Int. J. Peptide Protein Res. **43**, 1994, 97–106
Printed in Belgium – all rights reserved

INTERNATIONAL JOURNAL OF
PEPTIDE & PROTEIN RESEARCH
ISSN 0367-8377

Relevance of sequence statistics for the properties of extremophilic proteins

GERALD BÖHM and RAINER JAENICKE

Institute for Biophysics and Physical Biochemistry, University of Regensburg, Regensburg, Germany

Received 1 May, accepted for publication 21 June 1993

The amino acid composition of proteins from mesophilic and extremophilic organisms is commonly assumed to reflect the mechanisms of molecular adaptation to extremes of physical conditions. In this context, halophilic behaviour has been attributed to significantly increased numbers of aspartic and glutamic acid residues. However, extending the analysis to a statistically relevant set of related proteins, dihydrofolate reductase from *Halobacterium volcanii*, as an example, shows that the increase in negative charge is found to be less significant than other exchanges of amino acids (e.g., Ala, Asn, Arg, Lys, Phe, Ser). Thus, the high water binding capacity of negatively charged residues cannot be unambiguously correlated with the anomalous stability of halophilic proteins.

A similar caveat holds for generalizations regarding the thermal stability of proteins. In this case, D-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* was compared with a number of mesophilic and moderately thermophilic homologs. Again, 'traffic rules of stabilization', in terms of amino acid changes in going from mesophilic to thermophilic proteins, cannot be given. © Munksgaard 1994.

Key words: extremophilic proteins; halophily; molecular modeling; sequence statistics; thermophily

The molecular adaptation of proteins to extremes of physical conditions is still an unsolved problem. This holds for all relevant variables defining the limits of growth in the biosphere, i.e., temperature, pH, hydrostatic pressure and water activity (1, 2). In the case of halophiles (which require high salt concentration for metabolic activity and growth), one reason for this shortcoming is that no high-resolution structure of a 'halophilic protein'[1] has yet been solved. Thus, hypotheses have been based primarily on the amino acid composition, on one hand, and global molecular parameters, obtained from solution studies, on the other (3, 4). Applying ultracentrifugal analysis and neutron scattering, halophilic malate dehydrogenase (h-MDH) from *Halobacterium marismortui* was shown to exhibit preferential solvation and anomalous salt binding, allowing the protein to compete with salt for hydration, even at 5 M salt (water activity $a_w \approx 0.6$) (5). In interpreting the data, excess water binding capacity was attributed to an anomalous preference of negatively over positively charged groups in the halophilic proteins compared with a non-halophilic homolog. In addition, ion binding was proposed to modulate the rigidity of the molecule and the structure of external loops, contributing to a further increase in hydration (6).

As in the case of halophilism, thermophilic adaptation has been correlated with preferential alterations of the amino acid composition in terms of 'traffic rules' governing the gross traffic of amino acid changes in going from mesophilic to thermophilic sequences and *vice versa*. According to Argos *et al.* (7) and Menéndez-Arias & Argos (8), the replacement of lysine by arginine, and serine by alanine (predominantly in helical regions), as well as shifts from glycine to alanine, serine to threonine, and isoleucine to valine, should lead to increased thermal stability of proteins. As taken from the marginal difference in the free energies of stabilization of mesophilic and thermophilic proteins ($\Delta\Delta G_{stab} \approx 50$ kJ/mol), one would expect that thermal stability is achieved by the cumulative effect of small improvements at many locations within the protein molecule (1). Thus, a sound analysis of the detailed adaptive mechanisms of stabilization for a given system requires comparative studies of homologous proteins

DHFR, dihydrofolate reductase; GAPDH, D-glyceraldehyde-3-phosphate dehydrogenase

[1] For sake of shortness, 'halophilic proteins' or 'thermophilic proteins' will be used for proteins from halophilic or thermophilic organisms.

with varying thermal stabilities. In this context it has been shown that structural differences may be too small to be detectable, even at maximum resolution of crystal structure analysis (9, 10).

In order to investigate the significance of alterations in the amino acid composition of mesophilic *versus* extremophilic proteins, the following calculations present statistical analyses of the distributions of the 20 natural amino acids in dihydrofolate reductases and D-glyceraldehyde-3-phosphate dehydrogenase of varying stability. In summarizing the result, simple 'traffic rules' in terms of shifts in the amino acid composition are insufficient to explain the mechanisms of halophilic and thermophilic adaptation, or to predict the stabilization/destabilization of proteins.

MATERIALS AND METHODS

Dihydrofolate reductase (DHFR) sequences and D-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) sequences were selected using the HUSAR database system at the German Cancer Research Institute (DKFZ) in Heidelberg, Germany, release of 4 October, 1990. The criteria for the selection of the sample data set were as follows: (i) only complete sequences were allowed; (ii) sequences were taken from proteins of different organisms or at least different cell compartments; (iii) the difference in length of the polypeptide chains must not exceed 10% of the size of the respective extremophilic protein under investigation.

DHFR sequences used were as follows: *Bacillus subtilis* (11), *Bacteriophage T4* (12), *Bos taurus* (13), *Cricetulus sp.* (14), *Citrobacter freundii* (15), *Enterobacter aerogenes* (15), *Enterococcus faecium* (16), *Escherichia coli* (17), *Gallus gallus* (18), *Halobacterium volcanii* (19), *Homo sapiens* (20), *Herpesvirus saimiri* strain 11 (21), *Herpesvirus saimiri* type 2 (21), *Lactobacillus casei* (22), *Mesocricetus auratus* (23), *Mus musculus* (24), *Neisseria gonorrhoe* (25), *Pneumocystis carinii* (26), *Saccharomyces cerevisiae* (27), *Salmonella typhimurium* (28), *Staphylococcus aureus* (29), *Sus scrofa* (30).

The following GAPDH sequences were selected: *Bacillus coagulans* (31), *Bacillus stearothermophilus* (32), *Bacillus subtilis* (33), *Caenorhabditis elegans* (34), *Drosophila melanogaster* (35), *Escherichia coli* (36), *Gallus gallus* (37), *Homarus americanus* (38), *Homo sapiens* (35), *Methanobacterium bryantii* (39), *Methanobacterium formicicium* (39), *Methanothermus fervidus* (40), *Pisum sativum* (41), *Pyrococcus woesei* (42), *Rattus norvegicus* (35, 43), *Saccharomyces cerevisiae* (44), *Sinapis alba* (45), *Spinacia oleracea* (41), *Sus scrofa* (38), *Thermotoga maritima* (46), *Thermus aquaticus* (47), *Ustilago maydis* (48)[2], *Zea mays* (49), *Zea mays* (chloroplast) (50), *Zygosaccharomyces rouxii* (51), *Zymomonas mobilis* (52).

The statistical significance of the amino acid composition of a protein was calculated by a new method: The average composition of the proteins included in the statistical analysis, except the protein under investigation, is determined for each of the 20 natural amino acids. The data (with their standard deviations) are normalized with respect to the length of the polypeptide chains. Then, the deviation of the protein under consideration from the average composition is calculated and expressed in units of the standard deviation; positive or negative signs stand for higher or lower occurrences than average, respectively.

