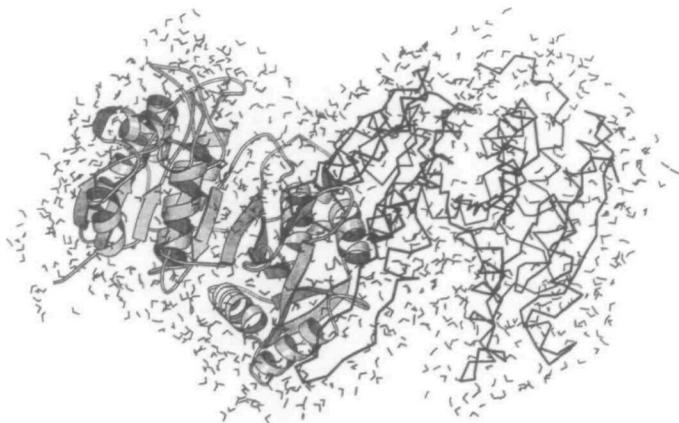
EcIPMDH:  FAGGSAPDIAGKNIANPIAQLISLALLLRYSLDADAACAIERAINRALE
TiIPMDH:  FVHGSAPEIAGKGIANPTAAILSRMMLHEHAPGLVEARLKVSDAVKALE
EcICD :  RTHGTAPKFAQQQKVMFGSITLGSABMLRHGHWTE---AALIVVHGEGATH

EcIPMDH:  EGI-RTGDLARGA-----AAVSTDEMGDIARYVAEGV
TiIPMDH:  ETP-PPDLGGSAG-----TEAFTATVLRHLA
EcICD :  AKTVYQFERLMDGAKLLKCFEQQDAIIRHM

```

Fig. 2. Sequence alignment of the reference proteins (TiIPMDH, isocitrate dehydrogenase from *E.coli* (EcICD)) and the sequence of EcIPMDH. The SCRs are highlighted in the alignment.

(TiIPMDH, isocitrate dehydrogenase from *E.coli* and a *B.subtilis/E.coli* chimeric IPMDH). A structural alignment of these reference proteins (Figure 1) was performed using the FSSP database as a starting point, and the alignment of the sequences was carried out according to this structural alignment. Finally, the EcIPMDH sequence was aligned to this sequence alignment by a simple Needleman–Wunsch algorithm (Figure 2). After identifying the structurally conserved regions and building a model for one subunit, a forcefield-based refinement was performed at the splice sites, and the side-chain conformations of the 'mutated' residues were corrected by a conformational search using the rotamer library of Ponder and Richards (1987). Because IPMDH is a homodimer, subunit–subunit interactions were handled with special care; subsequent refinements were performed on the assembled dimeric structure, including a solvation shell of 7 Å around the dimer. After global energy minimization, a molecular dynamics simulated annealing procedure was performed on the whole system, containing 21 868 atoms. Finally, the whole



**Fig. 3.** Homology model of the EcIPMDH dimer after the final energy minimization. One of the subunits is represented as a Richardson-type model (figure drawn using the program Molscript; Kraulis, 1991). Only water molecules in the inner 3.0 Å solvation shell are shown for clarity.

system was energy minimized. The final structure is presented in Figure 3.

The quality of our model structure was checked by several methods. The 3-D profile method revealed no serious errors in the structure (Figure 4), and the profile window plots of TtIPMDH and EcIPMDH mostly agreed well. There were no large atomic overlaps left in the structure, and the low final value of the forcefield energy ( $-80\,000$  kcal/mol) showed that the protein can exist in this conformation. However, the reliability of the model was low in two regions where long insertions (four residues) occurred, in the so-called arm-like region (Gln148–Phe162) and in the C-terminal part (Thr348–Tyr363). In the case of small insertions, the model is likely to have the correct conformation.

#### Analysis of the model

The model was compared with the X-ray structure of TtIPMDH and analyzed in terms of stabilizing interactions. Our goal was to pinpoint any characteristic differences between the structures of EcIPMDH and TtIPMDH that may underlie the differences in thermal stability.

**Electrostatic interactions.** The correlation between thermal stability and the presence of surface ion pairs was first suggested by Perutz and Raidt (1975), and has been demonstrated experimentally for a number of proteins (Walker *et al.*, 1980; Kelly *et al.*, 1993; Tomschy *et al.*, 1994; Chan *et al.*, 1995; Korndörfer *et al.*, 1995; Spassov *et al.*, 1995).

