Thermal Stability and Protein Structure†

Patrick Argos,* Michael G. Rossmann, Ulrich M. Grau,‡ Herbert Zuber, Gerhard Frank, and Jon Duri Tratschin

ABSTRACT: Amino acid sequences have been compared for thermophilic and mesophilic molecules of ferredoxin, glyceraldehyde-3-phosphate dehydrogenase, and lactate dehydrogenase. It is shown that Gly, Ser, Ser, Lys, and Asp in mesophiles are generally substituted by Ala, Ala, Thr, Arg, and Glu, respectively, in thermophiles. These exchanges suggest that thermal stability can be achieved by the addition of many small changes throughout the molecule without significant change in the backbone conformation. Their overall effect is primarily to increase internal and decrease external hydrophobicity as well as to favor helix stabilizing residues in helices. These substitutions minimize interruption of function or internal residue packing arrangements. Although the analysis has been confined to the above-mentioned molecules, the observed stabilizing principles may be more generally applicable.

Organisms tend to adapt their specific proteins to function efficiently within their normal environmental temperature (Low & Somero, 1976; Feeney & Osuga, 1976; Hochachka & Somero, 1973). This generally implies that proteins have a limited temperature range within which structural integrity is maintained. Outside this thermal span, denaturation occurs with corresponding loss of function, such as enzymic activity. The thermal stability of a protein can be changed intrinsically by alteration of amino acids or extrinsically by addition of suitable stabilizing effectors (e.g., cations, coenzymes, membranes, and peptides). This paper deals entirely with the effect of intrinsic alterations to thermal stability. The structure of the three-dimensional polypeptide backbone of proteins is highly conserved (Rossmann et al., 1975; Kretsinger, 1975), despite alterations in the primary sequence. Thus, with the use of the known tertiary structure as well as the amino acid sequences of thermophilic and mesophilic variants, it should be possible to determine some of the principles with which a protein achieves structural tolerance within defined temperature limits.

A variety of attempts have been made to understand the nature of thermal stability in proteins [cf. Zuber (1976)]. The energy profile for the unfolding of thermophilic and mesophilic proteins has been discussed by Nojima et al. (1977, 1978). Bull & Breese (1973) attempted to correlate the midpoint of the stable temperature range for proteins to their average amino acid residue volume. These results were reinterpreted and extended by Stellwagen & Wilgus (1978), who proposed that the ratio of surface area to volume for a given protein domain or subunit is a critical factor of thermal stability. However, only Perutz & Raidt (1975) carefully compared the amino acid substitutions in ferredoxins and hemoglobins for organisms of different temperature preferences, in light of the known three-dimensional polypeptide backbone structure. They concluded that thermal stability is enhanced by the formation of a few hydrogen bonds and salt bridges. These concepts have been elegantly reviewed by Perutz (1978) within the larger framework of electrostatic effects on protein hydration, folding, assembly, and catalysis.

Only a few tertiary structures of proteins have been determined where amino acid sequences are known for species of widely different thermal environments. Ferredoxin is one such protein, although the variation of thermal stability is not extreme. Another example is that of glyceraldehyde-3-phosphate dehydrogenase where the structures of the mesophilic lobster (Buehner et al., 1974; Moras et al., 1975) and thermophilic *Bacillus stearothermophilus* (Biesecker et al., 1977) proteins have been independently determined. Although Wonacott & Biesecker (1977) suggest a change in hydrogen bonding in the vicinity of the active center, this structural alteration does not involve an amino acid change and remains to be validated by a careful comparison of the electron densities. In addition, they also note the possibility of additional salt bridges for stabilizing the thermophile. More strikingly, the amino acid sequence differences for these two GAPDH[1] molecules involve 130 substitutions within 331 positions. It would thus seem possible that many small alterations would in net effect produce the desired thermal stability for this 140000 molecular weight tetramer. A more careful study of GAPDH is now possible with the advent of the sequence of an extreme thermophile, namely, *Thermus aquaticus* (Harris & Walker, 1977).

Recently some of us (Tratschin, Frank, and Zuber, unpublished experiments) have determined the partial sequence of *B. stearothermophilus* LDH, with some 250 out of 330 amino acids. The variation of *B. stearothermophilus* LDH to dogfish LDH is similar to that of *B. stearothermophilus* GAPDH and lobster GAPDH. An attempt has been made here to determine some of the principles of thermal stability by combining the information contained in these various proteins.

