6536 *Biochemistry* **1981**, *20*, 6536–6543

Structural Stability of Halophilic Proteins[†]

J. K. Mohana Rao and Patrick Argos*

ABSTRACT: An examination of halobacterial amino acid exchanges as they appear in the known *Spirulina platensis* [2Fe–2S] ferredoxin tertiary structure indicated that most of the additional acidic residues of the halophiles occurred on the external surface of the alga structure; however, further negative charges were not placed in the ferredoxin active site region. A statistical investigation of the amino acid compositions of seven halophile and nonhalophile protein counterparts indicated that the bulkiness of amino acids used by halophiles is considerably reduced and that the overall hydrophobicity of halophilic and nonhalophilic molecules was essentially the same. It is suggested that the principal mode of structural stabilization for halophilic proteins is effective competition with the cytoplasmic salt for water through utilization of many external carboxyl groups of glutamic and aspartic acids. A reduction in residue bulkiness would prevent inactivation in the presence of the high molarity, antichaotropic KCl. Halophilic functionality is preserved through avoidance of additional negative charge at the active site surface.

Extreme halophilic bacteria are unique in the realm of living organisms as their cellular interiors are approximately 4 M in KCl (Christian & Waltho, 1962; Ginzberg et al., 1970; Lanyi & Silverman, 1972; Ginzburg, 1978a,b). The particularly harsh cytoplasmic surroundings would require proteins with specially selected amino acids to achieve structural and functional stability. The structure of the three-dimensional polypeptide backbone of proteins is highly conserved (cf. Rossmann & Argos, 1978; Argos & Rossmann, 1979; Matthews et al., 1981) despite significant alterations in the primary sequence. Thus, with the utilization of the known tertiary structure of nonhalophilic proteins as well as the amino acid sequences and compositions of halophilic and nonhalophilic variants, it should be possible to determine some of the principles with which a protein achieves structural tolerance in a high ionic strength milieu.

The stability and activity of halophilic enzymes require high salt concentrations in the 1–4 M range (cf. Lanyi, 1974) which would tend to inactivate their nonhalophilic counterparts (von Hippel & Schleich, 1969; Jaenicke et al., 1971; Herskovitz et al., 1977). Halophilic proteins and enzymes follow the salt rank of the Hofmeister series by displaying functional and structural integrity at high concentrations of antichaotropic salts and destabilization in the presence of chaotropic salts (Higa & Cozzulo, 1965; Lanyi & Stevenson, 1970; Lieberman & Lanyi, 1971; Keradjopoulos & Wulff, 1974; Herskovitz et al., 1977; Prakash & Nandi, 1977; Pundak & Eisenberg, 1981). The amino acid compositions of halophilic proteins show an increased usage of glutamic and aspartic acids relative to their nonhalophilic counterparts (cf. Lanyi, 1974). It has also been noted from a comparison of ribosomal protein compositions that there is a lowered frequency of hydrophobic amino acids in the halophilic system and a higher occurrence of borderline hydrophobic residues as serine and threonine (Lanyi, 1974). Pertinent reviews on halophilism include those of Lanyi (1974), Bayley & Morton (1978), and Kushner (1978). These observations and others have suggested three principal modes for stabilization in halophilic proteins: (1) a charge screening of the molecule induced by the excessive number of acidic residues and effective after addition of at most 0.5–1 M salt (cf. Baxter, 1959; Soo-Hoo & Brown, 1967; Lanyi, 1974), (2) the coerced close packing of nonpolar marginally hydrophobic groups instigated by higher salt concentration (Lanyi, 1974, 1979), and (3) maintenance of a hydrated protein surface through utilization of the carboxyl groups in glutamates and aspartates (Weber et al., 1978; Pundak & Eisenberg, 1981).

Only one protein tertiary structure (*Spirulina platensis* ferredoxin) has been determined where amino acid sequences are known for several nonhalophilic species (Matsubara et al., 1980) and a few halophiles (Hase et al., 1977, 1978, 1980). Amino acid compositions for halophilic proteins and their nonhalophilic counterparts have been reported for seven different molecular types. In the present study, a statistical investigation of the compositional differences and their relationship to certain amino acid physicochemical characteristics as well as an examination of the halophilic amino acid exchanges as they appear in the known chloroplast-type ferredoxin structure suggested that (1) the overall hydrophobicity of residues used by halophiles is similar to that of nonhalophiles, (2) most of the additional negative charges of the halophiles occur at residues on the external surface of the *S. platensis* structure, (3) the bulkiness of amino acids used by halophiles is significantly reduced, (4) halophiles avoid the introduction of further acidic residues at the active-site surface, (5) the halobacterial glutamates and aspartates are primarily found about the hydrophobic β barrel and exposed turn regions of the *S. platensis* structure, and (6) the halophilic negative charges frequently show local clustering. These results point to a highly hydrated halophilic protein surface as a mode of structural integrity in a saline milieu.

Data

The three-dimensional architecture of a [2Fe–2Se] ferredoxin isolated from a blue-green alga, *Spirulina platensis*, has been determined through X-ray crystallographic techniques by Tsukihara and his group (Fukuyama et al., 1980; Tsukihara et al., 1981). An examination of the ferredoxin peptide tracing (Fukuyama et al., 1980) resulted in the following secondary structural assignments: α helix, residues 25–35; β strands, residues 3–9, 13–20, 51–57, 59–62, 80–84, and 86–91; turn regions, remaining residues in the total of 98 amino acids. Dr. Tomitake Tsukihara kindly supplied the atomic coordinates of the *S. platensis* ferredoxin structure and the accessibility of the side chains to solvent water molecules as evaluated by

---

[†] From the Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907. *Received May 22, 1981.* P.A. acknowledges support from National Institutes of Health (No. GM27682) and American Cancer Society (No. FRA173) grants.