We counted the ion pairs in our model of EcIPMDH and the X-ray structure of TtIPMDH. We found that there were 50 ion pairs per subunit in the EcIPMDH model while there were 72 ion pairs in the thermophilic structure (Table I), the net difference being 22 to the advantage of the *Th.thermophilus* enzyme (including the His–Glu and His–Asp ion pairs). In all, 23 conserved ion pairs occurred in both IPMDHs, 27 ion pairs were found only in EcIPMDH and 49 ion pairs in TtIPMDH had no equivalents in EcIPMDH (Table I).

We also examined the distribution of ion pairs in the two structures. We defined charge clusters as networks of adjacent ion pairs (see Materials and methods) and classified these clusters by the number of their constituting ion pairs (Table II). As can be seen from the table, there was a clear tendency for larger charge clusters to form in TtIPMDH than in its mesophilic counterpart. In the thermophilic protein, the number of isolated ion pairs was considerably less, while the number

of intermediate-sized charge clusters was higher than in the *E.coli* enzyme. The presence of large charge clusters probably results in cooperative strengthening of the affected ion pairs, as implied by an earlier experimental study of an ion triad in barnase (Horovitz *et al.*, 1990). A similar phenomenon was observed on a hyperthermophilic glyceraldehyde-3-phosphate dehydrogenase (Korndörfer *et al.*, 1995; Szilágyi and Závodszy, 1995). These observations suggest that this mechanism, i.e. the formation of cooperative clusters of ion pairs, may be one general strategy of thermostabilization for proteins from thermophilic sources.

The presence of large charge clusters involves a higher number of ion triads in TtIPMDH than in EcIPMDH. Presumably, the most stabilizing ion triads are those that connect distant parts of the polypeptide chain. One such ion triad in TtIPMDH is Glu37–Lys2–Glu63, which binds the N-terminal end of the chain to an  $\alpha$ -helix and a  $\beta$ -sheet, and is itself part of a larger ion cluster. The N-terminal part of EcIPMDH is four residues longer than that of TtIPMDH and lacks this stabilizing ion pair, possibly providing an easy point of attack for thermal unfolding. Another example is the ion triad Glu29–Lys309–Glu305, which connects the N- and C-terminal parts of the molecule. Other ion triads in the thermophilic enzyme with no equivalents in the mesophilic structure include Glu30–Lys310–Glu306, Glu113–Arg124–Glu120 and Arg229–Glu212–His179. These interactions probably contribute to the enhanced thermostability of EcIPMDH and are good targets for mutagenesis studies.

Subunit–subunit interactions may be especially important for protein stability. We found 10 intersubunit ion pairs in TtIPMDH and only four in the EcIPMDH model structure (Lys195a–Asp251b, Glu159a–Arg169b and their symmetry-related pairs). Site-directed mutagenesis studies performed on residues at the subunit–subunit interface (Kirino *et al.*, 1994) have proved that the hydrophobic interaction in this region is important for stabilization of the dimeric structure; our results suggest that electrostatic interactions may also play a role. Site-directed mutagenesis studies are now in progress in our laboratory to test this hypothesis.

In several cases, the ionic interactions in TtIPMDH are apparently replaced by one or more hydrogen bonds in EcIPMDH. Ion triads in which such a situation occurs include Glu37–Lys2–Glu63, Glu17–Lys21–Glu24, Glu51–Glu55–Arg58, Glu299–Arg309–Glu312, His300–Glu299–Arg176 and His343–Glu321–Lys317. In some cases, electrostatic stabilization in TtIPMDH or EcIPMDH is replaced by the rigidifying effect of proline residues in the other protein, e.g. the Glu51–Arg58–Glu55 ion triad in TtIPMDH, or the Arg335–Asp338 ion pair in EcIPMDH. These observations suggest that there are several ways of stabilizing substructures in proteins, and different ways may be preferred in proteins adapted to different environmental conditions. Site-directed mutagenesis studies are planned to test whether electrostatic interactions have a stronger stabilizing effect than hydrogen bonds or prolines at the sites where these differences have been observed.