In Table I are given references to the protein sequences used here and their associated "optimal" temperatures of stability and function. The latter are assumed to correspond to normal environmental temperatures and are taken for bacteria from Buchanon & Gibbons (1974) or as body temperatures for pig and chicken. While the exact mean thermal stable range for a given enzyme is debatable, the general temperature trend used here is clearly reasonable. Indeed, grouping the proteins

† From the Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907 (P.A., M.G.R., and U.M.G.), and the Institut für Molekularbiologie und Biophysik, ETH-Hönggerberg, CH-8093 Zürich, Switzerland (H.Z., G.F., and J.D.T.). *Received July 13, 1979; revised manuscript received September 20, 1979.* The work was supported by a National Science Foundation grant (No. PCM78-16584) and a National Institutes of Health grant (No. GM 10704) to M.G.R., by a National Science Foundation grant (No. PCM77-20287) and an American Cancer Society grant (No. FRA173) to P.A., and by a grant of the Swiss National Foundation (Project 3.1640.73 and 3.005.076) to H.Z.

‡ Recipient of a Deutsche Forschungsgemeinschaft postdoctoral fellowship.

[1] Abbreviations used: GAPDH, glyceraldehyde-3-phosphate dehydrogenase; LDH, lactate dehydrogenase.

0006-2960/79/0418-5698$01.00/0

Table I: Amino Acid Sequences Used to Represent the Ferredoxin, GAPDH, and LDH Proteins

protein	species	ref	assumed midpoint of temp stability (°C)
ferredoxin	*C. thermosaccharolyticum*	Tanaka et al. (1973)	55
	C. tartarivorum	Tanaka et al. (1971)	46
	P. elsdenii	Azari et al.[a]	37
	C. acidi-urici	Rall et al. (1969)	41
	C. pasteurianum	Tanaka et al. (1966)	37
	M. aerogenes	Tsunoda et al. (1968)	37
GAPDH	*T. aquaticus*	Harris & Walker (1977)	71
	B. stearothermophilus	Harris & Walker (1977)	65
	yeast	Jones & Harris (1972)	37
	pig	Harris & Perham (1968)	37
	lobster	Davidson et al. (1967)	20
LDH	*B. stearothermophilus*	Tratschin, Frank, and Zuber (unpublished experiments)	65
	chicken M_4	Eventoff et al. (1977)	37
	pig M_4	Eventoff et al. (1977); Kiltz et al. (1977)	37
	dogfish M_4	Eventoff et al. (1977); Taylor (1977)	20

[a] P. Azari, M. Glantz, J. Tsunoda, and K. T. Yasunobu, unpublished results cited in Yasunobu & Tanaka (1973).

simply into "high" and "low" temperatures gave similar results (vide infra).

Methods

Matrices of amino acid changes were constructed for each pair of known sequences for a given protein. Let a_{ij} be an element of this matrix with i denoting a given residue (e.g., Ala, Arg, etc.) belonging to the mesophilic protein and j denoting a residue in the homologous thermophilic protein. Then, a_{ij} is defined as the number of times a residue of type i has changed to type j in the aligned sequences. If there is a trend in amino acid exchange to achieve thermal stability, then this preference should be more pronounced the greater the temperature difference, ΔT, between the "optimal" temperatures. Thus, a "temperature-corrected" matrix can be constructed where

$$b_{ij} = a_{ij}/\Delta T$$

if a linear temperature relation is assumed to a first order. Any assumptions more complex than linearity were outside the observational range with the limited available data.

The coefficients, b_{ij}, may be similar for all pairwise comparisons. If this is the case, a more reliable set of coefficients, c_{ij}, can be determined by taking a weighted average such that

$$c_{ij} = \frac{\sum_n \omega_n b_{ij}}{\sum_n \omega_n}$$

where ω_n is the weight to be attributed to a given comparison and n is the number of comparisons. Since greater weight should be given to comparisons between sequences whose difference in optimal temperature is large, a useful weight is $\omega_n = (\Delta T_n)^2$. Thus

$$c_{ij} = \frac{\sum_n (\Delta T_n) a_{ij}}{\sum_n (\Delta T_n)^2}$$

The sum $G_{ij} = c_{ij} + c_{ji}$ will now represent a weighted estimate of the "gross traffic" between residues i and j per degree Celsius, while the difference $D_{ij} = c_{ij} - c_{ji}$ represents a tendency of a residue type i in the mesophilic sequence to become a residue of type j in the thermophilic sequence per degree Celsius. An analogy would be the traffic between states within America. If i stands for Minnesota and j stands for Florida, then a_{ij} is the number of cars traveling from Minnesota to Florida while a_{ji} is the converse. The total "traffic" on north–south roads between these two states would be $a_{ij} + a_{ji}$ and the net southernly motion is $a_{ij} - a_{ji}$. The task here is to determine whether there is a significant drift toward the warmer climate. The standard error in the net amino acid "traffic" can be determined from all those terms in the "net-traffic" matrix, D, corresponding to nonzero terms in the "gross traffic" matrix, G. Hence, the significance of drift will be $E_{ij} = D_{ij}/\sigma(D_{ij})$. It must be emphasized that this is a most stringent criterion since it neglects all those cases with an effective zero traffic.

