Protein Engineering vol.9 no.3 pp.265–271, 1996

Analysis of protein conformational characteristics related to thermostability

Enrique Querol[1], Josep A.Perez-Pons[1] and Angel Mozo-Villarias[2,3]

[1]Institut de Biologia Fonamental and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, E-08193 Barcelona and [2]Departament de Ciències Mèdiques Bàsiques, Facultat de Medicina, Universitat de Lleida, E-25198 Lleida, Spain

[3]To whom correspondence should be addressed

The thermal stability of proteins was studied, 195 single amino acid residue replacements reported elsewhere being analysed for several protein conformational characteristics: type of residue replacement; conservative versus non-conservative substitution; replacement being in a homologous stretch of amino acid residues; change in hydrogen bond, van der Waals and secondary structure propensities; solvent-accessible versus inaccessible replacement; type of secondary structure involved in the substitution; the physico-chemical characteristics to which the thermostability enhancement can be attributed; and the relationship of the replacement site to the folding intermediates of the protein, when known. From the above analyses, some general rules arise which suggest where amino acid substitutions can be made to enhance protein thermostability: substitutions are conservative according to the Dayhoff matrix; mainly occur on conserved stretches of residues; preferentially occur on solvent-accessible residues; maintain or enhance the secondary structure propensity upon substitution; contribute to neutralize the dipole moment of the caps of helices and strands; and tend to increase the number of potential hydrogen bonding or van der Waals contacts or improve hydrophobic packing.

Keywords: biotechnology/protein/protein engineering/protein structure/protein thermostability

Introduction

The enhancement of protein thermostability by rational design is one of the important goals of protein engineering [see Gupta (1993) for a review]. Using thermostable enzymes in industrial applications offers the benefits of increased rates of reaction, higher substrate solubility, decreased media viscosity, longer enzyme shelf-lives at normal storage temperatures and lowered risk of microbial contamination when reactions are carried out at higher temperatures. However, although a number of successful examples of stabilization of proteins exist, the mechanism of their thermostabilization is far from being understood; general methods of increasing protein stability are lacking. Two general approaches have been followed to analyse the stability of proteins. First, the comparison between homologous proteins from thermophiles and mesophiles has provided some insight into the reasons why related proteins performing the same functions could have very different stability (Argos *et al.*, 1979; Imanaka *et al.*, 1986; Querol and Parrilla, 1987; Menendez-Arias and Argos, 1989; Mrabet *et al.*, 1992). Second, the analysis by protein engineering of the contribution of the different interactions that take place in a protein has resulted in some general rules about possible ways to increase the stability of a protein (Fersht and Serrano, 1993; Nosoh and Sekiguchi, 1993). There is also the possibility of combining both procedures (Serrano *et al.*, 1993).

The first approach has the limitation that it is often difficult to identify the important determinants involved in a specific case since, in general, their sequences have diverged significantly. It is important to examine proteins that are at the beginning of divergence, since their structures are still almost identical and the effect of single differences can be easily analysed. However, at present few native protein isoforms, thermostable and mesostable, can be compared in this manner, exceptions being tyrosinase isoenzymes from *Neurospora crassa* and some bacterial and blue–green algal ferredoxins (Perutz and Raidt, 1975; Ruegg *et al.*, 1982). Most protein isoforms present such a number of substitutions that their use is precluded in this approach. Following the second approach, protein engineering, several strategies have been proposed, for example: (i) reducing the difference in entropy between folded and unfolded proteins, which in practice means reducing the number of possible conformations in the unfolded state (Matthews *et al.*, 1987), (ii) stabilizing the dipoles of α-helices (Nicholson *et al.*, 1988), (iii) increasing the number of hydrophobic interactions and packing ratio in the interior core (Yutani *et al.*, 1987) and (iv) reducing the area of water-accessible hydrophobic surface (Wigley *et al.*, 1987). Nevertheless, simple patterns that characterize stabilizing interactions and quantitatively predict the effects of amino acid substitutions have eluded identification.

An additional problem is the fact that in interpreting the stabilization in terms of specified local interactions, one has to consider that the free energy of stabilization of globular proteins in solution represents a marginal difference in large numbers. Thus, important contributions coming from small rearrangements of side chains or structural elements may escape detection. The difference between mesophilic and thermophilic enzymes is in the $\Delta\Delta G$ range ~5–7 kcal/mol, which can be accomplished by a few hydrogen bonds or by two salt bridges inside the protein globule (Perutz, 1978; Fersht and Serrano, 1993). The marginal energy for stabilization of proteins is clearly a result of natural selection. There appears to be a selective pressure for the marginal stability, probably to facilitate such processes as polypeptide folding and the flexibility required by the 'native' conformation in the protein function. On the other hand, there is no selective pressure to make proteins stable at temperatures above those which they encounter *in vivo*.

There are cases in which the thermal damage on a protein can be unambiguously related to a type of amino acid residue by chemical reasons. For example, Klibanov and co-workers have studied the irreversible thermal inactivation of ribonuclease A and α-amylase. Outstanding mechanisms of irreversible

inactivation of these proteins are thiol-catalysed disulphide interchange, β-elimination of disulphides, deamidation of asparagine and glutamine, formation of incorrect scrambled structures, oxidation of cysteines and hydrolysis of peptide bonds at aspartic residues (Ahern and Klibanov, 1985; Zale and Klibanov, 1986; Volkin and Klibanov, 1987; Tomazic and Klibanov, 1988a).