It should be noted that our definition of ion pairs, based on a conformational search procedure, is more sophisticated than the usual ways of defining ion pairs using a simple distance criterion. In addition, we chose a relatively large distance criterion (6.0 Å) for the search procedure. As a consequence, our definition allows for weak electrostatic interactions. If we use more strict criteria for the definition of ion pairs, the size of ion clusters will be smaller, and less difference is revealed

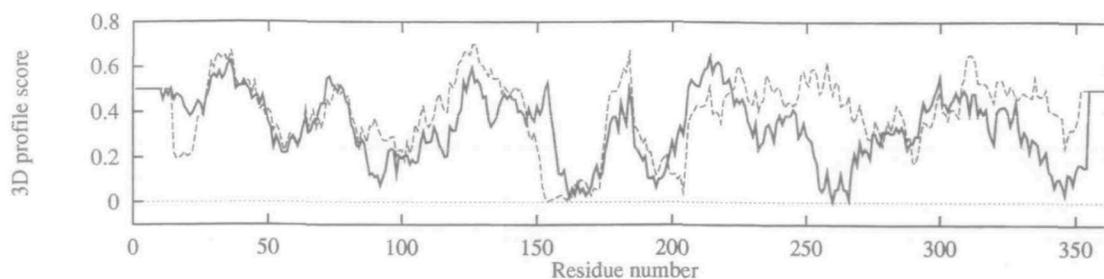


Fig. 4. 3-D profile window plots for the known structure of TtIPMDH (dotted line) and the model of EcIPMDH (solid, thicker line). A window size of 20 residues was applied. The plots are suitable to establish the quality of the structures.

Table I. Ion pairs in the EcIPMDH model and TtIPMDH

Ion pairs	
Occurring only in EcIPMDH (27 ion pairs)	E91-R92; D133-R310; D133-R324; E153-K160; E159-R152; E171-K103; E207-R169; E215-R177; E220-K189; D236-K235; D241-R186; D241-R187; E281-R109; D314-K185; D316-K185; D316-R310; E323-R310; E323-R324; E331-R328; D338-R335; E350-R341; H6-D69; D44-H53; E66-H101; E165-H168; H168-E171; H188-D218
Occurring in both IPMDHs (23 ion pairs)	D28-R31; D51-K80; D126-R124; D133-K112; D133-R186; D133-R187; E171-R174; E173-R169; E173-R177; E181-R184; E181-K185; D194-R206; E211-R169; D218-R184; D227-K195; D251-R109; D251-R138; E281-R99; E323-K112; E323-R186; E332-R328; D349-K25; D194-H223
Occurring only in TtIPMDH (49 ion pairs)	E17-K21; E17-R24; E28-R24; E30-K310; E37-K2; E51-R58; E55-R58; E55-K59; E62-R58; E62-K59; E63-K2; E63-K59; E87-R82; D98-K95; D98-R164; D98-R264; E113-R124; E120-R114; E120-R124; E121-R124; D127-R309; E133-R164; E142-R144; E142-R156; E161-R156; E163-R164; E171-R167; E193-R196; E193-K197; E200-R196; E200-K197; E201-K197; E201-R204; D208-K178; E212-R229; D245-R104; D245-R132; E299-K175; E299-R176; E299-R309; E306-K310; E312-R309; D313-R309; D313-K317; E321-R342; H179-E212; E200-H213; E299-H300; E321-H343

Table II. Number of ion pair clusters of various sizes in the EcIPMDH model and X-ray structure, and in TtIPMDH

Number of ion pairs in the cluster	EcIPMDH model	EcIPMDH X-ray	TtIPMDH
42	-	-	1
40	-	1	-
18	1	-	-
8	-	-	1
5	1	-	2
4	2	-	1
3	-	2	-
2	2	2	2
1	15	10	4
Total number of intrasubunit ion pairs	50	60	72

between TtIPMDH and EcIPMDH (data not shown). Therefore, the contribution of electrostatic interactions to the enhanced thermostability of TtIPMDH is difficult to assess.

**Hydrogen bonds.** There has been considerable uncertainty for many years as to whether hydrogen bonds contribute to protein stability. Recent experimental studies have shown that hydrogen bonds do have a stabilizing effect in many cases (Green *et al.*, 1992; Serrano *et al.*, 1992; Shirley *et al.*, 1992).

We compared the number of hydrogen bonds in the *Th. thermophilus* enzyme and the *E. coli* IPMDH model. Because the two polypeptide chains are of different length, the numbers of hydrogen bonds were normalized by the number of atoms in both structures. Only a very small difference was found in the normalized number of hydrogen bonds: the thermophilic structure contains slightly more hydrogen bonds per atom than its mesophilic counterpart. It is likely that the contribution of hydrogen bonds to the enhanced thermostability difference between EcIPMDH and TtIPMDH is insignificant.

