Protein Engineering vol.8 no.8 pp.779–789, 1995

Structural basis for the extreme thermostability of D-glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*: analysis based on homology modelling

András Szilágyi and Péter Závodszky[1]

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

[1]To whom correspondence should be addressed

D-Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from a hyperthermophilic eubacterium, *Thermotoga maritima*, is remarkably heat stable (T_m = 109°C). In this work, we have applied homology modelling to predict the 3-D structure of *Th.maritima* GAPDH to reveal the structural basis of thermostability. Three known GAPDH structures were used as reference proteins. First, the rough model of one subunit was constructed using the identified structurally conserved and variable regions of the reference proteins. The holoenzyme was assembled from four subunits and the NAD molecules. The structure was refined by energy minimization and molecular dynamics simulated annealing. No errors were detected in the refined model using the 3-D profile method. The model was compared with the structure of *Bacillus stearothermophilus* GAPDH to identify structural details underlying the increased thermostability. In all, 12 extra ion pairs per subunit were found at the protein surface. This seems to be the most important factor responsible for thermostability. Differences in the non-specific interactions, including hydration effects, were also found. Minor changes were detected in the secondary structure. The model predicts that a slight increase in α-helical propensities and helix–dipole interactions also contribute to increased stability, but to a lesser degree.

Keywords: glyceraldehyde-3-phosphate dehydrogenase/homology modelling/protein structure prediction/thermophiles/thermostability

Introduction

D-Glyceraldehyde-3-phosphate dehydrogenase (GAPDH; EC 1.2.1.12) from a variety of mesophilic and thermophilic species has been used to elucidate the mechanisms of thermophilic adaptation (Jaenicke, 1981, 1991; Olsen, 1983; Hensel *et al.*, 1987). Wrba *et al.* (1990) isolated the enzyme from *Thermotoga maritima*, an organism that grows optimally at 80°C, with an upper limit of growth at 90°C. The physico-chemical and folding properties of this enzyme have been studied extensively (Wrba *et al.*, 1990; Rehaber and Jaenicke, 1992, 1993). The melting point of this protein is 109°C according to differential scanning microcalorimetric measurements (Wrba *et al.*, 1990). Heat inactivation studies have revealed that the half-life of the holoenzyme exceeds 120 min, even at 100°C. The results of other experiments suggest that the gross structure of the enzyme must be very similar to the structure of GAPDHs from mesophilic sources (Wrba *et al.*, 1990). The anomalous temperature dependence of CD and fluorescence spectra indicates that the packing of the thermophilic aromatic side chains in their hydrophobic environment should be closer than in the mesophilic counterparts. With increasing temperature, further tightening seems to occur (Wrba *et al.*, 1990). Hydrogen–deuterium exchange reveals significantly increased rigidity of the thermophilic enzyme when compared with mesophilic GAPDHs at 25°C. However, the flexibilities of the thermophilic and mesophilic proteins are almost the same at the respective physiological temperatures (Wrba *et al.*, 1990), which supports the idea of 'corresponding states' of protein pairs adapted to different environmental conditions (Jaenicke and Závodszky, 1990). Comparison of the *Th.maritima* GAPDH sequence with sequences of GAPDHs from other species has led to the conclusion that 'traffic rules' suggested previously for amino acid exchanges of thermophilic adaptation (Argos *et al.*, 1979; Menéndez-Arias and Argos, 1989) can be applied only partially to this hyperthermophilic enzyme (the most significant exchange refers to Lys → Arg; Schultes *et al.*, 1990).

To explain the structural basis for the extreme thermostability of *Th.maritima* GAPDH, the 3-D structure of the enzyme is required. The 3-D structure is also the starting point for designing site-directed mutagenesis experiments to identify the structural elements responsible for the extreme thermostability and for undertaking computational modelling studies to predict the effects of possible mutations.

In this work, we have built a model for the *Th.maritima* GAPDH holoenzyme by homology modelling. To find possible explanations for the enhanced thermostability, some structural features of the model were compared with those of the known GAPDH structures, especially with *Bacillus stearothermophilus* GAPDH. Another aim of this analysis was to assist mutagenesis studies by finding target residues that may be particularly important for increased thermostability.

The heat stability of GAPDH from *B.stearothermophilus*, a moderate thermophile, has been discussed previously by Walker *et al.* (1980). Some buried salt bridges, increased hydrophobic contacts and an extra hydrogen bridge were observed between neighbouring subunits in the core of the enzyme. GAPDH from *Th.maritima* shows very high sequence similarity to the *B.stearothermophilus* enzyme (65% sequence identity; the sequence identity between GAPDHs from other species is also high, being ~50%), but its melting point is higher by ~20°C. The aim of this work was to find some possible answers to the question of which mechanisms are responsible for this extra stabilization, i.e. to identify the changes necessary to transform a moderately thermophilic enzyme into a hyperthermophilic enzyme.

Materials and methods

Construction of a model

The modelling procedure was performed basically as described by Greer (1990). The molecular graphics program InsightII (Biosym Inc.) was used on a Silicon Graphics Personal IRIS 4D25 workstation.

Reference proteins. The atomic coordinates of the reference proteins were obtained from the Protein Data Bank (Bernstein *et al.*, 1977). The reference proteins were the three GAPDH holoenzymes with known structures: human GAPDH (PDB entry 3GPD by H.C.Watson and J.C.Campbell), GAPDH from lobster (Moras *et al.*, 1975; PDB entry 1GPD) and GAPDH from *B.stearothermophilus* (Skarzinsky *et al.*, 1987; PDB entry 1GD1).

Finding the structurally conserved regions (SCRs) and structurally variable regions (SVRs). The first step in homology modelling is to find the SCRs and the SVRs of the protein family using a 3-D structural alignment of the reference structures. A number of sophisticated methods are available to perform this task (see, for example, Bachar *et al.*, 1993, and references therein). In our case, the structural and sequential similarities of the reference proteins are high (>50% sequence identity between any pair of reference proteins). Therefore a simple procedure proved to be adequate.

The structural alignment was performed at the subunit level by taking the three possible pairwise combinations and proceeding as follows. First, an initial sequence alignment was generated using the Needleman–Wunsch algorithm (Needleman and Wunsch, 1970) with the Dayhoff PAM250 matrix (Barker and Dayhoff, 1972). Based on this alignment, the two structures were superimposed in space by a least-squares superposition of C_α traces. Then the sequential alignment was corrected by visual inspection of the superimposed structures. This procedure was repeated iteratively until the structural and sequential alignments became consistent with each other. In the final step, only C_αs closer than 1.7 Å were superimposed. These regions were defined as SCRs and the remaining regions as SVRs. Thereafter, all three reference protein sequences were aligned based on the pairwise alignments. The SCRs for all three reference proteins were defined as the largest common subset of SCRs identified for any pair of reference proteins (Figure 1).

Construction of a rough model for the subunit. The sequence of *Th.maritima* GAPDH (Schultes *et al.*, 1990) was aligned with the three previously aligned reference protein sequences. The Needleman–Wunsch algorithm (Needleman and Wunsch, 1970) with the Dayhoff PAM250 matrix (Barker and Dayhoff, 1972) was used again. Based on this alignment, the SCRs and SVRs were assigned in the *Th.maritima* GAPDH sequence.

In building the model of the *Th.maritima* GAPDH subunit, the SCRs were constructed first. The main-chain atom positions in each SCR were copied from the corresponding region of the reference protein with the most similar sequence in that particular region. In this way, the number of required side-chain replacements when constructing the side chains was minimized. Side-chain conformations were either copied from the chosen reference protein (in the case of residue identity) or adjusted in such a way that common torsional angles were copied and the remaining parts were constructed in an extended conformation (in the case of differing residues).