Let S_x be the sequence of size N of a protein from a class k of proteins of identical function, taken from the organism x of K different organisms. For each of the 20 amino acids A_i, the average probability function ν_i is determined according to

$$\nu_i = K^{-1}\Sigma\{N^{-1}\Sigma A(S_x)_i\}.$$

In addition, the corresponding standard deviations σ_i are calculated. For an extremophilic protein sequence Z_x of the size N_x, the significance index R_i for each amino acid is determined by

$$R_i = \pm \sigma_i^{-1}\{\nu_i - N_x^{-1}\Sigma A(Z_x)_i$$

RESULTS

Dihydrofolate reductase and halophilism

As demonstrated in Table 1, the selected dihydrofolate reductases are similar in size; the eukaryotic proteins show a higher molecular mass than their prokaryotic homologs. On the other hand, the amino acid composition characterized by the four traditional classes of amino acids (polar, apolar, positively charged, and negatively charged) demonstrates significant differences of the proteins and suggests low evolutionary relationships between the organisms. The detailed amino acid composition is shown in Table 2; again, the data reflect the low relationship of the proteins. However, the multiple alignment of the sequences shows that key positions of the sequences are conserved, thus suggesting similar structures and enzyme mechanisms (53).

Table 1 also includes the ratio of negatively over positively charged amino acids in the given sequences; the presently accepted model of halophilism is based on the assumption that the necessary water binding capacity comes from a high amount of negatively charged residues, thus allowing the protein to compete with the electrolyte for hydration. This shift is also observed for h-DHFR of *Halobacterium volcanii*. However, this does not imply that the altered ratio is statistically significant, nor does it mean that it is responsible for the anomalous stability properties of the protein. Compared with the *E. coli* enzyme, the increase in negative charges

[2] Not citable. The database entry 'g3p-ustma' of the SWISSPROT database has a wrong reference citation.

TABLE 1

Sequence-based statistical comparison of dihydrofolate reductases from 22 different organisms[a]

Organism	Amino acids	Molecular mass (Da)	Pos.	Neg.	Polar	Apolar	% Pos.	% Neg.	% Polar	% Apolar	Neg./ Pos.	Charged/ Apolar[b]
Bacillus subtilis	168	19.205	19	25	59	65	11.31	14.88	35.12	38.69	1.316	1.585
Bacteriophage T4	193	21.714	21	23	66	83	10.88	11.92	34.20	43.01	1.095	1.325
Bos taurus	186	21.473	26	26	52	82	13.98	13.98	27.96	44.09	1.000	1.268
Cricetulus sp.	187	21.661	27	26	53	81	14.44	13.90	28.34	43.32	0.963	1.309
Citrobacter freundii	159	18.029	16	26	45	72	10.06	16.35	28.30	45.28	1.625	1.208
Enterobacter aerogenes	159	17.975	16	27	43	73	10.06	16.98	27.04	45.91	1.688	1.178
Enterococcus faecium	167	19.584	21	26	48	72	12.57	15.57	28.74	43.11	1.238	1.319
Escherichia coli	159	18.000	15	25	46	73	9.43	15.72	28.93	45.91	1.667	1.178
Gallus gallus	189	21.651	26	25	57	81	13.76	13.23	30.16	42.86	0.962	1.333
Halobacterium volcanii	162	17.981	15	30	45	72	9.26	18.52	27.78	44.44	2.000	1.250
Homo sapiens	186	21.323	25	25	59	77	13.44	13.44	31.72	41.40	1.000	1.416
Herpesvirus saimiri strain 11	187	21.721	26	23	56	82	13.90	12.30	29.95	43.85	0.885	1.280
Herpesvirus saimiri type 2	197	22.847	29	24	58	86	14.72	12.18	29.44	43.65	0.828	1.291
Lactobacillus casei	162	18.325	17	22	52	71	10.49	13.58	32.10	43.83	1.294	1.282
Mesocricetus auratus	186	21.536	27	26	54	79	14.52	13.98	29.03	42.47	0.963	1.354
Mus musculus	186	21.464	26	27	53	80	13.98	14.52	28.49	43.01	1.038	1.325
Neisseria gonorrhoe	162	17.732	17	19	46	80	10.49	11.73	28.40	49.38	1.118	1.025
Pneumocystis carinii	206	23.885	29	24	69	84	14.08	11.65	33.50	40.78	0.828	1.452
Saccharomyces cerevisiae	211	24.248	29	28	64	90	13.74	13.27	30.33	42.65	0.966	1.344
Salmonella typhimurium	162	18.034	16	19	51	76	9.88	11.73	31.48	46.91	1.188	1.132
Staphylococcus aureus	161	18.462	16	19	61	65	9.94	11.80	37.89	40.37	1.188	1.477
Sus scrofa	186	21.456	29	26	56	75	15.59	13.98	30.11	40.32	0.897	1.480

[a] Data show that the homologous DHFRs have approximately the same size, the eukaryotic proteins being slightly larger than the prokaryotic ones. The ratio of negatively versus positively charged residues is increased in the case of h-DHFR from *Halobacterium volcanii*, in agreement with the general theory on halophilic adaptation.

[b] Ratio of charged plus polar residues to apolar residues.

is five. We were able to demonstrate that the additional charges are clustered rather than randomly distributed over the surface of the protein, suggesting that they have a specific function apart from increased water binding (Böhm & Jaenicke, submitted).

The amino acid compositions summarized in Table 2 illustrate a wide range of possibilities which still form the unique DHFR topology. Obviously, the high variabilities in the sequences cause only minor structural alterations. Thus, the data are well suited for a comparative investigation of sequence statistics.

Table 3 indicates the average composition for the 20 common types of amino acids for all 22 dihydrofolate reductases in the primary database. Alanine and lysine show the highest degree of variability (highest standard deviation and variance). For a measurement of the statistical significance of the amino acid exchanges, the algorithm described under Materials and Methods was applied. The results, summarized in Table 4, indicate for the various types of amino acids in each of the sequences the statistical significance of the increase (positive sign) or decrease (negative sign) of the respective amino acids compared with the database of the other 21 proteins.

The most significant changes for the h-DHFR from *Halobacterium volcanii* are the increased arginine and the decreased lysine contents. Large differences are also found for alanine, phenylalanine, serine and asparagine. In the sequence of their statistical significance, aspartic and glutamic acid only follow on rank seven and eight, thus indicating that from the point of view of statistics the increase in negative charges is of minor importance. Table 4 also demonstrates that non-halophilic sequences exhibit divergences for several (charged and uncharged) amino acids which are of the same order of magnitude as the changes in the h-DHFR sequence. This confirms that sequence-based statistics are not suitable for the prediction of halophilic properties of a protein.

