The crystal structure of Dps, a ferritin homolog that binds and protects DNA. Nat Struct Biol 5:294–303

ARTICLE *in* NATURE STRUCTURAL BIOLOGY · MAY 1998

DOI: 10.1038/nsb0498-294 · Source: PubMed

The crystal structure of Dps, a ferritin homolog that binds and protects DNA

R.A. Grant[1], D.J. Filman[1], S.E. Finkel[2], R. Kolter[2] and J.M. Hogle[1,3]

The crystal structure of Dps, a DNA-binding protein from starved *E. coli* that protects DNA from oxidative damage, has been solved at 1.6 Å resolution. The Dps monomer has essentially the same fold as ferritin, which forms a 24-mer with 432 symmetry, a hollow core and pores at the three-fold axes. Dps forms a dodecamer with 23 (tetrahedral) point group symmetry which also has a hollow core and pores at the three-folds. The structure suggests a novel DNA-binding motif and a mechanism for DNA protection based on the sequestration of Fe ions.

Aerobic organisms have evolved a variety of mechanisms to protect their DNA from oxidative damage caused by reactive oxygen species that may be present in the environment or generated during metabolism. These include highly regulated enzymatic systems which recognize and repair damaged DNA and which prevent damage by breaking down reactants that produce free radicals. The oxidative stress response has been well characterized biochemically and genetically in bacterial systems[1]. Recently the 19,000 M_r *E. coli* protein Dps (PexB) has been shown to play an important role in protecting DNA from oxidative damage[2–5]. Under conditions of either nutritional or oxidative stress *E. coli* produces high levels of Dps, which binds DNA with no apparent sequence specificity[2] and (in stationary phase cells) induces significant compaction of the chromosomal DNA (A. Minsky, pers. comm.). These properties make Dps a member of the class of bacterial nucleoid-associated or 'histone-like' proteins which includes HU, HN-S, IHF and Fis[6]. However, in contrast to the other class members, which bind DNA as dimers, electron microscopy has shown that Dps forms a roughly spherical oligomer (estimated to be a dodecamer) ~90 Å in diameter and that Dps in complex with plasmid DNA forms hexagonally packed two-dimensional arrays[2].

A diverse group of proteins with sequence homology to Dps has been identified in other distantly related prokaryotes. These include two other functionally similar DNA-binding proteins, DpsA from *Synechococcus sp.*[7] and MrgA (metallo-regulated) from *Bacillus subtilis* [8]. In addition, the Dps family includes a functional ferritin[9], a novel pilin[10], a bromoperoxidase[11], and other proteins of unknown function[12–14]. Several members of the family have been shown to have the same electron microscopic appearance as Dps[9,12,13]. One of these has also been reported to form hexagonal paracrystalline arrays[12]. Although Dps itself has no sequence homology with any ferritin, the observation of homology between members of the Dps family and the bacterioferritins[7,12] has led to the suggestion that Dps and its homologs are divergent members of the bacterioferritin/ferritin superfamily.

The ferritins are a widely distributed group of iron storage proteins[15]. The structures of several ferritins have been solved by X-ray crystallography[15], including horse spleen apoferritin[16] (a heteropolymer consisting of primarily L chain with ~15% H chain), examples of vertebrate L and H chain homopolymers[17,18], and the heme-binding bacterioferritin from *E. coli*[19] (Bfr). In each case the ferritin monomer is a four helix bundle which oligomerizes such that 24 ferritin monomers associate to form a large, roughly spherical particle with 432 (octahedral) symmetry. The ferritin particle measures a ~125 Å in diameter. It has a 80 Å diameter hollow core which can store up to 4,000 iron atoms in a dense structure composed primarily of the mineral ferrihydrite. It is thought that iron ions enter the core through hydrophilic pores at the three-fold axes of the 24-mer. Ferritins have a ferroxidase activity (associated only with the H chains in the vertebrate ferritins) which is linked to the ability to efficiently incorporate ferrous iron into the mineralized core.

We report here the 1.6 Å resolution X-ray crystal structure of the Dps dodecamer. The structure proves that Dps is a structural analogue of ferritin. The Dps monomer has essentially the same protein fold (four helix bundle) as the ferritin monomer, and the packing of Dps monomers in the dodecamer closely resembles the packing of ferritin monomers around the two-fold and three-fold symmetry axes of the ferritin 24-mer. Like the ferritin oligomer, the Dps dodecamer is a hollow ball with pores at the three-fold axes. In contrast to the 432 symmetry of ferritin, the Dps dodecamer has 23 (tetrahedral) point group symmetry. Dps lacks the small C-terminal α-helix that makes the fourfold interactions in ferritin. Instead of making fourfold interactions, the C-terminal end of the Dps monomer makes a second kind of three-fold interaction not present in the ferritin oligomer. This results in a smaller, lower symmetry oligomer that maintains many of the structural features of the ferritin particle, including unusual clusters of acidic residues at the three-fold axes and a significant net negative charge inside the hollow core.