0006-2960/81/0420-6536$01.25/0 © 1981 American Chemical Society

Table I: Amino Acid Sequences of Chloroplast-Type [2Fe–2S] Ferredoxins[a]

	-22												-10										1									10										20									
S. platensis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	T	Y	K	V	T	L	I	N	E	A	E	G	I	N	E	T	I	D	C	D	D	D	T	Y	I	L	D	A
H. halobium	P	T	V	E	Y	L	N	Y	E	T	L	D	D	Q	G	W	D	M	D	D	D	D	L	F	E	K	A	A	D	A	G	L	D	G	E	D	Y	G	T	M	E	V	A	E	G	E	Y	I	L	E	A
H. Dead Sea	P	T	V	E	Y	L	N	Y	E	V	V	D	D	N	G	W	D	M	Y	D	D	D	V	F	G	E	A	S	D	M	D	L	D	D	E	D	Y	G	S	L	E	V	N	E	G	E	Y	I	L	E	A

	30										40										50										60										70						
S. platensis	A	E	E	A	G	L	D	L	P	Y	S	C	R	A	G	A	C	S	T	C	A	G	T	I	T	S	G	T	I	D	Q	S	D	Q	S	F	L	D	D	D	Q	I	E	A	G	Y	V
H. halobium	A	E	A	Q	G	Y	D	W	P	F	S	C	R	A	G	A	C	A	N	C	A	S	I	V	K	E	G	E	I	D	M	D	M	Q	Q	I	L	S	D	E	E	V	E	E	K	D	V
H. Dead Sea	A	E	A	Q	G	Y	D	W	P	F	S	C	R	A	G	A	C	A	N	C	A	A	I	V	L	E	G	D	I	D	M	D	M	Q	Q	I	L	S	D	E	E	V	E	D	K	N	V

					80										90								98							105
S. platensis	-	L	T	C	V	A	Y	P	T	S	D	C	T	I	K	T	H	Q	E	E	G	L	Y	-	-	-	-	-	-	-
H. halobium	R	L	T	C	I	G	S	P	A	A	D	E	V	K	I	V	Y	N	A	K	H	L	D	Y	L	Q	N	R	V	I
H. Dead Sea	R	L	T	C	I	G	S	P	D	A	D	E	V	K	I	V	Y	N	A	K	H	L	D	Y	L	Q	N	R	V	I

[a] The primary structures have been aligned by Matsubara et al. (1980). The numbering system is that of *S. platensis* ferredoxin. The primary sequence of *S. platensis* has been determined by Wada et al. (1974) and Matsubara et al. (1976), *H. halobium* by Hase et al. (1977, 1978), and *H.* Dead Sea by Hase et al. (1980).

the method of Gō & Miyazawa (1980). In the present work, if the ratio of the observed accessible surface area to the theoretically possible accessibility was greater than or equal to 0.20, the side chain was considered to be an external residue; a violation of the criterion led to an internal assignment for the side group.

A total of 26 primary structures are known for chloroplast-type ferredoxins, 13 from vascular plants, 3 from green algae, 2 from red algae, 6 from blue-green algae, and 2 from halophilic bacteria (*Halobacterium halobium* and *Halobacterium* Dead Sea). The organisms covered by the known sequences range from photosynthetic eukaryotes (e.g., red algae and plants) to prokaryotes (e.g., blue-green algae and the halobacteria). The 26 homologous primary sequences are listed and referenced by Matsubara et al. (1980) according to a 101-position designation. The halobacterial ferredoxins also contain an additional 22 residues at their homologous N termini and a further 6 amino acids at their C termini. The amino acid sequences for *S. platensis* (Wada et al., 1974; Matsubara et al., 1976) and the two halobacteria ferredoxins (Hase et al., 1977, 1978, 1980) are shown in Table I as aligned by Matsubara et al. (1980). The sequence numbering system used in the present work is that of *S. platensis.*

The amino acid compositions for several halophilic and nonhalophilic protein counterparts have been determined. These include 70S ribosomal protein fractions from *Escherichia coli* (Spahr, 1962) and *H. cutirubrum* (Visentin et al., 1972), glutamate dehydrogenases from calf (Moon & Smith, 1973) and *H.* Dead Sea (Leicht et al., 1978), malate dehydrogenases from *Bacillus subtilis* (Yoshida, 1965) and *H.* Dead Sea (Mevarech et al., 1977), [2Fe–2S] ferredoxins from *S. platensis* (Wada et al., 1974; Matsubara et al., 1976), *H. halobium* (Hase et al., 1977, 1978), and *H.* Dead Sea (Hase et al., 1980), RNA polymerases from *E. coli* (Nicholson, 1971) and *H. cutirubrum* (Louis & Fitt, 1972), cytoplasmic bulk protein from *Pseudomonas fluorescens* and *H. salinarium* (Reistad, 1970), and cytoplasmic bulk protein from *Sarcina lutea* and *Halococcus* No. 24 (Reistad, 1970). For the ferredoxin case, the average amino acid composition of *H. halobium* and *H.* Dead Sea was used; similarly the mean composition of the α and β subunits of *H. cutirubrum* polymerase was taken. The amount of asparagine and glutamine was not known for all the proteins and was therefore excluded from the statistical investigations. Since the amino acid compositions for the various proteins were given in different dimensions, the values for each amino acid in a particular protein were normalized to 1.0 to achieve a similar scale for the compositions. Such a common scale would facilitate calculation of standard deviations.

The parameters used to characterize the twenty amino acids include the Chou-Fasman[1] helix (P_α), sheet (P_β), and turn (P_t) conformational preference values as determined by Levitt (1978), the hydrophobicity index experimentally measured by Nozaki & Tanford (1971), bulkiness and polarity listed by Jones (1975), the free energy of transfer calculated by Janin (1979), and the side chain charge assigned as (+1) for lysine and arginine and (−1) for glutamate and aspartate. Normalization procedures similar to those used for the composition were then applied to all the amino acid characteristics except for the charge assignments and the Chou–Fasman parameters already normalized to 1.0 by definition (Chou & Fasman, 1974a,b).

