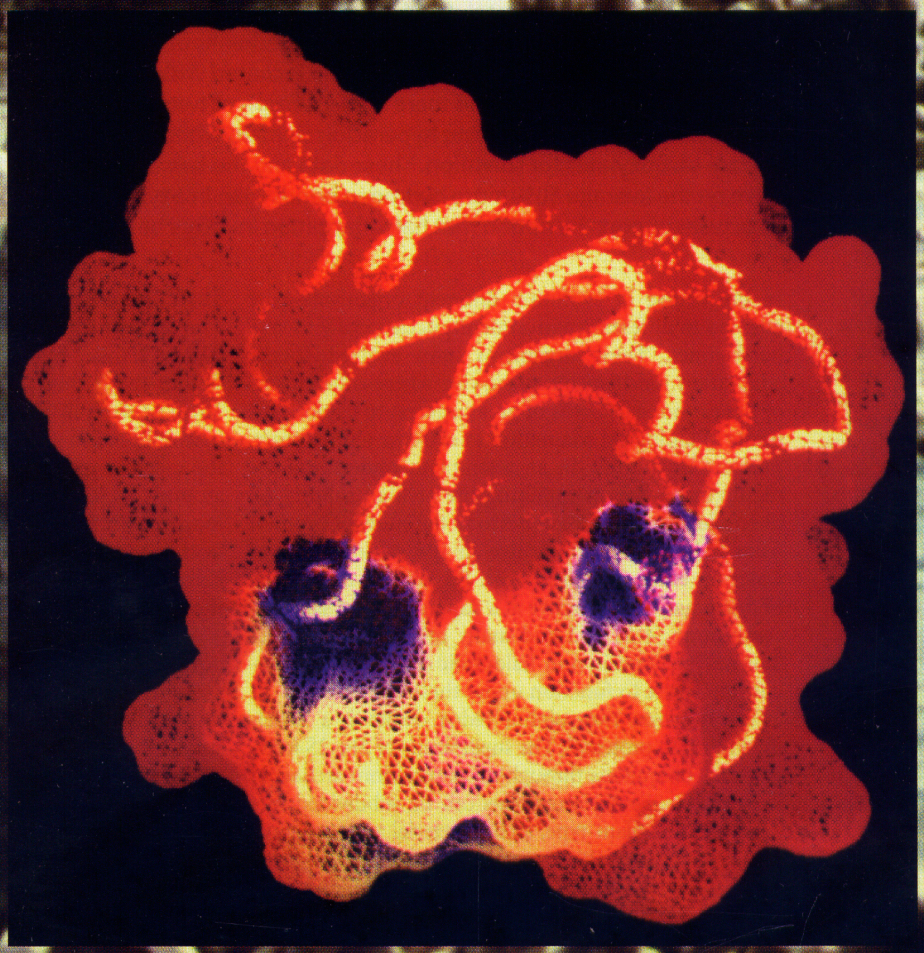
nature
Structural
biology

may 1996
volume 3 no. 5



Surviving the salty sea

RNA and protein folding parallels

The attraction between proteins

Engineering inhibitors of allergies

Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin

Felix Frolow[1], Michal Harel[1], Joel L. Sussman[1,2], Moshe Mevarech[3] and Menachem Shoham[4]

Haloarcula marismortui **is an archaebacterium that flourishes in the world's saltiest body of water, the Dead Sea. The cytosol of this organism is a supersaturated salt solution in which proteins are soluble and active. The crystal structure of a 2Fe-2S ferredoxin from *H. marismortui* determined at 1.9 Å is similar to those of plant-type 2Fe-2S ferredoxins of known structure, with two important distinctions. The entire surface of the protein is coated with acidic residues except for the vicinity of the iron–sulphur cluster, and there is an insertion of two amphipathic helices near the N-terminus. These form a separate hyperacidic domain whose postulated function to provide extra surface carboxylates for solvation. These data and the fact that bound surface water molecules have on the average 40% more hydrogen bonds than in a typical non-halophilic protein crystal structure support the notion that haloadaptation involves better water binding capacity.**

[1]Department of Structural Biology, The Weizmann Institute of Science, Rehovot 76100, Israel
[2]Departments of Biology & Chemistry, Brookhaven National Laboratory, Upton., New York 11973, USA
[3]Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Ramat Aviv 69978, Israel
[4]Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106-4935, USA

Correspondence should be addressed to M.S.

In the course of evolution different organisms have adapted to different environments, occasionally to very harsh ones. These organisms have evolved special adaptation mechanisms to survive where other forms of life cannot exist. It is of great interest to explore the limits of biological adaptation to extreme temperature, pressure or salinity. In an attempt to understand the requirements for maintaining life at high salinity on the molecular level, we undertook the study of proteins from halobacteria which live in the saltiest body of water on earth, the Dead Sea[1]. Proteins in the cytosol of this organism are active at a supersaturated salt concentration[2] (4 M KCl and many other salts) whereas 'normal' proteins in non-halophilic cells would precipitate and cease to function under these conditions.

