J. Biochem. **88**, 1895-1898 (1980)

Thermostability and Aliphatic Index of Globular Proteins

Atsushi IKAI

Department of Biophysics and Biochemistry, Faculty of Science,
The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113

Received for publication, September 5, 1980

A statistical analysis shows that the aliphatic index, which is defined as the relative volume of a protein occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine), of proteins of thermophilic bacteria is significantly higher than that of ordinary proteins. The index may be regarded as a positive factor for the increase of thermostability of globular proteins.

Since the first purification of an enzyme from a thermophilic bacterium, many have tried to find structural factors that keep such enzymes active at high temperatures. Studies have ranged from the comparison of amino acid composition of two groups of proteins, one from thermophilic and the other from mesophilic organisms, to the analysis of the three dimensional structure of proteins of thermophilic bacteria (*1-3*). A study on the amino acid substitutions observed in glyceraldehyde-3-phosphate dehydrogenases recently allowed Argos *et al.* to pinpoint the increase of alanine in a thermostable protein (*4*). No clear-cut differences in the structural parameters so far chosen for a more general comparison have been recognized between the two groups of proteins.

Recently, I have looked into the possible significance of aliphatic side chains in maintaining the thermostable structure of proteins. There were two reasons for choosing aliphatic side chains. First, initial scanning of the amino acid composition of several thermostable proteins invariably showed a high content of either one or several of Ala, Val, Ileu, and Leu while the content of aromatic amino acids was more varied. Second, the hydrophobicity of aliphatic amino acids is a potentially attractive measure of the stability of proteins at high temperatures as well as against denaturants such as urea (*vide infra*). In this short communication, I would like to report that there is a good correlation between the "aliphatic index" and thermostability of proteins.

The aliphatic index of a protein was calculated according to the following formula:

$$\text{Aliphatic Index} = x_A + a x_V + b(x_I + x_L)$$

where x_A, x_V, x_I, and x_L are mole percent ($100\times$ mol fraction) of alanine, valine, isoleucine, and leucine. As coefficients a and b, respectively, I took the relative volumes of aliphatic side chains to that of alanine side chains. These values were $a=2.9\pm0.1$ and $b=3.9\pm0.1$, based on the volume of amino acid residues in proteins (*5*). According to this convention, the average volume of amino acid residues is 4.8 to 4.9, therefore, the total volume of a protein is roughly 480 to 490. In some sets of data from which the tryptophan content is lacking, no correction was made for the slight increase in the mole percent of all the other amino acid residues, the differences usually being small. Amino acid compositions of proteins from thermophilic and mesophilic organisms

were taken from the references cited in the legend to Fig. 1. Amino acid compositions of 208 non-homologous proteins compiled by Reeck in the "Handbook of Biochemistry and Molecular Biology" were also used (*6*).

In Fig. 1, the distribution of the aliphatic index values of 208 non-homologous proteins and that of 34 proteins from thermophilic bacteria are shown. The former distribution has a sample mean at $\bar{X}_1=78.8$, with a sample standard deviation of $S_1=14.5$, and for the latter, $\bar{X}_2=92.6$, and $S_2=10.6$. The difference in the two population means has a 95% confidence interval of $\mu_2-\mu_1=13.8\pm5.1$ by Studen's t test. So the difference between the two populations is highly significant and a difference of 10 in the aliphatic index corresponds, for example at 25°C, to a difference of 5 to 7 kcal/mol of protein in terms of hydrophobic free energy.

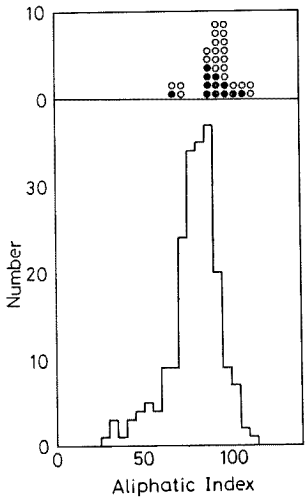

Fig. 1. Distribution of aliphatic index for 34 proteins from thermophilic bacteria (above) and 208 proteins of mesophilic organisms (below). Of the above, ● represents proteins from B-stearothermophilus, a moderate thermophile. Proteins from thermophilic bacteria include phosphoglycerate kinase (105.3; *9*), ATPase (90.6; *8*), formyltetrahydrofolate synthetase (99.7; *8*), 3 lactate dehydrogenases (93.8, 96.6, 97.9; *10*), amino peptidase (105.6; *8*), aldolase (93.0; *8*), glucose 6-phosphate isomerase (86.9; *8*), acetylkinase (94.4; *11*), 2 glyceraldehyde 3-phosphate dehydrogenases (100.7, 103.2; *8*), triosephosphate isomerase (96.8; *10*), 2 enolases (92.0, 73.7; *8*), α-amylase (66.1; *8*), cytochrome c_{552} (91.3; *12*), alkaline proteinase (69.4; *8*), ribosomal proteins (average) (86.9; *8*), tyr-tRNA synthetase (90.5; *8*), methylene tetrahydrofolate dehydrogenase (110.4; *8*), mitochondrial ATPase F_1 (98.0; *13*), TF_1 (99.0; *13*), TF_0 (111.2; *13*), 2 superoxide dismutases (85.2, 87.7; *14*), isocitrate dehydrogenase (87.6; *8*), malate dehydrogenase (96.4; *15*), glutamine synthetase (85.8; *16*), malate synthetase (99.4; *17*), thermolysin (71.3; *8*), elongation factor EF-Tu (90.3; *18*), EF-Ts (93.4; *18*), EF-G (97.6; *18*), where the numbers in parentheses are (aliphatic index; reference number for the amino acid composition).