Results

The comparison of lobster and *T. aquaticus* GAPDH is given here as an example since it represents the largest difference in optimal temperatures within the available sample. The matrix a_{ij} is shown in Table II from which the net traffic ($D_{ij} = a_{ij} - a_{ji}$) was calculated. The latter elements form an antisymmetric matrix. The root mean square value, σ, of that matrix (not including the terms for which $a_{ij} + a_{ji}$ were zero) was 1.46 amino acids. With this result, the ratios E_{ij} (the number of times a given exchange is greater than the corresponding standard error) were calculated, and those greater than 1.9 are given in Table III.

The ratios E_{ij} are shown in Table III for each particular protein and for the weighted average among the three proteins GAPDH, LDH, and ferredoxin. The top five exchanges of the overall results were greater than three standard errors and the next five were greater than 1.9 standard errors. Furthermore, there is reasonable agreement between the independent results from each protein. Of the top ten substitutions found in the overall statistics, eight out of ten with $E_{ij} > 1.9$ appeared in GAPDH, seven out of eight appeared in LDH, and three out of five appeared in ferredoxin. The ferredoxin results do not give equally good agreement since (1) they include only about one-sixth of the residues compared to those in GAPDH, reducing the number of possible exchanges and the reliability of the statistics, and (2) differences in optimal temperatures are smaller. The *B. stearothermophilus* LDH sequence is only about two-thirds complete and thus had the effect of reducing the significance of the LDH results with respect to GAPDH, although the LDH results are nevertheless

Table II: Matrix a_{ij} between Two Aligned GAPDH Sequences[a]

residues in lobster GAPDH	residues in *T. aquaticus* GAPDH																			
	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	17		2	1				1	1	1	3		1			1			1	2
Arg		7							1		1									
Asn	2		7				1								1					
Asp		1	1	12			4	1				1					1			
Cys		1			1								1			1				1
Gln	1		1			1	1			1	1	1								
Glu				2			8	2				3				1				1
Gly	5	1				1		18	1		1	1	1				1			
His							1		2		1								1	
Ile	2									8	1	1								5
Leu				1		1			2	1	10			2						1
Lys	3	4	2	1			3				1	12			1	1				
Met									1		3	1	2				1		1	1
Phe										2	4		1	4			2	1		1
Pro	1									1	1				9					
Ser	5			1				1			1	1	1			8	4			1
Thr	1	1	1	2					1	1	1		1				11			
Trp											1							2		
Tyr	1								1								1		5	1
Val	4	1					1	1		7	1	1		1		1	1		2	17

[a] Exchange of residues between *T. aquaticus* and lobster GAPDH. For example, the table shows that there are five positions in the aligned sequences where a glycine in lobster GAPDH changes to an alanine in *T. aquaticus* GAPDH. Terms along the diagonal show the number of times that a given amino acid is not exchanged.

Table III: Exchanges with $E_{ij} \geq 1.9$[a]

exchange, cold → hot	weighted average of GAPDH, LDH, and ferredoxin	GAPDH		LDH	ferredoxin
		weighted average	lobster → *T. aquaticus* alone	weighted average	weighted average
Gly → Ala	1 (4.3)	1 (5.0)	1 (2.7)		
Ser → Ala	2 (4.1)	2 (3.2)	1 (2.7)	6 (2.0)	1 (2.2)
Ser → Thr	3 (3.9)	5 (2.6)	1 (2.7)	4 (2.6)	2 (2.1)
Lys → Arg	4 (3.4)	6 (2.2)	1 (2.7)	1 (3.1)	
Asp → Glu	5 (3.0)	4 (3.0)			
Ser → Gly	6 (2.9)			3 (2.8)	5 (1.9)
Lys → Ala	7 (2.8)	9 (1.9)	5 (2.1)	5 (2.1)	
Val → Ala	8 (2.5)	9 (1.9)		7 (1.9)	
Asp → Asn	9 (2.5)	3 (3.1)			
Val → Ile	10 (1.9)			2 (2.9)	
Met → Leu		7 (2.0)	5 (2.1)		
Phe → Leu		7 (2.0)			
Ala → Leu			5 (2.1)		
Ile → Phe				7 (1.9)	
Ala → Glu					2 (2.1)
Gln → Glu					4 (2.0)

[a] The first number is the rank within a column and the second number, set in parentheses, is the ratio E_{ij} of the net weighted exchange per degree Celcius to the standard error.

more reliable than the ferredoxin statistics. It should be noted that ferredoxin is a small monomer whereas LDH and GAPDH are large tetramers. Hence, the mode of achieving thermal stability might be different in these proteins. Nevertheless, there is reasonable agreement between the three different proteins, suggesting that thermal stability may be attained by similar principles in these, and possibly other, proteins.