Janin (1979) has recently noted that the Nozaki–Tanford hydrophobicity values are not good measures for the degree of residue buriedness within protein interiors. Nozaki & Tanford (1971) compared empirically the hydration potentials of isolated residues in an aqueous environment or in solutions containing ethanol or dioxane while Janin (1979) has determined the buried/accessible surface area ratio of amino acids in known tertiary structures. For example, Nozaki & Tanford assign a low hydrophobicity value to threonine and serine which can easily hydrogen bond with water. Yet within the protein interior, these two amino acids can also bond with the main chain or side groups; Janin assigns them a relatively large degree of buriedness. The difference in the two indicators is further illustrated by virtue of only a 0.35 cross correlation between them. If a protein were to adapt to a high salt environment, the Janin values should provide a better measure of any change in the utilization of amino acids that are preferentially exposed or buried to achieve structural stabilization. The Janin transfer free energies were all made positive in sign by adding the absolute value of the most negative value to each of the transfer energies which would place the Janin indexes on the same scale as the Nozaki–

NONHALOPHILIC FERREDOXINS

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A																				
C																				
D	-20																			
E	62	-18	-93																	
F	-14		19	3																
G	-14		-44	44	-7															
H	2		-12	-22		-2														
I	3		10	-28	-46	4														
K	14			-15		-42	2	-1												
L	-8		46	-7	-3			-10	-4											
M			-38	-11	-7			-16	-7	-3										
N	-3		17	-2	-9	9	2		-32											
P	10		35	9								16								
Q	-22		3	52		-3		4	-2	-2	44	12								
R	1		3	2		2								-4						
S	105		8	29	12	-12	2	-7	-3			31		48						
T	38		53	34	18	18		1	6	7		20		2		32				
V	38	-26	6	30		4		42	8	-7	-6	2				3	-79			
W			7							-48		7								
Y	-10		12	13	34	19	-48	-12	-2	-24		-2	-2	-2	-4	38	-16	-2		

HALOPHILIC FERREDOXINS

FIGURE 1: Frequency of amino acid exchanges in comparing halobacterial and nonhalophilic chloroplast-type ferredoxins. The single letter code is used for the amino acids. A positive, nonzero F_{ij} matrix element would show the number of times a nonhalophilic amino acid type designated in the vertical column would change to a halophilic residue type denoted in the horizontal row. The negative elements (F_{ji}) of the antisymmetric array are also given.

Tanford hydrophobicities after normalization.

Methods

Matrices of amino acid changes were constructed for each halophile–nonhalophile pair of known ferredoxin sequences. The procedure used is similar to that introduced by Argos et al. (1979) to examine thermophilic proteins. Let a^k_{ij} be an element of the matrix with i denoting a given residue (e.g. Ala, Arg, etc.) belonging to the nonhalophilic protein, j designating a residue in the homologous halophilic protein, and k referring to the particular pairwise comparison. If there is a trend in amino acid exchange to achieve halophilic stability, then a^k_{ij} should be one of the larger matrix elements.

The matrix arrays for each halophilic–nonhalophilic comparison were appropriately summed resulting in a new matrix with elements $d_{ij} = \sum_k a^k_{ij}$. An antisymmetric matrix with elements $F_{ij} = d_{ij} - d_{ji}$ [called the "net-traffic" matrix by Argos et al. (1979)] would represent the overall tendency of a residue type i in the nonhalophilic ferredoxin sequences to become a residue type j in the homologous halophilic sequences. The significance of the exchanges can be tested by calculation of the standard error determined from all positive, nonzero terms in the matrix which would be the most stringent criterion. The matrix representing the significance of amino acid substitutions would contain elements $E_{ij} = F_{ij}/\sigma(F_{ij})$.

The F matrix was also determined for particular secondary structural regions of the *S. platensis* structure, for example, the amino acid exchanges occurring in only β strands. Such calculations assume not only homology in the various ferredoxin primary sequences but also similar tertiary structures. This assumption would appear quite reasonable as (i) the various sequences can be aligned (Matsubara et al., 1980), (ii) several examples exist where proteins with similar functions also maintain their polypeptide backbone topology depsite considerable alterations in their primary structure (cf. Rossmann & Argos, 1978; Argos & Rossmann, 1979; Matthews et al., 1981), (iii) the halobacterial charged residues are largely found to correspond to residues on the external surface of *S. platensis* ferredoxin (see below), and (iv) a conserved halophilic aspartate substitution (Asp-7) for an internal hydrophobic *S. platensis* residue (Leu-7) is complemented by another conserved halophilic lysyl substitution (Lys-89) for an internal hydrophobic *S. platensis* amino acid (Ile-89) such that the halophilic acidic and basic residues would be allowed charge interaction in the nonhalophilic ferredoxin structure (see below).

The mean of a given physical parameter normalized to 1.0 was calculated for each halophilic and nonhalophilic ferredoxin sequenced by averaging the parameter values for all the amino acids in the particular primary structure. The standard deviation could then be examined for the 24 nonhalophilic ferredoxin sequences and compared to the two halophilic parametric means as a test for significant change.

In the compositional studies, a weighted average $M_{j,l}$ was calculated by

$$M_{j,l} = \frac{\sum_i w_{i,j} P_{i,l}}{\sum_i w_{i,j}}$$

where $w_{i,j}$ is the normalized compositional value of the ith amino acid type in protein j and $P_{i,l}$ is the lth characteristic for the ith amino acid type. The $M_{j,l}$ values and corresponding standard deviations were examined for the halophilic and nonhalophilic compositional groups.

Results

The amino acid exchange F matrix shown in Figure 1 was determined from a pairwise comparison of each of the two aligned halobacterial ferredoxin sequences with each of the similarly aligned 24 nonhalophilic primary structures (Matsubara et al., 1980). Standard errors were determined from

Table II: Significant Amino Acid Exchanges as well as Deletion (−) or Addition (+) of Amino Acid Types in Moving from Nonhalophilic to (→) Halophilic Ferredoxins[a]

amino acid exchanges (nonhalophile → halophile)	amino acid increase (+) or decrease (−) for halophiles
All Residues	
Ser → Ala (3.3)	Thr − (3.0)
Thr → Val (2.5)	Ala + (1.7)
Asp → Glu (2.0)	Glu + (1.6)
β Strands	
Thr → Val (3.2)	Thr − (3.0)
α Helix	
Asp → Glu (2.6)	Asp − (2.8)
Glu → Ala (2.5)	Gln + (2.4)
	Ala + (2.2)
Turns	
Ser → Ala (3.7)	Ser − (2.8)
Gln → Glu (2.0)	Asn + (2.2)
External Residues	
Ser → Ala (4.1)	Ser − (2.5)
Asp → Glu (3.8)	Glu + (2.3)
Internal Residues	
Thr → Val (2.6)	Thr − (2.3)
Val → Ile (2.2)	