It might be argued that the low evolutionary relationship between the selected proteins in the database does not allow a sound comparison. Therefore, a subset of 10 prokaryotic sequences from the complete database of 22 was selected and investigated by the previously described method. The resulting data, which are summarized in Table 5, indicate the same tendencies for the subset as have been described before for the whole database; here, the significance of the negatively charged amino acids ranks at positions six and nine in the subset, instead of seven and eight as in the complete database. The conservation of glycine in the subset is slightly higher than in the complete database; however,

TABLE 2

Relative amino acid composition of the 22 dihydrofolate reductases selected for this work[a]

Organism	Ala	Val	Leu	Ile	Pro	Met	Phe	Trp	Gly	Ser	Thr	Cys	Tyr	Asn	Gln	Asp	Glu	Lys	Arg	His
Bacillus subtilis	3.6	4.8	7.1	6.5	5.4	3.0	7.1	1.2	7.1	8.3	4.2	1.8	4.2	4.8	1.8	7.7	7.1	7.7	3.6	3.0
Bacteriophage T4	5.7	9.3	8.8	6.2	4.7	2.6	4.1	1.6	6.7	6.7	7.8	0.5	4.7	2.6	3.1	6.7	5.2	6.2	4.7	2.1
Bos taurus	3.2	9.1	8.6	5.9	7.5	2.7	5.4	1.6	5.9	4.8	2.7	0.5	3.2	5.9	3.8	4.8	9.1	9.1	4.8	1.1
Cricetulus sp.	3.2	7.5	9.1	5.9	6.4	4.3	5.3	1.6	6.4	5.3	3.2	0.5	3.2	4.3	4.3	4.3	9.6	9.6	4.8	1.1
Citrobacter freundii	8.2	7.5	6.3	6.9	6.3	3.1	3.8	3.1	6.3	5.7	3.1	1.3	2.5	3.8	2.5	8.2	8.2	4.4	5.7	3.1
Enterobacter aerogenes	8.8	7.5	6.9	6.9	5.7	3.1	3.8	3.1	6.3	6.3	2.5	1.3	2.5	3.1	1.9	8.8	8.2	4.4	5.7	3.1
Enterococcus faecium	3.6	5.4	9.0	9.0	3.6	4.2	6.0	2.4	6.6	4.2	4.8	0.0	3.0	3.0	4.2	6.6	9.0	7.8	4.8	3.0
Escherichia coli	8.2	6.9	6.9	7.5	6.3	3.1	3.8	3.1	6.3	5.7	3.8	1.3	2.5	3.8	2.5	8.2	7.5	3.8	5.7	3.1
Gallus gallus	5.8	7.4	9.0	6.3	5.8	2.6	4.2	1.6	5.3	6.3	3.7	0.5	4.2	4.2	3.7	5.8	7.4	9.5	4.2	2.1
Halobacterium volcanii	13.0	8.0	6.8	4.3	4.9	2.5	2.5	2.5	5.6	9.3	4.3	0.0	3.7	0.6	2.5	8.6	9.9	1.2	8.0	1.9
Homo sapiens	2.7	7.5	10.2	4.8	6.5	3.2	4.8	1.6	7.0	6.5	3.8	0.5	3.2	5.4	3.8	4.8	8.6	9.1	4.3	1.6
Herpesvirus saimiri strain 11	4.3	7.5	10.2	5.3	5.3	4.3	5.3	1.6	3.7	5.9	3.7	1.6	3.2	5.3	4.3	4.8	7.5	10.7	3.2	2.1
Herpesvirus saimiri type 2	4.1	7.6	10.7	5.6	5.1	4.1	5.1	1.5	3.6	6.6	3.6	1.5	3.0	5.1	4.1	4.6	7.6	11.7	3.0	2.0
Lactobacillus casei	9.3	9.9	8.6	3.1	4.3	1.2	4.9	2.5	6.2	2.5	8.6	0.0	3.1	1.9	5.6	9.3	4.3	5.6	4.9	4.3
Mesocricetus auratus	3.2	7.5	9.1	5.9	6.5	3.2	5.4	1.6	6.5	5.4	3.2	0.5	3.2	4.3	4.3	4.3	9.7	9.7	4.8	1.6
Mus musculus	3.2	7.5	9.7	5.9	7.0	3.2	4.8	1.6	6.5	5.9	3.2	0.5	3.2	4.3	3.8	4.8	9.7	8.1	5.9	1.1
Neisseria gonorrhoe	13.6	7.4	7.4	7.4	4.9	3.1	3.7	1.9	8.0	3.7	5.6	2.5	3.1	1.9	1.9	3.7	8.0	4.3	6.2	1.9
Pneumocystis carinii	3.4	6.3	9.2	7.3	3.4	2.9	5.3	2.9	5.8	9.2	5.3	0.5	3.4	3.9	2.4	5.8	5.8	8.7	5.3	2.9
Saccharomyces cerevisiae	4.3	5.2	8.1	7.6	7.1	3.3	5.7	1.4	5.7	6.2	4.3	1.4	2.8	5.2	3.3	5.2	8.1	8.5	5.2	1.4
Salmonella typhimurium	10.5	5.6	13.6	4.9	6.2	1.9	2.5	1.9	6.8	4.9	4.3	1.2	3.1	4.3	3.7	6.2	5.6	2.5	7.4	3.1
Staphylococcus areus	2.5	8.1	8.7	7.5	4.3	2.5	5.6	1.2	6.8	4.3	7.5	0.0	2.5	6.2	5.6	6.2	5.6	6.2	3.7	5.0
Sus scrofa	2.7	7.5	8.6	5.9	6.5	3.2	4.3	1.6	5.9	6.5	3.8	1.1	3.2	5.4	3.2	4.8	9.1	10.8	4.8	1.1

[a] The values demonstrate a high variability with respect to the composition of the various homologous proteins, thus indicating a low evolutionary relationship of the proteins.

this aspect does not seem to be relevant in the present context. The lower values in Table 5 (prokaryotic organisms) compared to Table 4 (prokaryotic *and* eukaryotic organisms) stress the close relationship between the prokaryotic organisms; the differences between each of the entries in Tables 4 and 5 also indicate the sensitivity of the absolute values with respect to the underlying database. Again, it is obvious that the pairwise comparison of the sequence composition of halophilic and non-halophilic proteins may cause erroneous results because a large database reveals different results compared with a small one.

Glyceraldehyde-3-phosphate dehydrogenase and thermophilism

A similar conclusion may be drawn if the comparison algorithm is applied to thermophilic and mesophilic proteins. In this case, sequence data of 26 GAPDHs from several archaea, bacteria and eukarya were investigated. Data are summarized in Table 6. However, there is a fundamental difference between halophilism and thermophilism: Temperature adaptation is ambivalent, in that there is no well defined separation of the various ranges of thermal adaptation. Thus, in determining the significance of the amino acid composition of a specific protein, one has to consider that 'thermophilic behavior' does not refer to a well defined property.

As in the case of DHFR, the data in Table 6 dem-

TABLE 3

Average amino acid composition of the 22 dihydrofolate reductases[a]

Amino acid	Average occurrence (%)	Standard deviation	Variance	Standard error of estimate
Alanine	5.77	3.42	11.73	0.73
Valine	7.33	1.27	1.64	0.27
Leucine	8.75	1.61	2.58	0.34
Isoleucine	6.22	1.29	1.69	0.28
Proline	5.62	1.12	1.25	0.24
Methionine	3.06	0.74	0.55	0.16
Phenylalanine	4.71	1.11	1.24	0.24
Tryptophan	1.96	0.64	0.41	0.14
Glycine	6.13	1.00	1.00	0.21
Serine	5.92	1.62	2.63	0.35
Threonine	4.41	1.64	2.68	0.35
Cysteine	0.87	0.66	0.44	0.14
Tyrosine	3.22	0.56	0.31	0.12
Asparagine	4.05	1.42	2.01	0.30
Glutamine	3.46	1.07	1.15	0.23
Aspartic acid	6.11	1.68	2.83	0.36
Glutamic acid	7.76	1.60	2.58	0.34
Lysine	7.26	2.87	8.26	0.61
Arginine	5.04	1.20	1.46	0.26
Histidine	2.35	1.05	1.12	0.23

[a] Data are given in percent, relative to the size of the respective protein.