The results obtained by taking an overall weighted average between all three proteins (Table III) are shown diagrammatically in Figure 1. It is immediately apparent that there is a considerable preference for alanine in the thermophiles. Furthermore, the thermal stabilizing process could occur in a series of steps: for example, exchanging a serine first to glycine and then to alanine. The preference for arginine in thermophiles is consistent with the increase in thermal stability of pig H_4 LDH after modifying its external lysine residues with methyl acetimidate to generate arginine-like amino acids (Tuengler & Pfleiderer, 1977).

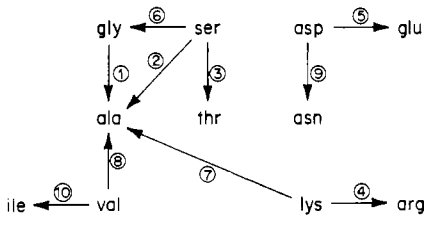

FIGURE 1: Direction of observed preferred exchanges. Arrows point from the mesophilic to the thermophilic protein. Numbers indicate the ranking of the significance for the given exchange.

Properties of Preferred Exchanges. In this section an attempt will be made to understand the principles which govern the nature of the observed preferred exchanges (Table III and Figure 1). In Table IV are listed the alterations in a variety of physical and structural characters for the top ten amino acid exchanges determined here. The characters were selected with an eye to those properties which might have a bearing on the stability of the protein. They were (1) α-helix preference,

Table IV: Physical Characteristics of Preferred Exchanges[a]

exchange, cold → hot	H	$\Delta P(\alpha)$	S	$\Delta P(\beta)$	E	I	ΔHy	ΔPo	ΔBu	MBC/C
(1) Gly → Ala	8	+0.73	8	−0.02	1	19	+0.77	0	+8.1	1
(2) Ser → Ala	15	+0.47	3	−0.05	9	15	+0.80	0	+2.0	1
(3) Ser → Thr	8	0.00	13	+0.26	12	6	0.00	0	+6.3	1
(4) Lys → Arg	6	−0.27	6	+0.22	20	2	−0.75	+3.0	−1.4	1
(5) Asp → Glu	6	+0.40	2	+0.03	20	2	0.00	0	+1.9	1
(6) Ser → Gly	7	−0.28	5	−0.03	8	9	+0.03	−1.7	−6.0	1
(7) Lys → Ala	5	+0.06	4	+0.13	17	0	−0.70	−49.0	−4.2	2
(8) Val → Ala	16	+0.40	7	−0.60	1	17	−1.00	−0.1	−10.1	1
(9) Asp → Asn	4	−0.14	0	+0.04	7	1	−0.57	−46.3	+1.1	1
(10) Val → Ile	0	+0.06	18	−0.04	3	7	+1.28	0	−0.2	1

[a] Difference in physical characteristic, C, is calculated as C(hot) − C(cold). H, S, E, and I are the number of times the specific exchanges are found in helix, sheet, externally, and internally for all GAPDH and LDH pairwise comparisons where $\Delta T > 0$. $\Delta P(\alpha)$, $\Delta P(\beta)$, ΔHy, ΔPo, and ΔBu are changes in the α-helix-forming ability, β-sheet-forming ability, hydrophobicity, polarity, and bulkiness, respectively. MBC/C is the minimum number of base changes per codon required for the exchange. Top five exchanges are those with $E_{ij} \geq 3.0$.

$P(\alpha)$, (2) β-sheet preference, $P(\beta)$, (3) hydrophobicity, Hy, (4) bulkiness, Bu, and (5) polarity, Po. In addition, the minimum base change per codon is listed, since this is a measure of the mutational difficulty in achieving a desirable change. The bulkiness and polarity scales were taken from Zimmerman et al. (1968), the hydrophobicity scale was derived by Nozaki & Tanford (1971), and the $P(\alpha)$ and $P(\beta)$ values were those calculated by Levitt (1978) and are similar to those originally published by Chou & Fasman (1974) and by Fasman et al. (1976). The limits of the various scales are $0.52 \leq P(\alpha) \leq 1.47$, $0.64 \leq P(\beta) \leq 1.49$, $0.0 \leq$ Hy ≤ 3.15, $3.4 \leq$ Bu ≤ 21.7 and $0 \leq$ Po ≤ 52.0. The differences Δ in character C, listed in Table IV, are given as $\Delta = C$(hot) − C(cold).