The two-iron two-sulphur redox cluster is a widely used motif in proteins[3]. The superfamily of 2Fe-2S ferredoxins is ubiquitous throughout nature from archaeons to humans. 2Fe-2S ferredoxins can be subdivided into three families: plant-type, archaebacterial and vertebrate 2Fe-2S ferredoxins. Plant-type 2Fe-2S ferredoxins usually are about 98 residues long, and they fold into a four-stranded mixed β-sheet, a helix that packs against the sheet and a long loop that contains three of the four cysteine ligands to the iron atoms[4]. Archaebacterial 2Fe-2S ferredoxins are homologous to plant-type ferredoxins but they are 30 residues longer. For example, the degree of sequence identity of HmFd to the plant-type 2Fe-2S ferredoxin from heterocyst *Anabaena* is 33%. Vertebrate 2Fe-2S ferredoxins are different; they exhibit no apparent homology to corresponding archaeal and plant-type ferredoxins but they do contain the four cysteine ligands[5]. Here we report the crystal structure of a 2Fe-2S ferredoxin from the extreme halophile *H. marismortui* (HmFd) at 1.9 Å resolution (Table 1).

Structure description

The structure consists of a mixed four-stranded β-sheet flanked by five α-helices; α_1, α_2 and α_5 on one side of the β-sheet, and α_3 and α_4 on the other (Fig. 1*a*). The sheet is twisted to form a barrel-like structure in combination with the intervening loops. Three of the cysteine residues that serve as ligands to the iron atoms, Cys 63, 68 and 71 are located within the long loop connecting helix α_3 with strand β_3. The fourth iron ligand, Cys 102, is located immediately following helix α_4. The iron-sulphur cluster is located ~6 Å from the surface of the protein.

A striking feature of the HmFd structure is the preponderance of acidic residues on the surface of the protein. Thirty-four carboxylates cover the entire surface of the protein except for the vicinity of the iron-sulphur cluster (Fig. 1*b*). HmFd has the highest negative charge density of any protein structure in the Protein Data Bank (Table 2). Carboxylates are predominantly located on the surface-exposed side of the helices and in the loop regions connecting secondary structural elements. For example, five out of the seven residues in helix α_4 are acidic. Likewise, glutamates 50 and 53 are on the solvent-exposed surface of α_3. These carboxylates are highly solvated with up to five water molecules in the first hydration shell. The negative charges are shielded from each other mostly by intervening solvent molecules and occasionally by salt bridges. Some of these carboxylates are also hydrogen-bonded to either the amide nitrogen or the carbonyl oxygen atom of the very same residue. Some parts of the primary structure contain contiguous stretches of acidic residues; these are located in loops connecting elements of secondary structure. For example, the tripeptide Asp 20-Asp 21-Asp 22 is the last part of a

Table 1 Statistics for data collection, phase determination and refinement

Data collection	Resolution (Å)	Reflections	Completeness	R_{sym}[1]	R_{iso}[2]
native at room temp.	2.4	5677	0.96	0.08	
$K_2Pt(CN)_4$ derivative at room temp.	3.2	2287	0.98	0.08	0.13
native at 100° K	1.9	9760	0.99	0.04	

$K_2Pt(CN)_4$ heavy-atom sites and phasing statistics

Site	x	y	z	Relative occupancy	B (Å²)	Phasing power (3.2Å)	Fig. merit (3.2Å)
1	.1571	.3395	.4688	.21	17.		
2	.1142	.3023	.4597	.26	25.		
3	.1652	.2752	.4663	.29	44.	1.82	.55

Native anomalous scattering

site	x	y	z	B(Å²)	Phasing power (3.2 Å)
Fe 1	.1432	.1794	.1472	6.0	
Fe 2	.1071	.1903	.1538	6.0	1.9

Figure of merit following phase combination and solvent flattening at 3.2 Å: 0.81.
Refinement statistics *versus* native data at 100° K:
R=0.195 for all 9760 reflections to 1.9 Å resolution (no cutoff applied), r.m.s. bond lengths 0.018 Å, r.m.s. bond angles 3.29°.

[1] $R_{sym} = \frac{\Sigma|I - I_{av}|}{\Sigma I_{av}}$ I, intensity; I_{av} average intensity.

[2] $R_{iso} = \frac{\Sigma|I_{der} - I_{nat}|}{\Sigma I_{nat}}$ I_{nat}, intensity of native; I_{der} intensity of derivative.

loop connecting α_1 with α_2. The three carboxylates are pointing away from each other and are accessible to the solvent. Likewise, six out of the eight residues in the stretch 29–36 are acidic. These residues form the loop connecting α_2 with β_2. They form a network of solvent accessible carboxyl groups shielded from each other by intervening water molecules.

The six basic residues of HmFd, all located within the C-terminal half of the molecule, are involved in a total of four salt bridges: Glu 4–Lys 112, Glu 53–Arg 64, Asp 79–Arg 99, and Glu 92–Arg 126. In addition there is an indirect salt bridge between Glu 92 and Lys 96, mediated by an intervening water molecule. The Glu 53–Arg 64 salt bridge seems to be important for function. Arg 64, immediately following the first cysteine iron ligand in the iron-sulphur cluster which is conserved in 93% of all plant-type and halobacterial 2Fe-2S ferredoxins. This arginine side chain forms a salt bridge with Glu 53 of helix α_3. This salt bridge sequesters the iron-sulphur cluster from access to solvent.