Next, the SVRs were constructed. The corresponding region in each reference protein was examined. The region with the same length (number of residues) and the most similar sequence was used, i.e. its conformation was copied and its side chains were adjusted in the same way as described for SCRs (see above). If such a suitable fragment was not found from among the reference proteins, a loop search procedure (Jones and Thirup, 1986) was applied to a subset of the Protein Data Bank which contained all available structures of at least 2.0 Å resolution.

Refinement of the subunit structure. After adding hydrogen atoms to the system (assuming a pH of 7.0), the distorted peptide bonds at SCR–SVR boundaries (splicing points) had to be corrected. Sometimes the peptide bonds were too long, and the value of the omega torsion angles differed from the optimal value (180°) at a number of splicing points. Correction was made by energy minimization using the Discover (Biosym Inc.) consistent valence forcefield (CVFF) without cross terms and harmonic bond stretching potentials. Only the SVRs and the two residues flanking them in the two neighbouring SCRs were included in the minimization. A 10 Å cut-off distance was applied and omega angles at the splicing points were forced to be *trans* by a force constant of 200 kcal/rad^2/mol. Minimization was carried out by the steepest descents algorithm until the maximum derivative fell below 1.0 kcal/Å/mol.

Assembly of the holoenzyme. To be able to model the contacts between subunits in the tetramer and substrate binding, the holoenzyme was assembled. We created three additional copies of the subunit model and fitted the four subunits together according to the known, although approximate, 222 symmetry of the enzyme. The NAD coenzymes (one per subunit), sulfate ions (two per subunit, occupying the suspected substrate binding sites) and water molecules (173 per subunit) present in the *B.stearothermophilus* GAPDH holoenzyme crystal structure were also transferred to the model. The entire system contained 23 154 atoms.

The partial atomic charges of the NAD and sulfate molecules were assigned by the MOPAC program with the MNDO semi-empirical method. A single self-consistent field calculation was carried out (the MOPAC keyword was 1SCF).

Refinement by energy minimization. For energy minimizations and subsequent molecular dynamics calculations, the Discover CVFF was used without cross terms and with harmonic bond stretching potentials. A 14 Å cut-off distance was applied throughout and the absence of bulk water was compensated for by a distance-dependent dielectric constant $\varepsilon(r) = r$ (Å). The main-chain atoms in the SCRs, except for their N- and C-terminal residues, were held fixed during all calculations.

All minimizations were performed by the steepest descents algorithm until the maximum derivative fell to <10.0 kcal/Å/mol, and then by the conjugate gradient algorithm until the maximum derivative fell to <1.0 kcal/Å/mol.

First, a series of minimizations was performed during which only the most strongly overlapping atoms were allowed to move, so as to relax the most severe contacts within the molecule.

Employing the symmetry of the tetramer, the subsequent steps of the refinement were performed on only one selected subunit and those parts of the remaining three subunits that were closer than 7 Å to the selected subunit. Water molecules and substrates belonging to the selected subunit were also included.

Minimization was carried out in three phases. In the first phase, only the variable regions and water molecules were allowed to relax. In the second phase, the side chains in the SCRs that had been replaced in the course of the model building were also included in the minimization. Finally, the minimization was extended to the residues that had not been replaced.

As a last step, a minimization was performed in which the

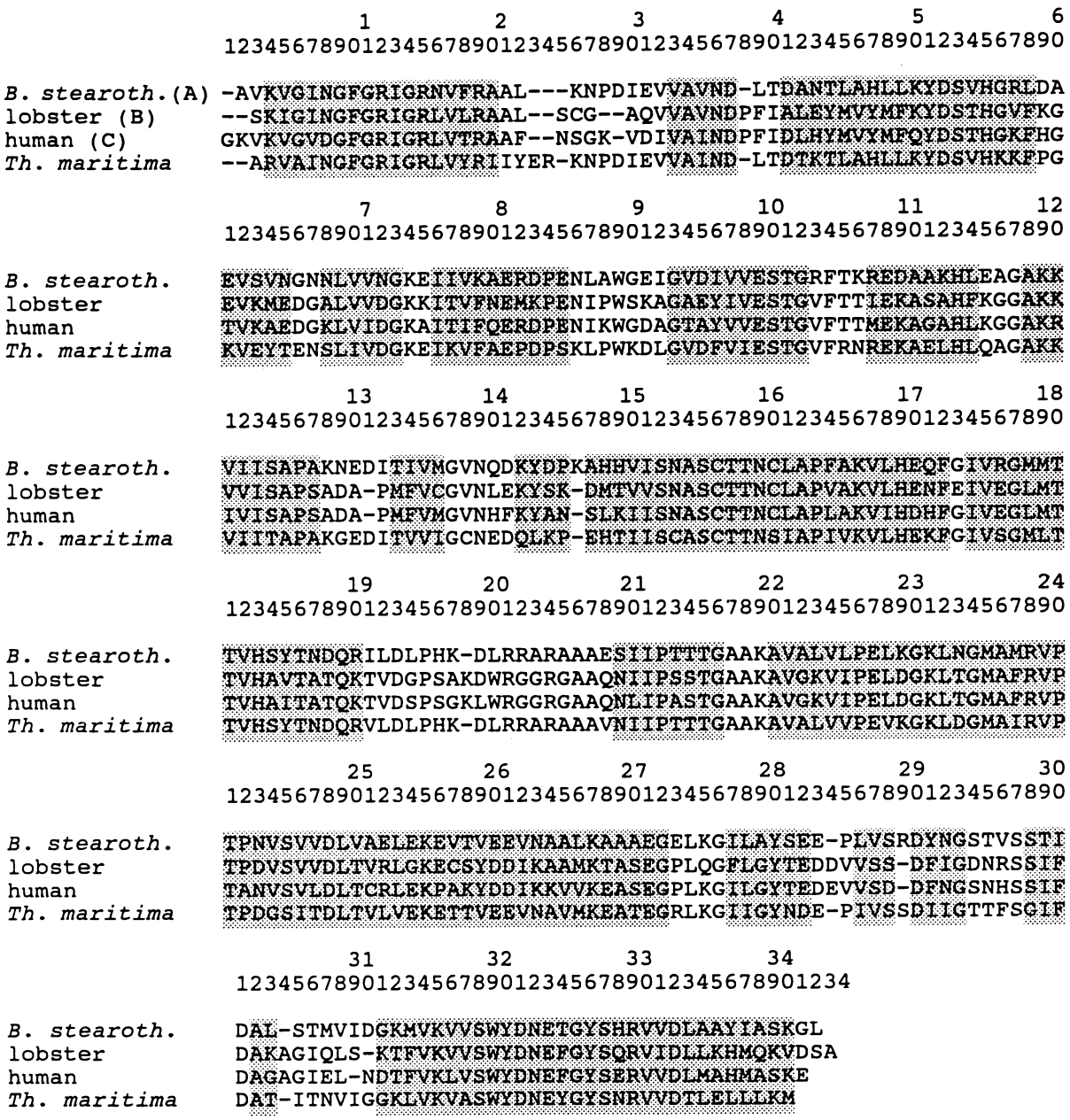

Fig. 1. Sequence alignment of the reference proteins (human, lobster and *B.stearothermophilus* GAPDHs) and the sequence of *Th.maritima* GAPDH. The SCRs are highlighted in the alignment.

entire subunit and those parts of the remaining subunits that were in contact with it were allowed to relax, together with water and substrate molecules.

Altogether, the minimization procedure required ~10 days of continuous computer processing time.

Further refinement by molecular dynamics simulated annealing. Energy minimization forces the structure into the nearest local minimum of the potential energy surface. To allow the structure to overcome energy barriers and to find another energy minimum which corresponds to a structure having less strain and a more favourable free energy, a molecular dynamics simulation was performed. To ensure better performance, we applied molecular dynamics simulated annealing, as outlined in Results.

The forcefield and all other parameters were the same as with the energy minimizations. Only the selected subunit and those parts of the other subunits that were closer than 7 Å to it were simulated. The main-chain atoms in the SCRs were held fixed except for the two residues at both ends. The waters, NAD molecules and sulfate ions were tethered with a force constant of 120 kcal/Å/mol to prevent them from 'boiling off'. The time step was 0.4 fs (a larger time step resulted in serious integration errors at high temperatures). A time period of 7.5 ps was simulated altogether, which required ~30 days of computer processing time.