Shown also in Table IV are the number of occurrences with which a given exchange has been found in helical (H) or sheet (S) regions and whether the position was on the molecular surface (external, E) or within it (internal, I). Assignments for GAPDH are taken from Olsen et al. (1975) and for LDH are taken from Eventoff et al. (1977). Ferredoxin was omitted from these statistics since there are no explicit published data and, furthermore, it has little secondary structure. Thus, for a given exchange to have been beneficial, the products $H\Delta P(\alpha)$ (helix formation), $S\Delta P(\beta)$ (sheet formation), $(I - E)\Delta$Hy (internal hydrophobicity), $(E - I)\Delta$Po (external polarity), and $(I - E)\Delta$Bu (internal packing) would have to be large and positive. Presumably, any given exchange need not be beneficial to every physical characteristic, but there should be an overall improvement (Table V).

Inspection of Table IV shows that only one exchange (Lys → Ala) requires more than one base change per codon; that is, 9 out of 10 are relatively easy mutational events. Table IV also demonstrates that, in general, when an exchange occurs more frequently in a helix, the exchange favors helix formation in the thermophile (e.g., Ser → Ala, Asp → Glu, and Val → Ala). Preferences are also expressed in sheet regions but to a lesser extent (e.g., Ser → Thr). Hence, thermal stability is achieved more by stabilizing helix formation than sheet formation. Since helix and sheet formation represent mostly opposing requirements, the correlation between frequency of occurrence and preference demonstrated in Table IV is striking. An outcome of these observations is greater ease in predicting secondary structure from amino acid sequences in thermophiles than in mesophiles (Argos et al., 1976; Argos, unpublished results) and furthermore suggests that a weighting scheme which favors thermophiles might give improvement in the determination of the Chou and Fasman parameters.

The increase or decrease in hydrophobicity is extremely marked whenever the exchange is primarily internal or external, respectively. Thus, this characteristic is probably the most important stabilizing property as indeed might be anticipated (Veronese et al., 1976; Amelunxen & Singleton, 1976; Ljungdahl et al., 1976). Improvement in the internal packing arrangement of residues is less marked while increase of external polarity exhibited little consistency (Table V). However, there is a trend to increased bulkiness and, hence, surface area in those exchanges with $E_{ij} \geq 3.0$. There is also a tendency to decreased side-chain flexibility (Janin et al., 1978). Each of the top five exchanges with $E_{ij} \geq 3.0$ showed an excess of beneficial effects in achieving structural stability; simultaneously, each physical characteristic contributed to the overall molecular integrity. Furthermore, the five most favored exchanges all had a minimum base change per codon of only 1. However, the next five ($3.0 > E_{ij} \geq 1.9$) are less marked in their effects. The strength of this analysis lies in the similarity of the results from procaryotic and eucaryotic organisms, an independence of phylogenetic relationships, and a strong correlation between preferred exchanges and their stabilizing effects.

Table V: Benefit of Exchange to Stability[a]

exchange, cold → hot	(1) helical region	(2) sheet region	(3) internal hydrophobicity	(4) external polarity	(5) internal packing
(1) Gly → Ala	++		++		++
(2) Ser → Ala	++		+		+
(3) Ser → Thr		+			−
(4) Lys → Arg	−	+	++	++	+
(5) Asp → Glu	+				−
(6) Ser → Gly	−			+	−
(7) Lys → Ala		+	++	−−	++
(8) Val → Ala	++	−	−−	+	−−
(9) Asp → Asn	−−		+	−−	−
(10) Val → Ile		−−	+		+

[a] Criteria used:

	(1) $H\Delta P(\alpha)$	(2) $S\Delta P(\beta)$	(3) $(I-E)$ × ΔHy	(4) $(E-I)$ × ΔPo	(5) $(I-E)$ × ΔBu
++	>5	>5	>10	>50	>50
+	>0.4	>0.4	>0	>0	>0
−	<0.4	<0.4	<0	<0	<0
−−	<5	<5	<10	<50	<50

Note: top five exchanges are those with $E_{ij} \geq 3.0$.