[a] The data are given for comparisons of all aligned residues as well as those in specific conformational states of the *S. platensis* structure. Turn residues were assigned as those not appearing in helices of β strands. The ratio of the matrix element or column summation with the respective standard deviation is shown in parentheses.

the positive nonzero elements of the matrix. The elements (both negative and positive) in each column were added, and a standard deviation was determined for the 20 resulting sums. This procedure allowed detection of any significant increase or decrease of a particular amino acid type in moving from nonhalophilic to halophilic structures. Similar matrices and standard errors were determined for only those residues designated in the α-helical, β-strand, and non-α and non-β (turn) conformations assigned from the *S. platensis* structure. The calculations were also repeated for residues denoted external or internal in the known ferredoxin structure. Table II lists for an all-residue comparison and for each of the structural regions the most significant amino acid exchanges as well as the residue types that are principally decreased or increased in moving from nonhalophiles to halophiles. The cutoff was 2.0σ for the residue substitutions and 1.6σ and 2.0σ for the column summations in all-residue and structurally constrained statistics, respectively. It is noteworthy that all the exchanges but one (Thr → Val) would require only one minimum base change per codon. The prinicpal amino acid substitutions point to a halophilic depletion in threonine and additions of alanine and glutamic acid. The avoidance of threonine is largely confined to β-strand residues and internal positions. The attraction toward glutamate is generally in external groups along with an avoidance of serine in external and turn regions.

Averages of the amino acid characteristics given under Data were calculated for each of the halophilic and nonhalophilic ferredoxins in order to attach physical meaning to the substitutions. The characteristic means were found for only amino acids in specific structural states and for the all-residue case which would include the extra N- and C-terminal regions of the halobacteria. A further categorization of the data included all residues and only uncharged residues to differentiate any effects caused by the plethora of acidic amino acids in the halophiles. Table III lists the results. Entries in Table III were made under the conditions that the two halophile amino acid parametric means had to be larger (or smaller) than all nonhalophile averages and were at least 1 standard deviation larger (or smaller) than the mean of the characteristic averages for all nonhalophiles.

It is clear that halophiles show a strong preference for addition of acidic charge, a fact that has not escaped previous notice (cf. Larsen, 1962; Brown, 1964; Lanyi, 1974). However, this charge attraction is frequently found in the ferredoxin β-strand structures. The bulkiness of β-sheet residues is also considerably reduced, and in α helices there is a movement toward helix-stabilizing amino acids in halophilic ferredoxins. It is also noteworthy that the hydrophobicity index increases

Table III: Average Amino Acid Characteristics for Ferredoxin Halophilic and Nonhalophilic Residues in Various Structural States (Residue Conditions)[a]

		nonhalophiles					
residue conditions	parameter	mean value	max value	min values	SD	no. of residues compared	halophiles
all	charge	−0.15	−0.17	−0.11	0.02	97	−0.20 (2.5), −0.22 (3.5)
all, uncharged only	P_t	1.02	1.04	0.97	0.02	71	0.96 (3.0), 0.95 (3.5)
β sheet, all	polarity	0.97	1.36	0.65	0.17	36	1.41 (2.6), 1.41 (2.6)
	bulkiness	0.99	1.01	0.95	0.02	36	0.91 (4.0), 0.91 (4.0)
	charge	−0.08	−0.14	−0.06	0.03	36	−0.22 (4.7), −0.32 (8.0)
β sheet, uncharged only	transfer free energy	1.17	1.22	1.12	0.03	27	1.24 (2.3), 1.26 (3.0)
	bulkiness	1.02	1.06	0.97	0.03	27	0.93 (3.0), 0.96 (2.0)
α helix, all	P_α	1.09	1.11	1.03	0.03	10	1.12 (1.0), 1.12 (1.0)
	polarity	1.24	1.86	1.11	0.20	10	0.77 (2.3), 0.77 (2.3)
	charge	−0.30	−0.40	−0.30	0.10	10	−0.20 (1.0), −0.20 (1.0)
α helix, uncharged only	P_α	1.00	1.03	0.94	0.04	7	1.05 (1.3), 1.05 (1.3)
turn, all	hydrophobicity	0.80	0.85	0.74	0.04	50	0.89 (2.3), 0.88 (2.0)
turn, uncharged only	hydrophobicity	0.89	0.97	0.76	0.05	38	1.03 (2.8), 1.01 (2.4)
	P_t	1.05	1.08	0.97	0.04	38	0.96 (2.3), 0.96 (2.3)
internal, all	hydrophobicity	1.01	1.10	0.92	0.05	40	1.11 (2.0), 1.11 (2.0)
internal, uncharged only	hydrophobicity	1.08	1.18	0.97	0.06	34	1.20 (2.0), 1.21 (2.2)

[a] The characteristic averages listed under halophiles are given in order for *H. halobium* and *H.* Dead Sea ferredoxins, with the accompanying values (shown in parentheses) expressing the number of standard deviations removed from the nonhalophile mean.

Table IV: Counts of Charge Transition between the *S. platensis* and the Two Halobacterial Primary Sequences[a]

residue state	– → – –	0 → – –	0 → 0 –	– → 0 –	– → 00
external	10	11	1	1	8
internal	1	2	2	0	2
β strand, external	4	4	0	0	3
β strand, internal	0	1	2	0	1
α helix, external	2	0	0	0	1
α helix, internal	0	0	0	0	0
turn, external	4	7	1	1	4
turn, internal	1	1	0	0	1

[a] A neutral residue is designated as (0) and acidic amino acids as (–). The structural state of the residues is determined from the *S. platensis* tertiary architecture. An exemplary transition 0 → – – would refer to a neutral amino acid in the alga structure and an acidic residue in the corresponding position of each of the halobacterial sequences.

for internal and turn residues as does the transfer free energy for strand amino acids.