After the simulated annealing process, another energy minimization was carried out with the same parameters as described earlier.

Analysis of the model

Firstly, we evaluated the quality of the model to reveal errors in the structure. Next, we analysed the model to reveal the structural specificities underlying the extreme thermostability

Fig. 2. 3-D alignment (spatial superposition) of subunits from two reference proteins (the lobster and the *B.stearothermophilus* GAPDH subunit). Solid lines, SCRs; dashed lines, SVRs.

and to find locations in the structure where site-directed mutagenesis could be used to justify our hypotheses based on the modelling study. We used the *B.stearothermophilus* GAPDH holoenzyme structure as a reference, i.e. all calculations and investigations performed on our model were also carried out on the *B.stearothermophilus* GAPDH structure for comparison. We chose the *B.stearothermophilus* GAPDH structure because its resolution is the highest (1.8 Å) among the reference proteins, and the structural background of the moderate thermostability of this enzyme had been discussed previously by Walker *et al.* (1980).

Evaluation of the quality of the model. Structures obtained by homology modelling always have errors and inaccuracies. Therefore, methods are necessary to recognize such errors in the model structure.

First, the number of contacts (overlaps between non-bonded atoms) and the forcefield energy were calculated. To increase sensitivity, we used the 3-D profile method (Bowie *et al.*, 1991) employing 'profile window plots' as described by Lüthy *et al.* (1992).

Estimation of the role of non-specific interactions. We used the method of Oobatake and Ooi to predict the thermodynamics of protein unfolding from the 3-D structure (Ooi *et al.*, 1987; Ooi and Oobatake, 1988, 1991; Oobatake and Ooi, 1992). The method allows the calculation of the unfolding enthalpy, entropy and free energy as a function of temperature, based on the difference between the accessible surface areas of seven atomic groups in the folded and unfolded states, using parameters derived from experimental data. The method gives a relatively good estimation of the thermodynamic quantities, although it cannot take specific interactions into account explicitly (these are mainly electrostatic, e.g. buried salt bridges, helix–charge interactions, etc.). However, it is suitable for comparative calculations and the estimation of the role of non-specific interactions in protein stability.

Hydrogen bonds and ion pairs. A hydrogen bond between a donor and an acceptor heavy atom was defined if they were closer than 3.0 Å to each other. The ion pairs were defined using a search procedure, as described in Results.

Secondary structure definition. The secondary structure of the proteins was defined by the main-chain hydrogen bonding pattern using the DSSP program (Kabsch and Sander, 1983). Only the SUMMARY column of the DSSP output file was used.

α-Helical propensities and helix–dipole interactions. α-Helical propensities were calculated by summing the helix formation parameters (the equilibrium constants for adding a residue of a given type to the end of a helix; Scheraga, 1978) for the residues which are in helices in both our model and *B.stearothermophilus* GAPDH.

Helix–dipole interactions were identified by examining the residues in the first and last turns in the helices. Negatively charged residues at the N-terminus or positively charged residues at the C-terminus of a helix were assumed to interact favourably with the helix–dipole moment, while opposite charges at the same locations were assumed to decrease the stability of the helix.

Results

Construction of the model

We constructed the 3-D structural model of *Th.maritima* GAPDH using the three GAPDH holoenzymes with known structures as reference proteins: human GAPDH, GAPDH from lobster and GAPDH from *B.stearothermophilus*.

Figure 1 shows the sequential alignment of the reference proteins and the *Th.maritima* protein, with SCRs highlighted. Figure 2 shows the 3-D alignment of two reference proteins.

In the rough model of the subunit, most of the N- and C-terminal parts of the chain were taken from the eukaryotic reference proteins, while the middle part of it was taken from the *B.stearothermophilus* enzyme (Figure 3). This was based on the results of sequence similarity calculations.

In the case of the Ile18–Val28 loop (see Figure 1), the fragment containing the last seven residues of the loop was constructed using the corresponding part of the *B.stearothermophilus* protein (because of the perfect sequence identity). This was followed by a limited loop search procedure on the Protein Data Bank to construct a model for the region containing the first four residues of the loop. Based on sequence similarity

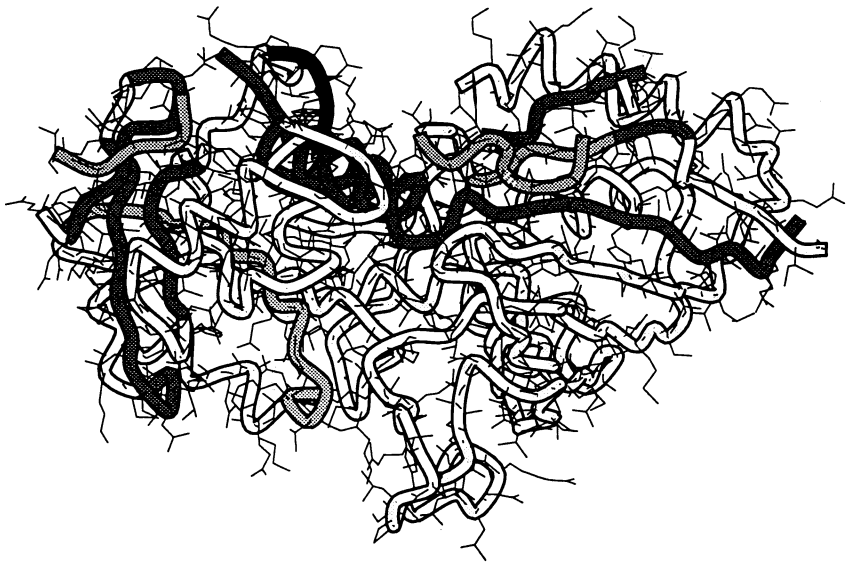

Fig. 3. Rough model of the *Th.maritima* GAPDH subunit. The side chains in this model correspond to the sequence of *Th.maritima* GAPDH. The peptide backbone is represented as a ribbon which is shaded according to the origin of the respective protein fragments. White, *B.stearothermophilus* GAPDH; light grey, lobster GAPDH; dark grey, human GAPDH; black, constructed by limited loop search on the Protein Data Bank.

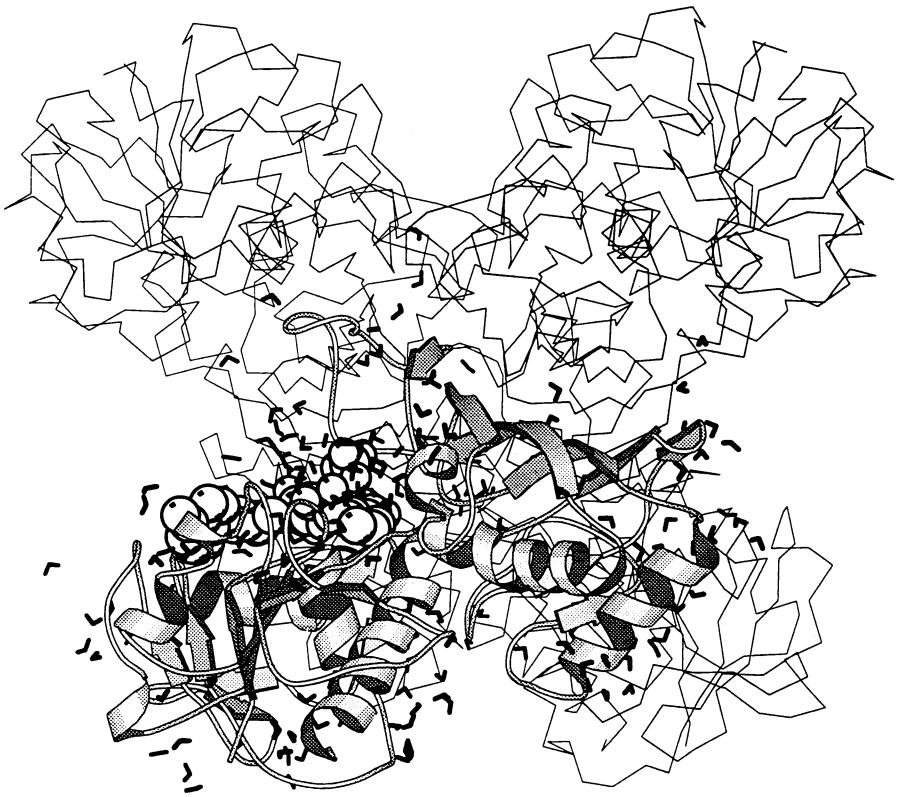

Fig. 4. The homology model of the *Th.maritima* GAPDH holoenzyme after energy minimization. The subunit that was extensively minimized is represented as a Richardson-type model (figure drawn with the program Molscript; Kraulis, 1991). A NAD molecule and two sulfate ions are represented as space-filling models. Water molecules are also shown.