Then we can notice two things in the upper distribution in Fig. 1. First, four proteins with low aliphatic index values are rather isolated from the rest. These proteins are α-amylase, thermolysin, enolase, and alkaline proteinase, the first three being metallo-proteins. These proteins, in contrast to their low contents of aliphatic residues, have high contents of serine plus threonine residues, 13.4, 16.1, 13.8, 18.6 mol% in that respective order. The presence of bound metal ions and/or the high content of hydrogen bond forming residues suggest that the thermostability of these proteins can be attributed to factors other than the aliphatic index.

Similar treatment of the aromatic index did not show a good correlation with the thermostability of proteins and will not be presented here.

In Fig. 2 the aliphatic indices of representative proteins that appeared in Fig. 1 are plotted against their native molecular weight. First we can notice

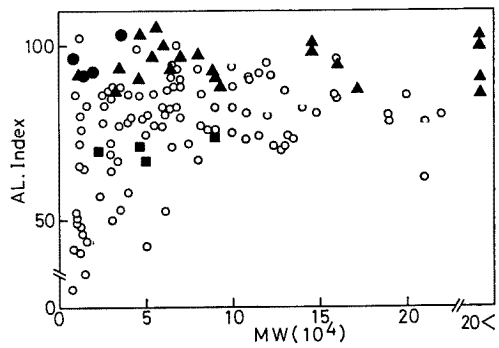

Fig. 2. Aliphatic index *vs.* molecular weight of proteins. Symbols are ○ proteins of mesophilic origin, ● same as before but with proven thermostability, ▲ proteins of thermophilic origin, and ■ same as before but with high contents of serine and threonine.

TABLE I. Aliphatic indices of proteins.

Proteins	Thermophilic origin	Mesophilic origin
Phosphoglycerate kinase	105	95, (86)[a]
Glyceraldehyde 3-phosphate dehydrogenase	101, 103	89, 83, 90, (86)
Triosephosphate isomerase	97	87
Malate synthetase	99	86
Superoxide dismutase	88, 85	(80)
Cytochrome *c*	91	(67)
Glucose 6-phosphate isomerase	87	(78)
Lactate dehydrogenase	98, 97	88, (101)
Isocitrate dehydrogenase	88	63
Aldolase	93	80, (92, 104)
Ribosomal proteins	87	83
Elongation factors	90, 93, 98	86, 89, 85
Enolase	93, 74	(91)
α-Amylase	66	64

[a] Values in parentheses represent those for enzymes of vertebrate origin.

that the aliphatic index seems to have a wider range of variation in smaller proteins. Because of this, the comparison of amino acid compositions in terms of aliphatic index is more meaningful for proteins of molecular weights of less than 100,000. Second, there are several proteins of mesophilic origin that have high aliphatic index values. Four of them (α-lactalbumin, β-lactoglobulin, myoglobin, and acyl carrier protein) have been shown to be remarkably heat stable. Such exceptional thermostability has not been reported for others (thioredoxin, indol-3-glycerolphosphate synthase, L-glycerol-3-phosphate dehydrogenase, and systathionine-γ-synthetase).

In Table I, the values of aliphatic index are given for analogous enzymes (but not necessarily homologous to each other) of thermophilic and mesophilic origin. Again it is clear that the aliphatic index is often significantly higher for the former proteins. This result reduces the possibility that the proteins dealt with in this paper are biased to high values of aliphatic index irrespective of their origin.

From the observations that have been made with Fig. 1, Fig. 2, and Table I, the pros and cons of aliphatic index can be summarized as follows.

1. The aliphatic index, as defined in this paper, is significantly higher for proteins of thermophilic origin, particularly for those with molecular weights of less than 100,000.
2. Proteins of mesophilic origin but with high aliphatic index values are often thermostable.
3. For proteins like α-amylase, enolase, thermolysin, and alkaline proteinase, with low aliphatic indices, I found high contents of serine+threonine (hydrogen bond forming amino acids).
4. There are proteins with high aliphatic index values regardless of the origin. In these cases, a high aliphatic index alone cannot account for the differences in their thermostability.