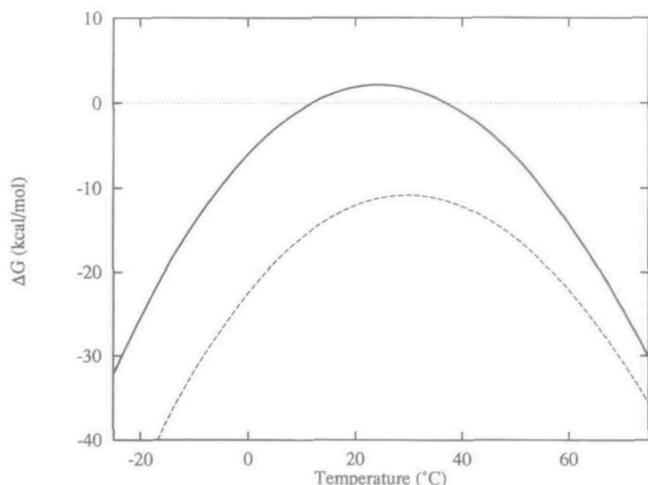
Table III. Buried polar and apolar surface areas ( $\text{\AA}^2$ ), and their ratios, in the EcIPMDH model and X-ray structure, and in TtIPMDH

	EcIPMDH model	EcIPMDH X-ray	TtIPMDH
Buried polar surface area	13 661	13 395	11 612
Buried apolar surface area	27 700	26 266	25 740
The ratio of the buried polar and apolar surface areas	0.49	0.51	0.45

**Hydrophobic interactions.** The importance of good packing and good hydrophobic bonding for the stabilization of proteins is a well-established fact (Chothia, 1974; Lee, 1993).

To estimate the contribution of hydrophobic interactions to the increased stability of the thermophilic IPMDH, we calculated the buried polar and apolar surface areas of both structures (Table III). The ratio of the buried polar and apolar surface areas is smaller for the *Th. thermophilus* structure than for the *E. coli* enzyme, which indicates an increased hydrophobicity of the protein core.

We also compared the accessible surface areas of different types of side chain in both the EcIPMDH model and TtIPMDH, looking for buried hydrophilic residues that are mutated to hydrophobic residues, as well as exposed hydrophobic residues that are mutated to hydrophilic residues in the other protein. In TtIPMDH, five buried hydrophilic residues are found that have hydrophobic equivalents in EcIPMDH, while in the latter, nine buried hydrophilic residues are found that have hydrophobic equivalents in TtIPMDH. However, all the extra buried hydrophilic residues in EcIPMDH contribute to hydrogen bonds. Site-directed mutagenesis should answer the question as to whether replacement of these residues in EcIPMDH increases its thermostability. There is almost no difference



**Fig. 5.** Free energy change upon unfolding as a function of temperature, calculated by the method of Oobatake and Ooi (1993) from the known 3-D structure of TtIPMDH (dashed line) and the model of EcIPMDH (solid line).

in the number of exposed hydrophobic residues where the equivalent in the other structure is mutated to a hydrophilic side chain.

**Contribution of non-specific interactions to stability.** Non-specific interactions include hydrophobic and solvation effects, as well as an average direct interaction between side chains and the entropic effect of reducing the freedom of side-chain rotations upon folding. A good way to approximate this effect is to assume that thermodynamic quantities are proportional to the accessible surface area and to use atomic parameters to estimate the hydration and chain parts of thermodynamic quantities, as in the method of Oobatake and Ooi (1993). This method predicts the thermodynamics of protein unfolding from the 3-D structure with reasonable accuracy. However, it does not take specific interactions into account explicitly. Therefore the results are suitable to estimate the role of non-specific interactions (hydration and entropic effects) in protein stability.

We calculated the unfolding free energy for both the EcIPMDH model and TtIPMDH by the method of Oobatake and Ooi (1993). The calculated free energy curves (as a function of temperature) are shown in Figure 5. The curve for TtIPMDH is shifted along the temperature axis towards higher temperatures by  $\sim 8^\circ\text{C}$  with respect to the curve for EcIPMDH. Given that the difference between the melting points of the two proteins is  $\sim 20^\circ\text{C}$ , as measured by scanning microcalorimetry (unpublished data from this laboratory), we can conclude that non-specific interactions are responsible for a significant part of the heat stability difference. This is in accordance with the results of site-directed mutagenesis experiments (Kirino *et al.*, 1994), which proved the thermostabilizing effect of increased hydrophobic contacts at the subunit-subunit interface in TtIPMDH.