Early attempts to predict the thermal stability of proteins yielded a small number of general suggestions for replacements, for example Lys → Arg, Asp → Glu and Gly → Ala, already recommended in the pioneering work of Argos *et al.* (1979). In a previous paper (Querol and Parrilla, 1987), we reported the comparison of some proteins presenting single or very few replacements related to thermostability and the secondary structure involved. The advent of site-directed mutagenesis methodology has yielded an increasing number of cases of thermostabilization, many of them quite empirical. In the present work, we started to create a database comprising 122 references containing 195 single amino acid residue replacements, in 164 different positions on protein sequences, which have been unambiguously related to thermostability enhancement. Several references reported studies made on the same mutants, therefore our analysis deals only with a non-overlapping subset of the database. Thereafter, to gain insight into those characteristics leading to thermostabilization, the following properties were analysed: type of residue replacement; conservative versus non conservative according to the Dayhoff matrix (Dayhoff, 1978); change in hydrogen bond tendency and secondary structure propensity; external versus internal replacement; type of secondary structure involved in the replacement; when determined, the physico-chemical characteristics to which the thermostability enhancement can be attribed; replacement inside highly-homologue stretches of amino acid residues (indeed, the specific replacement was not a conservative one); replacement in the active or substrate binding site; and the relationship of the replacement site to the folding intermediates of the protein, when known. From the above analyses, some general rules can be suggested. It is also worth mentioning that in the attempt to define such rules one faces the fact that they should be able to predict thermal sensitivity. However, in reporting their results, most authors usually do not describe cases of thermosensitivity since they are considered negative results in their search for more thermostable proteins. In order to understand the mechanisms of enhanced thermoresistence, it is reasonable to expect that mutations attempted on the same sites as those described in our database but leading to thermosensitivity should provide hints for further conformational exploration and thermodynamic analysis on such mechanisms. At any rate, the main purpose of this work was to set a database and general rules which could be completed or improved as new substitution data leading both to thermostability and thermosensitivity appear.

Materials and methods

Data from thermostable and mesostable isoproteins

The wild-type and thermostable mutant proteins so far analysed from the literature are the following: α-amylase; α_1-antitrypsin; calbindin; carboxyl esterase; chloramphenicol acetyltransferase; cytochrome *c*; ferredoxin; β-galactosidase; glucose dehydrogenase; glyceraldehyde phosphate dehydrogenase; 3-isopropylmalate dehydrogenase; kanamycin nucleotidyltransferase; lactate dehydrogenase; T4, hen egg-white, bird and human lysozymes; luciferase; malate dehydrogenase; neutral protease; oligo-1,6-glucosidase; protein HU; λ cro and arc repressors; ribonucleases A, T1 and barnase; RNA polymerase subunit; subtilisins; superoxide dismutase; triosephosphate dehydrogenase; tryptophan synthase; tyrosinase; xylanase; and xylose isomerase (Perutz and Raidt, 1975; Yutani *et al.*, 1977, 1982, 1987; Perutz, 1978; Welpy *et al.*, 1981; Ruegg *et al.*, 1982; Hecht *et al.*, 1984; Imanaka *et al.*, 1986; Ahern *et al.*, 1987; Matthews *et al.*, 1987; Perry and Wetzel, 1987; Wigley *et al.*, 1987; Matsumura *et al.*, 1988a,b; Nicholson *et al.*, 1988, 1991; Stearman *et al.*, 1988; Tomazic and Klivanov, 1988b; Wend *et al.*, 1988; Das *et al.*, 1989; Makino *et al.*, 1989; Nagao *et al.*, 1989; Pakula and Sauer, 1989, 1990; Pantoliano *et al.*, 1989; Serrano and Fersht, 1989; Shirley *et al.*, 1989; Suzuki *et al.*, 1989; Takagi and Imanaka, 1989; Zagorski *et al.*, 1989; Dao-Pin *et al.*,1990, 1991; Declerck *et al.*, 1990; Eijsink *et al.*, 1990, 1992a,b; Erwin *et al.*, 1990; Igarashi *et al.*, 1990; Wilson *et al.*, 1990, 1992; Eijsink *et al.*, 1991; Hallewell *et al.*, 1991; Hickey *et al.*, 1991; Lim and Sauer, 1991; Narhi *et al.*, 1991; Quax *et al.*, 1991; Vriend *et al.*, 1991; Zhang *et al.*, 1991; Heinz *et al.*, 1992; Joyet *et al.*, 1992; Kallwass *et al.*, 1992; Lim *et al.*, 1992; Margarit *et al.*, 1992; Mrabet *et al.*, 1992; Arase *et al.*, 1993; Haezebrouck *et al.*, 1993; Hardy *et al.*, 1993; Ishikawa *et al.*, 1993; Kajiyama and Nakano, 1993; Kotik and Zuber, 1993; Meng *et al.*, 1993; Pjura *et al.*, 1993; Schuman *et al.*, 1993; Serrano *et al.*, 1993; Stonehouse and Stockley, 1993; Amaki *et al.*, 1994; Goward *et al.*, 1994; Masul *et al.*, 1994; Heringa *et al.*, 1995).

Most of the thermostable forms have been obtained by site-directed mutagenesis; few of them emerge from the analysis of native isoproteins, in which the mesostable and thermostable forms differ only in one amino acid residue. It deserves to be mentioned that in the present analysis we have considered both thermostability and thermoresistence, according to the data from the above-reported references. To consider a mutant as enhancing thermostability, an increase of $\geq 1°C$ in T_m was considered as the threshold.

Computer graphics and predictive structural analyses from the amino acid sequence

When the structural characteristics hereafter analysed were not reported in the references, they were visualized in the PDB (Protein Data Bank). Molecular graphics were performed on a Crimson Elan, from Silicon Graphics. The structures of the proteins were visualized and analysed by means of the TURBO FRODO program (Roussel *et al.*, 1994). Multiple sequence alignments and comparisons were performed by means of the Clustal method (Higgins and Sharp, 1989) from the LaserGene package (DNASTAR) or GCG software package (Genetics Computer Group, University of Wisconsin, Madison, WI). When no X-ray or 2-D NMR structural data were available, protein structural characteristics (secondary structure, hydropathic profiles, structural class of the protein, etc.) were predicted by means of computer programs reported previously (Clotet *et al.*, 1994). Nevertheless, as most of the proteins analysed correspond to molecules whose structures are known by crystallography, or have been modelled on a homologue, the structural characteristics have only been predicted for five proteins: carboxyl esterase, isopropyl malate dehydrogenase, kanamycin nucleotidyl transferase, luciferase and tyrosinase, with 10 thermostable mutants so far analysed.