The asymmetric distribution of acidic charges in the putative halophilic structure prompted a more detailed investigation. Amino acids in the *S. platensis* sequence and in the corresponding positions of the halobacterial primary structures were designated as negative (–) if they were glutamate or aspartate or as neutral (0). Transitions involving acidic charge from the nonhalophile to the two halophiles would be of types (a) – → – –, (b) 0 → – –, (c) 0 → 0 –, (d) – → 00, and (e) – → 0 –. Counts were made for such substitutions in the *S. platensis* secondary structural regions as well as for the external and internal categories. The results are given in Table IV. The negative transitions tend to be favored by the two halophiles in β structures, turn regions, and external positions. These results are illustrated in the stereodiagram of Figure 2 which marks the halophilic acidic positions within the *S. platensis* structure. Local clusters of carboxyl groups are also in evidence (Figure 2), for example, *S. platensis* positions 55, 57, and 87. The active-site regions as well as the barrel backside are devoid of acidic residues. The former avoidance would point to functional preservation at the Fe–S complex. The additional N- and C-terminal segments of the halophilic sequence may provide a structural fold to protect the uncharged barrel face; alternatively, this may be a binding region for the ferredoxin reductase. The occurrence of the acidic charges at the surface of the putative halophilic ferredoxin structure would support the assumption of a common fold for the halophiles and nonhalophiles. Further evidence can be found at the *S. platensis* amino acids Leu-7 and Ile-89 which are internal residues in the hydrophobic β barrel (Figure 2) and whose C_α atoms are separated by only 4.9 Å. The conserved halobacterial amino acids at the respective positions are Asp and Lys. An internal electrostatic interaction would be implied for the halophilic structure.

The normalized amino acid compositions for seven halophilic and nonhalophilic protein counterparts listed under Data were also investigated. The compositional differences for each amino acid and their standard errors are given in Table V. Though the changes are only marginally significant, the number of acidic residues increases in halophiles while the number of basic amino acids decreases. Glycine, alanine, and tryptophan all show gains while phenylalanine, isoleucine, leucine, methionine, and threonine exhibit losses. The increase in alanine and glutamic acid and the decrease in threonine are consistent with the ferredoxin trends.

The physical effect of such alterations were examined through calculation of the compositionally weighted, amino acid characteristic averages for each protein. The means of each parameter for the halophile and nonhalophile sets, their differences, and corresponding standard errors are listed in Table VI. The same procedures were applied to the compositions by using only uncharged amino acids (i.e., Lys, Arg, His, Glu, and Asp deleted) to test for charge biases in the

Table V: Average Amino Acid Compositions (Normalized to 1.0) for Seven Halophile and Nonhalophile Protein Counterparts Listed under Data[a]

amino acid	mean nonhalophilic scaled composition	mean halophilic scaled composition	Δ	σ	Δ/σ
Ala	1.71	1.84	0.13	0.44	0.3
Cys	0.22	0.16	−0.06	0.36	−0.2
Asp	1.66	2.28	0.62	0.55	1.1
Glu	1.98	2.44	0.46	0.39	1.2
Phe	0.58	0.44	−0.14	0.17	−0.8
Gly	1.46	1.84	0.36	0.57	0.6
His	0.32	0.34	0.02	0.19	0.1
Ile	1.02	0.76	−0.26	0.29	−0.9
Lys	1.11	0.72	−0.39	0.54	−0.7
Leu	1.53	1.18	−0.35	0.33	−1.1
Met	0.40	0.27	−0.13	0.21	−0.6
Pro	0.73	0.70	−0.03	0.14	−0.2
Arg	1.00	0.81	−0.19	0.41	−0.5
Ser	1.01	1.17	0.16	0.74	0.2
Thr	1.12	0.98	−0.14	0.33	−0.4
Val	1.44	1.38	−0.06	0.34	−0.2
Trp	0.08	0.14	0.06	0.12	0.5
Tyr	0.61	0.57	−0.04	0.40	−0.1

[a] The symbol Δ refers to the scaled compositional differences (halophile minus nonhalophile) and σ to their standard errors calculated from the standard deviations within the halophilic and nonhalophilic compositional sets.

Table VI: Mean of the Weighted Averages of Various Amino Acid Characteristics Calculated from the Seven-Protein Halophilic and Nonhalophilic Data Sets[a]

parameter	nonhalophile	halophile	Δ	σ	Δ/σ
	All Amino Acids				
hydrophobicity	0.90	0.81	−0.09	0.06	−1.5
transfer free energy	0.98	0.98	0.00	0.03	0.0
charge	−0.08	−0.18	−0.10	0.05	−2.0
polarity	1.28	1.38	0.10	0.09	1.1
bulkiness	0.96	0.91	−0.05	0.03	−1.7
	Uncharged Amino Acids				
hydrophobicity	1.04	0.95	−0.09	0.11	−0.8
transfer free energy	1.22	1.21	−0.01	0.02	−0.5
bulkiness	1.01	0.94	−0.07	0.06	−1.2

[a] The statistics were repeated by utilizing all amino acids and only neutral residues (deletion of Lys, Arg, His, Glu, and Asp). P_α, P_β and P_t were excluded as they are not relevant when only compositions are used. The symbol Δ refers to the halophile minus the nonhalophile characteristic means and σ to the standard error calculated from the standard deviations within the halophilic and nonhalophilic compositional sets.