and side-chain orientations, the Lys106–Lys109 fragment of the protein pseudoazurin from *Alcagenes faecalis* (PDB entry 1PAZ) was selected as a good model for this region.

In the course of model building, 113 residues had to be replaced (34% of all residues). A residue was replaced by a longer residue in 39 cases. Most of these residues were located at the protein surface, which shows that the inner core of this protein family is well conserved.

The forcefield energy of the initial rough model of the subunit was very high ($>10^{15}$ kcal/mol). This was reduced to ~2000 kcal/mol during the energy minimizations with omega-forcing, performed to correct the distorted peptide bonds. Large distortions were eliminated, while new smaller distortions were introduced into the structure. This is the consequence of an intrinsic limitation of energy minimization, i.e. energy barriers cannot be overcome by this procedure.

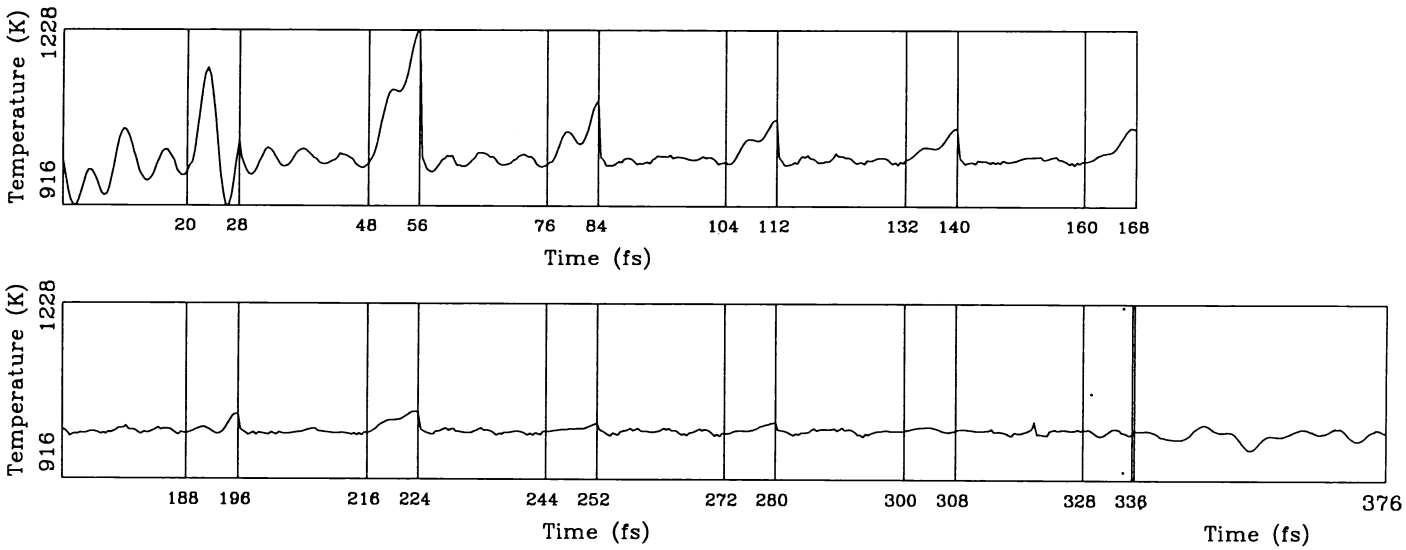

Fig. 5. Record of the temperature versus time in a molecular dynamics experiment designed to establish the time needed by our system to partially equilibrate during molecular dynamics. The results were used to set up a cooling schedule for simulated annealing (see text).

The forcefield energy of the assembled holoenzyme before refinement (with main-chain atoms in the SCRs fixed as described in Materials and methods) was very high ($>10^{15}$ kcal/mol), and there were numerous atomic overlaps in the structure. Our strategic energy minimization process reduced the value of the energy and the number of atomic overlaps. At the end of the minimization, the forcefield energy was –8190 kcal/mol. The energy-minimized structure of the holoenzyme is shown in Figure 4.

Simulated annealing using molecular dynamics. Simulated annealing (Kirkpatrick *et al.*, 1982) is a very effective optimization method which finds minima on an energy surface; in some cases it can even find the global minimum. Although mathematicians have worked out the theoretical basis of the simulated annealing algorithm, and numerous examples of applications and tests of it are known (see, for example, Laarhoven and Aarts, 1987), there are relatively few examples of the application of simulated annealing in protein structure prediction (see Laughton, 1994) and the effectiveness and limitations of the method are not well known. It is clear, however, that care must be taken when adjusting the parameters of the algorithm, known as a cooling schedule.

Molecular dynamics simulated annealing is a long molecular dynamics calculation which is started at high temperature and then the temperature is decreased in finite steps until a final temperature is reached. Important parameters of this cooling schedule are the initial and the final temperatures, as well as the rule for decreasing the temperature and the number of dynamic steps to be performed at each temperature. The simulated annealing algorithm works well only if the temperature is decreased slowly and the system can spend sufficient time at each temperature to get close to the state of thermodynamic equilibrium.

Reaching the state of complete thermodynamic equilibrium in such a large and complicated system as a protein molecule is a very long process, in the order of seconds. However, during molecular dynamics simulated annealing, it is important to ensure that the system reaches 'partial' equilibrium at each temperature, i.e. dynamic processes that have a relaxation time comparable with the length of simulation have enough time to equilibrate. Molecular dynamics simulations can be performed on a time scale in the order of 10 ps at present, i.e. such a 'partial' equilibration involves mainly the vibrational and rotational (in the case of surface side chains) degrees of freedom.

To establish the time necessary for 'partial' equilibration, i.e. the minimum time period to be simulated at any temperature during the simulated annealing process, we performed a preliminary calculation. The calculation was based on the fact that a system coupled to a heat bath and equilibrated will retain its equilibrium state when isolated from the heat bath, but if it has not reached equilibrium yet, then its energy and temperature will change upon isolation. Therefore, in our preliminary calculation we performed short (20 fs) simulations with coupling to a heat bath (Berendsen *et al.*, 1984) at 1000 K. Then we decoupled the system from the heat bath (i.e. no velocity scaling was applied) and continued the simulation for a further 8 fs. We repeated this process 12 times, and then performed a further 40 fs simulation in the isolated state. The values of temperature and potential energy were recorded continuously during all simulations.

The results of this preliminary calculation are shown in Figure 5. Initially, strong fluctuations in both the potential energy and the temperature were observed. This fluctuation reduced slowly. From the sixth phase (140–168 fs) onwards, both the potential energy and the temperature were constant in the 20 fs-long 'isothermal' phases. It was also observed that in the 8 fs-long 'isolated' phases, the temperature always increased and the potential energy decreased, indicating that equilibrium had not yet been reached. This tendency remained up to the 10th phase, but it was no longer observed after the 11th phase (280 fs). In the last 40 fs-long 'isolated' phase, only small fluctuations were observed without a definite change in the average value. This showed that processes that had a relaxation time comparable with the time of the simulation had equilibrated. The calculation established that a time period of 300 fs is sufficient for the system to reach 'partial' equilibrium.