Controls. A number of checks verified that the above results were not a consequence of unforeseen bias. The first test examined whether the observed trends attributed to thermal

stability could also be observed in proteins with the same nominal optimal temperatures. For this purpose those proteins, listed in Table I with optimal temperatures of 37 °C, were arbitrarily grouped into high and low temperatures. All possible combinations were used with a weighting factor of $\Delta T^2 = 1$. The results showed no $E_{ij} > 2.0$, and none of the eight $E_{ij} > 1.0$ coincided with those listed in Table III.

It was noted that the largest temperature differences in the available data, used in computing the results in Table III, were dependent on eucaryotic (cold) to procaryotic (hot) exchanges. It might therefore be possible that the above conclusions merely reflect procaryotic and eucaryotic sequence differences. Hence, a series of proteins were investigated with homologous sequences from eucaryotic and procaryotic organisms of similar optimal temperatures. These include an equal number of procaryotic and eucaryotic sequences for cytochrome *c*, serine proteases, dihydrofolate reductase, and triosephosphate isomerase (Dayhoff, 1979). In total there were nine each of procaryotic and eucaryotic sequences, giving approximately 4500 individual amino acid comparisons, a number roughly equal to that used in establishing the results in Table III.

The controls for the eucaryotic–procaryotic prejudice showed that all exchanges were less than three standard deviations; however, eleven were between 1.9 and 2.9σ. Seven of the eleven showed a minimum base change per codon equal to 2, and in an analysis the same as described above (see Tables IV and V) no consistent stabilizing effects were found. The only exchange in common with the significant five in Tables IV and V was Ser → Ala, but its positioning in the control structures did not enhance stability. In contrast, the top five exchanges in Tables IV and V had five exchanges with $E_{ij} \geq 3.0$, uniformly favored stability, and had only one minimum base change per codon. Furthermore, the exchanges generated by the generally larger number of alanines in the procaryotic compositions did not correlate to those required for thermal stability.

It is possible that the conclusions drawn here are biased by the dominance of data from dehydrogenases. Unfortunately, no data are currently available to perform a suitable control.

Possible Types of Exchanges. An attempt will now be made to determine whether the exchanges discovered experimentally (Table III and Figure 1) are unique in their ability to increase thermal stability or whether other exchanges might occur in a different set of mesophilic–thermophilic transitions. In the following derivation it will, however, be assumed that the physical parameters found important for the current set of exchanges will also be of importance in others. The alteration of any one physical character, while definite, must not be excessive. Mutational ease (only one minimum base change per codon) is an obvious example of this principle. Similarly, if the change in $P(\alpha)$ or $P(\beta)$ is too great, then a residue which was originally helical could become nonhelical in character. A residue which has increased its bulk slightly might be able to better fill the gaps within a hydrophobic cavity, but, if it increases too much, then the whole structure is distorted and destroyed. Such criteria could represent easy evolutionary steps. Individual changes which have a large stabilizing or destabilizing effect might occur, particularly in artificial laboratory selection of mutants (Grütter et al., 1979) or chemical modifications (Tuengler & Pfleiderer, 1977; Torchilin et al., 1979), but are improbable during more natural forms of selection. Thus, the following criteria were used to select beneficial and yet structurally feasible exchanges: (1) minimum base change per codon = 1; (2) $\Delta P(\alpha) < 0.75$ and $\Delta P(\beta) < 0.75$; (3) ΔHy < 1.5; (4) ΔPo < 47.0; (5) ΔBu < 11.0; (6) the only allowed charge changes were to neutrality; and (7) cysteine and tryptophan residues are excluded (since their infrequent occurrence implies an inability to find a significant sample of exchanges).

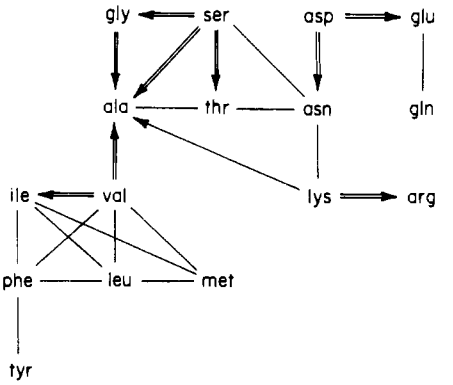

FIGURE 2: The top ten observed exchanges are shown with arrowheads in the direction of thermophilic preference. The 23 theoretically predicted easy exchanges are shown without arrowheads. Hence, the nine exchanges common to both sets have a double-tailed arrow.

