Structure, Vol. 11, 375–385, April, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S0969-2126(03)00048-0

Crystal Structure of Halophilic Dodecin: A Novel, Dodecameric Flavin Binding Protein from *Halobacterium salinarum*

Boris Bieger,[1,3] Lars-Oliver Essen,[2,*] and Dieter Oesterhelt[1,*]

[1]Department of Membrane Biochemistry
Max Planck Institute for Biochemistry
Am Klopferspitz 18a
D-82152 Martinsried
Germany

[2]Department of Chemistry
Philipps University
Hans-Meerwein-Strasse
D-35032 Marburg
Germany

Summary

A novel, 68 amino acid long flavoprotein called dodecin has been discovered in the proteome of *Halobacterium salinarum* by inverse structural genomics. The 1.7 Å crystal structure of this protein shows a dodecameric, hollow sphere-like arrangement of the protein subunits. Unlike other known flavoproteins, which bind only monomeric flavin cofactors, the structure of the dodecin oligomer comprises six riboflavin dimers. The dimerization of these riboflavins along the *re*-faces is mediated by aromatic, antiparallel π staggering of their isoalloxazine moieties. A unique aromatic tetrade is formed by further sandwiching of the riboflavin dimers between the indole groups of two symmetry-related Trp36s. So far, the dodecins represent the smallest known flavoproteins. Based on the structure and the wide spread occurrences in pathogenic and soil eubacteria, a function in flavin storage or protection against radical or oxygenic stress is suggested for the dodecins.

Introduction

Protein adaptation to extreme environmental conditions is a prerequisite for many prokaryotic organisms to thrive in peculiar places such as hot vents, the deep sea, or rocks. The ability to survive at very high salt concentrations, called halophily, is unique compared to extremophilies like thermo- or barophily, because it is the chemical composition of the environment but not a fundamental thermodynamic control quantity like temperature that differs radically from mesophilic conditions. In haloarchaea such as *Halobacterium salinarum*, the proteins of the cytosol face KCl at almost saturating concentrations [1]. Very early, it was found that these halophilic proteins have a high proportion of acidic residues in their primary structure [2], which were implicated in the structural stabilization of halophilic proteins due to their high water binding capacity [3]. X-ray crystallography and NMR studies on a small set of halophilic proteins [4–6] confirmed this notion. However, the progress toward a concise structural understanding of halophily is hampered by the unique difficulties encountered with the crystallization of soluble halophilic proteins. For example, high salt concentrations very often fail to salt-out the protein of interest, but are needed to keep the halophilic protein in its native state. Furthermore, the requirement of salt is often incompatible with conventional crystallization strategies.

Among the prokaryotes, haloarchaeal organisms also stand out by the unusually large number of genes that were imported by horizontal gene transfer from eubacterial organisms. For the genome sequence of *Halobacterium sp. NRC-1* [7] it was estimated that 384 genes, about 15% of the genomic content, are not of archaeal origin [8]. This tremendous uptake of eubacterial genes cannot only be explained on the basis of habitat sharing, as the hypersaline environment hosts much less eubacterial than archaeal life. Therefore, it is likely that a significant number of these genes is needed to support the lifestyle of Haloarchaea, namely aerobic, chemoorganotrophic growth.

Recently, we solved the crystal structure of one of these lifestyle proteins of *H. salinarum*, a 12-mer ferritin, which was discovered as a major cytosolic protein during a pilot project in "inverse" structural genomics (S. Offermann, B.B., L.-O.E., and D.O., submitted). This protein was suggested to contribute to the protection against reactive oxygenic species by the sequestration of free cytosolic iron species [9]. Now, using the same approach, we have solved the 1.7 Å crystal structure of another protein associated with growth in an extreme environment, the dodecin of *Halobacterium salinarum*. With 68 amino acids and one riboflavin bound per polypeptide, it is the smallest known flavoprotein to date. Although its function is unknown so far, its unique dodecameric coassembly with flavins suggests a novel role for countering the action of light, radicals, or oxygenic species.

Results and Discussion

Discovery by Inverse Structural Genomics

A light orange-colored crystal form was generated during the course of an inverse structural genomics experiment on halophilic proteins where the proteome of the halophile *Halobacterium salinarum* was first fractionated and then systematically screened for crystallizable proteins (S. Offermann, B.B., L.-O.E., and D.O., submitted). Microsequencing of the octahedrally shaped crystals by Edman degradation and comparison of the partial amino acid sequence with the genome sequence of *H. salinarum* (see www.halolex.mpg.de; D.O. et al., unpublished) showed that the crystallized protein was derived from the open reading frame OE3073R. The hypothetical

*Correspondence: essen@chemie.uni-marburg.de (L.-O.E.), oesterhe@biochem.mpg.de (D.O.)

[3]Present address: Switch Biotech AG, Floriansbogen 2-4, D-82061 Neuried, Germany.

Key words: flavin cofactor; halophily; *Halobacterium salinarum*; structural genomics; X-ray crystallography