The free energy curve for TtIPMDH is also shifted downwards with respect to the EcIPMDH curve which, together with the shift towards higher temperatures, is indicative of the increased hydrophobic stabilization in TtIPMDH. It is notable that the free energy for unfolding of TtIPMDH is negative in the whole temperature range according to this calculation. However, we should bear in mind that IPMDH is a dimer, and that the stability of dimeric proteins depends on protein concentration. This means that a concentration-dependent

entropic term should be added to the calculated curve. Because this term is the same for both proteins (assuming equal concentrations), the relative positions of the two curves are not affected. The negative value of the calculated free energy curve for TtIPMDH may also indicate that specific interactions play a greater role in the stabilization of the TtIPMDH dimer than in the stabilization of EcIPMDH. A similar situation was observed in our earlier study on glyceraldehyde-3-phosphate dehydrogenase (Szilágyi and Závodszy, 1995), suggesting that this mechanism may be a general strategy of thermostabilization for proteins from thermophilic sources.

**Secondary structure and stability of helices.** No significant difference was found in the length of the regular secondary structural elements ( $\alpha$ -helices,  $\beta$ -sheets) in the two IPMDHs.

An increase in the number of residues promoting the formation of  $\alpha$ -helices has been observed previously in proteins isolated from thermophiles (Argos *et al.*, 1979). The stability of helices is governed by a number of factors, including the intrinsic helical propensities of various side chains, interactions between side chains, helix-dipole interactions and helix capping. The contribution of all these effects to helix stability has been studied extensively using synthetic oligopeptides and protein engineering techniques (see the review by Fersht and Serrano, 1993).

We compared the intrinsic stabilities (stabilities in the absence of tertiary interactions) of helical regions in TtIPMDH and EcIPMDH using the program AGADIR to calculate the average helicities per residue for both polypeptide chains. The results are shown in Figure 6. As can be seen from the diagram, the calculated helicity for most regions of the polypeptide chain that are in  $\alpha$ -helices in the folded chains is higher in TtIPMDH than in EcIPMDH, i.e. the intrinsic stability of most helices is larger in the thermophilic IPMDH variant. We also examined the effect of the various interactions that are important for the stabilization of  $\alpha$ -helices by examining the types of residues that occur at the N- and C-termini of helices and at internal positions, respectively (see Fersht and Serrano, 1993). These considerations showed that the higher stability of TtIPMDH helices is mainly the result of additional helix-dipole interactions and capping effects, relative to EcIPMDH (data not shown). The effect of increased helix stability on the overall stability of the whole molecule has to be confirmed by site-directed mutagenesis studies.

#### *Comparison of the model of EcIPMDH with its X-ray structure*

The X-ray structure of EcIPMDH was solved after we completed and analyzed our model (Wallon, G., Kryger, G., Lovett, S., Oshima, T., Ringe, D. and Petsko, G.A., 1996, submitted to *Structure*). We received the preliminary coordinates of the X-ray structure from the authors (Gerlind Wallon and Gregory A. Petsko at Brandeis University, MA); thus we had the chance to compare our homology model with the X-ray structure, assess its accuracy and see how our conclusions concerning thermostabilizing interactions were affected.

The overall r.m.s. deviation for  $C_\alpha$ s between our model and the X-ray structure was 4.5 Å. This seems to be quite large. However, the conformational differences between the two structures are mostly localized to a few short regions where large insertions occur in EcIPMDH with respect to the reference structures used for model building (Figure 7). These regions are (i) part of the so-called arm-like region (residues Gly151-Ala161), with a  $C_\alpha$  r.m.s. deviation of 4.0 Å, and (ii) a loop

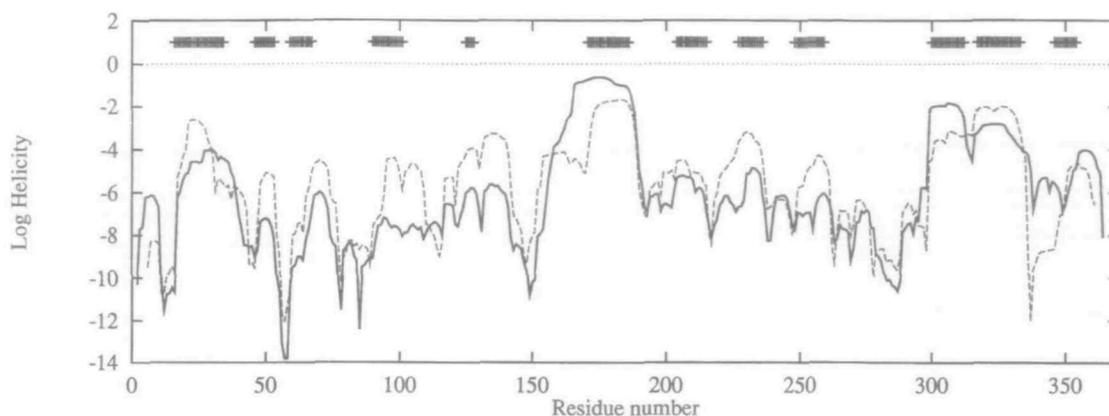


Fig. 6. Average helicities per residue calculated from the sequence of EcIPMDH (solid line) and TtIPMDH (dashed line) by the program AGADIR. The values are plotted on a logarithmic scale for ease of comparison. (+) at the top of the graph indicate the residues in  $\alpha$ -helices in the EcIPMDH model.