In the analysis the following criteria were undertaken. For the conservative character of the replacement the PAM250

Weight Table was utilized (Dayhoff, 1978). Those maintaining the positive value of the PAM250 table were considered conservative replacements, because replacements involving difference values, say from 5 to 3, could not be considered as distinct; all of them are positive and commonly found in nature. In the case of secondary structure propensity, the table of conformational parameters of Chou and Fasman was utilized (Privilege and Fasman, 1989), considering as propensity enhancement (or decrease) those replacements showing propensity differences ≥ 0.3. Values below that figure were considered as maintaining the propensity. Still, those replacements maintaining and enhancing the propensity were pooled in the table, since it is unlikely both cases would perturb the secondary structure.

The solvent accessibility and secondary structure are the most explicitly indicated data by authors in the references. When not explicitly reported, solvent accessibility was obtained from Connolly surfaces (rolling sphere of 1.4 Å). For those five proteins lacking crystallographic structure, the accessibility was predicted by the algorithms of Emini *et al.* (1985) and Kyte and Doolittle (1982) by means of the LaserGene computer package (DNASTAR). Most data on the flexibility or rigidity of the stretch of residues where the mutant is come from authors' references, whenever they reported them, or from *B* factors in some cases. The fact that many proteins are not solved at better than 2 Å precludes obtaining suitable factors in many cases.

Results

It should be noted in what follows no formal statistical conclusions are meant to be drawn from the reported data but indications about the tendencies found in the references. The reported references related to enhanced protein thermostabilization contain 195 one-residue replacements at 164 different positions on the protein sequences. In this analysis those references reporting substitutions leading to additional or new disulphide bridges have not been taken into account, although some of them are included in the reference list. This is a strategy that although successful for stabilization in some cases (Perry and Wetzel, 1984; Pantoliano *et al.*, 1987; Matsumura *et al.*, 1989a,b; Nishikawa *et al.*, 1990; Takagi *et al.*, 1990; Clarke and Fersht, 1993; Wakarchuk *et al.*, 1994), may lead to folding problems or yields a non-functional protein.

Most of the replacements analysed herein are represented one to two times, a few of them several times, for example those reported in the early work of Argos and co-workers (Argos *et al.*, 1979) and subsequently by many authors as leading to enhanced stability, i.e. Lys → Arg (10 times) or Gly → Ala (eight times). As is now widely assumed, there are no privileged residue replacements leading to thermostabilization. From the above 195 substitutions, 121 represent a different type of residue replacement. Since the whole space of directional changes has 380 substitutions (20×19), a large set of the potential substitution space has been explored by nature or by protein engineers. As could be expected, most replacements, 146 out of 195, are conservative according to the Dayhoff matrix of replacements (Dayhoff, 1978) (Table I). As in the type of replacement, a large set of the Dayhoff space of conservative non-identical replacements has been explored.

The tendency to conserve the propensity for the secondary structure could be expected (Table I). Upon checking, it was found that 162 out of 195 replacements lead to maintenance of (82 out of 195) or an increase in (80/195) this propensity, analysed according to a classical scale of secondary structure propensities (Privilege and Fasman, 1989). These findings agree with results from Mark and van Gunsteren, (1992) for HEW lysozyme and Caflisch and Karplus (1994) for barnase, which simulated the thermal denaturation of both proteins by molecular dynamics. These authors reported that the protein unfolds with partial destruction of regular secondary structures, with water molecules replacing the hydrogen bonds of secondary structural elements.

Considering the specific amino acid replacements independently of their sequential and structural context, most of the replacements, 121 out of 195, lead to a residue with the same (66/195) or enhanced (55/195) potential for hydrogen bonding lying on its side chain (Table I).

The type of secondary structures involved in the replacements are shown in Table II. As can be seen, most of them lie in regular secondary structures, 66 out of 164 positions in α-helices and 33 out of 164 in β-strand. Some substitutions in the N- and C-caps of α-helices have been largely reported to lead to enhanced thermostability. There are 16 substitutions in the N-cap and six in the C-cap of helices, and six in the N-cap and three in the C-cap of strands. There are 28 out of 164 substitutions on turns and 36 out of 164 on loop/coil regions. These results have to be treated with caution as they probably reflect a biased decision of the researchers. The lack of strong rules for thermostabilization makes some empirical rules useful. For example, for a long time there has been a general consensus that compensation of the dipole of N- and C-caps of α-helices can lead to stabilization (Nicholson *et al.*, 1988). Not surprisingly, most substitutions are designed on the cap. However, it is remarkable that most replacements lie on regular secondary structure-conserved stretches, not in loops, which have usually been assumed by protein engineers to be regions that accept mutations without major structural perturbations. This would reflect the widely assumed consideration that protein structures have not been fully optimized by evolution and maintain to some extent the potential to improve their characteristics.

The putative amphipathicity of the regular secondary structures in which the mutants lie has also been checked. Of the 109 point mutations, 23 lie on different amphipathic α-helix or β-strand secondary structures (19 on helices and four on strand). However, this result may also reflect the biased selection of researchers. We have checked whether the internal

Table I. Some propensities related to thermostable replacements

Conservative (Dayhoff)		Secondary structure propensity		Potential for hydrogen bonding	
Conserved	Non-conserved	Maintain or increase	Decrease	Maintain or increase	Decrease
146/195	49/195	162/195	33/195	121/195	74/195

replacements would primarily correspond to internal residues from amphipathic structures and found the following: 15 external, 16 internal amphipathic-helix residues, and four external, two internal amphipathic-strand residues, results that do not suggest any special tendency.

All the above results could be expected from the point of view of the conservation of protein structure and function. However, there are striking results more difficult to explain. For example, considering the rate of external versus internal replacements; 121 out of 164 substitutions lie on solvent-accessible regions whilst 43 out of 164 are in buried regions (Table III). This result confirms our previously reported analysis on a small set of proteins (Querol and Parrilla, 1987).