TABLE 4

Statistical significance, R, *of deviations from the average amino acid composition for the 22 dihydrofolate reductases*[a]

Organism	Ala	Val	Leu	Ile	Pro	Met	Phe	Trp	Gly	Ser	Thr	Cys	Tyr	Asn	Gln	Asp	Glu	Lys	Arg	His
Bacillus subtilis	−0.7	−2.3	−1.1	+0.3	−0.2	−0.1	+2.6	−1.3	+1.1	+1.6	−0.1	+1.5	+1.9	+0.5	−1.7	+1.0	−0.4	+0.2	−1.3	+0.6
Bacteriophage T4	0.0	+1.7	0.0	0.0	−0.9	−0.7	−0.5	−0.7	+0.6	+0.5	+2.4	−0.5	+3.2	−1.1	−0.3	+0.4	−1.8	−0.4	−0.3	−0.3
Bos taurus	−0.8	+1.5	−0.1	−0.2	+1.9	−0.5	+0.6	−0.6	−0.2	−0.7	−1.1	−0.5	0.0	+1.4	+0.3	−0.8	+0.9	+0.7	−0.2	−1.3
Cricetulus sp.	−0.8	+0.1	+0.2	−0.3	+0.7	+1.8	+0.6	−0.6	+0.3	−0.4	−0.8	−0.5	0.0	+0.2	+0.8	−1.1	+1.2	+0.9	−0.2	−1.3
Citrobacter freundii	+0.7	+0.2	−1.7	+0.6	+0.6	+0.1	−0.9	+2.1	+0.2	−0.2	−0.8	+0.6	−1.3	−0.2	−0.9	+1.3	+0.3	−1.0	+0.5	+0.8
Enterobacter aerogenes	+0.9	+0.2	−1.2	+0.6	0.0	+0.1	−0.9	+2.1	+0.2	+0.2	−1.2	+0.6	−1.3	−0.7	−1.6	+1.8	+0.3	−1.0	+0.5	+0.8
Enterococcus faecium	−0.7	−1.6	+0.1	+2.5	−2.0	+1.7	+1.2	+0.7	+0.5	−1.1	+0.2	−1.4	−0.4	−0.8	+0.7	+0.3	+0.8	+0.2	−0.2	+0.6
Escherichia coli	+0.7	−0.3	−1.2	+1.1	+0.6	+0.1	−0.9	+2.1	+0.2	−0.2	−0.4	+0.6	−1.3	−0.2	−0.9	+1.3	−0.1	−1.3	+0.5	+0.8
Gallus gallus	0.0	+0.1	+0.2	+0.1	+0.2	−0.6	−0.4	−0.6	−0.9	+0.3	−0.4	−0.5	+2.0	+0.1	+0.2	−0.2	−0.2	+0.8	−0.7	−0.2
Halobacterium volcanii	+2.4	+0.6	−1.3	−1.6	−0.6	−0.8	−2.3	+0.8	−0.6	+2.4	−0.1	−1.4	+0.9	−2.9	−1.0	+1.6	+1.4	−2.4	+3.0	−0.5
Homo sapiens	−0.9	+0.2	+1.0	−1.1	+0.8	+0.2	+0.1	−0.6	+0.9	+0.3	−0.4	−0.5	0.0	+1.0	+0.3	−0.8	+0.5	+0.7	−0.6	−0.7
Herpesvirus saimiri strain 11	−0.4	+0.1	+0.9	−0.7	−0.2	+1.8	+0.6	−0.6	−2.9	0.0	−0.4	+1.2	0.0	+1.0	+0.8	−0.8	−0.2	+1.3	−1.6	−0.2
Herpesvirus saimiri type 2	−0.5	+0.2	+1.3	−0.5	−0.5	+1.4	+0.3	−0.7	−3.2	+0.4	−0.5	+1.0	−0.3	+0.7	+0.6	−1.0	−0.1	+1.7	−1.8	−0.3
Lactobacillus casei	+1.1	+2.3	−0.1	−3.0	−1.2	−3.0	+0.2	+0.8	0.0	−2.5	+3.2	−1.4	−0.2	−1.7	+2.2	+2.1	−2.5	−0.6	−0.1	+2.1
Mesocricetus auratus	−0.8	+0.2	+0.2	−0.2	+0.8	+0.2	+0.6	−0.6	+0.3	−0.3	−0.7	−0.5	0.0	+0.2	+0.8	−1.1	+1.3	+0.9	−0.2	−0.7
Mus musculus	−0.8	+0.2	+0.6	−0.2	+1.3	+0.2	+0.1	−0.6	+0.3	0.0	−0.7	−0.5	0.0	+0.2	+0.3	−0.8	+1.3	+0.3	+0.8	−1.3
Neisseria gonorrhoe	+2.7	+0.1	−0.9	+1.0	−0.6	0.0	−0.9	−0.2	+2.1	−1.5	+0.7	+2.9	−0.2	−1.7	−1.6	−1.5	+0.2	−1.1	+1.0	−0.5
Pneumocystis carinii	−0.7	−0.8	+0.3	+0.9	−2.3	−0.2	+0.6	+1.6	−0.3	+2.3	+0.6	−0.6	0.3	−0.1	−1.0	−0.2	−1.3	+0.5	+0.3	+0.5
Saccharomyces cerevisiae	−0.5	−1.8	−0.4	+1.1	+1.4	+0.4	+0.9	−0.9	−0.5	+0.2	−0.1	+0.9	−0.7	+0.8	−0.1	−0.5	+0.2	+0.5	+0.1	−0.9
Salmonella typhimurium	+1.5	−1.5	+4.1	−1.0	+0.5	−1.8	−2.3	−0.2	+0.7	−0.6	−0.1	+0.6	−0.2	+0.2	+0.2	0.0	−1.5	−1.8	+2.2	+0.7
Staphylococcus aureus	−1.0	+0.6	0.0	+1.0	−1.2	−0.8	+0.8	−1.2	+0.7	−1.0	+2.1	−1.4	−1.4	+1.7	+2.3	+0.1	−1.5	−0.4	−1.1	+3.0
Sus scrofa	−0.9	+0.2	−0.1	−0.2	+0.8	+0.2	−0.4	−0.6	−0.2	+0.3	−0.4	+0.3	0.0	+1.0	−0.2	−0.8	+0.9	+1.3	−0.2	−1.3

[a] Values are given in units of standard deviations; positive and negative signs measure high and low abundances of the respective amino acid.

TABLE 5

Statistical significance R *of the deviations from the average amino acid composition for 10 prokaryotic sequences of dihydrofolate reductases*[a]