There are 190 possible pairwise combinations of amino acids of which 75 require a minimum base change per codon of 1. None of the above criteria taken individually eliminates more than 22 combinations from the set of 75. Yet taken together there remain only 23 combinations (Figure 2). Nine of these combinations are the same as the top ten listed in Table III and shown in Figure 1. A further four are also shown in Table III as changes with $E_{ij} > 1.9$ for any one of the proteins individually.

The remaining substitutions not found in the observed list (Figure 2) are mainly among the hydrophobic residues Val, Ile, Phe, Leu, Met, and Tyr. Quite possibly these exchanges do occur but their stabilizing effect is only marginal since they are all rather hydrophobic. Hence, their preference would not be easily detected with the limited amount of data available for this study.

Discussion

Thermal stability is largely achieved in GAPDH, LDH, and ferredoxin by an additive series of very small improvements at many locations within the molecule. The most significant changes increase internal hydrophobicity and the helix-forming ability of residues in helices. In addition, stabilization is achieved by increasing the sheet-forming tendency of residues in β sheets and by slight increases in bulk for internal residues to obtain better packing organization. These optimal changes can be determined experimentally and proposed from theory.

The statistical procedure which has been outlined above is useful in elucidating broad structural principles of protein thermostability and protein thermophilicity. However, an alternative approach to the determination of structural elements important for thermal stability and other thermophilic properties lies in the comparison of homologous amino acid sequences of thermophilic and mesophilic proteins from phylogenetically closely related organisms (Zuber, 1978). This procedure has the possibility of recognizing residues which provide thermophilic properties in a given protein and which might be partially obscured by a statistical method. This alternative approach was able to recognize (Zuber, 1978, 1979) "thermophilic" arginines in LDH and "thermophilic" glutamic acids in ferredoxin consistent with the results shown in Table III. However, it was also possible to recognize (Zuber, 1978) "thermophilic" histidines and lysines in ferredoxin in agreement with the observations of Perutz & Raidt (1975) but

missing from the results in this paper. It is to be expected that, as more sequence data become available for a larger number of phylogenetically related thermophilic and mesophilic proteins, the results of the alternative approach used by Zuber and his colleagues (Zuber, 1978, 1979) will converge further with those of the statistical method reported here.

Acknowledgments

We are grateful for helpful discussions with Drs. S. Wodak and T. Richmond and to Sharon Wilder and Maria Victoria Manterola for help in the preparation of the manuscript.

References

Amelunxen, R. E., & Singleton, R., Jr. (1976) in *Enzymes and Proteins from Thermophilic Microorganisms* (Zuber, H., Ed.) pp 107–120, Birkhäuser Verlag, Basel.

Argos, P., Schwarz, J., & Schwarz, J. (1976) *Biochim. Biophys. Acta 439*, 261–273.

Biesecker, G., Harris, J. I., Thierry, J. C., Walker, J. E., & Wonacott, A. J. (1977) *Nature (London) 266*, 328–333.

Buchanan, R. E., & Gibbons, N. E., Eds. (1974) *Bergey's Manual of Determinative Bacteriology*, 8th ed., Williams and Wilkins Co., Baltimore, MD.

Buehner, M., Ford, G. C., Moras, D., Olsen, K. W., & Rossmann, M. G. (1974) *J. Mol. Biol. 90*, 25–49.

Bull, H. B., & Breese, K. (1973) *Arch. Biochem. Biophys. 158*, 681–686.

Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry 13*, 211–222.

Davidson, B. E., Sajgô, M., Noller, H. F., & Harris, J. I. (1967) *Nature (London) 216*, 1181–1185.

Dayhoff, M. O., Ed. (1979) *Atlas of Protein Sequence and Structure*, Vol. V, Suppl. 3, National Biomedical Research Foundation, Silver Spring, MD.

Eventoff, W., Rossmann, M. G., Taylor, S. S., Torff, H. J., Meyer, H., Keil, W., & Kiltz, H. H. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2677–2681.

Fasman, G. D., Chou, P. Y., & Adler, A. J. (1976) *Biophys. J. 16*, 1201–1238.

Feeney, R. E., & Osuga, D. T. (1976) *Comp. Biochem. Physiol. A 54*, 281–286.

Grütter, M. G., Hawkes, R. B., & Matthews, B. W. (1979) *Nature (London) 277*, 667–669.

Harris, J. I., & Perham, R. N. (1968) *Nature (London) 219*, 1025–1028.

Harris, J. I., & Walker, J. E. (1977) in *Pyridine Nucleotide-Dependent Dehydrogenases* (Sund, H., Ed.) pp 43–61, W. de Gruyter, Berlin.