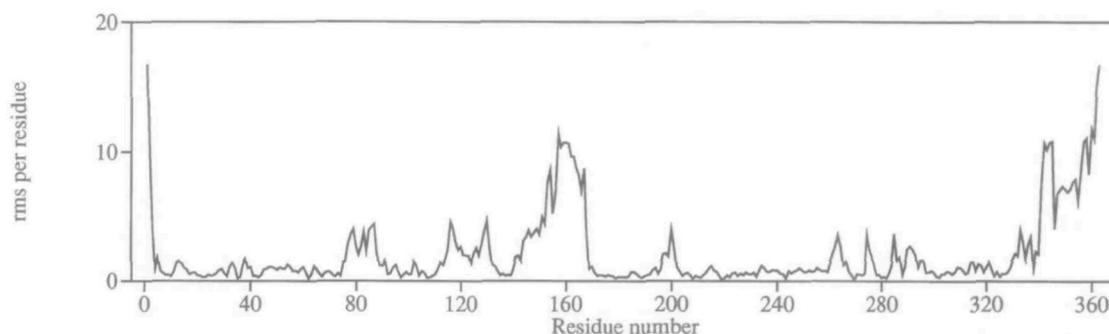


Fig. 7. R.m.s. deviation (in  $\text{\AA}$ ) between the  $C_\alpha$  traces of the model and X-ray structures of EcIPMDH.

region before the C-terminal helix (residues Ala340–Ala344) which causes a slight dislocation of the helix, with a  $C_\alpha$  r.m.s. deviation of 4.5  $\text{\AA}$  for the Ala340–Val363 region. We had no structural information for the N-terminal four residues (Met1–Asn4) during model building; this region also differs in our model from the X-ray structure. Omitting these parts (Met1–Asn4, Gly151–Ala161, Ala340–Ala363) from the structural superposition results in a much better r.m.s. deviation value of 1.44  $\text{\AA}$ . In the case of small insertions, the model structure is close to the X-ray structure.

Because of local conformational differences, we have somewhat underestimated the number of intrasubunit ion pairs in EcIPMDH (50 in our model versus 60 in the X-ray structure; see Table II). This is mainly caused by small structural deviations in our model from the X-ray structure and reflects the fact that ion pairs are sensitive to even small displacements of atoms. In the Leu272–Phe277 region, there is a single residue insertion in EcIPMDH with respect to TtIPMDH, and the conformation of the polypeptide backbone in our model differs slightly from that in the X-ray structure in this region (Figure 8). This deviation in turn causes the  $C_\delta$  of Glu274 (the inserted residue) in the model to be 2.5  $\text{\AA}$  away from its position in the X-ray structure. We missed three ion pairs because of this displacement. Another critical situation is observed with Glu361, which is in the overhanging C-terminal part of EcIPMDH, i.e. we had no structural information from homology for this residue during model building. Again, three ion pairs were missed as a consequence of the inaccurate prediction of this overhanging three-residue region. Because of the missing ion pairs, ion pair clusters (networks) are broken at some points, and the model predicts smaller clusters for EcIPMDH than observed in the X-ray structure. However, our

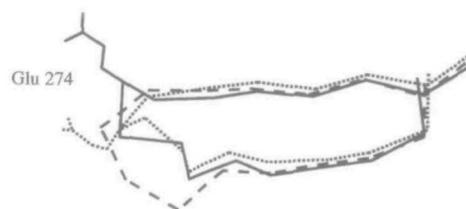


Fig. 8. Structure of the Leu272–Phe277 region in TtIPMDH (dashed line), the EcIPMDH model (dotted line) and the X-ray structure of EcIPMDH (solid line). Only the  $C_\alpha$  trace and the atoms of Glu274 in EcIPMDH are shown.

conclusion, that the number of isolated ion pairs is higher and that of intermediate-sized clusters is smaller in EcIPMDH than in TtIPMDH, still holds.