Based on these data, our cooling schedule was set as follows. (i) The simulation starts at 1000 K, which enables the system to overcome high energy barriers without unfavourable structural changes. (ii) The simulation ends at 300 K, which is close to

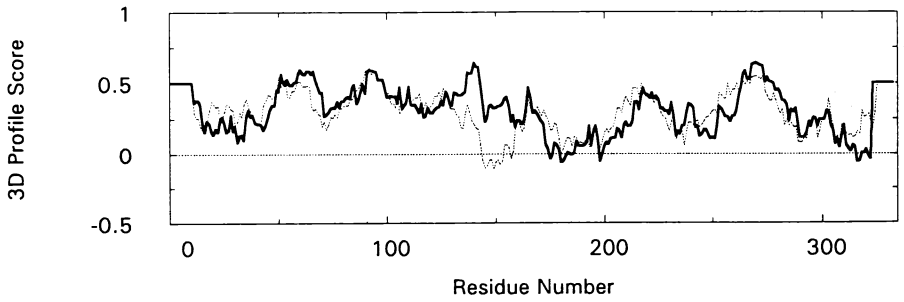

Fig. 6. 3-D profile window plots for the known structure of *B.stearothermophilus* GAPDH (dotted line) and the model of *Th.maritima* GAPDH (solid, thicker line). A window size of 20 residues was applied. The plots are suitable to establish the quality of the structures.

the physiological temperature and prevents the system from getting into energy minima that are too deep and do not provide enough entropy for the system to ensure a low value of free energy. (iii) The temperature is decreased by 5% in every phase of the process. This rate is slow enough to get close to equilibrium, and this non-linear way of temperature reduction follows the increasing time requirement of relaxation with decreasing temperature. (iv) The simulation is performed for 300 fs at each temperature.

After the 7.5 ps simulated annealing process, another energy minimization was carried out. The resulting forcefield energy was –9230 kcal/mol which is >1000 kcal/mol less than before the simulation. This shows that simulated annealing is a suitable method to find better energy minima for the system.

Analysis of the model

Evaluation of the quality of the model. We calculated some quantities for the model and for the corresponding part of the *B.stearothermophilus* GAPDH X-ray structure that are known to correlate with the quality of protein structures.

The forcefield energy value of our energy minimized model was –9230 kcal/mol, while it was –2252 kcal/mol for the *B.stearothermophilus* GAPDH X-ray structure. There were no serious overlaps between atoms in either of the structures. The number of weaker atomic overlaps was nearly the same.

For a more sensitive evaluation, we applied the 3-D profile method (Bowie *et al.*, 1991), as described in Materials and methods. The total 3-D profile score was 104.06 for our model and 104.5 for the *B.stearothermophilus* GAPDH subunit structure. The profile window plots are shown in Figure 6. The two curves are very similar, which demonstrates that the two structures are of similar quality.

Ion pairs. Buried ion pairs are easy to define using a simple distance rule, i.e. two oppositely charged residues are defined as an ion pair if the oppositely charged atoms are closer than 4.0 Å. However, the situation is more complicated in the case of surface residues because of the flexibility of the side chains and the screening effect of the solvent. The two side chains may form an ion pair with a certain probability. A simple distance rule is not sufficient in this case because in a homology model, energy minimization in the absence of bulk water may cause artefacts, and in a crystal structure, the crystalline environment may modify the conformation of surface side chains. Instead of performing time-consuming simulations on our model in the presence of bulk water, or introducing artefacts into the reference proteins by *in vacuo* simulations, we applied the following search procedure to define 'possible' ion pairs.

First, we took all residue pairs with opposite charges and a C_α distance <16.0 Å. These pairs are candidates for ion pairs; other pairs cannot form ion pairs for geometrical reasons. Next, all rotameric states were set for both side chains in every pair. The rotamer conformations were taken from the Ponder and Richards (1987) rotamer library. For every pair, all combinations of rotameric state and the original conformations were examined, and the forcefield energy and the number of atomic overlaps were calculated. Conformational states with any atomic overlap or positive forcefield energy were refused; the remaining states were accepted as energetically allowable states. Finally, a residue pair was defined as having a reasonable probability of forming an ion pair if the oppositely charged atoms were closer than 4.0 Å in at least 10% of these energetically allowable states. The 10% limit was chosen to maximize the agreement with the results of earlier distance-based definitions of ion pairs on known crystal structures of proteins.

Table I. Extra ion pairs found in the known structure of *B.stearothermophilus* GAPDH and the model of *Th.maritima* GAPDH

B.stearothermophilus GAPDH	*Th.maritima* GAPDH	
Asp54–Arg52	Asp67–Arg20b	Glu163–Arg266
Glu80–Lys107	Glu70–Arg2	Asp181–Arg231
Asp104–Lys107	Glu70–Lys72	Asp226–Lys212
Glu110–Lys107	Asp78–Lys81	Glu261–Lys164
Glu245–Arg169	Asp86–Lys85	Glu264–Lys268
Glu103–Lys101	Glu103–Arg102	Glu326–Lys330
Arg169 (o)–Glu245 (p)	Glu106–Arg102	Asp36–Lys38
Arg169 (o)–Asp301 (p)	Glu140–Lys138	Glu58–Lys56
Glu201 (o)–Arg281 (p)	Glu163–Lys164	Asp236–His50
Glu245 (o)–Arg169 (p)		
Asp301 (o)–Arg169 (p)		

Only those salt bridges that are not found in the other structure are shown. (o) and (p) are subunit identifiers.

We performed this procedure on both the X-ray structure of *B.stearothermophilus* GAPDH and our homology model of *Th.maritima* GAPDH. The results were compared.

We found that there are 18 potential intrasubunit ion pairs in the *Th.maritima* GAPDH model per subunit that are not present in *B.stearothermophilus* GAPDH, and there are six potential intrasubunit ion pairs in *B.stearothermophilus* GAPDH that are not present in the *Th.maritima* GAPDH model (Table I). Thus, the net difference in the number of intrasubunit ion pairs per subunit is 12 to the advantage of the *Th.maritima* enzyme. There are 14 ion pairs that are present in both proteins.

The situation is different for the intersubunit ion pairs: there are five ion pairs between the green subunit and the other

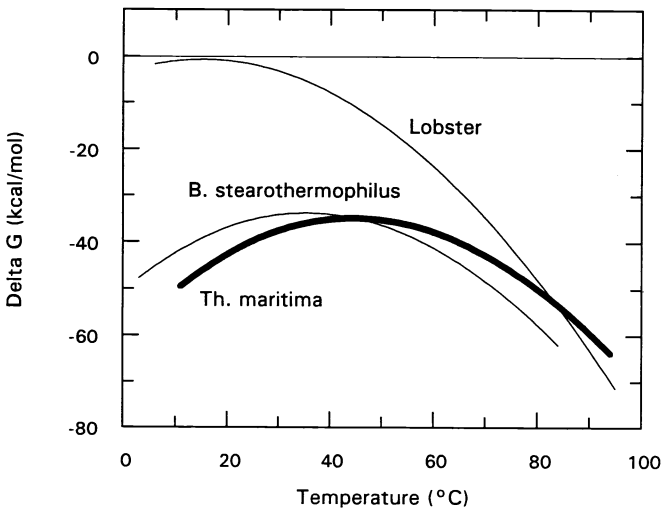

Fig. 7. Free energy change upon unfolding as a function of temperature calculated by the method of Oobatake and Ooi (1992) from the known 3-D structures of GAPDHs from lobster (thin line), *B.stearothermophilus* (thin line) and the model of *Th.maritima* GAPDH (thick line).

subunits in *B.stearothermophilus* GAPDH that have disappeared in the *Th.maritima* enzyme (Table I).