In contrast to HmFd, malate dehydrogenase from *H. marismortui* (HmMDH)—the only published three-dimensional structure of a soluble halophilic protein—contains many more salt bridges which distinguish it from its non-halophilic homologues[6]. These extra salt bridges are likely to provide stabilization for this protein. While HmFd acts as a monomer in solution, HmMDH is a tetramer with salt bridge clusters at the subunit interfaces. Since in HmFd the folded state is stabilized by the iron-sulphur cluster it is conceivable that HmFd does not need further stabilization in high salt, and therefore does not have many salt bridges. This would explain the observation that HmFd remains active at an NaCl concentration as low as 0.4 M[7]. In contrast, HmMDH loses its activity below 2.5 M NaCl[2].

Why do halophilic proteins in general and HmFd in particular have such a high preponderance of acidic residues? The most compelling reason may be that the extraordinary capacity of carboxylates to bind water molecules[8] is an advantage for any protein that has to compete with a multitude of small cations for free water. There indeed are many water molecules in the crystal structure of HmFd, a total of 237 with an average of 1.9 water molecules per residue of HmFd, or 3.6 water molecules per 100 Å² of accessible surface area (Table 3). Carboxylates are the most hydrated side-chain moieties in the crystal structure of HmFd. The number of water molecules in the first hydration shell around carboxylates varies between 2 and 6. In contrast, lysine residues have only 1–2 water molecules within their first hydration shell. Likewise, polar uncharged residues, on the average, are bound to fewer water molecules than aspartates or glutamates. Similar results have been reported for the water binding capacity of individual amino acids in solution[8]. However, a high degree of hydration is not unusual for protein crystal structures determined at low temperature. A good comparison can be made with hen egg white lysozyme, a protein of

Table 2 The most negatively charged water-soluble proteins in the Protein Data Bank

Protein	PDB code	Net charge	No. residues	Charge/residue	Accessible surface area (Å²)	Net charge density × 10³ (Å⁻²)
Pepsin	4PEP	-39	326	-0.12	13592	-2.9
HmMDH[1]	1HLP	-32	303	-0.11	22725[2]	-1.4
Troponin C	5TNC	-30	162	-0.19	10001	-3.0
HmFd	1DOI	-28	128	-0.22	6532	-4.3
Calmodulin	4CLN	-23	148	-0.16	10289	-2.2

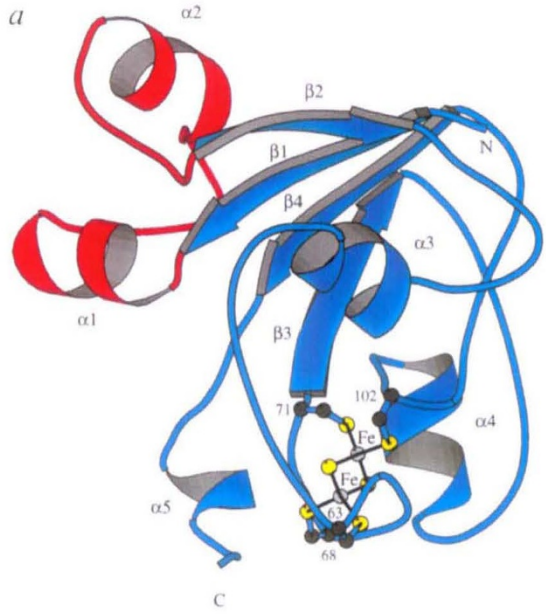
The accessible surface area was calculated with program DSSP[19].
[1]Malate dehydrogenase from *H. marismortui*.
[2]Calculated for the monomer, HmMDH is a tetramer in solution.

the same size as HmFd with X-ray data collected at the same temperature and resolution as HmFd[9] (hen egg white lysosyme; PDB code 5LYT). The degree of hydration in both these proteins is about the same, but HmFd has 40% more hydrogen bonds per water molecule (Table 3). This increase in water layer tightness holds for both protein–water hydrogen bonds as well as water–water hydrogen bonds. In addition to crystallographically observed water molecules, the multitude of carboxylates in HmFd presumably bind many more water molecules with retention times that would be too low to be observed in the time course of an X-ray diffraction experiment.

Shielding of a halophilic protein from its environment is expected to be provided by bound cations. Indeed, during the refinement of the HmFd structure, several of the solvent molecular peaks were showing surplus residual density after being assigned as water molecules and their temperature factors reached the minimum value allowed by the refinement program X-PLOR (2.0 Å^2). Six of these peaks were reassigned as K^+ ions (present in the crystallization liquor) and as such the refinement converged smoothly.