Hochachka, P. W., & Somero, G. N. (1973) *Strategies of Biochemical Adaptation*, W. B. Saunders, Philadelphia, PA.

Janin, J., Wodak, S., Levitt, M., & Maigret, B. (1978) *J. Mol. Biol. 125*, 357–386.

Jones, G. M. T., & Harris, J. I. (1972) *FEBS Lett. 22*, 185–189.

Kiltz, H. H., Keil, W., Griesbach, M., Petry, K., & Meyer, H. (1977) *Hoppe-Seyler's Z. Physiol. Chem. 358*, 123–127.

Kretsinger, R. H. (1975) in *Calcium Transport in Contraction and Secretion* (Carafoli, E., Ed.) pp 469–478, North-Holland Publishing Co., Amsterdam.

Levitt, M. (1978) *Biochemistry 17*, 4277–4285.

Ljungdahl, L. G., Sherod, D. W., Moore, M. R., & Andreesen, J. R. (1976) in *Enzymes and Proteins from Thermophilic Microorganisms* (Zuber, H., Ed.) pp 237–248, Birkhäuser Verlag, Basel.

Low, P. S., & Somero, G. N. (1976) *J. Exp. Zool. 198*, 1–12.

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

Nojima, H., Ikai, A., Oshima, T., & Noda, H. (1977) *J. Mol. Biol. 116*, 429–442.

Nojima, H., Hon-nami, K., Oshima, T., & Noda, H. (1978) *J. Mol. Biol. 122*, 33–42.

Nozaki, Y., & Tanford, C. (1971) *J. Biol. Chem. 246*, 2211–2217.

Olsen, K. W., Moras, D., Rossmann, M. G., & Harris, J. I. (1975) *J. Biol. Chem. 250*, 9313–9321.

Perutz, M. F. (1978) *Science 201*, 1187–1191.

Perutz, M. F., & Raidt, H. (1975) *Nature (London) 255*, 256–259.

Rall, S. C., Bolinger, R. E., & Cole, R. D. (1969) *Biochemistry 8*, 2486–2496.

Rossmann, M. G., Liljas, A., Brändén, C. I., & Banaszak, L. J. (1975) *Enzymes, 3rd Ed. 11*, 61–102.

Stellwagen, E., & Wilgus, H. (1978) *Nature (London) 275*, 342–343.

Tanaka, M., Nakashima, T., Benson, A., Mower, H., & Yasunobu, K. T. (1966) *Biochemistry 5*, 1666–1681.

Tanaka, M., Haniu, M., Matsueda, G., Yasunobu, K. T., Himes, R. H., Akagi, J. M., Barnes, E. M., & Devanathan, T. (1971) *J. Biol. Chem. 246*, 3953–3960.

Tanaka, M., Haniu, M., Yasunobu, K. T., Himes, R. H., & Akagi, J. M. (1973) *J. Biol. Chem. 248*, 5215–5217.

Taylor, S. S. (1977) *J. Biol. Chem. 252*, 1799–1806.

Torchilin, V. P., Maksimenko, A. V., Smirnov, V. N., Berezin, I. V., Klibanov, A. M., & Martinek, K. (1979) *Biochim. Biophys. Acta 567*, 1–11.

Tsunoda, J. N., Yasunobu, K. T., & Whiteley, H. R. (1968) *J. Biol. Chem. 243*, 6262–6272.

Tuengler, P., & Pfleiderer, G. (1977) *Biochim. Biophys. Acta 484*, 1–8.

Veronese, F. M., Grandi, C., Boccù, E., & Fontana, A. (1976) in *Enzymes and Proteins from Thermophilic Microorganisms* (Zuber, H., Ed.) pp 147–155, Birkhäuser Verlag, Basel.

Wonacott, A. J., & Biesecker, G. (1977) in *Pyridine Nucleotide-Dependent Dehydrogenases* (Sund, H., Ed.) pp 140–156, W. de Gruyter, Berlin.

Yasunobu, K. T., & Tanaka, M. (1973) *Syst. Zool. 22*, 570–589.

Zimmerman, J. M., Aliezer, N., & Simha, R. (1968) *J. Theor. Biol. 21*, 170–201.

Zuber, H., Ed. (1976) *Enzymes and Proteins from Thermophilic Microorganisms*, Birkhäuser Verlag, Basel.

Zuber, H. (1978) in *Biochemistry of Thermophily* (Friedman, S. M., Ed.) pp 267–285, Academic Press, New York.

Zuber, H. (1979) in *Dahlem Workshop Report on "Strategy of Life in Extreme Environments"*, Abakon Verlagsgesellschaft, Berlin (in press).