The number of intersubunit ion pairs is six in the EcIPMDH X-ray structure, while it is four in our model. We missed the Arg152a–Glu207b pair (and its symmetry-related pair) because of the deviation between the arm-like regions (Gly151–Ala161) in the model and the X-ray structure. However, our conclusion concerning the stabilization of TtIPMDH by intersubunit ion pairs still holds because TtIPMDH has 10 intersubunit ion pairs.

The buried polar and apolar surfaces and their ratio, calculated from our EcIPMDH model, are in good agreement with those calculated from the X-ray structure (see Table III). This confirms our conclusion concerning the stabilization of TtIPMDH by stronger hydrophobic interactions in the protein core.

There is no difference in secondary structure (the lengths of helical and  $\beta$ -strand regions) between our model and the X-ray structure. The model somewhat underestimates the number of hydrogen bonds (normalized to the number of

atoms): according to the X-ray structure, there is practically no difference between EcIPMDH and TtIPMDH, while our model predicts a slightly lower value. However, the observation that electrostatic interactions in TtIPMDH are replaced by hydrogen bonds in EcIPMDH at many sites is confirmed by the X-ray structure.

In summary, our model agrees well with the X-ray structure. Significant deviations between the model and the X-ray structure were only found where large insertions occur in EcIPMDH relative to the reference structures used for model building. All the quantities considered important for thermostability were predicted by the model with reasonable accuracy; some specific interactions were also predicted correctly. A few specific interactions were missed because of subtle conformational differences between the model and the X-ray structure. This points to the importance of finding more accurate modeling techniques that can reliably predict the conformations of regions with a low structural homology to known structures.

The modeling proved useful in determining the major tendencies of changes in structural features that characterize the transition from the thermophilic IPMDH to the mesophilic variant, and provided a good basis for designing site-directed mutagenesis studies.

### Acknowledgements

We thank Professor Gregory A.Petsko and Ms Gerlind Wallon from Brandeis University, MA, for providing us with the preliminary coordinates of EcIPMDH prior to publication. This work was supported in part by grants from the Hungarian Science Foundation (OTKA grant nos F6303, T006307 and T5206).

### References

- Amelunxen, R.E. and Murdock, A.L. (1978) *Crit. Rev. Microbiol.*, **6**, 343–393.
- Argos, P., Rossmann, M.G., Grau, U.M., Zuber, H., Frank, G. and Tratschin, J.D. (1979) *Biochemistry*, **18**, 5698–5703.
- Barker, W.C. and Dayhoff, M.O. (1972) In Dayhoff, M.O. (ed.), *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Washington, DC, pp. 101–110.
- Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) *J. Chem. Phys.*, **81**, 3684–3690.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.*, **112**, 535–542.
- Bowie, J.U., Lüthy, R. and Eisenberg, D. (1991) *Science*, **253**, 164–170.
- Chan, M.K., Mukund, S., Kletzin, A., Adams, M.W. and Rees, D.C. (1995) *Science*, **267**, 1463–1469.
- Chothia, C.H. (1974) *Nature*, **248**, 338–339.
- Daniel, R.M. (1986) *Protein Structure, Folding and Design*. Liss, New York.
- Fersht, A.R. and Serrano, L. (1993) *Curr. Opin. Struct. Biol.*, **3**, 75–83.
- Green, S.M., Meeker, A.K. and Shortle, D. (1992) *Biochemistry*, **31**, 5717–5728.
- Greer, J. (1990) *Proteins: Struct. Funct. Genet.*, **7**, 317–334.
- Holm, L., Ouzounis, C., Sander, C., Tuparev, G. and Vriend, G. (1992) *Protein Sci.*, **1**, 1691–1698.
- Horovitz, A.H., Serrano, L., Avron, B., Bycroft, M. and Fersht, A.R. (1990) *J. Mol. Biol.*, **216**, 1031–1044.
- Hurley, J.H. and Dean, A.M. (1994) *Structure*, **2**, 1007–1016.
- Hurley, J.H., Dean, A.M. and Koshland, D.E. (1991) *Biochemistry*, **30**, 8671–8678.
- Imada, K., Sato, M., Tanaka, N., Katsube, Y., Matsuura, Y. and Oshima, T. (1991) *J. Mol. Biol.*, **222**, 725–734.
- Jaenicke, R. (1981) *Annu. Rev. Biophys. Bioengng.*, **10**, 1–67.
- Jaenicke, R. (1991) *Eur. J. Biochem.*, **202**, 715–728.
- Jaenicke, R. and Závodszy, P. (1990) *FEBS Lett.*, **268**, 344–349.