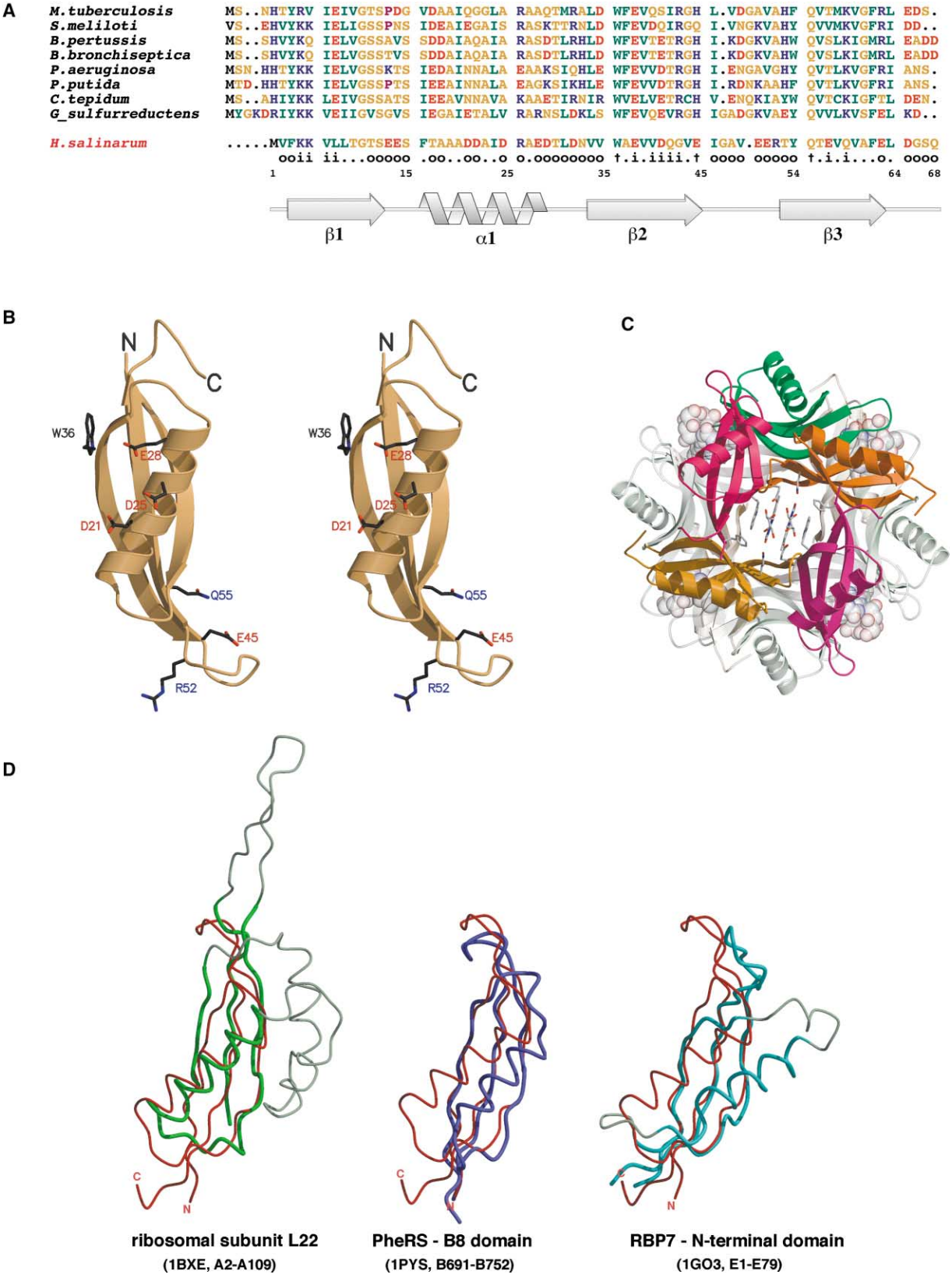

Figure 1. Overall Structure of the *H. salinarum* Dodecin

(A) Alignment of the amino acid sequences of dodecin orthologs. The alignment was performed by CLUSTALX [37] and includes the dodecin sequences found in the genomes of *Halobacterium salinarum* (OE3073R), *Sinorhizobium meliloti* (CAC48468), *Bordetella pertussis* (unfinished

gene product of OE3073R was found to correspond closely to Vng1446h of the published genome sequence of *Halobacterium sp. NRC-1* [7]. However, the latter included 10 additional residues at the N terminus due to improper assignment of the translational start site.

The gene product of OE3073R is a very small, hydrophilic protein of 68 amino acids and hitherto unknown function (calculated MW of 7438 Da). Orthologs of this gene were found by BLAST searches [10] in other, evolutionary distant eubacterial organisms such as *Mycobacterium tuberculosis*, *Chlorobium tepidum*, and *Sinorhizobium meliloti* (Figure 1A), but in none of the currently known archaeal und eukaryotic genomes. The corresponding ortholog in the *M. tuberculosis* strain H37Rv is among the best expressed cytosolic proteins as recently shown by a study on mycobacterial proteomes (EBP database: www.mpiib-berlin.mpg.de/2D-PAGE/EBP-PAGE/index.html). The occurrence in only one representative of the euryarchaeotes, *H. salinarum*, favors the hypothesis that Haloarchaea gained the OE3073R gene from a eubacterial organism by horizontal gene transfer during evolution [8].

Overall Structure and Crystal Packing

The crystal structure of OE3073R was solved by multiple isomorphous replacement at 1.7 Å resolution. The folding unit of OE3073R has a simple $\beta1$-$\alpha1$-$\beta2$-$\beta3$ topology and comprises an amphipathic α helix (F16-T30) that is partly enwrapped by the three-stranded, antiparallel β sheet. Despite the topological simplicity of this 45 $\times$ 20 $\times$ 17 Å large domain (Figure 1B), the OE3073R fold has no direct structural homolog in the protein database with which it shares sequence similarity. However, a similarity search using the DALI web server [11] showed structural resemblance with several proteins that comprised truncated variants of the RNP module (Figure 1D). Canonical RNP domains are small RNA binding domains of a length of about 80 amino acids and a $(\beta\alpha\beta)_2$ topology [12]. The truncated forms miss parts of the C-terminal secondary structure elements as exemplified by the B8 domain of the phenylalanine-tRNA synthetase [13]. This domain not only has the same topology as the OE3073R monomer (rmsd 2.7 Å for 54 C_α positions), but also comprises a twisted, three-stranded β sheet that is kinked by a β bulge at the same position in strand 2 (L727–L730) as in the OE3073R structure (V39–Q42). The location of the α helix along the surface of the twisted β sheet appears to be variable among the truncated RNP modules (Figure 1D). Whereas the α helix is displaced by about 6 Å between the OE3073R $\beta\alpha\beta\beta$ fold and the B8 domain, its location is almost matched by the ribosomal subunit L22 with a displacement of 1.5 Å [14]. Two major insertions in the L22 subunit, a short helix-turn-helix motif and an extended β hairpin (Figure 1D), indicate that the $\beta\alpha\beta\beta$ fold is apparently a minimal scaffold that mediates different context-dependent functions.

The OE3073R monomers assemble to dodecamers with 23-cubic point group symmetry (Figure 1C). Hereby, four protein trimers coassemble with six riboflavin dimers that reside along the 2-fold symmetric axes of the dodecamer. The dodecin oligomer has a total molecular weight of 89.3 kDa and adopts the overall shape of a hollow sphere that has an inner and outer diameter of 23 and 60 Å, respectively. Due to its ability to coassemble with bound flavin cofactors to a dodecamer (see below), we name the gene product of OE3073R "dodecin" (*dodec*amer and flav*in*).

The crystal contacts between the dodecin particles employ only bridging via water molecules, but no direct protein-protein interactions. A major contribution to the contacts is made by a single magnesium ion that is octahedrally coordinated to the side chain of E14 on the outer surface. Three of its five coordinating water molecules are hydrogen bonded to the side chains of a symmetry-related dodecamer. Such a lack of protein-protein interactions for crystal lattice formation was also noted in previous crystal structures of halophilic proteins [5]. Obviously, electrostatic repulsion between the highly negatively charged protein surfaces enforces this type of lattice formation, because the inner and outer surfaces of dodecin carry large excesses of acidic groups on their surfaces, a general hallmark of halophilic proteins [3].

Quaternary Structure

The three-stranded β sheets of the subunits line the interior of the dodecin dodecamer, while the α helices face the outer side of the dodecin particle (Figure 2A). The dodecin subunits are heavily involved in contacts within the dodecamer, as 42% of the monomer surface becomes buried in intersubunit contacts (1835 Å^2 of a total of 4410 Å^2). Furthermore, each subunit is involved in 18 intersubunit hydrogen bonds, of which 12 belong to main chain-main chain interactions. Due to the 23-cubic symmetry there are two kinds of 3-fold symmetric faces in the dodecin particle: the dominant protein-protein contacts are found along the 3-fold symmetric face I where 972 Å^2 of each subunit (22%) are occluded upon trimerization. Here, the three-stranded β sheet of each monomer extends to a five-stranded, antiparallel β sheet

genome), *Bordetella bronchiseptica* (unfinished genome), *Pseudomonas aeruginosa* (B83641), *Pseudomonas putida* (NP 742256), *Chlorobium tepidum* (NP 663103), *Mycobacterium tuberculosis*, and *Geobacter sulfurreducens* (unfinished genome). The secondary structure of the *H. salinarum* dodecin is drawn beneath the alignment. Red, blue, green, and orange amino acid symbols indicate acidic, basic, apolar, and polar residues, respectively. Crosses correspond to riboflavin-contacting residues, and (o) and (i) to residues that face the outer or inner surface, respectively.