In comparison with the *B.stearothermophilus* enzyme, the majority of the extra ion pairs (14 out of 18) found in our model are the consequence of mutations rather than local conformational changes.

Most of the extra ion pairs are located at or near the protein surface. Of the 18 potential extra intrasubunit ion pairs, 12 are 'short range', i.e. the two members of the ion pair are not more than 15 residues from each other in the sequence.

It is observed that some of the new ion pairs form clusters consisting of positive and negative charges that are arranged in a favourable configuration. Such clusters are formed, for example, around Arg102 and Arg266.

Estimation of the role of non-specific interactions. The plots of the free energy of unfolding as a function of temperature for our model, as well as GAPDHs from *B.stearothermophilus* and lobster predicted by the method of Oobatake and Ooi (see Materials and methods), are shown in Figure 7. It can be seen that the three curves are all similar to each other but are shifted relative to each other along the temperature axis, so that the positions of their maxima follow the order of thermostability of the three proteins. The curve for the *Th.maritima* enzyme model is shifted towards higher temperatures by ~9°C relative to the curve of *B.stearothermophilus* GAPDH. This shows that non-specific interactions are likely to be partly responsible for the increased thermostability of *Th.maritima* GAPDH. As can be seen from the diagram, the value of the free energy of unfolding is negative at any temperature for all three proteins, which means that they would not be stable at any temperature. We should not forget, however, that the unfolding thermodynamics of a tetrameric protein depend on protein concentration. This means that an entropy term should be added to the calculated unfolding free energies which would cause the free energy values to be more positive. The results may also indicate that specific interactions may play a more significant role in the stabilization of thermophilic proteins than in mesophilic proteins, because the parameters for the calculation were fitted using experimental data from mesophilic proteins.

Hydrogen bonds. The main-chain hydrogen bonding pattern will be discussed in a later section as a part of the secondary structure. We examined the side-chain–side-chain and side-chain–backbone hydrogen bonds, and compared the results for our model and the *B.stearothermophilus* GAPDH X-ray structures. Only a little rearrangement of the hydrogen bonding pattern is observed, with no significant change in the total number of hydrogen bonds. The ratio of short-range and long-range hydrogen bonds (depending on how far the two residues involved are from each other in the sequence) is nearly the same in the two molecules.

Secondary structure. If our model is compared with the *B.stearothermophilus* GAPDH X-ray structure, a slight change can be observed in the secondary structure. This is not surprising because some parts of the secondary structural elements are in SVRs. An increase in the number of residues in α-helices (83 versus 73) was observed. The new helical regions originated mainly from older 3_{10} helices. In addition, the number of residues in β-strands decreased (from 95 to 77). Most of these residues were transferred into parts categorized as 'disordered'. The number of residues in 3_{10} helices decreased dramatically (from 25 to six). The 3_{10} helices were transformed mainly into α-helices and β-turns.

Helix–dipole interactions. A change in helix–dipole interactions was observed in only two cases. Lys107 in *B.stearothermophilus* GAPDH, which is at the N-terminus of a helix and interacts unfavourably with the helix–dipole moment, is replaced by a leucine residue in the *Th.maritima* enzyme. Gln166 in the *B.stearothermophilus* enzyme, which is at the C-terminal end of a helix, is replaced by a lysine residue in the *Th.maritima* enzyme which could interact favourably with the helix–dipole moment.

α-Helical propensities. Seven common helical regions were found in *B.stearothermophilus* GAPDH and our model of *Th.maritima* GAPDH. An increase in α-helical propensity was observed in five helices and a decrease in the remaining two helices. However, the overall increase was very small (the total score was 66.6 for our model versus 65.4 for the *B.stearothermophilus* enzyme).

Discussion

Construction of the model

The construction of a homology model for one subunit of *Th.maritima* GAPDH was relatively straightforward because of the high degree of sequence and structural similarities of this enzyme to other known GAPDH structures. To ensure a correct refinement and to obtain a model suitable for analysis, the holoenzyme was assembled from four subunits and co-enzyme molecules, and water molecules were added. In this model, subunit interactions are expected to be treated correctly in all types of energy calculations. Even if we make use of the symmetry of the tetramer and simulate only a part of the whole system during refinement, the large size of the system (23 154 atoms) makes the refinement procedure time consuming and calls for compromises when deciding how to include solvent effects in the simulation and when adjusting the cut-off distance and the duration of the molecular dynamics simulation.

An important part of the applied refinement procedure is the molecular dynamics simulated annealing process. Simulated annealing is a very efficient optimization method, but molecular dynamics is definitely not the most effective way to perform

it. The main advantages of this method are its algorithmic simplicity and that it is free of bias (apart from a possible bias caused by the applied forcefield). We would like to call attention to the importance of carefully adjusting the parameters of the simulated annealing protocol (the cooling schedule). A basic requirement for a correct simulated annealing process is that thermodynamic equilibrium must be at least partially approximated at any temperature during the cooling process. Our preliminary molecular dynamics calculation (described in detail in Materials and methods), in which we performed a series of short molecular dynamics simulations with coupling to a heat bath, followed each time by another short simulation without temperature control, is suitable to determine the time required by the system to equilibrate (regarding the dynamic processes that have a relaxation time comparable with the time of simulation). The knowledge of this time value is essential to allow the determination of a correct cooling schedule.

Evaluation of the quality of the structure. The fact that no errors could be detected by the methods we applied to evaluate the quality of the model justifies the (experimentally also well supported) hypothesis that the structure of *Th.maritima* GAPDH shows a very high degree of similarity to GAPDHs from other sources. The increased thermostability is obviously the result of delicate changes at many locations in the structure of the enzyme.

Analysis of the model

Ion pairs. The most remarkable difference between our model structure for *Th.maritima* GAPDH and the known *B.stearothermophilus* GAPDH structure is the presence of 12 additional intrasubunit potential ion pairs per subunit in the hyperthermophilic enzyme. The correlation between thermal stability and the presence of surface ion pairs was first suggested by Perutz and Raidt (1975). In addition, the involvement of surface ion pairs in increasing thermostability was observed by Walker *et al.* (1980) in the case of hyperthermophilic GAPDH from *Thermus aquaticus*. Additional solvent-accessible salt bridges were also observed in the X-ray structure of malate dehydrogenase (an enzyme very similar to GAPDH) isolated from *Thermus flavus*, a moderate thermophile (Kelly *et al.*, 1993). Although recent experimental studies by site-directed mutagenesis or chemical synthesis have demonstrated that individual surface salt bridges generally have only marginal effects on protein stability (Dao-pin *et al.*, 1991; Sali *et al.*, 1991) because of the screening influence of the solvent and thermodynamic effects, a large number of salt bridges in a protein can clearly have a significant stabilizing effect. In some cases, even a single salt bridge can have a large stabilizing effect if other factors also favour the conformation in which the salt bridge is present. As the majority of the salt bridges observed in our model of *Th.maritima* GAPDH result from a change in sequence (relative to *B.stearothermophilus* GAPDH) and not the local conformation (in which case the imperfections of the structure prediction and possible refinement errors might throw doubt upon the results), it is very likely that these ion pairs contribute considerably to the enhanced stability of the enzyme. Our search procedure which we developed to find ion pairs was designed to reduce the possible artefacts that might arise from the fact that our energy minimization was performed in the absence of bulk solvent.

The formation of charge clusters, consisting of favourably arranged positive and negative charges, is an interesting phenomenon. This way of arranging charged residues on the surface of a protein may help to stabilize individual salt bridges that are part of these structures. Fersht and co-workers (Horovitz *et al.*, 1990) have analysed a tripartite salt bridge in barnase, in which Arg110 interacts with both Asp8 and Asp12. They have demonstrated by mutant thermodynamic cycles that the two salt bridges work cooperatively, i.e. each Asp–Arg interaction is strengthened by the presence of the other interaction. This cooperative mechanism may also work in *Th.maritima* GAPDH.