The coordination around the bound K^+ ions is octahedral except for site 6 which has eight ligands (Table 4). The ligands are carboxyl, threonine and amide side-chain oxygen and nitrogen atoms, backbone amide nitrogen and carbonyl oxygen atoms as well as water molecules. The potassium–ligand distances exhibit a greater variability than generally observed for hydrogen bond distances, ranging from 2.3–4.5 Å. Two potassium sites are located on the interface with neighbouring molecules.

With the enormous number of negative charge on the surface of this protein, the question presents itself as to how crystal contacts are formed with neighbouring molecules that are equally negatively charged. There are no intermolecular side chain–side chain interactions in the crystalline lattice of HmFd. The



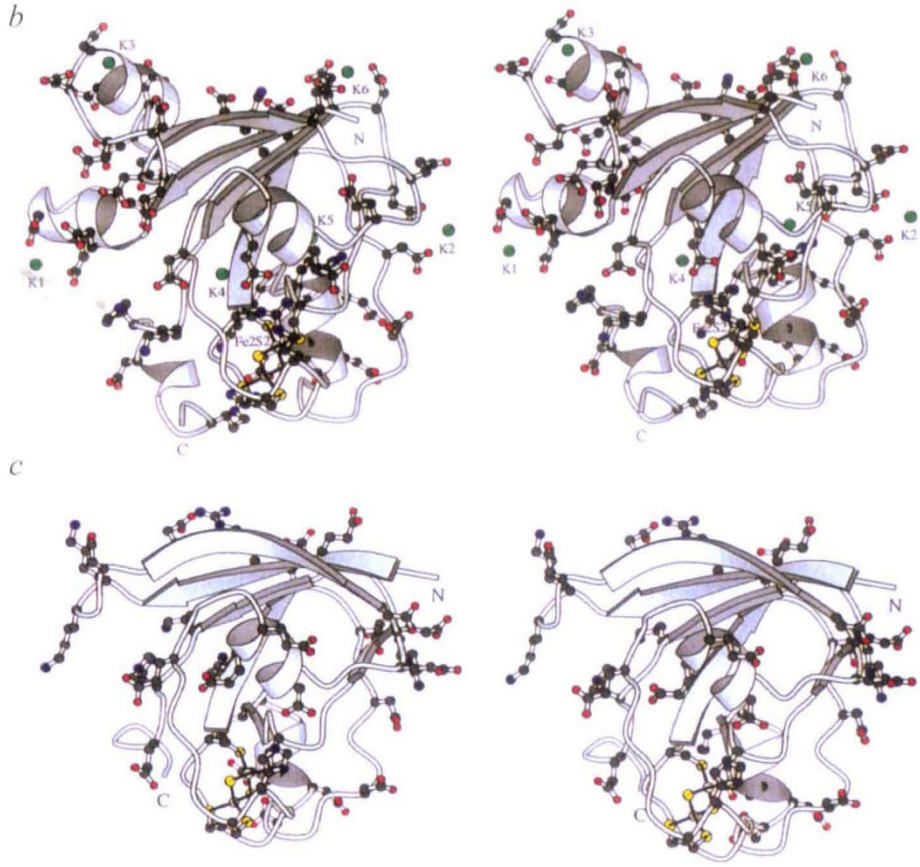


Fig. 1 *a*, Ribbon drawing of the HmFd structure. The secondary structure consists of a mixed four-stranded β-sheet flanked by helices α_1, α_2 and α_5 on one side and by helices α_3 and α_4 on the other. The 2Fe-2S cluster is located on the bottom of this barrel-like structure, covalently bound to cysteine residues 63, 68, 71 and 102. The N and C termini are labelled. The hyperacidic N-terminal domain is shown in red. *b*, Ribbon drawing of the HmFd structure in stereo with all charged residues depicted in ball-and-stick representation (34 aspartic and glutamic acid residues, six lysines and arginines and one histidine) in the same orientation as in (*a*). Carboxylates are located all over the surface of the protein except in the vicinity of the iron-sulphur cluster which is largely uncharged. The six bound potassium ions are shown as green spheres, labelled K1 through K6. The iron-sulphur cluster is labelled Fe2S2. *c*, Same as (*b*) but for *Anabaena* heterocyst 2Fe-2S ferredoxin (PDB code 1FRD), a typical representative of plant-type 2Fe-2S ferredoxins. This protein has fewer charged residues: 18 aspartic and glutamic acid residues, seven lysines and arginines and one histidine. These figures were made with the program MOLSCRIPT[21].

Table 3 Bound water molecules in the crystal structures of HmFd and lysozyme at 100 °K

Protein	No. residues	Resolution (Å)	No. water molecules	No. water molecules/ residue	No. H-bonds/ water[1]	$B_{average}$[2]	Accessible suface area (Å^2)[3]
HmFd	128	1.9	237	1.85	1.83	28.1	6513
lysozyme[4]	129	1.9	237	1.84	1.28	23.1	6752

[1]Hydrogen bond distance ≤3.2 Å
[2] Average temperature factor for the bound water molecules.
[3]Solvent accessible surface area was calculated in program DSSP[19].
[4]PDB code 5LYT.

only direct contacts with neighbouring molecules are backbone interactions of the carbonyl oxygen of Asp 13 with the amide nitrogen of Leu 32 of a neighbouring molecule, and the amide nitrogen of Ala 54 with the carbonyl oxygen of Ala 67 of a neighbouring molecule. All other crystal contacts are mediated by solvent moieties, either K^+ ions or one or more water molecules on the interface. It is possible that some of the current solvent peaks refined as water molecules are actually K^+ ions but at this resolution (1.9 Å) it is impossible to distinguish between a water molecule and a less than fully occupied K^+ ion site.