Another interesting result comes from the homology/similarity analyses of the amino acid sequence stretches in which the mutated residue lies (we do not refer to the conservation of the specific replaced residue, which has been analysed above). This was done with the Clustal algorithm (Higgins and Sharp, 1989) on those proteins having a number of sequences to allow this analysis (in fact, it was the whole set of proteins used in this work except lysozymes and repressors); 113 regions in total (see Table IV). Most of the replacements match conserved stretches of residues, especially those lying on a regular secondary structure. It is well known that regular secondary structure stretches are highly conserved in protein sequences, but to our surprise most substitutions lying on loops, turns and coil regions correspond to conserved regions too. Thermostability is certainly related to protein structure and conserved external regions are indeed important for structure.

It is usually assumed that replacements leading to enhanced protein thermostability correspond to rigid regions. We checked this for those cases in which these data (*B*-factors) are either reported by the authors or can be obtained from the Protein Data Bank (Table V). From the set of mutant positions we obtained the data for 75 substitutions. Of these, 37 replacements match regions having flexibility over the average and 38 regions lower than the average. Although these data are not exhaustive for the whole data set of 164 sequence positions, they suggest that one should be cautious in considering increasing rigidity as a sure predictor for thermostablity enhancement.

A survey of the physico-chemical basis of the thermostability enhancement, when indicated by the authors, shows that they are diverse. This list is biased by the fact that site-directed mutants are chosen by researchers to accomplish, as far as possible, some physico-chemical expectations. Table VI summarizes these results. According to the reported record (135 out 195; 60 mutations did not show available data), the outstanding causes of stabilization can be grouped and described as follows: improved hydrogen bonding (18); optimized hydrophobic packing (16); enhanced secondary structure propensity (12); helix dipole stabilization (10); replacements according to Argos (Menendez-Arias and Argos, 1989) (10); replace free thiols sensitive to oxidation (8); substitute hydrophobic residues on solvent-accessible areas (7); improved electrostatic interactions (6); strengthening intersubunit association (6); and others unspecified, leading to improvement of ΔG (20). Conversely, some reports describe loss of thermostability upon losing electrostatic interactions (Sali *et al.*, 1988; Serrano *et al.*, 1993), loss of hydrophobic contacts (Pakula and Sauer, 1990; Heringa *et al.*, 1995), loss of hydrogen bonding (Narhi *et al.*, 1991) and loss of helicity (Matthews *et al.*, 1987; Sali *et al.*, 1988; Dao-Pin *et al.*, 1990; Blaber *et al.*, 1993)

Finally, up to now the analysis has considered all the substitutions that report some degree of enhancement of thermostability. However, a close inspection of the most striking examples of thermostability enhancement (Perutz and Raidt, 1975; Hecht *et al.*, 1984; Matsumura *et al.*, 1988a; Stearman *et al.*, 1988; Makino *et al.*, 1989; Pakula *et al.*, 1989, 1990; Zagorski *et al.*, 1989; Wilson *et al.*, 1990; Hickey *et al.*, 1991; Narhi *et al.*, 1991; Joyet *et al.*, 1992; Lim *et al.*, 1992; Hardy *et al.*, 1993; Kotik and Zuber, 1993) deserves to be

Table II. Secondary structure involved

Structure	Total	On the N-cap	On the C-cap
α-Helix	66	16	6
β-Strand	33	6	3
Turn	28	–	–
Loop	20	–	–
Coil	16	–	–

Table III. Solvent accessibility

Accessible	Inaccessible
121/164	43/164

Table IV. Replacement on conserved stretches of multi-aligned homologous sequences

Region	On regular secondary structure	On loops, turns or coil
Highly conserved region	46	36
Medium conserved region	6	10
Non-conserved	7	8

Table V. Replacement in flexible/rigid regions

Flexible	37
Rigid	38
Unknown	113

Table VI. Physical and chemical contributions to ΔG to which thermostability enhancement can be attributed

Type	Contribution
Better hydrogen bonding	18
Better hydrophobic internal packing	16
Enhanced secondary structure propensity	12
Helix dipole stabilization	10
Argos' replacements	10
Removal of residues sensitive to oxidation or deamidation	10
Burying hydrophobic accessible area	7
Improved electrostatic interactions	6
Strengthening intersubunit association	6
Decrease chain strain	5
Salt bridge optimization	4
Better van der Waals	3
Better affinity for calcium	2
Improved ΔH upon substitution	1
Unspecified improved ΔG upon substitution	25
Unknown or not described	60

mentioned. We have defined this subset as those replacements leading to a thermostability enhancement of better than 5°C or, alternatively, those that duplicate the thermoresistance of the protein. There are 18 replacements in this subset. It is remarkable that only two of them involve the creation of a new disulphide bridge (Matsumura *et al.*, 1988a; Stearman *et al.*, 1988). The analysis of the other substitutions does not indicate strong rules. There are 12 replacements in α-helices (five of them in the N-cap); four in β-strands (one in the N-cap); three in turns and one in loop. In addition, 15 out of 18 substitutions lie on solvent-accessible regions; 12 out of 18 correspond to conserved substitutions (according to the Dayhoff matrix). There is an outstanding improvement in secondary structure propensity with respect to the whole set of 195 substitutions: 10 out of 18 enhance propensity; six maintain and two decrease. From this subset, only six cases have reported the local flexibility/rigidity, four being flexible and two rigid. In eight cases the authors reported the physico-chemical basis of the enhancement: two designed disulphide bridges, one stronger salt bridge, three increased secondary structure propensity, one improved subunit interaction and one due to entropic factors. There are instances reported in which mutations leading to increased configurational entropy lead to thermosensitivity (Alber *et al.*, 1987; Matthews *et al.*, 1987; Dao-Pin *et al.*, 1990; Blaber *et al.*, 1993).