(B) Stereo diagram of the protein fold of dodecin. Selected residues are highlighted.

(C) Quaternary arrangement of the dodecin particle as viewed along the 2-fold symmetry axis.

(D) Structural comparison of the dodecin monomer (red) with the ribosomal L22 protein of the 50S ribosomal subunit (green), the B8 domain of the β subunit of phenylalanine-tRNA synthetase (PheRS, blue), and an archaeal RBP7 homolog of the eukaryotic RNA polymerase II (cyan). Regions outside the truncated RNP motif are shown in gray.

Figures 1B–1D, 2, and 3 were prepared by MOLSCRIPT [35] and RASTER3D [36].

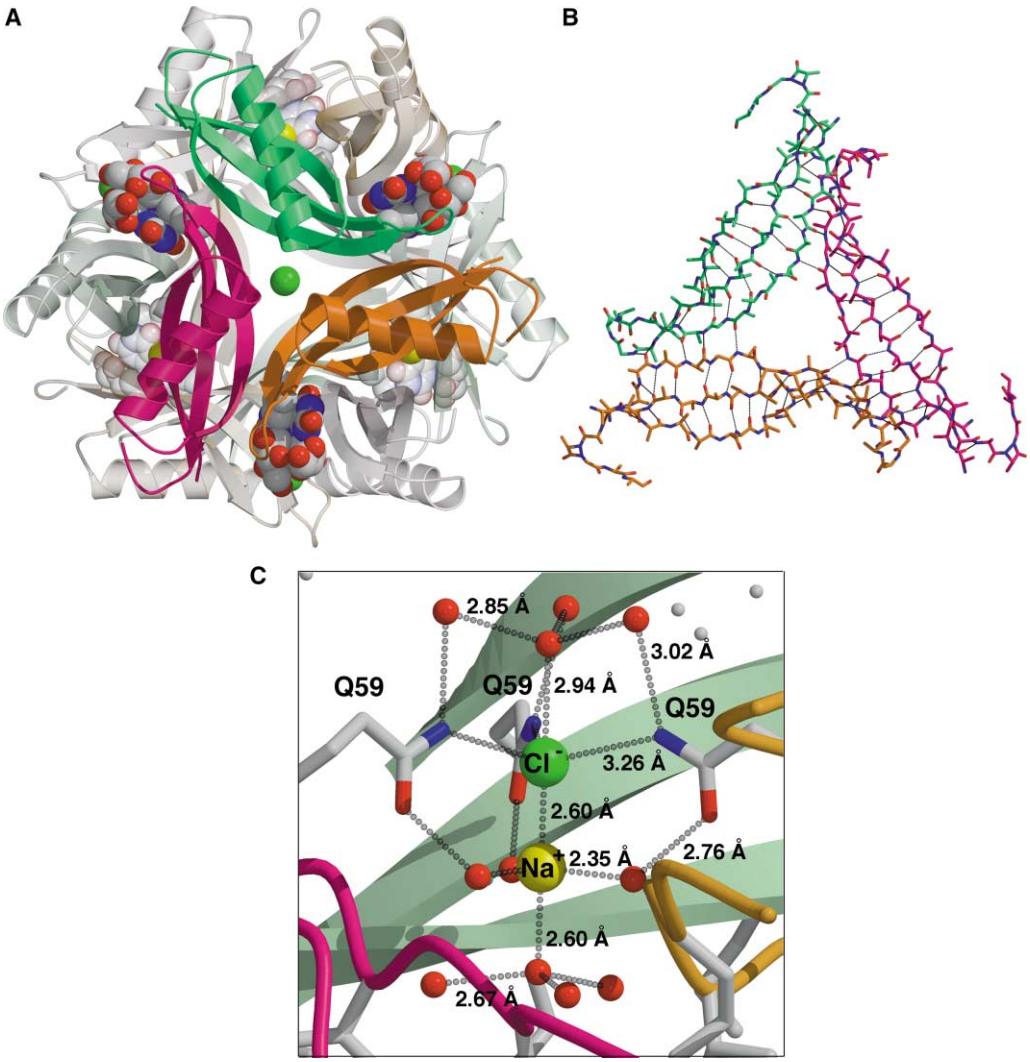

Figure 2. Protein-Protein and Protein-Ion Interactions along the Three-Fold Symmetry Axes

(A) View along the 3-fold symmetric face I on the trimeric subassembly of dodecin. The chloride that is bound in channel I is shown as a green sphere.

(B) Hydrogen-bonding pattern of the trimeric subassembly.

(C) The water/ion channel II in the 3-fold symmetric face II. The interatomic distances for the chloride and sodium ion ligands are indicated. The inclusion of a Na^+ ion instead of a water molecule as a fifth chloride ligand was inferred from its regular bipyramidal-trigonal coordination geometry and the high concentration in the crystallization buffer (>1 M). The B factors of the Cl^- (27.1 $Å^2$) and the Na^+ (30.2 $Å^2$) ions are similar to that of the amide group of Q59 (21.6 $Å^2$) and indicate full occupancies.

by local pairing of the β2 strand (V35–V39) of one monomer with the β2 strand (Q42–I46) of another monomer (Figure 2B). In the 3-fold symmetric face II, these β sheets are continued throughout the whole dodecamer by antiparallel main chain-main chain pairing between F3–K5 (β1) and L7–T9 (β1). Besides this network of polar interactions, there are a few stabilizing interactions in the dodecin particle that appear to be unique for the halophilic protein: the nonconserved residue R52 forms three salt bridges to the acidic residues D21, D25, and E28; three other, symmetric H bonds are made by the side chains of Q42 along the 3-fold axis.

In the inner compartment of the dodecin particle, 372 water molecules and 12 magnesium ions could be localized. The latter are only coordinated to water molecules and the side chains of D41 (2.1 Å). A chaperone-like function of the inner compartment as a folding cage can be excluded with certainty. Its volume is too small (6.3 nm^3) to accommodate other, unfolded proteins as done by the 12- and 24-meric small heat shock proteins, which also form hollow spheres of cubic symmetry [15, 16]. Likewise, the high sequence variability of residues exposed to the inner surface indicates that the inner compartment is probably not taking part in another yet unknown function, that is, sequestering of cations as done by the 12- and 24-meric ferritins with their larger, inner compartments, but merely plays a structural role like the similarly sized cavity in the bromoperoxidase dodecamer of *Corallina officinalis* [17].