Plant-type and halobacterial 2Fe-2S ferredoxins

Several homologous plant-type 2Fe-2S ferredoxin crystal structures are available from the cyanobacteria *Anabaena*[10,11], *Aphanothece sacrum*[12] and *Spirulina platensis*[13] and from the plant *Equisetum arvense*[14]. Comparison with these structures highlights distinctive structural features of HmFd. Two properties of HmFd stand out. First, with a net charge of –28 at neutral pH, HmFd is far more negatively charged than the others which are also acidic but have net charges ranging from –15 to –18 (Fig. 1*c*). Second, the longer HmFd contains an extra N-terminal domain made up of two amphipathic helices and intervening loops. This N-terminal domain contains 15 negative and no positive charges.

Amino acid sequences of plant-type 2Fe-2S ferredoxins can be compared to the two halobacterial proteins from *H. marismortui* and *Halobacterium halobium* by aligning the plant-type sequences with the core part of halophilic sequences, residues 39–128. There is a high degree of homology; fifty positions in the sequence are occupied by a single residue in at least 75% of the members of the family. Eleven residues are absolutely conserved, including the four cysteine residues bound to the iron-sulphur cluster. Interesting differences exist in the amino acid composition (Table 5). The most striking and statistically significant difference is in the serine and threonine content; halophilic 2Fe-2S ferredoxins have far fewer serine and threonine residues. This is compensated primarily by an increase in glutamic acid content. The changes in the abundance of polar residues seem to be correlated with their hydrogen bonding capacity to water molecules. Serine and threonine have a single atom capable of forming hydrogen bonds whereas asparagine, glutamine, aspartic and glutamic acid have two such atoms. Aspartic and glutamic acid are known to be better water binders than serine or threonine, with glutamate superior to aspartate. Water binding capacities of individual amino acids have been reported to be 7.5, 6.0, 2.0, 2.0, 2.0 and 2.0 molecules of water per molecule of amino acid for glutamate, aspartate, glutamine, asparagine, serine and threonine, respectively[8]. A protein that has to function in a supersaturated salt milieu has to be able to effectively compete with the multitude of inorganic ions for water binding in order to stay in solution. The best way for the protein to do this is to have many carboxylates, in particular glutamates, on the surface of the protein. Indeed, glutamate is the most abundant residue in the core region of HmFd. Why are not all surface residues in this region glutamates? Avoidance of charge repulsion precludes a negatively charged residue in some cases. For example, residue 62 is a serine in all sequences, including halophilic ones. A carboxylate

Table 4 Coordination of bound K^+ ions

Site	Ligand	Distance (Å)	Site	Ligand	Distance (Å)
1	D12 O	2.69	4	C71 O	2.56
	D12 Oδ1	3.01		A72 O	2.72
	D31 O*	2.70		T101 Oγ1	3.52
	D31 Oδ1*	3.07		N116 N	3.29
	H_2O 129	3.52		H_2O 8	3.60
	H_2O 218	2.87		H_2O 11	2.43
2	D81 Oδ2	2.97	5	E94 O	2.60
	Q85 O*	2.64		N97 Oδ1	2.89
	Q85 Oε1*	3.63		N97 Nδ2	4.00
	D107 Oδ1	3.08		H_2O 14	3.10
	H_2O 74	2.38		H_2O 239	2.31
	H_2O 226	2.74		H_2O 240	3.07
3	S28 O	2.97	6	T2 Oγ1	2.78
	M30 O	2.67		D109 O	3.14
	H_2O 44	*4.39*		D109 Oδ1	3.25
	H_2O 148	*4.49*		D109 Oδ2	3.12
	H_2O 153	*4.02*		E110 Oε1	2.80
	H_2O 174	2.72		H_2O 54	*4.06*
				H_2O 203	3.36
				H_2O 237	*4.08*

Distances shorter than 2.5 Å are underlined, distances longer than 4.0 Å are shown in italics. Protein atoms belonging to symmetry related molecules are labelled by an asterisk.

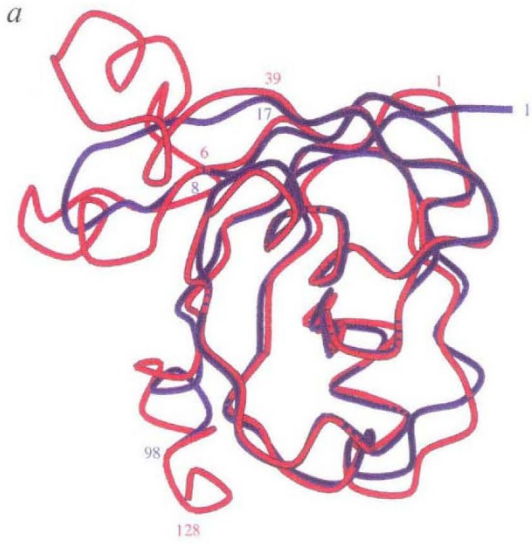
at this position would be repulsed by the nearby C-terminal carboxylate. Likewise, the side chain of Ser 89 is sandwiched in between the side chains of Glu 91 and Glu 92. A carboxylate in this position would cause charge repulsion.