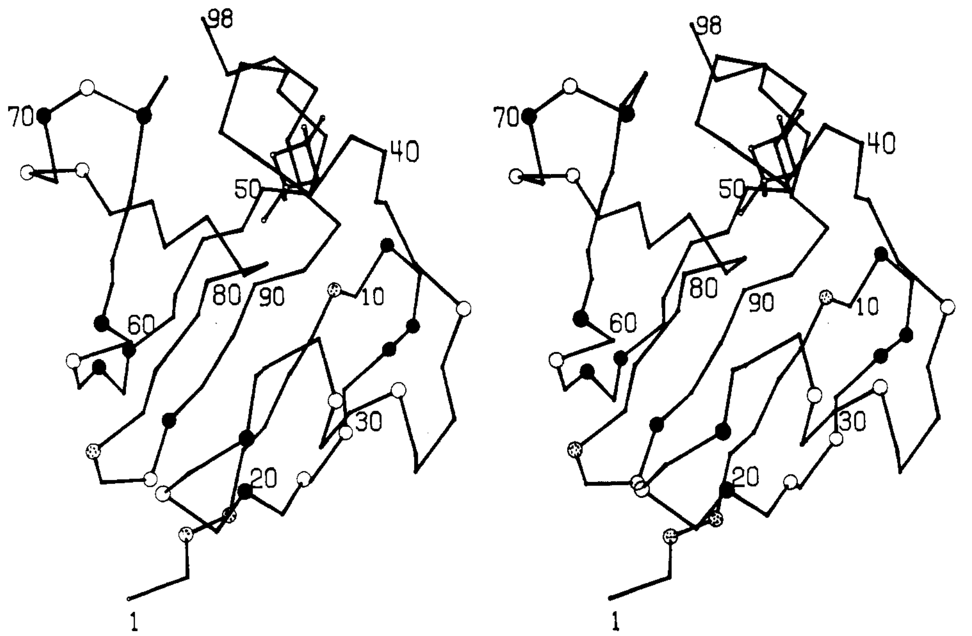

FIGURE 2: Stereodiagram of the *S. platensis* fold (Fukuyama et al., 1981). C_α positions of the transition type (–→––) are shown as large open circles while (0→––) are denoted as filled circles and (0→0–) as speckled regions. The iron–sulfur cage is also illustrated.

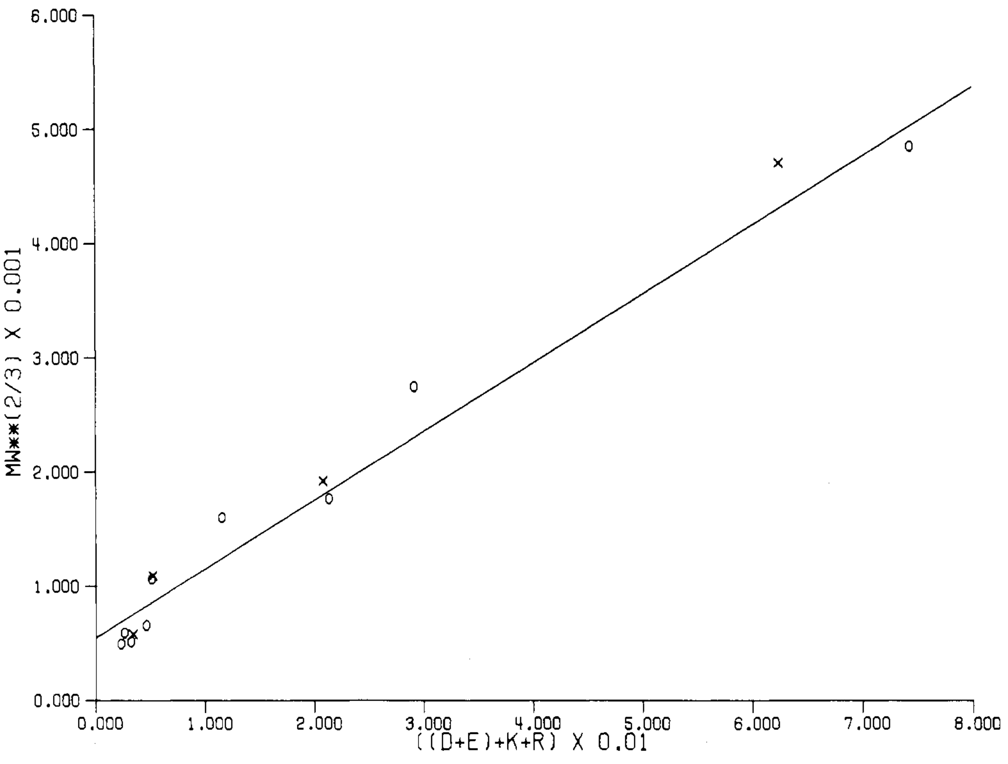

FIGURE 3: Plot of the molecular weight raised to the two-thirds power ($Mr^{2/3}$ or $MW^{2/3}$) vs. the number of (Glu + Asp + Lys + Arg) in various nonhalophilic proteins (O) or the number of (Glu + Asp) in four of the halophilic proteins (×) listed under Data. The halophilic compositions for bulk protein fractions were not used as unique molecular weights could not be determined. The nonhalophilic proteins whose primary sequences are given by Dayhoff (1972) or in references under Data include chicken lysozyme, porcine glyceraldehyde-3-phosphate dehydrogenase, bovine carboxypeptidase A, *S. aureus* nuclease, rabbit hemoglobin, *S. platensis* ferredoxin, equine cytochrome *c*, *B. subtilis* malate dehydrogenase, and bovine glutamate dehydrogenase. The line shown is the least-squares fit for the nonhalophile data with a standard deviation of 243. The halophile points showed a standard deviation of 245 from this line.

weighted averages (Table VI). Once again the results point to a halophilic increase in negative charge and polarity. The bulkiness of the halophilic residues shows a consistent decrease for both the all and uncharged-only residue cases. These statistics are consistent with the ferredoxin trends. The Nozaki–Tanford hydrophobicity mean shows a decrease for the halophiles, though with little significance when only the uncharged groups are used. The Janin transfer free energy is nearly the same for both halophiles and nonhalophiles. For the halophilic ferredoxins, the Nozaki–Tanford hydrophobicity values actually increase (Table III) while the Janin measure changes little.