- Jones, T.A. and Thirup, S. (1986) *EMBO J.*, **5**, 819–822.
- Kelly, C.A., Nishiyama, M., Ohnishi, Y., Beppu, T. and Birktoft, J.J. (1993) *Biochemistry*, **29**, 6264–6269.
- Kirino, H., Aoki, M., Aoshima, M., Hayashi, Y., Ohba, M., Yamagushi, A., Wakagi, T. and Oshima, T. (1994) *Eur. J. Biochem.*, **220**, 275–281.
- Korndörfer, I., Steipe, B., Huber, R., Tomschy, A. and Jaenicke, R. (1995) *J. Mol. Biol.*, **246**, 511–521.
- Kraulis, P.J. (1991) *J. Appl. Crystal.*, **24**, 946–950.
- Lee, B. (1993) *Protein Sci.*, **2**, 733–738.
- Lüthy, R., Bowie, J.U. and Eisenberg, D. (1992) *Nature*, **356**, 83–85.
- Miyazaki, K. and Oshima, T. (1993) *FEBS Lett.*, **332**, 37–38.
- Miyazaki, K. and Oshima, T. (1994) *Protein Engng.*, **7**, 401–403.
- Moriyama, H., Onodera, K., Sakurai, M., Tanaka, N., Kirino-Kagawa, H., Oshima, T. and Katsube, Y. (1995) *J. Biochem.*, **117**, 408–413.
- Muñoz, V. and Serrano, L. (1994) *Nature Struct. Biol.*, **1**, 399–409.
- Muñoz, V. and Serrano, L. (1995a) *J. Mol. Biol.*, **245**, 275–296.
- Muñoz, V. and Serrano, L. (1995b) *J. Mol. Biol.*, **245**, 297–308.
- Needleman, S.B. and Wunsch, C.D. (1970) *J. Mol. Biol.*, **48**, 443–453.
- Numata, K., Muro, M., Akutsu, N., Nosoh, Y., Yamagishi, A. and Oshima, T. (1995) *Protein Engng.*, **8**, 39–43.
- Onodera, K., Sakurai, M., Moriyama, H., Tanaka, N., Numata, K., Oshima, T. and Sato, M. (1994) *Protein Engng.*, **7**, 453–459.
- Oobatake, M. and Ooi, T. (1993) *Prog. Biophys. Mol. Biol.*, **59**, 237–284.
- Pace, C.N. (1990) *Trends Biochem. Sci.*, **15**, 14–17.
- Perutz, M.F. and Raidt, H. (1975) *Nature*, **255**, 256–258.
- Ponder, J.W. and Richards, F.M. (1987) *J. Mol. Biol.*, **193**, 775–791.
- Serrano, L., Kellis, J.T., Cann, P., Matouschek, A. and Fersht, A.R. (1992) *J. Mol. Biol.*, **224**, 783–804.
- Shenkin, P.S., Yarmush, D.L., Fine, R.M., Wang, H.J. and Levinthal, C. (1987) *Biopolymers*, **26**, 2053–2085.
- Shirley, B.A., Stanssens, P., Hahn, U. and Pace, C.N. (1992) *Biochemistry*, **31**, 725–732.
- Shrake, A. and Rupley, J.A. (1973) *J. Mol. Biol.*, **79**, 351–371.
- Spassov, V.Z., Karshikoff, A.D. and Ladenstein, R. (1995) *Protein Sci.*, **4**, 1516–1527.
- Suguna, K., Bott, R.R., Padlan, E.A., Subramanian, E., Sheriff, S., Cohen, G.H. and Davies, D.R. (1987) *J. Mol. Biol.*, **196**, 877–900.
- Szilágyi, A. and Závodszy, P. (1995) *Protein Engng.*, **8**, 779–789.
- Tomschy, A., Böhm, G. and Jaenicke, R. (1994) *Protein Engng.*, **7**, 1471–1478.
- Vriend, G. (1990) *J. Mol. Graph.*, **8**, 52–56.
- Vriend, G. and Sander, C. (1991) *Proteins: Struct. Funct. Genet.*, **11**, 52–58.
- Walker, J.E., Wonacott, A.J. and Harris, J.I. (1980) *Eur. J. Biochem.*, **108**, 581–586.
- Yaoi, T., Miyazaki, K. and Oshima, T. (1995) *Biochem Biophys. Res. Commun.*, **210**, 733–737.
- Zhang, T. and Koshland, D.E., Jr (1995) *Protein Sci.*, **4**, 84–92.

Received December 8, 1995; revised March 5, 1996, accepted March 20, 1996