Organism	Ala	Val	Leu	Ile	Pro	Met	Phe	Trp	Gly	Ser	Thr	Cys	Tyr	Asn	Gln	Asp	Glu	Lys	Arg	His
Bacillus subtilis	−1.4	−1.9	−0.5	+0.1	+0.2	+0.3	+2.5	−1.8	+0.9	+1.6	−0.4	+1.1	+3.1	+1.0	−1.1	+0.2	−0.1	+1.7	−1.7	−0.2
Citrobacter freundii	0.0	+0.3	−1.0	+0.3	+1.4	+0.5	−0.4	+1.3	−0.5	+0.1	−1.0	+0.4	−1.0	+0.3	−0.5	+0.5	+0.5	−0.2	+0.1	0.0
Enterobacter aerogenes	+0.2	+0.3	−0.6	+0.3	+0.5	+0.5	−0.4	+1.3	−0.5	+0.4	−1.4	+0.4	−1.0	−0.1	−1.0	+1.0	+0.5	−0.2	+0.1	0.0
Enterococcus feacium	−1.3	−1.3	+0.4	+1.8	−2.3	+2.3	+1.2	+0.1	0.0	−0.7	0.0	−1.2	0.0	−0.2	+0.7	−0.5	+1.1	+1.7	−0.6	−0.2
Escherichia coli	0.0	−0.1	−0.6	+0.7	+1.4	+0.5	−0.4	+1.3	−0.5	+0.1	−0.6	+0.4	−1.0	+0.3	−0.5	+0.5	+0.1	−0.5	+0.1	0.0
Halobacterium volcanii	+1.5	+0.6	−0.7	−1.4	−0.3	−0.4	−1.4	+0.3	−2.0	+2.5	−0.3	−1.2	+1.4	−2.2	−0.5	+0.8	+1.8	−2.2	+2.3	−1.6
Lactobacillus casei	+0.3	+2.5	+0.3	−2.6	−1.0	−2.6	+0.4	+0.3	−0.7	−1.8	+2.9	−1.2	+0.1	−1.0	+2.0	+1.3	−2.3	+0.4	−0.5	+1.4
Neisseria gonorrhoe	+1.7	+0.2	−0.4	+0.6	−0.3	+0.4	−0.5	−0.6	+3.4	−0.9	+0.4	+2.3	+0.1	−1.0	−1.0	−3.5	+0.4	−0.2	+0.5	−1.6
Salmonella typhimurium	+0.7	−1.1	+6.2	−0.9	+1.2	−1.3	−1.4	−0.6	+0.3	−0.3	−0.3	+0.4	+0.1	+0.7	+0.4	−0.8	−1.2	−1.3	+1.5	−0.1
Staphylococcus aureus	−1.8	+0.7	+0.3	+0.6	−1.0	−0.4	+0.9	−1.7	+0.4	−0.6	+1.6	−1.2	−1.0	+2.4	+2.1	−0.7	−1.1	+0.7	−1.5	+2.7

[a] Units are as in Table 4. Values demonstrate the close evolutionary relationship of the prokaryotic organisms; values for the h-DHFR from *Halobacterium volcanii* show that the differences in the occurrence of negatively charged residues are large, but not the most prominent exchanges in the overall compositions.

onstrate that the proteins selected have the same size. The multiple alignment of the sequences (53) indicates that the proteins are closely related and may be assumed to possess a similar fold and quaternary structure. The amino acid compositions in Table 6 support this view. In contrast to the given similarities, the wide range of the growth optima of the organisms (Table 6) implicates a large diversity of the thermal stabilities of their protein inventory. Assuming that composition-based rules for the molecular adaptation of proteins to high temperatures do exist, these rules should be highlighted by the sequence-statistics algorithm described in Materials and Methods.

In the present statistical analysis, GAPDH from the hyperthermophilic bacterium *Thermotoga maritima* was investigated. The most significant differences in composition are isoleucine (+2.8), threonine (+2.2), methionine (−1.6), serine (−1.6), alanine (−1.4) and glutamine (−1.4) (Table 7). However, when GAPDH from the hyperthermophilic archaeon *Pyrococcus woesei* is considered, the most significant differences are (again) isoleucine (+3.2), glutamic acid (+2.6), aspartic acid (−2.4), glycine (−2.2) and tyrosine (+1.9). Thermal stability (when discussed in terms of the amino acid composition) is obviously not a strictly conserved property when functionally related proteins with a considerable distance in evolution are compared.

A similar result is obtained for GAPDH from *Thermus aquaticus*, a closely related bacterium with an optimum growth temperature 10 °C below that observed for *Thermotoga*. In this case the most significant exchanges are leucine (+2.2), serine (−2.1), alanine (+1.9), cysteine (−1.7), histidine (+1.6) and valine (−1.5), which are completely different from the previously mentioned differences for *Thermotoga maritima* and *Pyrococcus woesei*. This demonstrates clearly that 'traffic rules of thermophilic adaptation' cannot be defined in terms of significant differences in the amino acid composition; instead, local three-dimensional interactions (which may result from a single exchange that cannot be highlighted by sequence statistics) are the only relevant aspect in the discussion of temperature adaptation.

It may be argued that the amino acid composition of thermophilic proteins is a misleading database since there is no definition of a binary property 'thermophilism' *versus* 'non-thermophilism'. Therefore, composition data have been investigated further in terms of the growth optimum of various organisms which may be taken as a crude approximation for the thermostability of the protein inventory. Figure 1 demonstrates that there are practically no significant preferences in the amino acid compositions of the thermophilic proteins compared to the mesophilic ones; this holds for all amino acids except glycine, isoleucine and serine. A significantly different composition would imply that the points in Fig. 1 at high temperatures should deviate from the distribution at low temperatures, whatever

TABLE 6

Statistical properties of the sequences of GAPDHs from 26 different organisms[a]

Organism	Amino acids	Molecular mass (Da)	% Pos.	% Neg.	% Polar	% Apolar	Neg./ Pos.	Charged/ Apolar	Optimal growth (°C)
Bacillus coagulans	335	35.675	9.85	11.64	30.75	47.76	1.182	1.094	41
Bacillus stearothermophilus	335	35.842	11.04	11.94	31.34	45.67	1.081	1.190	58
Bacillus subtilis	335	35.835	10.15	13.13	33.13	43.58	1.294	1.295	37
Caenorhabditis elegans	341	36.429	10.56	10.26	35.48	43.70	0.972	1.289	20
Drosophila melanogaster	332	35.318	10.84	10.24	34.34	44.58	0.944	1.243	20
Escherichia coli	331	35.510	11.78	12.08	34.74	41.39	1.026	1.416	37
Gallus gallus	333	35.786	11.11	9.91	34.83	44.14	0.892	1.265	40
Homarus americanus	333	35.719	11.11	11.71	33.33	43.84	1.054	1.281	20
Homo sapiens	335	36.055	10.75	9.85	36.72	42.69	0.917	1.343	37
Methanobacterium bryantii	338	37.029	10.95	14.50	34.02	40.53	1.324	1.467	38
Methanobacterium formicicum	338	37.223	11.54	15.09	33.14	40.24	1.308	1.485	41
Methanothermus fervidus	337	37.410	11.87	16.02	29.38	42.73	1.350	1.340	85
Pisum sativum	405	43.440	12.10	10.62	34.07	43.21	0.878	1.314	26
Pyrococcus woesei	334	37.429	12.87	14.07	29.34	43.71	1.093	1.288	102
Rattus norvegicus	333	35.838	11.11	10.21	35.44	43.24	0.919	1.313	37
Saccharomyces cerevisiae	331	35.602	11.18	11.78	34.44	42.60	1.054	1.348	35
Sinapis alba	337	36.795	12.76	12.46	32.05	42.73	0.977	1.340	20
Spinacia oleracea	451	48.129	11.09	11.31	33.04	44.57	1.020	1.244	20
Sus scrofa	332	35.711	10.84	11.14	34.94	43.07	1.028	1.322	40
Thermotoga maritima	332	36.296	12.35	13.25	30.72	43.67	1.073	1.290	80
Thermus aquaticus	331	35.894	11.78	12.39	29.00	46.83	1.051	1.135	70
Ustilago maydis	337	35.915	9.79	9.50	37.09	43.62	0.970	1.293	25
Zea mays	337	36.525	11.57	12.17	32.94	43.32	1.051	1.308	34
Zea mays (Chloroplast)	405	43.023	10.37	10.37	35.31	43.95	1.000	1.275	34
Zygosaccharomyces rouxii	333	35.603	10.51	11.41	35.44	42.64	1.086	1.345	25
Zymomonas mobilis	337	36.104	10.98	11.87	33.53	43.62	1.081	1.293	25

[a] Data refer to monomers; since there is a close relationship between the proteins, all GAPDHs may be assumed to be homotetramers in their native state. The last column indicates the optimal growth tmperature (53).