Discussion

The great increase of acidic charge at the halophilic ferredoxin surface would suggest surface hydration is the principal mode of halophilic structural stability and effective competition with the salt for water. One of the major ferredoxin amino acid substitutions was glutamic acid for aspartic acid (Table II). Kuntz (1971) has suggested that glutamate binds more water at physiologic pH than any other residue [see also Kuntz & Kauzmann (1974)]. The local clustering of carboxyl groups in the halophilic ferredoxin may allow optimal hydration through ordering of the water structure. External serines and threonines are avoided as the hydrogen bonding capacity of their side-group oxygens could disrupt the highly ordered aqueous structure. The acidic side groups are primarily found on the surface of the ferredoxin hydrophobic core or as part of flexible external loops (Figure 2), structural regions perhaps requiring the greatest stabilization. Helical structures are stabilized by use of helix-favoring residues. Additional negatively charge groups do not appear about the active-center region apparently to avoid disruption of function. The compositional studies also support these suggestions as halophilic molecules display an increase in acidic groups and decrease in threonine.

Janin (1976) has shown that the surface area of a protein accessible to water is proportional to its molecular weight raised to the two-thirds power ($M_r^{2/3}$). A plot of $M_r^{2/3}$ for various halophilic and nonhalophilic molecules vs. the number of (Glu + Asp)'s for halophiles or the number of (Glu + Asp + Lys + Arg)'s for nonhalophiles shows a strong linear relationship (Figure 3). This result suggests that the halophilic acidic groups are largely on the external molecular surface.

Empirical results lend support to the suggestions implied by the theoretical statistics. Bull & Breese (1968) have noted that the addition of NaCl dehydrates ovalbumin. Pundak and others (1980, 1981) have shown that highly purified halophilic malate and glutamate dehydrogenases (Leicht & Pundak, 1981) display above normal hydration in high molarity salt solutions. As the salt concentration is decreased, the dehydrogenases first compact, eventually unfold, and reach a hydration state similar to nonhalophilic proteins. The number of glutamates and aspartates cannot account for the observed hydration level in dehydrogenases, thus suggesting carboxyl clusters capable of binding highly ordered water (Eisenberg et al., 1978; Pundak & Eisenberg, 1981).

The correlation between protein structural integrity and the Hofmeister series would indicate hydrophobic interactions as principal stabilizing modes. Lanyi (1974) has thus proposed that the halophilic "need for salt may reflect the lack of an adequate number of nonpolar residues and the requirement to increase the hydrophobic character, through the salting-out effect, of borderline, nonpolar residues, such as the side chains of serine and threonine, and of the polypeptide backbone". Through a statistical investigation of the amino acid composition of 14 *H. cutirubrum* and *E. coli* ribosomal proteins, Lanyi (1974) noted a significant decrease in Val, Leu, Ile, and Phe and an increase in Ser and Thr for the halophilic proteins. Lanyi (1974, 1979) further states compositionally weighted averages of the Nozaki–Tanford hydrophobicities for various halophilic proteins generally appeared lower. In the present work, the hydrophobicity indexes actually increased for the halophilic ferredoxins. The mean value from the compositional studies decreased marginally for the halophiles when only uncharged amino acids were used (Table VI). The preferred Janin measure increased for the halophilic ferredoxin β-strand residues (Table III) changed little between the halophilic and nonhalophilic amino acid compositions (Table VI). Pundak et al. (1981) have noted that increasing temperature is linearly related to decreasing stability in halophilic glutamate and malate dehydrogenases and that these dehydrogenases reach normal hydration levels upon unfolding. It thus seems unlikely that hydrophobic interactions would play any major additional role in halophilic protein stabilization.

The parameter that displayed a consistent decrease for the halophiles was the residue bulkiness (Table VI). In the ferredoxins, the bulk changes were largest in the β-strand residues (Table III). These results correlate with the (Ser → Ala) and (Thr → Val) transitions for the ferredoxins (Table II) and the halophilic compositional increases in Gly and Ala. It is thus suggested that halophilic proteins would resist internal compaction in the presence of highly hydrated salts by reducing considerably their interior bulk. The removal of internal threonines would alleviate the necessity of rigidifying hydrogen bonds.

Conclusions

A theoretical investigation of halophilic primary sequences suggest that halobacterial proteins are simply large anions which effectively compete with the many salt ions of their enviroment for water molecules. The crystallization of *H.* Dead Sea [2Fe-2S] ferredoxin has been recently reported by Sussman et al. (1979). Knowledge of its tertiary architecture should reflect on the cogency of the suggestions made here.

Acknowledgments

We acknowledge the generosity of Dr. T. Tsukihara in providing the atom coordinates of the *S. platensis* ferredoxin structure. We are also grateful for many helpful and creative discussions with Drs. Shlomo Pundak and John Ericson. Marilyn Anthony was indispensible in the preparation of the manuscript.

References

Argos, P., & Rossmann, M. G. (1979) *Biochemistry 18*, 4951–4960.

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

Baxter, R. M. (1959) *Can. J. Microbiol. 5*, 47–57.

Bayley, S. T., & Morton, R. A. (1978) *CRC Crit. Rev. Microbiol. 6*, 151–205.

Brown, A. D. (1964) *Bacteriol. Rev. 28*, 296–329.

Bull, H. B., & Breese, K. (1968) *Arch. Biochem. Biophys. 128*, 488–496.

Chou, P. Y., & Fasman, G. D. (1974a) *Biochemistry 13*, 211–221.

Chou, P. Y., & Fasman, G. D. (1974b) *Biochemistry 13*, 222–245.

Christian, J. H. B., & Waltho, J. A. (1962) *Biochim. Biophys. Acta 65*, 502–506.

Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, Vol. 5, National Biomedical Research Foundation, Silver Spring.

Eisenberg, H., Haik, Y., Ifft, J. B., Leicht, W., Mevarech, M., & Pundak, S. (1978) in *Energetics and Structure of Halophilic Microorganisms* (Caplan, S. R., & Ginzburg, M., Eds.) pp 13–40, Elsevier, Amsterdam.

Fukuyama, K., Hase, T., Matsumoto, S., Tsukihara, T., Katsube, Y., Tanaka, N., Kakudo, M., Wada, K., & Matsubara, H. (1980) *Nature* (*London*) *286*, 522–523.

Ginzburg, B. Z. (1978a) in *Energetics and Structure of Halophilic Microorganisms* (Caplan, S. R., & Ginzburg, M., Eds.) pp 543–560, Elsevier, Amsterdam.