The protein shell of the dodecin oligomer almost com-

pletely shields the inner compartment from access to the bulk solvent. The entry of ions and small solutes into the inner compartment is hence restricted to two narrow channels that run through the centers of the 3-fold symmetric faces I and II. Interestingly, both channels are plugged by the binding of chloride ions. The first chloride ion resides at the outer side of channel I (Figure 2A). The binding of this chloride, which is fully hydrated by bitrigonal coordination to five water molecules, is apparently stabilized by ion-dipole interactions, because the N termini of three surrounding α helices closely encounter the chloride site with a distance of 5 Å (Figure 2C). A second chloride binding site was found at the inner entrance of channel II. Here, the chloride is electrostatically stabilized by direct coordination to a sodium ion (Cl^--Na^+: 2.60 Å). In addition, the chloride makes three H bonds to the side chains of the 3-fold-related residue Q59 (Cl^- Q59NE2: 3.26 Å). The sodium ion is bitrigonally coordinated to four water molecules (Figure 2C) and fills channel II together with the chloride and the water molecules WAT22, WAT36, and WAT100 in a single-file manner.

The Flavin Binding Site

A novel feature of the dodecin that was not observed before in any other flavoprotein is its binding site for dimeric riboflavin (Figure 3A). $2F_{obs} - F_{calc}$ simulated annealing omit electron density maps [18] clearly showed the binding of two 7,8-dimethyl isoalloxazine groups along the 2-fold symmetry axes of the particle. The isoalloxazine rings contact each other via their *re*-faces (Figure 3B). The N5 atom of each riboflavin cofactor points its lone electron pair into the interior compartment of the dodecin oligomer where it forms a hydrogen bond with WAT96 (3.2 Å). Together with the presence of methyl substituents at the C7 and C8 positions at the isoalloxazine ring, it can be excluded that a deazaflavin cofactor like F420 is bound to the dodecin decamer. During the refinement of the dodecin structure, the ribityl side chains clearly emerged at the N10 atom of the isoalloxazine moiety in difference electron densities. The conformation of the ribityl group is stabilized by an intramolecular H bond between the hydroxyl O4′ and the N1 atom of the isoalloxazine ring (2.9 Å). Further interactions include H bonds from the ribityl hydroxyls to the side chain of E45, the carbonyl oxygen of V35, and between the O3′ and O5′ hydroxyls of the 2-fold-related riboflavins (Figure 3A). No substituents were observed at the O5′ hydroxyl of the ribityl group that would indicate the presence of either flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN). However, due to the surface exposure of the ribityl groups, it cannot be strictly ruled out that FMN or FAD likewise bind to the dodecin particle with the residual part of the cofactors remaining disordered upon binding.

The observed riboflavin molecules are part of an aromatic tetrade, sandwiched between the indole groups of W36, which were derived from two neighboring dodecin subunits (Figure 3B). The coplanar side chain orientation of W36 is hereby stabilized by putative charge-transfer interactions to the isoalloxazine rings and a hydrogen bond to the conserved residue E38 (W36NE1–E38OE2: 2.8 Å). The only other polar interactions between the isoalloxazine rings and the protein moiety are two hydrogen bonds to the amide side chain of Q55. This unusual aromatic tetrade consisting of the π-staggered indole groups of W36 and the isoalloxazine moieties of riboflavin (Figure 3B) dominates the interactions between the trimers, as the only other stabilizing protein-protein contact along the 2-fold axes is a salt bridge between K5 and E57 (2.8 Å).

Dodecin is significantly smaller than the FMN binding protein of *Desulfovibrio vulgaris* [19], the smallest known flavoprotein thus far. The interfacial binding of riboflavin dimers is certainly a minimal solution, as each dodecin monomer contributes only 3 residues to the binding site, W36, E45, and Q55. This binding mode contrasts sharply to the binding sites of other flavoproteins, in which the flavin nucleotides occur only as monomers. Interestingly, flavin nucleotides have an intrinsic propensity to form dimers and higher aggregates in aqueous solution [20] or in the crystalline state. For example, in at least five X-ray structures of (iso)alloxazine derivatives in the Cambridge Structure Database (CSD, status February, 2001) π staggering and potential charge-transfer interactions were likewise observed between the electron-rich pyrimidine rings and the electron-deficient benzo rings (Figure 3C). These dimeric assemblies are not related by a 2-fold axis along the atoms N5 and N10, as in the dodecin particle (Figure 3C), but by a 2-fold axis perpendicular to the ring system. As the prevalent dimeric species of FMN in aqueous solution (Figure 3D) also adopts a different staggering interaction [21], it might be the specific dodecin-riboflavin interactions that impose the observed dimerization mode on the riboflavins.

Halophily and Comparison to Mesophilic Dodecins

The *H. salinarum* dodecin exhibits the prominent features of halophilic proteins that are adapted to the almost saturating KCl concentrations in the cytosol of Haloarchaea: first, the sequence of the *H. salinarum* dodecin comprises a huge compositional excess of aspartates and glutamates (24%) over basic residues (6%), which causes the very low pI of 3.8 for the *H. s.* dodecin (Table 1). Thirteen acidic residues are exposed on the outer surface, while 3 residues line the inner compartment. The densities with which these acidic residues are found on the outer and inner surfaces of the dodecin particle (total areas: 20,898 $Å^2$, 4591 $Å^2$) are high, on average 1 acidic residue per 134 $Å^2$ and 128 $Å^2$, respectively, of the protein surface area. For comparison, other halophilic proteins such as ferredoxin and malate dehydrogenase from *Haloarcula marismortui* [4, 5] or the dihydrofolate reductase from *Haloferax volcanii* [6] have surface densities of 153–231 $Å^2$ per acidic residue. Likewise, an in silico analysis of 185 homology models of the cytosolic *H. salinarum* proteins demonstrated that an acidic residue occupies on average 246 $Å^2$ of the surface of an halophilic protein, while for other, nonhalophilic organisms, the surface densities are significantly lower, with 350–400 $Å^2$ per acidic residue (L.-O.E., unpublished data). The up to 3-fold surplus