Discussion

One of the central questions in protein evolution is how the changes in the amino acid sequence of a protein result in changes in stability and structure and, therefore, in its function. It is generally assumed that small changes in the sequence of a protein should not have large effects on its structure or stability and that changes on the surface of a protein should have a smaller impact than those in buried regions. Consequently, globular proteins should keep more (though marginal) capacity to evolve by exploring the conformational space on the external area than in the core. Moreover, it is known that groups that are constrained by the folded conformation to interact productively can make much stronger non-covalent bonds than groups with greater motional freedom. Therefore, mutations within the core of a protein are likely to affect more than just one energy term, leading the protein molecule to prefer those being conformationally and functionally less expensive, which usually should correspond to those at the solvent–protein interface. However, an accurate thermodynamic prediction using present-day theories is difficult, especially a prediction based on thermodynamic scales. For example, from statistical mechanical considerations Pickett and Sternberg (1993) reported a scale of relative conformational entropies for each amino acid, taking glycine as a reference. These conformational entropy values are combined with values of hydrophobic tendencies in scales obtained either by Fauchère and Pliska (1983) or by Sharp *et al.* (1991), in order to obtain values of ΔG for each amino acid. This free energy variation, in principle, should express the relative tendency of each amino acid to be transferred to the interior of a protein and consequently reflect stability properties. A calculation of $\Delta\Delta G = \Delta G_{mut} - \Delta G_{nat}$, (when a given amino acid of free energy ΔG_{nat} is exchanged for another amino acid of free energy ΔG_{mut}) done on this basis for each of the transformations described in the reported mutagenesis database yielded values and, most important, signs of such values that indicate no correlation with the expected tendency deduced from the mutagenesis database (not shown). Moreover, most of the transitions which should show stronger changes in their tendencies (for example, Arg → Trp) appear in the database with a behaviour opposed to that expected from calculus based on these scales. This result is hardly surprising, since from the strict thermodynamic point of view it is not possible to consider for a given component of a system an entropy or a free energy contribution, nor can the entropy or free energy of the whole system be considered the sum of such contributions (Mark and van Gunsteren, 1994).

It has been reported that thermostabilization is mainly lead by hydrophobic interactions (Yutani *et al.*, 1977, 1987) and, conversely, loss of hydrophobic interactions result in thermosensitivity (Pakula and Sauer, 1990; Heringa *et al.*, 1995). There are, however, several reasons to question such a simplistic explanation. For example, there is no direct correlation between protein volume and thermal stability. Still, there are cases in which the stability of a protein was lower than predicted by merely considering the hydrophobicity of the substituting residues (Yutani *et al.*, 1987; Matsumura *et al.*, 1988a). Many local interactions and compensatory structural rearrangements other than the hydrophobic effect need to be considered: hydrogen bonding, entropy, torsional strain and van der Waals interactions. In fact, our finding that most stabilizing substitutions correspond to external regions of the protein may, in some cases, reflect a chemical justification: some thermal inactivation processes such as oxidative damage of free cysteines and sensitive methionines, hydrolysis of peptide bonds at aspartic residues and deamidation of asparagine and glutamine residues, are more likely to occur on solvent-accessible areas of the protein [see Ahern and Klibanov (1985) and Tomazic and Klibanov (1988a,b) for the analysis of thermal inactivation of some proteins]. Recently, Heringa *et al.* (1995) have identified strong clusters from the 3-D structure of protein families and suggested that they are involved in stability and folding. From their analysis five mutants with enhanced thermal stability can be found. It is remarkable that they all lie on solvent-accessible regions.

It is well known that α-helices have a dipole moment (equivalent to half a positive unit charge at the N-terminus) estimated to be approximately 3.5 D per peptide unit (Hol *et al.*, 1981). Although lower than for helices, β-sheets also have a dipole moment (approximately 0.5 D), the N-terminal end of the structures corresponding to the positive end of the dipole. Many substitutions in α-helices have been designed to stabilize the dipole moment, introducing a negative charge in the N-cap, and to a lesser extent they introduce a positive charge in the C-cap. Although not always explicitly detailed by the authors, an inspection of the reported cases shows that eight of the replacements in β-strands would accomplish enhanced thermal stability upon stabilizing their strand dipoles. From these considerations on secondary structure, it is not surprising that a decrease in helicity results in a reduced thermostability (Alber *et al.*, 1987; Matthews *et al.*, 1987; Dao-Pin *et al.*, 1990; Blaber *et al.*, 1993). Mention should be made of the molecular dynamics simulation results of Mark and van Gunsteren (1992), who found that thermal denaturation of proteins starts by the loss of interactions $i \rightarrow i + 4$ in their ends.

The possibility that the changes in thermal stability could be related to folding deserves to be explored. As reported by Caflisch and Karplus (1994), the effect of temperature involves structural rearrangements related to unfolding. Could the thermostability problem be considered closely related to the

folding of proteins? May thermostable substitutions yield hints on sites related to folding intermediates? The small set of proteins for which intermediates of folding have been studied precludes the possibility of definite answers. We have checked the reported data on recent folding studies of some of the proteins studied herein, specifically: T4 lysozyme, HEW lysozyme, barnase, dihydropholate reductase, cytochrome *c* and lactate dehydrogenase. Results from amide proton exchange experiments obtained from the above proteins yield 50 out of 73 replacements on sequence regions that are protected from exchange (most of the cases come from the widely analysed T4 lysozyme). It is remarkable that 35 of them correspond to solvent-accessible positions. It is likely that mutations which predominantly affect the unfolded state will tend to occur on the surface of the folded protein, since mutations in the core will be likely to affect intramolecular interactions in the folded state.

There are two general mechanisms by which amino acid substitutions can decrease the rate of thermal inactivation: (i) by increasing conformational stability in such a way that the unfolded fraction at the temperature of exposure is reduced and (ii) by increasing the reversibility of unfolding (even at the cost of a significant reduction in conformational stability). It should be noted that besides the tendency to describe changes in stability in terms of the folded structure of a protein, it is always possible that such changes might be due to alterations in the unfolded state. Mutations in general can alter the free energies of both the folded and unfolded states of the protein. However, as has been mentioned above, attempts to make quantitative predictions of thermal properties upon point substitutions based on thermodynamic scales lack the appropriate theory or computer tools to make such unambiguous calculations.

The seemingly puzzling results coming from our analysis or those reported by others (Gupta, 1993) could lead to the conclusion that no unambiguous rules can yet be established to explain or predict how substituted amino acids modify the thermostability of a protein. However, from the above results, some general principles can be stated to explore the possibility of improving the thermal stability of a protein even without any prior knowledge about its structure. Since rules that at least were able to identify those regions which seem to be of particular importance for temperature adaptation would be welcome, we suggest the following to enhance protein thermostability.