Ginzburg, M. (1978b) in *Energetics and Structure of Halophilic Microorganisms* (Caplan, S. R., & Ginzburg, M., Eds.) pp 561–582, Elsevier, Amsterdam.

Ginzberg, J., Sachs, L., & Ginzburg, B. Z. (1970) *J. Gen. Physiol. 55*, 187–207.

Gō, M., & Miyazawa, S. (1980) *Int. J. Pept. Protein Res. 15*, 211–224.

Hase, T., Wakabayashi, S., Matsubara, H., Kerscher, L., Oesterhelt, D., Rao, K. K., & Hall, D. O. (1977) *FEBS Lett. 77*, 308–310.

Hase, T., Wakabayashi, S., Matsubara, H., Kerscher, L., Oesterhelt, D., Rao, K. K., & Hall, D. O. (1978) *J. Biochem.* (*Tokyo*) *83*, 1657–1670.

Hase, T., Wakabayashi, S., Matsubara, H., Mevarech, M., & Werber, M. M. (1980) *Biochim. Biophys. Acta 623*, 139–145.

Herskovitz, T. T., Cavanagh, S. M., & San George, R. C. (1977) *Biochemistry 16*, 5795–5801.

Higa, A. I., & Cozzulo, J. J. (1965) *Biochem. J. 147*, 267–274.

Jaenicke, R., Kaberstein, R., & Teüscher, B. (1971) *Eur. J. Biochem. 23*, 150–159.

Janin, J. (1976) *J. Mol. Biol. 105*, 13–14.

Janin, J. (1979) *Nature* (*London*) *277*, 491–492.

Jones, D. D. (1975) *J. Theor. Biol. 50*, 167–183.

Keradjopoulos, D., & Wulff, K. (1974) *Can J. Biochem. 52*, 1033–1037.

Kuntz, I. D. (1971) *J. Am. Chem. Soc. 93*, 514–516.

Kuntz, I. D., Jr., & Kauzmann, W. (1974) *Adv. Protein Chem. 28*, 239–245.

Kushner, D. J. (1978) in *Microbial Life in Extreme Environments* (Kushner, D. J., Ed.) pp 317–380, Academic Press, London.

Lanyi, J. K. (1974) *Bacteriol. Rev. 38*, 272–290.

Lanyi, J. K. (1979) in *Strategies of Microbial Life in Extreme Enviroments* (Shilo, M., Ed.) pp 93–107, Dahlem Konferenzen, Berlin.

Lanyi, J. K., & Stevenson, J. (1970) *J. Biol. Chem. 245*, 4074–4080.

Lanyi, J. K., & Silverman, M. P. (1972) *Can. J. Microbiol. 18*, 993–995.

Larsen, H. (1962) *Bacteria 4*, 29–342.

Leicht, W., & Pundak, S. (1981) *Anal. Biochem.* (in press).

Leicht, W., Werber, M. M., & Eisenberg, H. (1978) *Biochemistry 17*, 4004–4010.

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

Lieberman, M. M., & Lanyi, J. K. (1971) *Biochim. Biophys. Acta 245*, 21–33.

Louis, B. G., & Fitt, P. S. (1972) *Biochem. J. 127*, 69–80.

Matsubara, H., Wada, K., & Masaki, R. (1976) in *Iron and Copper Proteins* (Yasunobu, K. T., Mower, H. F., & Hayaishi, O., Eds.) pp 1–15, Plenum Press, New York.

Matsubara, H., Hase, T., Wakabayashi, S., & Wada, K. (1980) in *The Evolution of Protein Structure and Function* (Sigman, D. S., & Brazier, M. A. B., Eds.) pp 245–246, Academic Press, New York.

Matthews, B. W., Grutter, M. G., Anderson, W. F., & Remington, S. J. (1981) *Nature* (*London*) *290*, 334–335.

Mevarech, M., Eisenberg, H., & Neumann, E. (1977) *Biochemistry 16*, 3781–3785.

Moon, K., & Smith, E. L. (1973) *J. Biol. Chem. 248*, 3082–3088.

Nicholson, B. H. (1971) *Biochem. J. 123*, 117–122.

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

Prakash, V., & Nandi, P. K. (1977) *J. Biol. Chem. 252*, 240–243.

Pundak, S. (1980) Ph.D. Thesis, Weizmann Institute of Science, Rehovot, Israel.

Pundak, S., & Eisenberg, H. (1981) *Eur. J. Biochem. 118*, 463–470.

Pundak, S., Aloni, H., & Eisenberg, H. (1981) *Eur. J. Biochem. 118*, 471–476.

Reistad, R. (1970) *Arch. Mikrobiol. 71*, 353–360.

Rossmann, M. G., & Argos, P. (1978) *Mol. Cell. Biochem. 21*, 161–182.

Soo-Hoo, T. S., & Brown, A. D. (1967) *Biochim. Biophys. Acta 135*, 164–166.

Spahr, P. F. (1962) *J. Mol. Biol. 4*, 395–406.

Sussman, J. L., Zipari, P., Harel, M., Yonath, A., & Werber, M. M. (1979) *J. Mol. Biol. 134*, 375–377.

Tsukihara, T., Fukuyama, K., Nakamura, M., Katsube, Y., Tanaka, N., Kakudo, M., Wada, K., Hase, T., & Matsubara, H. (1981) *J. Biochem.* (*Tokyo*) (in press).

Visentin, L. P., Chow, C., Matheson, A. T., Yaguchi, M., & Rollin, F. (1972) *Biochem. J. 130*, 103–110.

von Hippel, P. H., & Schleich, T. (1969) in *Structure and Stability of Biological Macromolecules* (Timasheff, S. N., Ed.) pp 417–574, Marcel Dekker, New York.

Wada, K., Hase, T., Tokunaga, H., & Matsubara, H. (1974) *FEBS Lett. 55*, 102–104.

Weber, M. M., Mevarech, M., Leicht, W., & Eisenberg, H. (1978) in *Energetics and Structure of Halophilic Microorganisms* (Caplan, S. R., & Ginzburg, M., Eds.) pp 427–445, Elsevier, Amsterdam.

Yoshida, A. (1965) *J. Biol. Chem. 240*, 113–117.