It is worth mentioning that the salt bridge between Arg194 and Asp293 in the *P* axis-related subunits, observed in *B.stearothermophilus* GAPDH by Walker *et al.* (1980), is also present in the *Th.maritima* enzyme. However, the double ionic bond between Arg281 from two *Q* axis-related subunits and Glu201 from the *R* axis- and *P* axis-related subunits is not found in *Th.maritima* GAPDH, although it is likely to contribute to the thermostability of the *B.stearothermophilus* enzyme (Walker *et al.*, 1980). Some other intersubunit ion pairs have also disappeared. This fact indicates that evolution may select different ways of producing thermostable variants of a given protein, and a stabilizing mutation present in one variant may not be found in another.

Non-specific interactions. It has been known for a long time that the core of proteins is well packed and hydrophobic interaction is important for the stabilization of proteins (Klapper, 1971; Chothia, 1974). A series of experimental studies has also justified the importance of good packing and good hydrophobic bonding in the protein interior (Erikson *et al.*, 1992, 1993; Lee, 1993).

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 of approximating 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 (Ooi *et al.*, 1987; Ooi and Oobatake, 1988, 1991; Oobatake and Ooi, 1992). The predicted curves of free energy change upon unfolding as a function of temperature (Figure 7) are shifted relative to each other along the temperature axis, and their order follows the order of thermostability. This observation suggests that non-specific interactions also contribute to the thermostabilization of *Th.maritima* GAPDH. This is in accordance with the results of spectroscopic measurements, which have proved that the aromatic side chains are more closely packed in their hydrophobic environments than in mesophilic GAPDHs, and the results of hydrogen–deuterium exchange, which have indicated that the enzyme has an increased structural rigidity (Wrba *et al.*, 1990).

Walker *et al.* (1980) emphasized the role of additional hydrophobic interactions between the S loops (residues 178–201) at the core of the tetramer in the stabilization of *B.stearothermophilus* GAPDH. In the *Th.maritima* enzyme, there is only one residue difference in the sequence of the S loop when compared with *B.stearothermophilus* GAPDH. Therefore this mechanism is very likely to work in the hyperthermophilic enzyme too.

Secondary structure. The observed differences in secondary structure between our model of *Th.maritima* GAPDH and the known structure of *B.stearothermophilus* GAPDH (increased α-helix content, decreased 3_{10} helix content and shorter β-

strands) can contribute to the stabilization of the hyperthermophilic enzyme. CD detects no difference in secondary structure (Wrba *et al.*, 1990), but this is consistent with our results because the difference predicted by us is small and the sensitivity of the CD measurement is limited; besides, the CD method is unable to distinguish between α-helices and 3_{10} helices.

The increase in α-helical propensities in thermostable proteins has been suggested by Olsen (1983) and Menéndez-Arias and Argos (1989). However, a large change in the length of α-helices and, consequently, α-helical content cannot be expected because this would cause a major change in the overall structure of the enzyme. The slight change predicted by us in the case of *Th.maritima* GAPDH may also be an artefact caused by inaccuracies of the forcefield and imperfect solvent modelling. These results could be justified only by the determination of the high-resolution X-ray structure. However, the contribution of this slight secondary structural change to thermostability should be small.

α-Helical propensities and helix–dipole interactions. An increase in the number of residues promoting the formation of α-helices has been observed previously in proteins isolated from thermophiles by comparing sequences of mesophilic and thermophilic proteins (Argos *et al.*, 1979). The stabilizing effect of increasing α-helical propensities in helical regions has also been observed in several experimental studies (O'Neil and DeGrado, 1990; Zhang *et al.*, 1991).

We have also found some increase in α-helical propensities when comparing the sequence of *Th.maritima* GAPDH with that of *B.stearothermophilus* GAPDH in helical regions. However, the overall effect is very small. Therefore, an increase in helix-forming tendencies cannot be a significant factor for extreme thermostability.

Helix–dipole interactions (Hol, 1985; Knowles, 1991) have been found to be stabilizing in a number of protein engineering studies (Nicholson *et al.*, 1988; Sali *et al.*, 1988; Serrano and Fersht, 1989; Eijsink *et al.*, 1992). When comparing the sequence of *Th.maritima* GAPDH with that of *B.stearothermophilus* GAPDH at the end of helical regions, we found very few mutations that could increase helix–dipole interactions. Therefore, helix–dipole interactions play an insignificant, if any, role in the extra thermostabilization of *Th.maritima* GAPDH.

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 certain cases (Jandu *et al.*, 1990; Scholtz *et al.*, 1991). When comparing side-chain–side-chain and side-chain–backbone hydrogen bonds in the model of *Th.maritima* GAPDH and in the known *B.stearothermophilus* GAPDH structure, we find that there is only minor rearrangement of the hydrogen bonds and the number of hydrogen bonds and their distribution according to length of the peptide chain between the two residues involved is almost the same. Therefore, no overall stabilizing effect can be attributed to the hydrogen bonds.

The surface loop region Ile18–Val28. In the loop region Ile18–Val28, there is a two-residue insertion in the *Th.maritima* GAPDH sequence compared with all homologues. The inserted residues are Glu20a and Arg20b. Looking at the conformation of this loop in our model, we can see that there is a significant change in comparison with the reference proteins used for the modelling. Due to this change, some side chains in this region have different conformations than in the homologous structures. Although the reliability of the model is low in this region, several interactions can be observed here that tend to increase the rigidity of the loop region (e.g. additional hydrogen bonds). The charges are arranged in a suitable configuration to form multiple ion pairs, although not all of them conform to the strict definition of an ion pair.

It is plausible to assume that this loop is important from the point of view of thermostability. Little is known about the role of loops in protein stability. Because they are usually variable within a protein family, one might assume that they have little significance with respect to stability. However, mutations in loop regions can slightly influence the stability of a protein (Hardy *et al.*, 1994). In isopropylmalate dehydrogenase, mutations in a loop region have also been shown to influence stability (T.Oshima, personal communication), which leads to the assumption that certain loops may behave as initiation sites for thermal unfolding. Therefore, residues in the loop Ile18–Val28 in *Th.maritima* GAPDH are good targets for site-directed mutagenesis studies.

Amino acid composition. Comparison of the sequence of *Th.maritima* GAPDH with sequences of GAPDHs from mesophilic sources has shown that the 'traffic rules' (amino acid exchanges earlier found to correlate with thermostability in statistical analyses; Argos *et al.*, 1979; Menéndez-Arias and Argos, 1989) can be applied only partially to the hyperthermophilic GAPDH from *Th.maritima* (Schultes *et al.*, 1990). The most significant exchange refers to Lys → Arg. Leu → Ile and Ser → Thr are also preferred.

When comparing the sequence of *Th.maritima* GAPDH with that of *B.stearothermophilus* GAPDH, these preferred exchanges cannot be observed. Apparently, any capacity to enhance thermostability by these preferred exchanges is exhausted in *B.stearothermophilus* GAPDH. To stabilize the molecule further, other mechanisms are required. In fact, the lysine content is somewhat increased (27 versus 23) and the arginine content decreased (14 versus 15) in the hyperthermophilic enzyme.

The amino acid composition of *Th.maritima* GAPDH is somewhat different from those of other thermophilic GAPDHs. The most surprising is its unusually low alanine content (25 versus 38–44 in other thermophilic GAPDHs), which is even lower than the alanine content of mesophilic GAPDHs, despite the fact that alanine has long been thought to be a preferred residue for the stabilization of α-helical regions in thermophilic proteins by increasing α-helical propensities. However, this effect may be partly compensated for by the increase in isoleucine content (30 versus 19–22 in other thermophilic GAPDHs), which has an even stronger helix-forming tendency.