(i) Look for conserved stretches of residues upon multiple alignment.
(ii) Make substitutions conservative according to the Dayhoff matrix.
(iii) Maintain or enhance the secondary structure propensity upon substitution.
(iv) Replace preferentially solvent accessible residues.
(v) Replace residues on N- and C-cap of α-helices (PDB or predicted) by introducing, if possible, a negative charge at the N-cap or a positive charge at the C-cap of the helix.
(vi) Replace residues on the N-cap and C-cap of β-strands introducing, if possible, a negative charge at the N-cap or a positive charge at the C-cap of the strand.
(vii) Substitute with residues which improve or increase the number of potential hydrogen bonding or van der Waals contacts.
(viii) Whenever possible consider thermal damage on specific amino acid residues: deamidation of Asn and Gln [the sequential motif Gly–Asn is a well known deamidation target (Tyler-Cross and Schirch, 1991)]; oxidation of methionines; splicing the peptidic bond besides aspartic residues and substitution of free thiol cysteines.

We also consider some suggestions elsewhere reported, e.g.:

(i) Argos' replacements (Lys → Arg, Asp → Glu, Gly → Ala) (Argos *et al.*, 1979).
(ii) Substitute with prolines in loops (Matthews *et al.*, 1987; Hardy *et al.*, 1993; Masul *et al.*, 1994).
(iii) Introduce an additional disulphide bridge (Perry and Wetzel, 1984; Pantoliano *et al.*, 1987; Matsumura *et al.*, 1989a,b; Nishikawa *et al.*, 1990; Takagi *et al.*, 1990; Clarke and Fersht, 1993; Wakarchuk *et al.*, 1994).
(iv) Introduce an additional metal binding site (Toma *et al.*, 1991; Haezebrouck *et al.*, 1993).
(v) Introduce new glycosylation sites (Olsen and Thomsen, 1991).

Obviously, the best strategy for increasing the stability of a specific protein by site-directed mutagenesis would require a thorough knowledge of its 3-D structure and of the molecular mechanisms responsible for its heat inactivation. There is still a limited number of proteins that fulfil these requirements. If the protein 3-D structure has not been determined, there are algorithms, many of them implemented as commercial software packages such as those mentioned in the Materials and methods section, to predict sufficient structural characteristics in order to apply the preceding general rules. In the future there is likely to be the possibility of simulating and predicting by molecular dynamics the effect of a site-directed mutation, especially the entropic effects. Some attempts have been made in this direction (Tapia *et al.*, 1991; Mark and van Gunsteren, 1992; Caflisch and Karplus, 1994).

Acknowledgements

This research was supported by grants BIO94–0912-CO2–01 and IN94–0347 from the CICYT (Ministerio de Educación y Ciencia, Spain) and by the Centre de Referència de R&D de Biotecnología de la Generalitat de Catalunya. The authors are grateful for technical assistance to Dr B. Oliva

References

Ahern,T.J. and Klibanov,A.M. (1985) *Science*, **228**, 1280–1284.
Ahern,T.J., Casal,J., Petsko,G.A. and Klivanov,A.M. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 675–679.
Alber, T., Dao-Pin,S., Nye,J.A., Muchmore,D.C. and Matthews,B.W. (1987) *Biochemistry*, **26**, 3754–3758.
Amaki,Y., Nakano,H. and Yamane,T. (1994) *Appl. Microbiol. Biotechnol.*, **40**, 664–668.
Arase,A., Yomo,T., Urabe,I., Hata,Y., Katsube,Y. and Okada,H. (1993) *FEBS Lett.*, **316**, 123–127.
Argos,P., Rossmann,M.G., Grau,U.M., Zuber,H., Frank,G. and Tratschin,J.D. (1979) *Biochemistry*, **18**, 5698–5703.
Blaber, M., Lindstrom,J.D., Gassner,N., Xu,J., Heinz,D.W. and Matthews,B.W. (1993) *Biochemistry*, **32**, 11363–11373.
Caflisch,A and Karplus,M. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 1746–1750.
Clarke,J. and Fersht,A.R. (1993) *Biochemistry*, **32**, 4322–4329.
Clotet,J., Cedano,J. and Querol,E. (1994) *Comp. Appl. Biosci.*, **10**, 495–500.
Dao-Pin,S., Baase,W.A. and Matthews,B.W. (1990). *Proteins: Struct. Funct. Genet.*, **7**, 198–204.
Dao-Pin,S., Söderlind,E., Baase,W.A., Wozniak,J.A., Sauer,U. and Matthews, B.W. (1991) *J. Mol. Biol.*, **221**, 873–887.
Das,G., Hickey,D.R., McLendon,D., McLendon,G. and Sherman,F. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 496–499.
Dayhoff, M. O. (ed.) (1978) *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Washington, DC, Vol. 5, Suppl. 3.
Declerck,N., Joyet,Ph., Gaillardin,C. and Masson,J.-M. (1990) *J. Biol. Chem.*, **265**, 15481–15488.
Eijsink,V.G.H., Vriend,G., van der Burg,B., Venema,G. and Stulp,B.K. (1990)