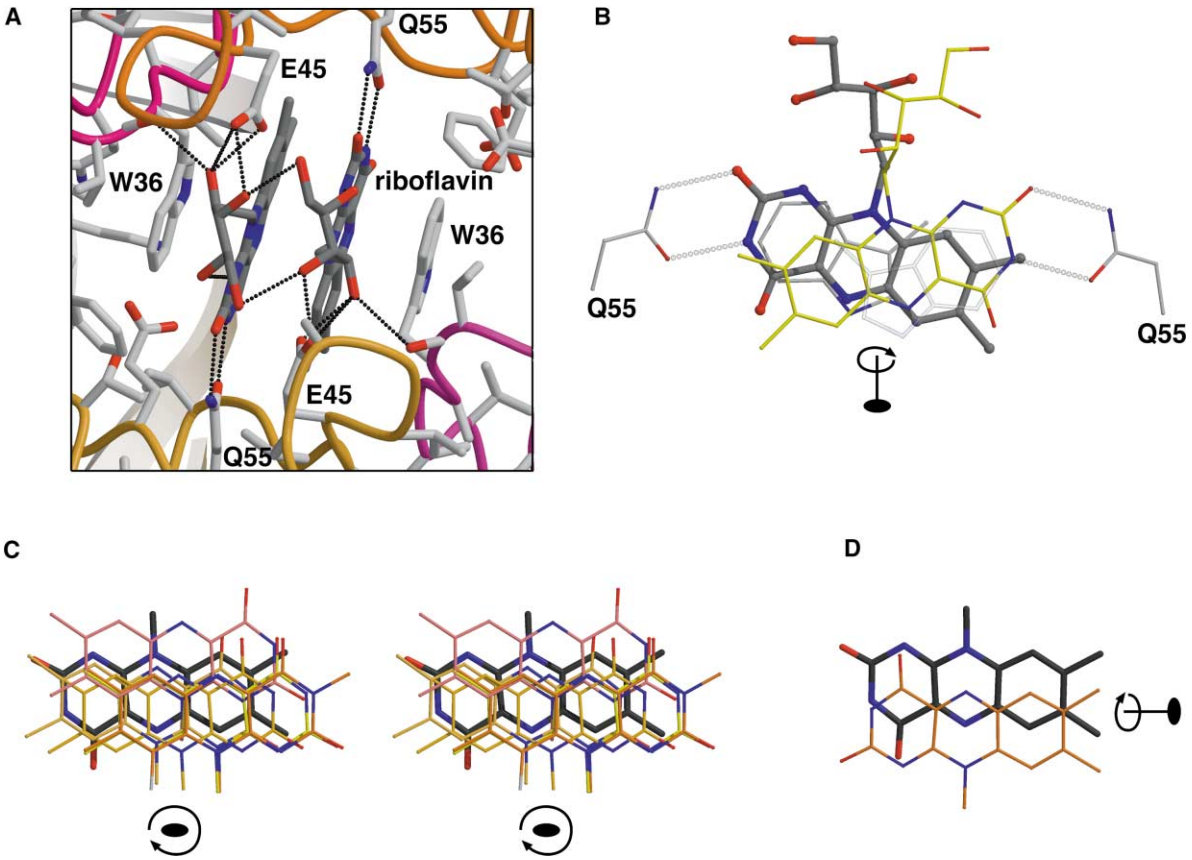

Figure 3. The Binding Site for Dimeric Flavins in the *H. salinarum* Dodecin Particle

(A) View along the 2-fold symmetry axis on the riboflavin binding site. Hydrogen bonds are shown as black, dotted lines.

(B) Schematic representation of the aromatic tetrade in the dodecin oligomer. The isoalloxazine moieties of the two riboflavins are off-rotated by 31° along an axis perpendicular to the ring centroid (12° relative to ring plane of W36).

(C) Dimerization of (iso)alloxazine derivatives in five crystal structures. The superposition used the CSD entries BACMIB10, BFLANH10, CPMIAL10, JUKMOX, and LUFAEP1.

(D) Dimerization of FMN in solution according to an NMR analysis by Kainosho and Kyogoku [21]. For clarity, substituents of the isoalloxazine rings were omitted in (C) and (D).

of acidic residues in halophilic proteins is thought to promote the tight binding of the hydration shell to the protein [3], as the carboxylates of glutamate and aspartate are strongly hydrated [22]. The alternative explanation for the prevalence of acidic residues on halophilic protein surfaces, electrostatic repulsion of polypeptides to avoid protein aggregation by the salting-out effect [23], might be ruled out for the dodecin as previously for the halophilic ferritin (S. Offermann, B.B., L.-O.E., and D.O., submitted) because the surface of the inner, occluded compartment has a similar negative charge density as the outer surface (Table 1; Figure 4A).

A second way that halophilic proteins adapt to their environment is the binding of cations and anions to their protein surfaces [3]. In the structure of the *H. salinarum* dodecin, the unusual Na^+-Cl^- ion pair inside channel II

Table 1. Exposure of Charged Residues on Dodecin Surfaces

			Surface Charge		Asp/Glu		Arg/Lys		His	
Organism	Residues	pI	Out	In	Out	In	Out	In	Out	In
H. salinarum	68	3.6	−132	−12	13	3	2	2	—	—
M. tuberculosis	70	5.8	−18	+12	6	2	3	3	3	—
S. meliloti	69	6.1	0	+12	4	3	5	4	2	—
B. pertussis	71	5.7	0	0	7	3	5	3	4	—
P. aeruginosa	71	6.6	+6	+12	5	3	3	4	5	—
C. tepidum	70	6.1	−12	−12	6	4	4	3	2	—
G. sulfurreducens	72	7.5	6	+12	7	3	6	4	1	—

Sequence statistics for the protein surfaces of dodecins. The assignment, whether a residue is exposed to the outer surface (out) or the inner surface (in), was based on the available structure of the *H. salinarum* dodecin. The estimation of overall surface charges for the intact 12-mer particles assumed fully ionized states for Asp, Glu, Lys, and Arg, and 0.5+ charges per histidine.