(arbitrary) boundary between 'high' and 'low' temperature is chosen; at least, tendencies for a different distribution of the composition at low and high temperatures should be observed if rules were deducible from composition data.

For serine, the value for the hyperthermophilic enzyme from *Thermus aquaticus* (3.9%) is found to be lower than in any of the mesophilic GAPDHs. The other hyperthermophilic organisms show values identical with *E. coli* (4.5%) which is definitely not a hyperthermophilic organism; thus an unambiguous rule based on serine cannot be defined. In the case of glycine, the data for the hyperthermophilic bacteria (*Thermotoga maritima* and *Thermus aquaticus*) are within the range of variability observed for the mesophilic proteins, whereas the hyperthermophilic archaea (*Pyrococcus woesi* and *Methanothermus fervidus*) deviate from this range. Therefore, a low glycine content cannot be taken as a criterion for thermophilism, although the low rotational energy barrier of the glycine backbone might support this idea. Here, more data for hyperthermophilic proteins from both evolutionary kingdoms, bacteria and archaea, would be required for a statistically reliable conclusion.

For isoleucine, the results are similar to those for serine: The GAPDHs from archaea (*Methanothermus fervidus* and *Pyrococcus woesi*) and from the bacterium *Thermotoga maritima* show a high isoleucine content, whereas the enzyme from the hyperthermophilic bacterium *Thermus aquaticus* does not differ from its mesophilic homologs. Obviously, the evolutionary aspect is not the dominant factor in this case. Again, a general rule for thermophilism cannot be deduced.

CONCLUSIONS

Data presented in this work have been used to demonstrate that in the case of halophilic DHFR from *Halobacterium volcanii*, and GAPDHs from a variety of hyperthermophiles, the amino acid compositions do not reveal unambiguous rules for the molecular adaptation to the respective extreme conditions. The information content in the amino acid compositions is insufficient to confirm the hypothesis that in halophilic DHFR increased water binding capacity is accomplished by additional negative charges; similarly, the discrimination between thermophilic and mesophilic GAPDHs cannot be generalized in terms of well defined shifts in specific amino acids. Proposed 'traffic rules of molecular adaptation' cannot be used to predict extremophilic

TABLE 7

Significance, R, *of the deviations from the average amino acid composition of the 26 GAPDH sequences, in units of standard deviation (cf. Table 4 and Materials and Methods)*

Organism	Ala	Val	Leu	Ile	Pro	Met	Phe	Trp	Gly	Ser	Thr	Cys	Tyr	Asn	Gln	Asp	Glu	Lys	Arg	His
Bacillus coagulans	+3.2	+2.7	+0.4	+1.0	−4.4	+0.1	−1.9	−1.0	−1.2	−1.1	−1.2	−0.9	+0.1	+1.4	−1.6	−2.0	+0.9	−2.2	+0.9	+0.9
Bacillus stearotherm.	+1.4	+2.7	+0.9	−0.7	−0.7	−0.3	−1.9	−1.0	−0.8	−1.1	−0.9	−0.9	−0.4	+1.8	−1.6	−2.4	+1.2	−0.6	+0.5	+0.9
Bacillus subtilis	+1.2	+0.1	+1.2	−1.3	−1.2	+0.5	−1.5	−1.0	−0.5	−0.1	−2.0	−0.9	+0.1	+3.4	−0.4	−0.8	+1.2	−0.6	−0.6	−0.7
Caenorhabditis elegans	−0.1	+1.2	−1.5	0.0	+0.6	−0.4	+0.7	−0.1	+0.5	+0.5	+0.4	−0.2	+1.1	−0.2	+0.7	+0.9	−1.3	+0.5	−1.4	−0.2
Drosophila melanogaster	+0.5	−0.1	−0.6	−0.1	−0.2	−0.3	+1.5	0.0	+1.1	+0.4	+0.6	−0.2	+0.2	+0.2	−1.0	−0.3	−0.7	+0.4	−0.9	−1.8
Escherichia coli	−0.1	−0.4	−0.6	−0.7	−1.6	+0.5	+0.2	0.0	+1.8	−1.6	+2.2	−0.2	−0.3	+0.5	−0.4	+0.8	−0.2	+0.8	−0.1	−0.6
Gallus gallus	+0.3	+0.5	−0.9	−0.7	−0.2	+0.9	+0.8	0.0	+1.0	−0.6	−0.5	+0.5	+0.2	+0.2	−0.4	−0.7	−0.7	+0.4	−0.6	+2.1
Homarus americanus	−0.1	+0.8	−1.1	−1.0	−0.2	+0.9	+1.5	0.0	+0.7	+0.9	−0.5	+1.3	+0.2	−2.1	+0.8	−0.4	+0.2	+1.0	−1.3	−1.2
Homo sapiens	−0.3	−1.3	−0.9	+0.2	−0.2	+0.9	+1.1	0.0	+1.6	−0.1	−0.2	−0.2	+0.1	+0.8	+0.8	−1.2	−0.5	+0.4	−1.0	+1.5
Methanobact. bryantii	−2.4	−1.7	+0.9	+0.1	+1.1	+2.8	−1.2	−2.1	+0.3	−0.4	−0.9	+0.5	+0.6	+0.4	+2.0	−0.1	+1.7	−0.3	+0.1	−0.7
Methanobact. formicicum	−2.1	−2.5	+0.6	+0.1	+1.1	+2.8	−0.9	−2.1	−0.3	−0.4	−0.9	+0.5	+0.6	+0.4	+1.3	−0.5	+2.3	+0.3	+0.1	−0.7
Methanothermus fervidus	−1.5	−1.0	−0.9	+3.1	+1.6	+1.8	−1.2	−2.1	−2.2	−1.6	−0.9	−0.2	+1.2	−0.5	+1.3	+1.4	+1.9	+0.3	+0.5	−0.2
Pisum sativum	−0.2	+0.5	+1.1	−1.0	−0.4	−1.9	+0.6	+1.1	−0.1	+1.4	+0.3	+0.6	−2.2	−0.5	+0.1	+0.9	−1.1	−0.3	+1.5	+0.1
Pyrococcus woesei	−1.1	−1.2	−0.6	+3.2	+0.2	−0.3	+0.8	0.0	−2.2	−1.6	−0.9	−1.7	+1.9	−0.1	+1.4	−2.4	+2.6	+1.3	+0.6	−0.7
Rattus norvegicus	−0.1	−0.5	−1.1	+0.2	+0.3	+0.9	+0.5	0.0	+1.0	−0.6	−0.5	+1.3	+1.3	+1.2	+0.2	−0.7	−0.5	+0.7	−0.9	−0.7
Saccharomyces cerevis.	−0.1	+0.6	−0.3	−0.4	−0.2	−0.7	−0.2	0.0	−0.4	+1.2	+1.0	−0.9	+1.3	−1.4	−0.4	+0.8	−0.3	+0.4	−0.5	+0.4
Sinapis alba	−1.1	+0.4	−0.4	+0.1	−0.7	−0.3	+1.1	+2.0	0.0	+0.1	+0.1	−0.9	−0.4	−0.5	−0.4	+0.6	+0.1	+1.9	−0.2	−0.7
Spinacia oleracea	+0.1	+0.2	+1.4	−1.1	+1.5	−1.7	+0.2	+0.7	−0.3	+1.1	−0.5	+2.0	−1.6	−0.7	+0.1	+1.1	−0.7	−0.8	+0.9	−0.7
Sus scrofa	−0.1	−0.5	−1.1	−0.1	−0.2	+0.5	+1.1	0.0	+1.4	−0.6	+0.3	+0.5	+0.2	−1.1	+0.2	+0.8	−0.7	+0.4	−0.9	+2.1
Thermotoga maritima	−1.4	+0.2	+1.0	+2.8	+0.8	−1.6	−0.8	−1.0	−1.1	−1.6	+2.2	−0.2	+0.2	−0.7	−1.6	−0.7	+1.2	+0.7	+0.6	−0.1
Thermus aquaticus	+1.9	−1.5	+2.2	+0.5	−0.2	−0.3	−1.2	0.0	−1.1	−2.1	+0.3	−1.7	+0.8	−1.4	−1.0	−0.3	+0.5	−0.5	+1.4	+1.6
Ustilago maydis	+0.8	−0.3	−1.2	+0.4	−0.7	−0.3	+0.4	+0.9	+0.3	+1.3	−0.6	+0.5	+0.7	+1.7	+0.2	−0.8	−0.9	−0.6	−1.0	−0.2
Zea mays	−0.7	0.0	−0.7	+0.1	−0.3	+0.1	+1.1	+2.0	+0.6	+0.3	+0.5	−0.9	+0.1	−1.1	−1.6	+1.0	−0.2	+0.9	−0.6	+0.4
Zea mays (chloroplast)	−0.2	−0.3	+1.3	−1.0	+1.5	−0.5	−0.5	+1.1	+0.9	+1.4	−0.3	+1.2	−2.2	−0.2	+2.1	+0.9	−1.2	−2.4	+1.9	−1.3
Zygosaccharomyc. rouxii	−0.1	−0.2	−0.9	−0.1	−0.2	+0.5	+0.1	0.0	+0.1	+1.7	+1.0	−0.9	+0.2	−0.7	−1.0	+0.4	−0.3	+0.4	−1.3	+0.9
Zymomonas mobilis	+0.4	−0.6	+1.2	−0.8	+0.2	+0.1	−0.5	−1.0	−0.9	+0.1	+4.0	−0.2	−2.2	−1.1	+0.2	+1.4	−0.5	−2.2	+2.6	+0.4