Protein Engng, **4**, 99–104.
Eijsink,V.G.H., van der Zee,J.R., van der Burg,B., Vriend,G. and Venema,G. (1991) *FEBS Lett.*, **282**, 13–16.
Eijsink,V.G.H., Vriend,G., van der Burg,B., van der Zee,J.R. and Venema,G. (1992a) *Protein Engng*, **5**, 165–170.
Eijsink,V.G.H., Vriend,G., van der Zee,J.R., van der Burg,B. and Venema,G. (1992b) *Biochem. J.*, **285**, 625–628.
Emini,E.A., Hughes,J., Perlow,D. and Boger,J. (1985) *J. Virol.*, **55**, 836–839.
Erwin,C.R., Barnett,B.L., Oliver,J.D. and Sullivan,J.F. (1990) *Protein Engng*, **4**, 87–97.
Fauchère,J.-L. and Pliska,V. (1983) *Eur. J. Med. Chem.-Chim. Ther.*, **18**, 369–375.
Fersht,A.R. and Serrano,L. (1993) *Curr. Opin. Struct. Biol.*, **3**, 75–83.
Goward,C.R., Miller,J., Nicholls,D.J., Irons,L.I., Scawen,M.D., O'Brien,R. and Chowdry,B.Z. (1994) *Eur. J. Biochem.*, **224**, 249–255.
Gupta,M.N. (ed.) (1993) *Thermostability of Enzymes*. Springer, Berlin.
Haezebrouck,P., De Baetselier,A., Joniau,M., Van Dael,H., Rosenberg,S. and Hanssens,I. (1993) *Protein Engng*, **6**, 643–649.
Hallewell,R.A. *et al.* (1991) *Biochem. Biophys. Res. Commun.*, **181**, 474–480.
Hardy,F., Vriend,G., Veltman,O.R., van der Vinne,B., Venema,G. and Eijsink,V.G.H. (1993) *FEBS Lett.*, **317**, 89–92.
Hecht,M.H., Sturtevant,J.M. and Sauer,R.T. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 5685–5689.
Heinz,D.W., Baase,W.A. and Matthews,B.W. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 3751–3755.
Heringa,J., Argos,P., Egmond,M.R. and de Vlieg,J. (1995) *Protein Engng*, **8**, 21–30.
Hickey,D.R. *et al.* (1991) *J. Biol. Chem.*, **266**, 11686–11694.
Higgins,D.G. and Sharp,P.M. (1989) *Comp. Appl. Biosci.*, **5**, 151–153.
Hol,W.G.J., Halie,L.M. and Sander,C. (1981) *Nature*, **294**, 532–536.
Igarashi,K., Fujita,N. and Ishihama,A. (1990) *Nucleic Acids Res.*, **18**, 5945–5948.
Imanaka,T., Shibazaki,M. and Takagi,M. (1986) *Nature*, **324**, 695–697.
Ishikawa,K., Kimura,S., Kanaya,S., Morikawa,K. and Nakamura,H. (1993) *Protein Engng*, **6**, 85–91.
Joyet,Ph., Declerck,N. and Gaillardin,C. (1992) *Biotechnology*, **10**, 1579–1583.
Kajiyama,N. and Nakano,E. (1993) *Biochemistry*, **32**, 13795–13799.
Kallwass, H.K.W., Surewicz,W.K., Parris,W., Macfarlane,E.L.A., Luyten, M.A., Kay,C.M., Gold,M. and Jones,J.B. (1992) *Protein Engng*, **5**, 769–774.
Kotik,M. and Zuber,H. (1993) *Eur. J. Biochem.*, **211**, 267–280.
Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.
Lim,W.A. and Sauer,R.T (1991) *J. Mol. Biol.*, **219**, 359–376.
Lim,W.K., Brouillette,C. and Hardman,J.K. (1992) *Arch. Biochem. Biophys.*, **292**, 34–41.
Makino,Y., Negoro,S., Urabe,I. and Okada,H. (1989) *J. Biol. Chem.*, **264**, 6381–6385.
Margarit,I., Campagnoli,S., Frigerio,F., Grandi,G., De Filippis,V. and Fontana,A. (1992) *Protein Engng*, **5**, 543–550.
Mark,A.E. and van Gunsteren,W.F. (1992) *Biochemistry*, **31**, 7745–7748.
Mark,A. and van Gunsteren,W.F. (1994) *J. Mol. Biol.*, **240**, 167–176.
Masui,A., Fujiwara,N. and Imanaka,T. (1994) *Appl. Environ. Microbiol.*, **60**, 3579–3584.
Matsumura,M., Becktel,W.J. and Matthews,B.W. (1988a) *Nature*, **334**, 406–410.
Matsumura,M., Yahanda,S., Yasumura,S., Yutani,K. and Aiba,S. (1988b) *Eur. J. Biochem.*, **171**, 715–720.
Matsumura,M., Signor,G. and Matthews,B.W. (1989a) *Nature*, **342**, 291–293.
Matsumura,M., Becktel,W.J., Lewitt,M. and Matthews,B.W. (1989b) *Proc. Natl Acad. Sci. USA*, **86**, 6562–6566.
Matthews,B.W., Nicholson,H. and Becktel,W.J. (1987) *Proc. Natl Acad. Sci. USA*, 84, 6663–6667.
Menendez-Arias,L. and Argos,P. (1989) *J. Mol. Biol.*, **206**, 397–406.
Meng,M., Bagdasarian,M. and Zeikus,G. (1993) *Biotechnology*, **11**, 1157–1161.
Mrabet,N.T. *et al.* (1992) *Biochemistry*, **31**, 2239–2253.
Nagao,T., Makino,Y., Yamamoto,K., Urabe,I. and Okada,H. (1989) *FEBS Lett.*, **253**, 113–116.
Narhi, L. O. *et al.* (1991) *Biotechnol. Appl. Biochem.*, **13**, 12–24.
Nicholson,H., Becktel,W.J. and Matthews,B.W. (1988) *Nature*, **336**, 651–656.
Nicholson,H., Anderson,D.E., Dao-Pin,S. and Matthews,B.W. (1991) *Biochemistry*, **30**, 9816–9828.
Nishikawa,S. *et al.* (1990) *Protein Engng*, **3**, 443–448.
Nosoh,Y. and Sekiguchi,T. (1993) In Gupta,M.N. (ed.) *Thermostability of Enzymes*. Springer, Berlin, pp. 182–203.