Halophilic adaptation

It is reasonable to assume that HmFd has as many carboxylates on the surface of the barrel as possible without disrupting its structure. This principle of haloadaptation may be called 'halophilic substitution'. However, that may not be enough to keep the protein in solution inside the *H. marismortui* cell. A second mechanism of haloadaptation presents itself by the introduction of an extra domain very rich in carboxylates. We postulate this 'halophilic addition' to be the major mechanism of haloadaptation in HmFd. The 34 residues of the extra domain, comprising helices α1 and α2 and interconnecting loops, are inserted in between strands β_1 and β_2, whereas in the non-halophilic counterparts these two antiparallel strands are connected by a short hairpin turn (Fig. 2). The stretch of residues 6–39 forms a separate domain with its own hydrophobic core made up of Leu 6, Val 11, Trp 16, Val 23 and Phe 24, and 15 extensively hydrated surface carboxylates. The introduction of this extra domain leads to a 26% increase in the solvent-accessible surface area, as compared to a non-halophilic plant-type 2Fe-2S ferredoxin (Table 6). It is unlikely that such a highly charged domain, located far from the iron-sulphur cluster, would be involved in the interaction with redox partner proteins. It is more plausible that the function of this domain is to provide more surface carboxylates and consequently more hydration. The numerous negatively charged carboxylates would prevent the protein from self aggregation and keep it extensively hydrated even in the face of mounting competition from a maximum number of inorganic cations in a supersaturated salt solution. Such a hydrophilic leader sequence may cause solvation in high salt much in the same way a hydrophobic leader sequence causes solvation in a lipid bilayer. Such a solvation domain does not necessarily have to be located at the N-terminus; it could in principle be inserted in between secondary structural elements anywhere in a protein, provided it would be located on the surface of the protein. Indeed, many halophilic proteins are larger than their non-halophilic counterparts[2]. The concept of a hydrophilic solvation domain may have important implications for engineering proteins to be active in salty water.

Methods

Crystallization and X-ray data collection. HmFd was crystallized from solutions of 3.8 M phosphate (a mixture of NaH_2PO_4 and K_2HPO_4) at pH 7.0, as previously reported[15]. Large crystals, up to 1 mm in the longest dimension, were grown by macroseeding. The space group is $P6_322$ with unit cell dimensions *a=b =59.2, c=124.2 Å* at 100° K. Room temperature data were collected on a Nonius CAD-4 diffractometer mounted on a sealed tube X-ray generator. Low temperature data were collected on a Siemens



b

```
                                                    < - β
                                    N K K Q D I D T
        1  <—  β  —> 8
AbFd    A S Y Q V R L I
HmFd      P T V E Y L I N Y E V V D D N G W D M Y D D D V F G E A S D M
          1 <—  β  —>  <—— α ——>                20      <—  α  —>30
                                  17– β —>      25 <—— α ——>        38
AbFd                              T I E I D E E T T I L D G A E E N G I E L P
HmFd    D L D D E D Y G S L E V N E G E Y I L E A A E A Q G Y D W P
      31          <— β —>       45     <— α —>              60
          40                    50<— β ->      <–β–>        65  <–
AbFd      F S C H S G S C S S C V G K V V E G E V D Q S D Q I F L D D
HmFd      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
        61                  70   <— β —>     <–β–>     85      <–
                                                                  98
                                                                  Y L A
        – α —>          <β>      80  <–β–>       <– β –>        95
AbFd    E Q M G – K G F A L L C V T Y P R S – N C T I K T H Q E P
HmFd    E E V E D K N V 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
        – α —>95      <–β–>       <–β–>     110<— β —>       120   <—α–>   128
```

Fig. 2 *a*, Superposition of Cα diagrams of HmFd (red) and the non-halophilic homologue from *Anabaena* (AbFd, blue). The latter lacks the helices α1, α2 and α5. The domain made up of helices α1 and α2 in HmFd is inserted in between strands β1 and β2 at residues 6 and 39. In non-halophilic 2Fe-2S ferredoxins, such as AbFd these two antiparallel β-strands are connected by a short reverse turn. The r.m.s. deviation between 78 corresponding Cα atoms in the two structures (core domain, residues 39–119 in HmFd and 17–95 in AbFd) is 0.93 Å for AbFd PDB entry 1FRD[11] and 0.95 Å for either of the two chemically identical molecules in PDB entry 1FXA[10]. The structures were superimposed with program ALIGN[22]. *b*, Alignment of the sequences of HmFd and AbFd[11], based on the respective crystal structures. Helical and β-strand regions are indicated by α and β, respectively. Parts of the AbFd sequence near both termini do not have counterparts in HmFd; these are indicated by straight lines above the HmFd sequence. The degree of sequence identity in the core domain is 33%.