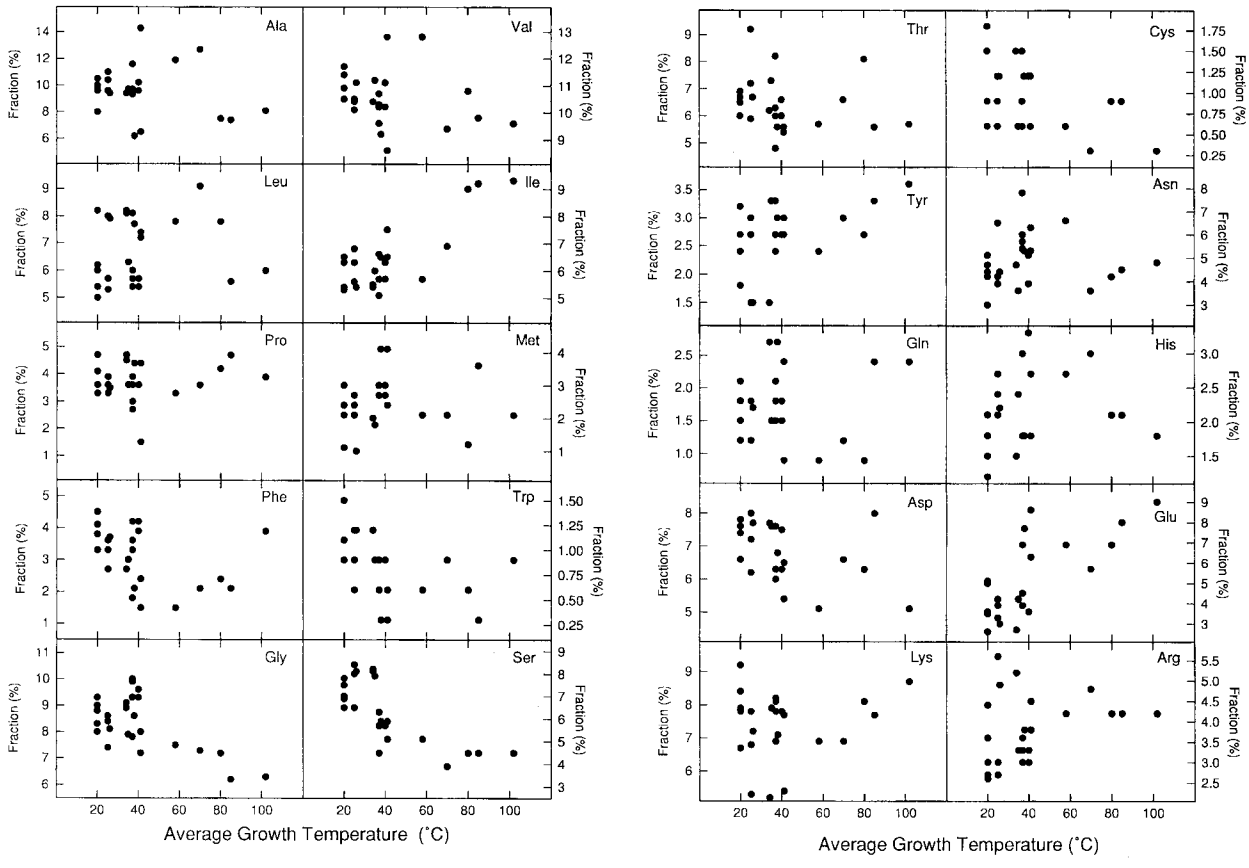

FIGURE 1

Amino acid occurrences in the database of GAPDH proteins, related to the optimal growth temperature of the respective sources.

behavior of proteins; rather, they may be useful to test hypotheses on sequence statistics without giving insight into possible mechanisms of protein stabilization under extreme conditions.

Tables 4, 5 and 7 show that the significance of the occurrence of a certain amino acid should be based on a sufficiently large set of data from related proteins, and not on two-species comparisons which consider neither the natural variability of amino acid compositions nor evolutionary aspects of multiple amino acid exchanges. Presently, the only relevant basis for the discussion of extremophilic adaptation is the three-dimensional structure of the respective protein provided either by crystallography or NMR, or knowledge-based structure prediction algorithms. Such high-resolution data may allow sound comparisons between mesophilic and extremophilic proteins, thus permitting the molecular forces underlying protein stability under extremes of physical conditions to be defined in an unambiguous way.

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG), grant No. Ja78/29-1 and No. Ja78/29-2, and the Fonds der Chemischen Industrie.