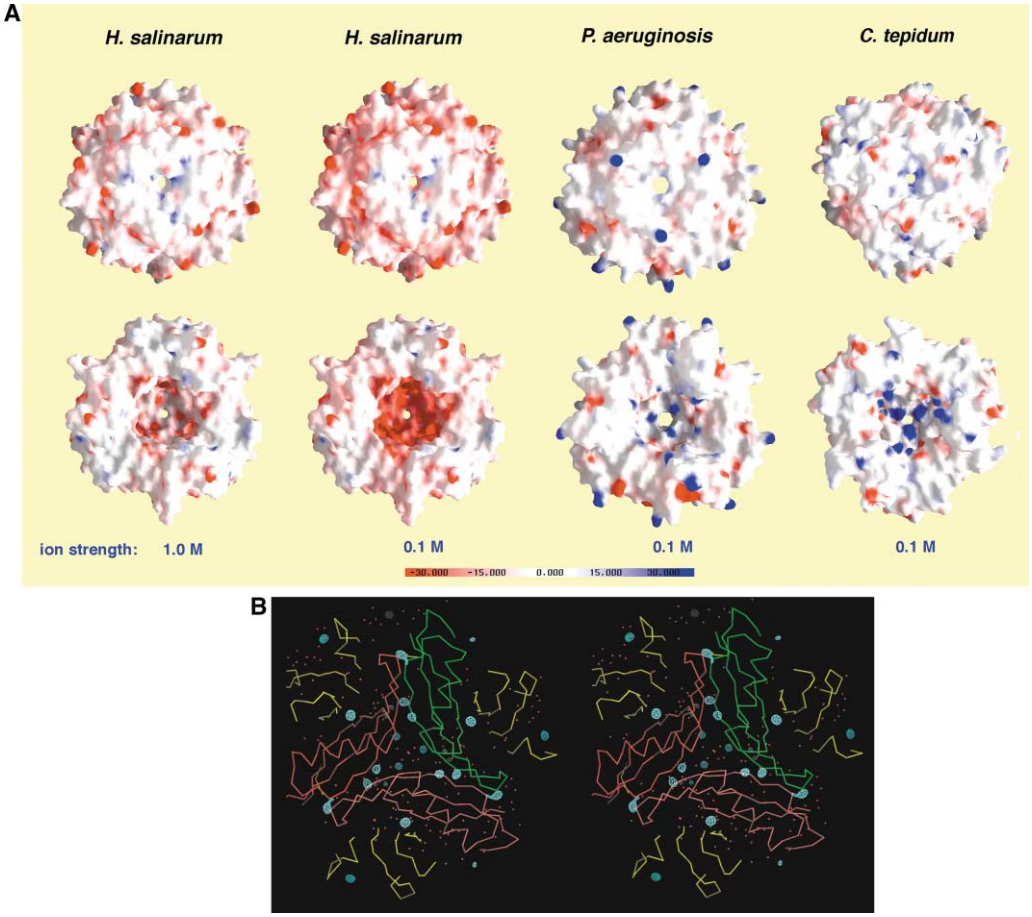

Figure 4. Comparison of the Surface Characteristics between Halophilic and Mesophilic Dodecins

(A) Electrostatic surface potential of the dodecin particle from *H. salinarum* and of the modeled dodecin particles from the mesophiles *P. aeruginosa* and *C. tepidum*. The upper and lower surfaces show a view along the 3-fold from the outside and inside of the particle, respectively. (B) Anomalous difference Fourier of the rubidium chloride soak calculated at 2 Å resolution (contouring level 5 σ). The electrostatic analyses were done by GRASP [38] using a probe radius of 1.4 Å and an ionic strength of 0.1 M and 1.0 M. For a better comparison between the surfaces of the inner compartment, a dodecin trimer was removed prior to calculation, although this does not fully reflect the electrostatic situation in a hollow sphere as given by the intact dodecin particle.

and the binding of a chloride on the entrance of channel I presumably stabilize the quaternary structure of the dodecin oligomer. This ion-mediated stabilization might be quite common among halophilic protein oligomers, as the "locking" of Cl^- or Na^+ ions in intersubunit faces was also observed in the X-ray structures of the tetrameric malate dehydrogenase of *H. marismortui* [24] or the 12-mer ferritin of *H. salinarum* (S. Offermann, B.B., L.-O.E., and D.O., submitted). Apart from the two magnesium sites, no further binding sites for cations could be crystallographically defined in the structure of native *H. salinarum* dodecin. However, anomalous and isomorphous difference Fourier maps of a dodecin crystal soaked in 1 M rubidium chloride showed four additional 5 σ peaks at 2 Å resolution (Figure 4B). Three of these putative alkali binding sites are located along the outer surface at the positions of WAT23, WAT34, and WAT56; one site is in the inner compartment close to WAT47. Therefore, the high charge density of the dodecin surfaces can be at least partly compensated by alkali ion binding.

An increased number of surface-exposed salt bridges might be the third form of halophilic adaptation. Each dodecin monomer is involved in four intersubunit salt bridges. Three of them involve the guanidinium group of R52 that is in the midst of a cluster of acidic residues from another monomer (D21, D25, and E28). This complex interaction is similar to several salt-bridging networks on the intersubunit interfaces of the *H. marismortui* malate dehydrogenase [4], which stabilize the tetrameric assembly of this enzyme [25]. The fourth salt bridge is between the side chains of K5 and E57. With the exception of K5, none of the residues involved in salt bridges is conserved in any other dodecin sequence where small and polar residues prevail at the equivalent positions (Figure 1A). Therefore, it is likely that these interactions contribute significantly to the stabilization of the *H. s*. dodecin oligomer at high salt concentrations.

Biological Implications

To date, the biological function of the dodecins is still elusive. No clear phenotype has been found so far for a dodecin-deficient *H. salinarum* strain (D.O., unpub-

lished; data not shown), thus indicating that this gene product is not required for general viability. According to the X-ray structure of the *H. salinarum* dodecin, several functions might be currently envisioned: first, a major biological role of the dodecins might be just the binding of flavin cofactors. Uncomplexed flavin cofactors are known to be involved in side reactions in the cytosol; for example, riboflavin and FMN induce the degradation of aromatic amino acids in several proteins by photosensitization [26]. The small size of the monomer and the unusual interfacial binding of flavin cofactors would make the dodecins, with 68 residues per bound cofactor, a mostly economic solution for sequestering flavins from free solution while still keeping them in a bioavailable state. By this, the dodecins might act as a buffer for flavin-like cofactors, which are released upon increased cellular demand.

A second function might be light harvesting or protection against UV radiation by the aromatic tetrade. However, excitational coupling, if any, between the riboflavins would have to occur over rather long distances, as the cofactor binding sites are 29 Å away from each other in the dodecin oligomer. Third, the dodecins and their riboflavin dimers could participate in some sort of redox reactions. However, due to the limited access to the reactive N10 atoms in the inner compartment, only ionic species or small compounds can be considered as substrates. Furthermore, the conformation of a fully reduced state of the flavin cofactors would be restrained by the structure of the dodecin particle, because a "butterfly"-like conformation along the N5 and N10 atoms of the isoalloxazine rings that is observed in several other flavoprotein structures could hardly be accommodated in the aromatic tetrade.

A role as an RNA binding protein as suggested by its truncated RNP fold is not supported, due to a lack of conserved sequence motifs in strands β1 and β3 [12]. Furthermore, the known RNA binding sites of RNP modules, which run along the surface of the β sheet, are buried in the dodecin complex. Consequently, such an RNA binding function might be only envisioned if the dodecin dodecamer undergoes reversible dissociation in vivo, that is, upon depletion of flavin cofactors. The purification of the halophilic dodecin under low salt conditions where many halophilic proteins lose their structural integrity indicated that the complex is rather stable. Biochemical and structural data on the dodecin from *Thermus thermophilus* also indicated that dodecin dodecamers are present if the complex is partially depleted of flavin cofactors (L.-O.E. and B. Meissner, unpublished data).

The dodecin gene is not ubiquitous in the prokaryotic world, but restricted to a few eubacteria and a haloarchaeon. A common characteristic of these organisms that bear a dodecin ortholog is an increased exposure to oxygen or radical stress. For example, the pathogenic eubacteria listed in Figure 1A face the action of reactive oxygen and nitrogen species that are released as a host defense by macrophages and neutrophiles [27], whereas nonpathogenic organisms have to tolerate radicals that are formed either by photosynthesis (*C. tepidum*) or metabolism (*G. sulfurreducens*). Interestingly, the second protein of *H. salinarum* that was crystallized and solved by the inverse structural genomics approach was a dodecameric ferritin (S. Offermann, B.B., L.-O.E., and D.O., submitted). This 23-symmetric protein oligomer apparently mediates protection against oxidative stress by sequestering free Fe^{2+} species in the cytosol. As for the dodecins, its occurrence in the proteome of *H. salinarum* is unique in the archaeal kingdom and might indicate that both genes were acquired as "lifestyle" genes [8] from eubacterial organisms, namely to support the aerobic growth of Haloarchaea under bright light conditions.