Table 5 Average amino acid composition in 54 plant-type and two halobacterial 2Fe-2S ferredoxins

Amino acid	Plant-type ferredoxins[1]	Halobacterial ferredoxins core domain (residues 39–120)[1]
A	8.6 (4.1–13.3)	13.0 (12.3–13.6)
C	5.1 (3.8–7.5)	4.9 (4.9–4.9)
D	10.7 (7.4–13.4)	8.6 (7.4–9.9)
E	9.6 (5.8–14.3)	13.6 (12.3–14.8)
F	2.4 (1.0–6.1)	1.2 (1.2–1.2)
G	7.1 (6.1–9.7)	6.2 (6.2–6.2)
H	1.3 (0.0–4.1)	1.2 (1.2–1.2)
I	5.2 (2.0–9.2)	7.4 (7.4–7.4)
K	4.0 (2.0–6.5)	4.3 (3.7–4.9)
L	7.6 (6.1 –10.4)	6.2 (4.9–7.4)
M	0.5 (0.0–2.1)	3.1 (2.5–3.7)
N	1.0 (0.0–5.1)	3.7 (2.5–4.9)
P	3.5 (2.0–6.3)	2.5 (2.5–2.5)
Q	4.5 (1.1–9.6)	3.7 (3.7–3.7)
R	1.3 (1.0–3.2)	2.5 (2.5–2.5)
S	7.6 (5.2–13.7)	4.3 (3.7–4.9)
T	8.2 (3.1–14.4)	1.2 (1.2–1.2)
V	7.1 (3.1–10.4)	7.4 (7.4–7.4)
W	0.2 (0.0–1.0)	1.2 (1.2–1.2)
Y	4.5 (2.1–7.3)	3.7 (3.7–3.7)
S+T	15.7 (9.2–21.4)	5.6 (4.9–6.1)
K+R	5.3 (3.1–8.4)	6.8 (6.2–7.4)
D+E	20.3 (13.5–24.5)	22.2 (22.2–22.2)
L+V+I+F+M	22.8 (18.9–27.4)	25.3 (24.6–25.9)

[1]Numbers indicate molar percentage. Values in parentheses indicate minimum and maximum amino acid content. The sequences were extracted from the SwissProt Data Bank (see Methods). Sequence analysis was performed in program SAPS[20].

area detector mounted on a Rigaku RU300 rotating anode X-ray generator operated at 40 kV and 250 mA, with a fine focus of 0.5 mm^2. A single crystal was used for each data set.

Phase determination. The structure was solved by a combination of the anomalous scattering of the iron-sulphur cluster and a single heavy-atom derivative using room temperature diffractometer data (Table 1). The resultant electron-density map was subjected to solvent flattening. All calculations were performed in program PHASES (W. Furey, University of Pittsburgh).

Table 6 Accessible surface area

Protein	Domain	Residues	Accessible surface area (Å^2)
HmFd	entire protein	1–128	6513
HmFd	core domain	1–5, 40–128	5150
AbFd[1]	entire protein	1–98	5220[2]

[1]2Fe-2S ferredoxin from the cyanobacterium *Anabaena*.
[2]Average value for the two molecules in the asymmetric unit, calculated from the coordinates of protein data bank entry 1FXA[11].
Solvent accessible surface area was calculated with the program DSSP[19].

Model building and refinement. The polypeptide chain was traced in the electron-density map initially with program FRODO[16] implemented on an Evans & Sutherland PS390 graphics system and later on with program O[17] on a Silicon Graphics Indigo2 workstation. The structure was subsequently refined with program X-PLOR[18] against area detector data collected at 100 °K to 1.9 Å resolution. The model consists of all 128 residues, the iron-sulphur cluster, 237 water molecules and six potassium ions. The criteria for deciding whether an isolated peak in the electron-density map was a solvent molecule or noise were threefold: a height of three standard deviations or more above the mean in the $\boldsymbol{F}_o$-$\boldsymbol{F}_c$ electron-density map, a distance of no more than 3.4 Å to any protein nitrogen or oxygen atom or to any other water molecule, and an isotropic temperature factor that refined to a value of no more than 60 Å^2. All solvent peaks were initially refined as water molecules. Six of these peaks showed residual electron density after refinement and their temperature factors dropped to the minimum allowed by the refinement program. Assigning these peaks as Na^+ ions still showed residual density which disappeared after their being assigned as K^+ ions. The refined temperature factors of these K^+ ions range from 10–29 Å^2. The coordinates have been deposited at the Brookhaven Protein Data Bank (accession code 1DOI).