Conclusions

In this work, we constructed a model for GAPDH from *Th.maritima* by homology modelling. To identify the structural features responsible for the extreme thermostability of the enzyme, we compared the model with the known structure of GAPDH from *B.stearothermophilus*, a moderate thermophile, the thermal stability of which was in part analysed previously (Walker *et al.*, 1980). The results of this comparison suggest how a moderately thermostable enzyme can be transformed into an extremely thermostable enzyme.

The appearance of numerous surface salt bridges (12 per subunit in the model) appears to be the most important factor

responsible for thermal stabilization. Non-specific interactions are also likely to be significant. Other effects seem to be negligible.

To justify our conclusions reached on the basis of the 3-D model, we plan to perform site-directed mutagenesis studies. The best targets for mutagenesis are the residues involved in salt bridges and residues in the Ile18–Val28 loop.

Acknowledgements

We would like to thank Dr Ferenc Vonderviszt for fruitful discussions on many questions concerning model construction and analysis. The comments of M.Oobatake on the prediction of unfolding thermodynamics are gratefully acknowledged.

Addendum

While this paper was under review, the crystal structure of *Th.maritima* GAPDH was published at 2.5 Å resolution (Korndörfer *et al.*, 1995). These authors also propose that extra ion pairs may be the most important thermostabilizing factor. A first comparison of the crystal structure with our homology model indicates that our prediction was successful. The r.m.s. deviation between the N-terminal domains is 1.40 and 2.21 Å for C_αs and all atoms, respectively. The corresponding r.m.s. deviations are 0.85 and 1.33 Å for the C-terminal domain. Eight extra ion pairs were predicted correctly. Five ion pairs were missed due to a rigid body rotation of 4.4° of the NAD binding domain relative to the catalytic domain, which could not be predicted. The changes in secondary structure follow the tendency predicted by us.

References

Argos,P., Rossmann,M.G., Grau,U.M., Zuber,H., Frank,G. and Tratschin,J.D. (1979) *Biochemistry*, **18**, 5698–5703.

Bachar,O., Fischer,D., Nussinov,R. and Wolfson,H. (1993) *Protein Engng*, **6**, 279–288.

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.

Chothia,C.H. (1974) *Nature*, **248**, 338–339.

Dao-pin,S., Sauer,U., Nicholson,H. and Matthews,B.W. (1991) *Biochemistry*, **30**, 7142–7153.

Eijsink,V.G.H., Vriend,G., Van den Burg,B., Van der Zee,J.R. and Venema,G. (1992) *Protein Engng*, **5**, 165–170.

Erikkson,A.E., Baase,W.A., Zhang,X.J., Heinz,D.W., Blaber,M., Baldwin,E.P. and Matthews,B.W. (1992) *Science*, **255**, 29–35.

Erikkson,A.E., Baase,W.A. and Matthews,B.W. (1993) *J. Mol. Biol.*, **229**, 747–769.

Greer,J. (1990) *Proteins: Struct. Funct. Genet.*, **7**, 317–334.

Hardy,F., Vriend,G., Van der Vinne,B., Frigerio,F., Grandi,G., Venema,G. and Eijsink,G.H. (1994) *Protein Engng*, **7**, 425–430.

Hensel,R., Laumann,S., Lang,J., Heumann,H. and Lottspeich,F. (1987) *Eur. J. Biochem.*, **170**, 325–333.

Hol,W.G.J. (1985) *Prog. Biophys. Mol. Biol.*, **45**, 149–195.

Horovitz,A.H., Serrano,L., Avron,B., Bycroft,M. and Fersht,A.R. (1990) *J. Mol. Biol.*, **216**, 1031–1044.

Jaenicke,R. (1981) *Annu. Rev. Biophys. Bioengng*, **10**, 1–67.

Jaenicke,R. (1991) *Eur. J. Biochem.*, **202**, 715–728.

Jaenicke,R. and Závodszky,P. (1990) *FEBS Lett.*, **268**, 344–349.

Jandu,S.K., Ray,S., Brooks,L. and Leatherbarrow,R.J. (1990) *Biochemistry*, **29**, 6264–6269.

Jones,T.A. and Thirup,S. (1986) *EMBO J.*, **5**, 819–822.

Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.

Kelly,C.A., Nishiyama,M., Ohnishi,Y., Beppu,T. and Birktoft,J.J. (1993) *Biochemistry*, **32**, 3913–3922.

Kirkpatrick,S., Gelatt,C.D.,Jr and Vecchi,M.P. (1982) *Science*, **220**, 671–680.

Klapper,M.H. (1971) *Biochim. Biophys. Acta*, **229**, 557–566.

Knowles,J.R. (1991) *Nature*, **350**, 121–124.

Korndörfer,I., Steipe,B., Huber,R., Tomschy,A. and Jaenicke,R. (1995) *J. Mol. Biol.*, **246**, 511–521.

Kraulis,P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.

Laarhoven,P.J.M. and Aarts,E.H.L. (1987) *Simulated Annealing: Theory and Applications*. D.Reidel Publishing Company, Dordrecht, The Netherlands.

Laughton,C.A. (1994) *Protein Engng*, **7**, 235–241.

Lee,B. (1993) *Protein Sci.*, **2**, 733–738.

Lüthy,R., Bowie,J.U. and Eisenberg,D. (1992) *Nature*, **356**, 83–85.

Menéndez-Arias,L. and Argos,P. (1989) *J. Mol. Biol.*, **206**, 397–406.

Moras,D., Olsen,K.W., Sabesan,M.N., Buehner,M., Ford,G.C. and Rossmann,M.G. (1975) *J. Biol. Chem.*, **250**, 9137–9162.

Needleman,S.B. and Wunsch,C.D. (1970) *J. Mol. Biol.*, **48**, 443–453.

Nicholson,H., Becktel,W.J. and Matthews,B.W. (1988) *Nature*, **336**, 651–656.

Olsen,K.W. (1983) *Int. J. Peptide Protein Res.*, **22**, 469–475.

O'Neil,K.T. and DeGrado,W.F. (1990) *Science*, **250**, 646–651.

Oobatake,M. and Ooi,T. (1992) *Prog. Biophys. Mol. Biol.*, **59**, 237–284.

Ooi,T. and Oobatake,M. (1988) *J. Biochem.*, **104**, 440–444.

Ooi,T. and Oobatake,M. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 2859–2863.

Ooi,T., Oobatake,M., Némethy,G. and Scheraga,H.A. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 3086–3090.

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.

Rehaber,V. and Jaenicke,R. (1992) *J. Biol. Chem.*, **267**, 10999–11006.

Rehaber,V. and Jaenicke,R. (1993) *FEBS Lett.*, **317**, 163–166.

Sali,D., Bycroft,M. and Fersht,A.R. (1988) *Nature*, **335**, 740–743.

Sali,D., Bycroft,M. and Fersht,A.R. (1991) *J. Mol. Biol.*, **220**, 779–788.

Scheraga,H.A. (1978) *Pure Appl. Chem.*, **50**, 315–324.

Scholtz,J.M., Marqusee,S., Baldwin,R., York,E.J., Stewart,J.M., Santoro,M. and Bolen,D.W. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 2854–2858.

Schultes,V., Deutzmann,R. and Jaenicke,R. (1990) *Eur. J. Biochem.*, **192**, 25–31.

Serrano,L. and Fersht,A.R. (1989) *Nature*, **342**, 296–299.

Skarzinsky,T., Moody,P.C.E. and Wonacott,A.J. (1987) *J. Mol. Biol.*, **193**, 171–187.

Walker,J.E., Wonacott,A.J. and Harris,J.I. (1980) *Eur. J. Biochem.*, **108**, 581–586.

Wrba,A., Schweiger,A., Schultes,V., Jaenicke,R. and Závodszky,P. (1990) *Biochemistry*, **29**, 7584–7592.

Zhang,X.J., Baase,W.A. and Matthews,B.W. (1991) *Biochemistry*, **30**, 2012–2017.

Received November 21, 1994; revised May 2, 1995; accepted May 29, 1995