Finally, the discovery of an unusual flavin binding protein by a procedure where crystals were first generated before the crystallized proteins were further characterized is an alternative to conventional approaches in structural genomics. Furthermore, dodecin is currently not only the smallest known flavoprotein with its 68 amino acids, but also the smallest fold that forms 23-symmetric protein shells. The latter feature might be interesting for engineering highly symmetric protein oligomers by fusing proteins of interest to this novel fold or by using this protein shell as a delivery vehicle for drugs or other bioactive compounds.

Experimental Procedures

Crystallization and Data Collection

The isolation of dodecin-containing samples from the cytosol of *H. salinarum* followed the procedure described in (S. Offermann, B.B., L.-O.E., and D.O., submitted); the elution point for the dodecin fractions was about 0.45 M NaCl during ion exchange chromatography on a MonoQ HR10/10 column (Amersham Pharmacia). Octahedrally shaped crystals of dodecin were grown at protein concentrations of 10–15 mg ml^{-1} by vapor diffusion at 18°C for 1–2 weeks using 30% PEG 400, 0.2 M $MgCl_2$, 0.1 M HEPES (pH 7.5), 1.0 M NaCl as crystallization buffer. The crystallized protein was initially identified as OE3073R by Edman degradation of single crystals and subsequent comparison of the partial amino acid sequence with the genome sequence of *H. salinarum* (www.halolex.mpg.de) using the BLAST option of the PEDANT genome interface [28]. The protein sequence confirmed the absence of an N-terminal methionine in the mature protein.

X-ray data were collected at 100 K with the mother liquor as cryoprotection buffer. Heavy metal derivatives were prepared by soaking the crystals in 10 mM $GdCl_3$ or trimethyl-lead acetate for 45–60 min. Diffraction data for the gadolinium derivative were collected on a CuK_α rotating anode (Rigaku RU-200, 80 mA, 50 kV) with focusing mirrors (Charles Supper) and a 30 cm imaging plate (MAR research). Synchrotron data were collected at 1.7 Å resolution from frozen crystals at beamline ID14/3, ESRF, Grenoble using a MAR 165 mm CCD detector and a wavelength of 0.933 Å. For the RbCl soak, a crystal was transferred to a solution where RbCl replaced the sodium chloride of the crystallization buffer. Data for the RbCl soak were collected on a MAR CCD detector at beamline BW6, MPG-ASF, Hamburg at the K-edge of rubidium ($\lambda = 0.816$ Å). Diffraction data of dodecin were indexed and integrated with DENZO (HKL research) either in space group $F4_132$ ($a = 142.6$ Å) or, due to better computational performance during phasing, in the lower symmetry space group $I4_122$.

Structure Determination and Refinement

The structure of dodecin was solved in space group $I4_122$ by multiple isomorphous replacement (MIR) using the programs SOLVE [29] and DM [30] to determine heavy atom positions and to perform phase refinement. An MIR map that was calculated at 1.8 Å resolution after phase extension allowed semiautomatic tracing and model building by wARP [31]. Subsequently, the dodecin structure was refined in the correct space group $F4_132$ using CNS 1.0 [32] until the refinement converged at an R factor/R_{free} of 0.180/0.208 for data

Table 2. Summary of Crystallographic Analysis

Data Collection Statistics

Data Set	Cell	Resolution (Å)	Measured, Unique Reflections	R_{merge}[a]	$I/\sigma(I)$[b]	Completeness
Native I, indexed in $F4_132$	a = b = c = 142.6 Å	20–1.7	116286, 14027	0.094 (0.542)	38.3 (2.6)	98.5 (99.9)
RbCl ($F4_132$)		24–2.0	76185, 8705	0.092 (0.293)	24.6 (5.4)	99.6 (96.7)
Native II, indexed in $I4_122$	a = b = 100.0 Å, c = 141.4 Å	30–2.4	70556, 13828	0.068 (0.209)	16.5 (3.4)	95.5 (48.5)
$Pb(Me)_3Ac$ ($I4_122$)		30–2.9	86273, 7684	0.114 (0.229)	14.9 (5.1)	92.8 (74.7)
$GdCl_3$ ($I4_122$)		30–3.4	25021, 4869	0.100 (0.240)	20.5 (7.6)	99.8 (99.2)

Phasing Statistics

Derivative	R_{iso}[c]	Sites	R_{Cullis}[d]	Figure of Merit (30–2.9 Å)[e]							
$Pb(Me)_3Ac$, A	0.21	1	0.64	10 Å	6.5 Å	5.1 Å	4.4 Å	3.9 Å	3.5 Å	3.2 Å	2.9 Å	Total
$GdCl_3$ (II), A	0.33	3	0.67	0.55	0.65	0.62	0.59	0.49	0.37	0.30	0.22	0.44

Refinement Statistics

Data Set	Resolution (Å)	Reflections ($F > 0$)	R factor, R_{free}[f]	No. of Atoms, Waters, Ions, Heterogens	Rmsd[g] Bond (Å)	Rmsd[g] Bond Angles, Dihedral Angles (°)
Native	15–1.7	13753	0.180 (0.193), 0.208 (0.214)	515, 98, 5, 27	0.022	2.4, 25.2

[a] $R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I_i(hkl)$; values in parentheses correspond to the highest resolution shell.
[b] As calculated with the program TRUNCATE [30].
[c] $R_{iso} = \Sigma ||F_{deriv}| - |F_{native}|| / \Sigma |F_{native}|$
[d] $R_{Cullis} = \Sigma ||F_{PH} \pm F_P| - F_{H(calc)}| / \Sigma |F_{PH} \pm F_P|$
[e] The figure of merit is defined as the estimated cosine of the phase error.
[f] R factor $= \Sigma |F_{obs} - F_{calc}| / \Sigma F_{obs}$; R_{free} was calculated with 9% of randomly selected data; values in parentheses correspond to the highest resolution shell.
[g] Rms deviations for bond angles and length in regard to Engh and Huber parameters [39].

between 15.0 Å and 1.7 Å (Table 2). As difference density peak heights cannot discriminate between water molecules, or magnesium or sodium cations, only those difference density peaks were assigned as cations where at least five coordination partners were present or, in case of Mg^{2+}, strict octahedral coordination geometry was apparent. The crystals have a solvent content of 65% and comprise one dodecin monomer (V2–Q68), one 5-ribityl-7,8-dimethyl-1*H*-benzo[*g*]pteridin-2,4-dion (riboflavin), two magnesium, one sodium, and two chloride ions, and 98 water molecules per asymmetric unit. The final model consists of 645 atoms and exhibits with an overall G factor of −0.1 good stereochemistry as analyzed by PROCHECK [33]. Homology modeling of the putative dodecin decamer of *P. aeruginosa* was performed by MODELLER4 [34].

Acknowledgments

This work is supported by a predoctoral fellowship (B.B.) of the Fonds der Chemischen Industrie. The authors are very grateful for protein sequencing by J. Kellermann, laboratory assistance by U. Heider, and the support of H. Bartunik at synchrotron beamline BW6, MPG-ASMB, Hamburg, and S. Arzt at ID14-3, ESRF, Grenoble.