Olsen,O. and Thomsen,K.K. (1991) *J. Gen. Microbiol.*, **137**, 579–585.
Pakula,A.A. and Sauer,R.T. (1989) *Proteins: Struct. Funct. Genet.*, **5**, 202–210.
Pakula,A.A and Sauer,R.T. (1990) *Nature*, **344**, 363–364.
Pantoliano,M.W., Ladner,R.C., Bryan,P.N., Rollence,M.L., Wood,J. and Poulos,T.L. (1987) *Biochemistry*, **26**, 2077–2082.
Pantoliano,M.W., Whitlow,M., Wood,J., Dodd,S.W., Hardman,K.D., Rollence, M.L. and Bryan,P.N. (1989) *Biochemistry*, **28**, 7205–7213.
Perry,I.J. and Wetzel,R. (1984) *Science*, **226**, 555–557.
Perry,I.J. and Wetzel,R. (1987) *Protein Engng*, **1**, 101–105.
Perutz,M.F. and Raidt,H. (1975) *Nature*, **255**, 256–259.
Perutz, M. F. (1978) *Science*, **201**, 1187–1189.
Pickett,S. and Sternberg,M. (1993) *J. Mol. Biol.*, **321**, 825–839.
Pjura,P., Matsumura,M., Baase,W.A. and Matthews,B.W. (1993) *Protein Sci.*, **2**, 2217–2225.
Privilege,P. and Fasman,G. (1989) In Fasman,G. (ed.) *Prediction of Protein Structure and the Principles of Protein Conformation*. Plenum Press, New York, 391–416.
Quax,W.J., Mrabel,N.T., Luiten,R.G.M., Schuurhuizen,P.W., Stanssens,P. and Lasters,I. (1991) *Biotechnology*, **9**, 738–742.
Querol,E. and Parrilla,A. (1987) *Enzyme Microbiol Technol.*, **9**, 238–244.
Roussel,A., Inisan,A.G. and Knoops-Mouthy,E. (1994) *TURBO FRODO (Version 5.0a) Manual*. BIOGRAPHICS, Marseille.
Ruegg,C., Ammer,D. and Lerch,K. (1982) *J. Biol. Chem.*, **257**, 6420–6426.
Sali,D., Bycroft,M. and Fersht,A.R. (1988) *Nature*, **335**, 740–743.
Schuman,J., Bohm,G., Schumacher,G., Rudolph,R. and Jaenicke,R. (1993) *Protein Sci.*, **2**, 1612–1620.
Serrano,L. and Fersht,A.R. (1989) *Nature*, **342**: 296–299.
Serrano,L., Day,A.G. and Fersht,A.R. (1993) *J. Mol. Biol.*, **233**, 305–312.
Sharp,K., Nicholls,A., Friedman,R. and Honig,B. (1991) *Biochemistry*, **30**, 9686–9697.
Shirley,B.A., Stanssens,P., Steyaert,J. and Pace,C.N. (1989) *J. Biol. Chem.*, **264**, 11621–11625.
Stearman,R.S., Frankel,A.D., Freire,E., Liu,B. and Pabo,C. (1988) *Biochemistry*, **27**, 7571–7574.
Stonehouse,N.J. and Stockley,P.G. (1993) *FEBS Lett.*, **334**, 355–389.
Suzuki,Y., Ito,N., Yuuki,T., Yamagata,H. and Udaka,S. (1989) *J. Biol. Chem.*, **264**, 18933–18938.
Takagi,M. and Imanaka,T. (1989) *FEBS Lett.*, **254**, 43–46.
Takagi,H., Takahashi,T, Momose,H., Inouye,M., Maeda,Y., Matsuzawa,H. and Ohta,T. (1990) *J. Biol. Chem.*, **265**, 6874–6878.
Tapia,O., Oliva,B., Nilsson.,O., Querol,E. and Avilés,F. X. (1991) *J. Mol. Engng*, **1**, 249–266.
Toma,S., Campagnoli,S., Magarit,I., Gianna,R., Grandi,G., Bolognesi,M., DeFilippis,V. and Fontana,A. (1991) *Biochemistry*, **30**, 97–106.
Tomazic,S.J. and Klibanov,A.M. (1988a) *J. Biol. Chem.*, **263**, 3086–3091.
Tomazic,S.J. and Klivanov,A.M. (1988b) *J. Biol. Chem.*, **263**, 3092–3096.
Tyler-Cross,R. and Schirch,V. (1991) *J. Biol. Chem.*, **266**, 22549–22556.
Volkin,D.B. and Klibanov,A.M. (1987) *J. Biol. Chem.*, **262**, 2945–2950.
Vriend,G., Berendsen,J.C.H., van der Zee,J.R., van der Burg,B. and Eisink,V.G.H. (1991) *Protein Engng*, **4**, 941–945.
Wakarchuk,W.W., Sung,W.L., Campbell,R.L., Cunningham,A., Watson,D.C. and Yaguchi,M. (1994) *Protein Engng*, **7**, 379–1386.
Welpy,J.K., Fowler,A.V. and Zabin,I. (1981) *J. Biol. Chem.*, **256**, 6811–6816.
Wend,B. *et al.* (1988) *Eur. J. Biochem.*, **175**, 439–445.
Wigley,D.B., Clarke,A.R., Dunn,C.R., Barstow,D.A., Atkinson,T., Chia,W.N., Muirhead,H. and Holbrook,J. J. (1987) *Biochim. Biophys. Acta*, **916**, 145–148.
Wilson,K., Vorgias,C.E., Tanaka,I., White,S.W. and Kimura,M. (1990) *Protein Engng*, **4**, 11–22.
Wilson,K., Malcolm,B. and Matthews,B.W. (1992) *J. Biol. Chem.*, **267**, 10842–10849.
Yutani,K., Ogasahara,K., Sugino,Y. and Matsushiro,A. (1977) *Nature*, **267**, 274–275.
Yutani,K., Ogasahara,K., Kimura,A. and Sugino,Y. (1982) *J. Mol. Biol.*, **160**, 387–390.
Yutani,K., Ogasahara,K., Tsujita,T. and Sugino,Y. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 4441–4444.
Zagorski,M.G., Bowie,J.U., Vershon,A.K., Sauer,R.T. and Patel,D.J. (1989) *Biochemistry*, **28**, 9813–9825.
Zale,S.E. and Klibanov,A.M. (1986) *Biochemistry*, **25**, 5432–5444.
Zhang,X.-J., Baase,W.A. and Matthews,B.W. (1991) *Biochemistry*, **30**, 2012–2017.

Received June 15, 1995; revised November 1, 1995; accepted December 22, 1995