Thus a high aliphatic index value seems to be a positive factor for increasing the thermostability of globular proteins. One can allow for the contribution of proline residues to the aliphatic index by adding $2.4x_p$ to the right-hand side of formula, without much influencing observations that have been already made.

Finally two questions must be dealt with. 1) Why aliphatic side chains and not the total hydrophobic residues? 2) Why volume related index rather than the hydrophobicity of aliphatic side chains?

A potential answer to the first question can be found in Ref. 7. The temperature dependence

of the solubility property of aliphatic amino acids is somewhat different from that of aromatic amino acids in a manner that suggests that the aliphatic hydrophobicity increases more rapidly with increasing temperature. (This attractive feature of aliphatic hydrophobicity was once used by Brandts in his study of protein denaturation (*7*).) Therefore, the result in this study that the aliphatic index has a better qualitative correlation with the thermostability of globular proteins is not unreasonable.

To answer the second question we must recall that the hydrophobic stabilization free energy of the native structure is both thermodynamical and geometrical in nature and its calculation requires knowledge of the temperature dependence of the hydrophobicity of each kind of residue and the degree of exposure of every one of the hydrophobic side chains in the native conformation. The aliphatic index is, on the other hand, purely numerical in nature and can be calculated very easily. There is also a possibility that the contribution of aliphatic side chains to the thermostability of proteins is not necessarily due to hydrophobicity alone.

More detailed statistical treatment of the data and discussion of the possible energetic significance of the aliphatic and the aromatic side chains in the stabilization and/or destabilization of globular proteins at higher temperatures will be presented elsewhere.

I want to express my deepest gratitude to Drs. Tairo Ohshima and Koyu Honnami for helping me amass the literature on proteins of thermophilic bacteria. Thanks are also due to Dr. Shinji Iijima for providing me with the data on amino acid composition of malate dehydrogenase of thermophilic bacteria before publication.

REFERENCES

1. Ljungdahl, L.G. & Sherod, D. (1976) in *Extreme Environments—Mechanisms of Microbial Adaptation* (Heinrich, M.R., ed.) pp. 147–187, Academic Press, New York
2. Singleton, R., Jr. (1976) in *Extreme Environments—Mechanisms of Microbial Adaptation* (Heinrich, M.R., ed.) pp. 189–200, Academic Press, New York
3. Wanacott, A.J. & Biesecker, G. (1977) in *Pyridine Nucleotide Dependent Dehydrogenases* (Sund, H., ed.) pp. 140–156, W. de Gruyter, Berlin
4. Argos, P., Rossmann, M.G., Grau, U.M., Zuber, H., Frand, G., & Tratschin, J.D. (1979) *Biochemistry* **18**, 5698–5703
5. Richards, F.M. (1977) *Annu. Rev. Biophys. Bioeng.* **6**, 151–176
6. Reeck, G. (1976) in *Handbook of Biochemistry and Molecular Biology* (Fasman, G., ed.) 3rd Ed., Proteins III, pp. 504–519, CRC Press, Cleveland
7. Brandts, J.F. (1964) *J. Am. Chem. Soc.* **86**, 4302–4312
8. See references in Ljungdahl, L.G. & Sherod, D. (1976) *Extreme Environments—Mechanisms of Microbial Adaptation* (Heinrich, M.R., ed.) pp. 147–187, Academic Press, New York
9. Nojima, H. (1979) Thesis for Doctor of Science, The Univ. of Tokyo
10. Frank, G., Haberstich, H.U., Shaer, H.P., Tratschin, J.D., & Zuber, H. (1976) in *Enzymes and Proteins from Thermophilic Microorganisms* pp. 375–389, Birkhauser Verlag, Basel
11. Nakajima, H. (1977) Thesis for Doctor of Science, The Univ. of Tokyo
12. Honnami, K. (1979) *J. Biochem.* **86**, 1689–1695
13. Kagawa, Y., Sone, N., Yoshida, M., Hirata, H., & Okamoto, H. (1976) *J. Biochem.* **80**, 141–151
14. Sato, S. & Nakazawa, K. (1978) *J. Biochem.* **83**, 1165–1171
15. Iijima, S., Saiki, T., & Beppu, T. (1980) *Biochim. Biophys. Acta* **613**, 1–9
16. Hachimori, A., Matsunaga, A., Shimizu, M., Samejima, T., & Nosoh, Y. (1974) *Biochim. Biophys. Acta* **350**, 461–474
17. Sundram, T.K., Chell, R.M., & Wildinson, A.E. (1980) *Arch. Biochem. Biophys.* **199**, 515–525
18. Nakamura, S., Ohta, S., Arai, K., Ohshima, T., & Kajiro, K. (1978) *Eur. J. Biochem.* **92**, 533–543