Sequence comparisons. The SwissProt entry codes for the two halobacterial 2Fe-2S ferredoxins are P00216 and P00217. The following 54 entries for plant-type 2Fe-2S ferredoxins were used in this analysis: P00220, P00222, P00223, P00224, P00225, P00226, P00227, P00229, P00230, P00231, P00232, P00233, P00234, P00235, P00236, P00237, P00238, P00239, P00240, P00241, P00242, P00243, P00244, P00245, P00246, P00247, P00248, P00249, P00250, P00251, P00252, P00253, P00254, P00255, P00256, P06517, P06543, P07484, P07838, P08451, P09735, P10770, P11051, P11053, P13106, P14936, P14937, P14938, P15788, P15789, P17007, P22341, P28610, P31965.

Received 5 January; accepted 25 March 1996.

Acknowledgements
We thank O. Dym for helpful discussions and assistance in the calculation and display of electrostatic surface potential. E. Abola and Dave Stumpf from the Protein Data Bank are thanked for help in the sequence analysis of highly charged proteins of known structure. This work was supported by grants from the U. S. Army Research Office.

1. Steinhorn, I. & Gat, J.R. The Dead Sea. *Scientific American* **249**, 102–109 (1983).
2. Eisenberg, H., Mevarech, M. & Zaccai, G. Biochemical, structural, and molecular genetic aspects of halophilism. *Adv. Prot. Chem.* **43**, 1–62 (1992).
3. Holden, H.M. *et al.* Structure-function studies of [2Fe-2S] ferredoxins *J. Bioenerg. Biomembr.* **26**, 67–88 (1994).
4. Tsukihara, T. *et al.* Structure-function relationship of [2Fe-2S] ferredoxins and design of a model molecule, *BioSystems* **15**, 243–257 (1982).
5. Coghlan, V.M. & Vickery, L.E. Expression of human ferredoxin and assembly of the [2Fe-2S] center in *Escherichia coli. Proc. Natl. Acad. Sci. USA* **86**, 835–839 (1989).
6. Dym, O., Mevarech, M. & Sussman, J.L. Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. *Science* **267**, 1344–1346 (1995).
7. Werber, M.M., & Mevarech, M. Purification and characterization of a highly acidic 2Fe-ferredoxin from *Halobacterium of* the Dead Sea. *Arch. Bioch. Biophys.* **187**, 447–456 (1978).
8. Kuntz, I.D. Hydration of macromolecules. IV. Polypeptide conformation in frozen solutions. *J. Am. Chem. Soc.* **93**, 516–518 (1971).
9. Young, A.C., Tilton, R.F. & Dewan, J. C. Thermal expansion of hen egg-white lysozyme. Comparison of the 1.9 Å resolution structures of the teragonal form of the enzyme at 100° K and 298° K. *J. Mol. Biol.* **235**, 302–317 (1994).
10. Rypniewski, W.R. *et al.* Crystallization and structure determination to 2.5 Å resolution of the oxidized [2Fe-2S] ferredoxin isolated from *Anabaena* 7120, *Biochemistry* **30**, 4126–4131 (1991).
11. Jacobson, B.L., Chae, Y.K., Markley, J.L., Rayment, I. & Holden, H. M. Molecular structure of the oxidized, recombinant, heterocyst [2Fe-2S] ferredoxin from *Anabaena* 7120 determined to 1.7 Å resolution. *Biochemistry* **32**, 6788–6793. (1993).
12. Tsukihara, T. *et al.* Structure of the [2Fe-2S] ferredoxin I from the blue-green alga Aphanothece sacrum at 2.2 Å resolution. *J. Mol. Biol.* **216**, 399–410 (1990).
13. Tsukihara, T.*et al.* X-ray analysis of a [2Fe-2S] ferredoxin from *Sprirulina platensis.* Main chain fold and location of side chains at 2.5 Å resolution. *J. Biochem.* **90**, 1763–1773 (1981).
14. Ikemizu, S., Bando, M., Sato, T., Morimoto, Y. & Tsukihara, T. Structure of [2Fe-2S] ferredoxin I from *Equisetum arvense* at 1.8 Å resolution. *Acta Crystallogr.* D**50**, 167–174 (1994).
15. Sussman, J.L., Zipori, P., Harel, M., Yonath, A. & Werber, M.M. Preliminary X-ray diffraction studies on 2Fe-ferredoxin from *Halobacterium* of the Dead Sea, *J. Mol. Biol.* **134**, 375–377 (1979).
16. Jones, A. A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272 (1978).
17. Jones T.A., Zou J.-Y., Cowan S.W., Kjeldgaard M. Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* A**47**, 110–119 (1991).
18. Brünger, A.T. X-PLOR Version 3.1 *A System for X-ray Crystallography and NMR.* (Yale University Press, New Haven and London, 1992).
19. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Bioploymers* **22**, 2577–2637 (1983).
20. Brendel, V., Bucher, P., Nourbakhsh, I.R., Blaisdell, B.E. & Karlin, S. Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl. Acad. Sci. USA* **89**, 2002–2006 (1992).
21. Kraulis, P.J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
22. Satow, Y., Cohen, G.H., Padlan, E.A. & Davies, D.R. Phosphocholine binding immunoglobulin Fab McP603. An X-ray diffraction study at 2.7 Å. *J. Mol. Biol.* **190**, 593–604 (1986).