Received: September 23, 2002
Revised: January 21, 2003
Accepted: January 31, 2003
Published: April 1, 2003

References

1. Ginzburg, M., Sachs, L., and Ginzburg, B.Z. (1970). Ion metabolism in a halobacterium. I. Influence of age of culture on intracellular concentrations. J. Gen. Physiol. *55*, 187–207.
2. Lanyi, J.K. (1974). Salt-dependent properties of proteins from extremely halophilic bacteria. Bacteriol. Rev. *38*, 272–290.
3. Madern, D., Ebel, C., and Zaccai, G. (2000). Halophilic adaptation of enzymes. Extremophiles *4*, 91–98.
4. Dym, O., Mevarech, M., and Sussman, J.L. (1995). Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. Science *267*, 1344–1346.
5. Frolow, F., Harel, M., Sussman, J.L., Mevarech, M., and Shoham, M. (1996). Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. Nat. Struct. Biol. *3*, 452–458.
6. Pieper, U., Kapadia, G., Mevarech, M., and Herzberg, O. (1998). Structural features of halophilicity derived from the crystal structure of dihydrofolate reductase from the Dead Sea halophilic archaeon, *Haloferax volcanii*. Structure *6*, 75–88.
7. Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N.S., Thorsson, V., Sbrogna, J., et al. (2000). Genome sequence of *Halobacterium* species NRC-1. Proc. Natl. Acad. Sci. USA *97*, 12176–12181.
8. Koonin, E.V., Makarova, K.S., and Aravind, L. (2001). Horizontal gene transfer in prokaryotes: Quantification and classification. Annu. Rev. Microbiol. *55*, 709–742.
9. Reindel, S., Anemüller, S., Sawaryn, A., and Matzanke, B.F. The *DpsA*-homologue of the archaeon *Halobacterium salinarum* is a ferritin. Biochim. Biophys. Acta *1598*, 140–146.
10. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
11. Holm, L., and Sander, C. (1995). Dali: a network tool for protein-structure comparison. Trends Biochem. Sci. *20*, 478–480.
12. Varani, G., and Nagai, K. (1998). RNA recognition by RNP proteins during RNA processing. Annu. Rev. Biophys. Biomol. Struct. *27*, 407–445.
13. Goldgur, Y., Mosyak, L., Reshetnikova, L., Ankilova, V., Lavrik, O., Khodyreva, S., and Safro, M. (1997). The crystal structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus* complexed with cognate $tRNA^{Phe}$. Structure *5*, 59–68.
14. Unge, J., Åberg, A., Al-Kharadaghi, S., Nikulin, A., Nikonov, S., Daydova, N.L., Nevskaya, N., Garber, M., and Liljas, A. (1998). The crystal structure of ribosomal protein L22 from *Thermus thermophilus*: insights into the mechanism of erythromycin resistance. Structure *6*, 1577–1586.
15. van Montfort, R.L.M., Basha, E., Friedrich, K.L., Slingsby, C., and Vierling, E. (2001). Crystal structure and assembly of a eukaryotic small heat shock protein. Nat. Struct. Biol. *8*, 1025–1030.
16. Kim, K.K., Kim, R., and Kim, S.-H. (1998). Crystal structure of a small heat-shock protein. Nature *394*, 595–599.
17. Isupov, M.N., Dalby, A.R., Brindley, A.A., Izumi, Y., Tanabe, T., Murshudov, G.N., and Littlechild, J.A. (2000). Crystal structure of dodecameric vanadium-dependent bromoperoxidase from the red algae *Corallina officinalis*. J. Mol. Biol. *299*, 1035–1049.
18. Brunger, A.T., Adams, P.D., and Rice, L.M. (1997). New applications of simulated annealing in X-ray crystallography and solution NMR. Structure *5*, 325–336.
19. Liepinsh, E., Kitamura, M., Murakami, T., Nakaya, T., and Otting, G. (1997). Pathways of chymotrypsin evolution suggested by the structure of the FMN-binding protein from *Desulfovibrio vulgaris* (Miyazaki F). Nat. Struct. Biol. *4*, 975–979.
20. Gibson, Q.H., Massey, V., and Atherton, N.M. (1962). The nature of compounds present in mixtures of oxidized and reduced flavin mononucleotides. Biochem. J. *85*, 369–383.
21. Kainosho, M., and Kyogoku, Y. (1972). High-resolution proton and phosphorous nuclear magnetic resonance spectra of flavin-adenine dinucleotide and its conformations in solution. Biochemistry *11*, 741–752.
22. Kuntz, I.D. (1971). Hydration of macromolecules. IV. Polypeptide conformation in frozen solutions. J. Am. Chem. Soc. *93*, 516–518.
23. Elcock, A.H., and McCammon, J.A. (1998). Electrostatic contributions to the stability of halophilic proteins. J. Mol. Biol. *280*, 731–748.
24. Richard, S.B., Madern, D., Garcin, E., and Zaccai, G. (2000). Halophilic adaptation: novel solvent protein interactions observed in the 2.9 and 2.6 Å resolution structures of the wild type and a mutant of malate dehydrogenase from *Haloarcula marismortui*. Biochemistry *39*, 992–1000.
25. Madern, D., Ebel, C., Mevarech, M., Richard, S.B., Pfister, C., and Zaccai, G. (2000). Insights into the molecular relationships between malate and lactate dehydrogenases: structural and biochemical properties of monomeric and dimeric intermediates of a mutant of tetrameric L-[LDH-like] malate dehydrogenase from the halophilic archaeon *Haloarcula marismortui*. Biochemistry *39*, 1001–1010.
26. Edwards, A.M., and Silva, E. (2001). Effect of visible light on selected enzymes, vitamins and amino acids. J. Photochem. Photobiol. B *63*, 126–131.
27. Nathan, C., and Shiloh, M.U. (2000). Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. Proc. Natl. Acad. Sci. USA *97*, 8841–8848.
28. Frishman, D., Albermann, K., Hani, J., Heumann, K., Metanomski, A., Zollner, A., and Mewes, H.W. (2001). Functional and structural genomics using PEDANT. Bioinformatics *17*, 44–57.
29. Terwilliger, T., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D *55*, 849–861.
30. CCP4 (Collaborative Computational Project 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D *50*, 760–763.
31. Lamzin, V.S., and Wilson, K.S. (1997). Automated refinement for protein crystallography. Methods Enzymol. *277*, 269–305.
32. Brünger, A., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grossekunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software package for macromolecular structure determinations. Acta Crystallogr. D *54*, 905–921.
33. Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.
34. Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. *234*, 779–815.
35. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both

detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.
36. Merrit, E.A., and Murphy, M.E.P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. Acta Crystallogr. D *50*, 869–873.
37. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. *24*, 4876–4882.
38. Nicholls, A. (1992). GRASP: graphical representation and analysis of surface properties. Columbia University, NY.
39. Engh, R.A., and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein-structure refinement. Acta Crystallogr. A *47*, 392–400.

Accession Numbers

Atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank under ID code 1MOG.