REFERENCES

1. Jaenicke, R. (1991) *Eur. J. Biochem.* **202**, 715–728
2. Jaenicke, R. (1991) *CIBA Foundation Symp.* **161**, 206–221
3. Eisenberg, H. & Wachtel, E.J. (1987) *Annu. Rev. Biophys. Biophys. Chem.* **16**, 69–92
4. Eisenberg, H., Mevarech, M. & Zacchai, G. (1992) *Adv. Prot. Chem.* **43**, 1–62
5. Zacchai, G. & Eisenberg, H. (1990) *Trends Biochem. Sci.* **15**, 333–337
6. Zacchai, G., Cendrin, F., Haik, Y., Borochow, N. & Eisenberg, H. (1989) *J. Mol. Biol.* **208**, 491–500
7. Argos, P., Rossmann, M.G., Grau, U.M., Zuber, H., Frank, G. & Tratschin, J.D. (1979) *Biochemistry* **18**, 5698–5703
8. Menéndez-Arias, L. & Argos, P. (1989) *J. Mol. Biol.* **206**, 397–406
9. Schumann, J. (1992) *Ph.D. Thesis*, University of Regensburg
10. Jaenicke, R. (1993) in *Modern Trends in Protein and Enzyme Research* (Liang, D.C., ed.), in press
11. Iwakura, M., Kawata, M., Tsuda, K. & Tanaka, T. (1988) *Gene* **64**, 9–20
12. Purohit, S. & Matthews, C.K. (1984) *J. Biol. Chem.* **259**, 6261–6266
13. Lai, P.-H, Pan, Y.-C.E., Gleisner, J.M., Peterson, D.L., Williams, K.R. & Blakley, R.L. (1982) *Biochemistry* **21**, 3284–3294
14. Dicker, A.P., Volkenandt, M., Schweitzer, B.I., Banerjee, D. & Bertino, J.R. (1990) *J. Biol. Chem.* **265**, 8317–8321

15. Hardies, S.C. & Garvin, L.D. (1991) *Mol. Biol. Evol.* **8**, 654–688
16. Peterson, D.L., Gleisner, J.M. & Blakley, R.L. (1975) *J. Biol. Chem.* **250**, 4945–4954
17. Smith, D.R. & Calvo, J.M. (1980) *Nucleic Acids Res.* **8**, 2255–2274
18. Kumar, A.A., Blankenship, D.T., Kaufman, B.T. & Freisheim, J.H. (1980) *Biochemistry* **19**, 667–678
19. Zusman, T., Rosenshine, I., Böhm, G., Jaenicke, R., Leskiw, B. & Mevarech, M. (1989) *J. Biol. Chem.* **264**, 18878–18883
20. Chen, M.J., Shimada, T., Moulton, A.D., Cline, A., Humphries, R.K., Maizel, J. & Nienhuis, A.W. (1984) *J. Biol. Chem.* **259**, 3933–3943
21. Trimble, J.J., Murthy, S.C.S., Bakker, A., Grassmann, R. & Desrosiers, R.C. (1988) *Science* **239**, 1145–1147
22. Andrews, J., Clore, G.M., Davies, R.W., Gronenborn, A.M., Gronenborn, B., Kalderon, D., Papadopoulos, P.C., Schafer, S., Sims, P.F.G. & Stancombe, R. (1985) *Gene* **35**, 217–222
23. Melera, P.W., Davide, J.P. & Oen, H. (1988) *J. Biol. Chem.* **263**, 1978–1990
24. Nunberg, J.H., Kaufman, R.J., Chang, A.C.Y., Cohen, S.N. & Schimke, R.T. (1980) *Cell* **19**, 355–364
25. Baccanari, D.P., Tansik, R.L., Paterson, S.J. & Stone, D. (1984) *J. Biol. Chem.* **259**, 12291–12298
26. Edman, J.C., Edman, U., Cao, M., Lundgren, B., Kovacs, J. & Santi, D.V. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8625–8629
27. Barclay, B.J., Huang, T., Nagel, M.G., Misener, V.L., Game, J.C. & Wahl, G.M. (1988) *Gene* **63**, 175–185
28. Fling, M.E., Kopfs, J. & Richards, C. (1988) *Plasmid* **19**, 30–38
29. Rouch, D.A., Meseroti, L.J., Loo, L.S.H., Jackson, C.A. & Skurray, R.A. (1989) *Mol. Microbiol.* **3**, 161–175
30. Smith, S.L., Patrick, P., Stone, D., Phillips, A.W. & Burchall, J.J. (1979) *J. Biol. Chem.* **254**, 11475–11484
31. Tesfay, H.S., Amelunxen, R.E. & Goldberg, I.D. (1989) *Gene* **82**, 237–248
32. Branlant, C., Oster, T. & Branlant, G. (1989) *Gene* **75**, 145–155
33. Viaene, A. & Dhaese, P. (1989) *Nucl. Acids. Res.* **17**, 1251–1261
34. Huang, X.Y., Barrios, L.A.M., Vonkhorporn, P., Hond, S., Albertson, D.G. & Hecht, R.M. (1989) *J. Mol. Biol.* **206**, 411–424
35. Tso, J.Y., Sun, X.H., Kao, T., Reece, K.S. & Wu, R. (1985) *Nucl. Acids Res.* **13**, 2485–2502
36. Branlant, G. & Branlant, C. (1985) *Eur. J. Biochem.* **150**, 61–66
37. Stone, E.M., Rothblum, K.N., Alevy, M.C., Kuo, T.M. & Schwartz, R.J. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1628–1632
38. Szorenyi, B., Sajgo, M. & Sabry, M. (1967) *Acta Biochim. Biophys. Acad. Sci. Hung.* **2**, 103–108
39. Fabry, S., Lang, J., Niermann, T., Vingron, M. & Hensel, R. (1989) *Eur. J. Biochem.* **179**, 405–413
40. Fabry, S. & Hensel, R. (1988) *Gene* **64**, 189–197
41. Brinkmann, H., Cerff, R., Salomon, M. & Soll, J. (1989) *Plant Mol. Biol.* **13**, 81–94
42. Zwickl, P., Fabry, S., Bogedain, C., Haas, A. & Hensel, R. (1990) *J. Bacteriol.* **172**, 4329–4338
43. Fort, P., Marty, L., Piechaczyk, M., El Sabrouty, S., Dani, C., Jeanteur, P. & Blanchard, J.M. (1985) *Nucl. Acids Res.* **13**, 1431–1442
44. Holland, J.P. & Holland, M.J. (1980) *J. Biol. Chem.* **255**, 2596–2605
45. Martin, W. & Cerff, R. (1986) *Eur. J. Biochem.* **159**, 323–331
46. Schultes, V., Deutzmann, R. & Jaenicke, R. (1990) *Eur. J. Biochem.* **192**, 25–31
47. Hecht, R.M., Garza, A., Lee, Y.H., Miller, M.D. & Pisegna, M.A. (1989) *Nucl. Acids Res.* **17**, 10123
48. Smith, T.L. & Leong, S. (1988) SWISSPROT sequence database entry "g3p-ustma"
49. Quigley, F., Brinkmann, H., Martin, W.F. & Cerff, R.J. (1989) *Mol. Evol.* **29**, 412–421
50. Quigley, F., Martin, W.F. & Cerff, R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2672–2676
51. Imura, T., Utatsu, I. & Toh-E, A. (1987) *Agric. Biol. Chem.* **51**, 1641–1647
52. Conway, T., Sewell, G.W. & Ingram, L.O. (1987) *J. Bacteriol.* **169**, 5653–5662
53. Böhm, G. (1992) *Ph.D. Thesis*, University of Regensburg

Address:

Dr. *Gerald Böhm*
Institut für Biophysik und Physikalische Biochemie
Universität Regensburg
Universitätsstrasse 31
D-93053 Regensburg
Germany

Tel: (+941)944 6453
Fax: (+941)944 6402
E-mail: GBOEHM@VAX1.RZ.UNI-REGENSBURG.DBP.DE