The crystal structure suggests a plausible model for Dps function. Given the remarkable structural similarity of Dps and ferritin, it is likely that Dps protects DNA from oxidative damage by sequestering iron ions that might otherwise generate free radicals by the Fenton reaction[15]. The Dps DNA-binding activity condenses the chromosome into a compact

[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Committee on Higher Degrees in Biophysics, Harvard University, 240 Longwood Avenue, Boston, Massachusetts 02115, USA.

Correspondence should be addressed to J.M.H. *email: hogle@hogles.med.harvard.edu*

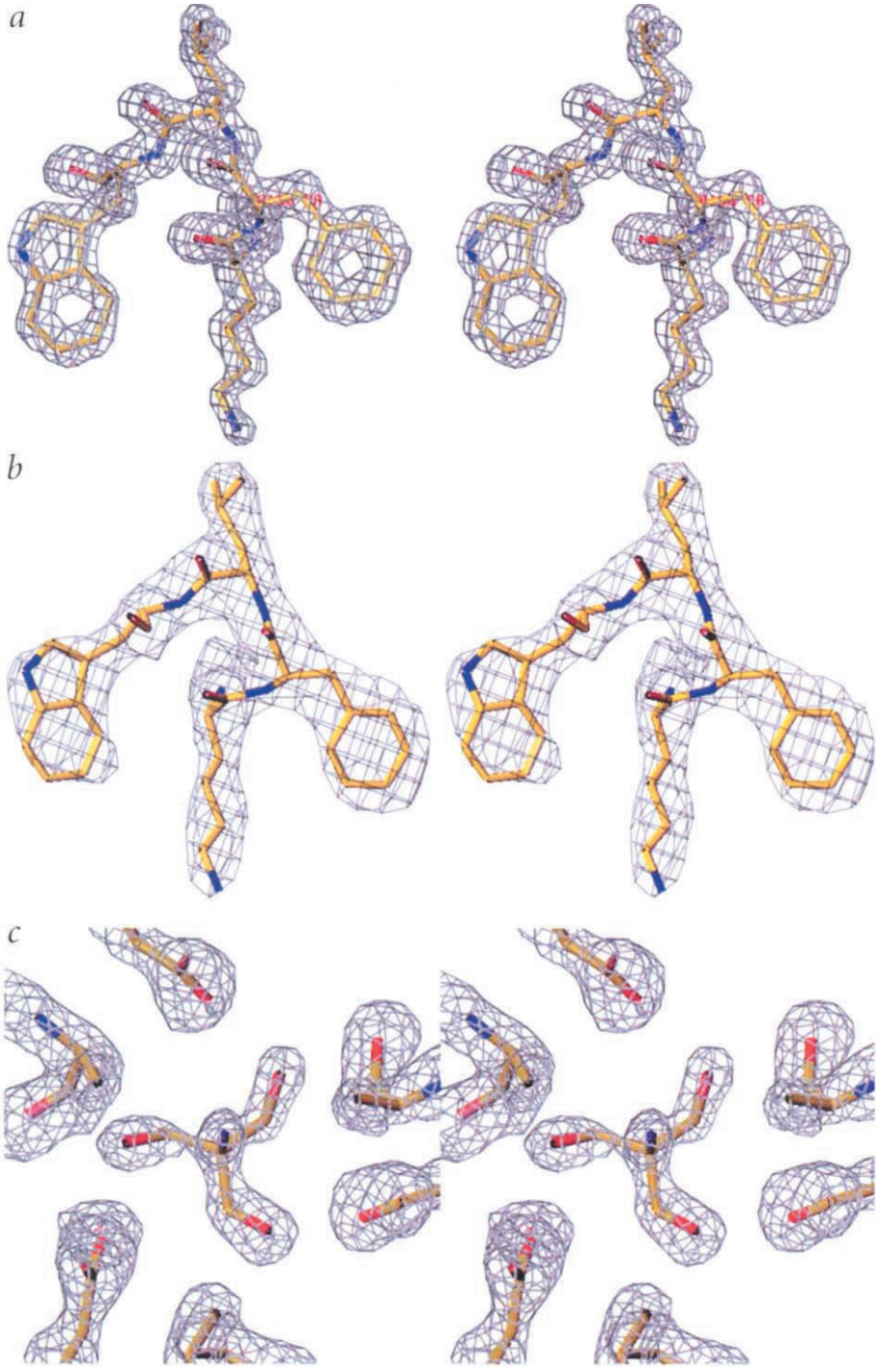

Fig. 1 Electron density ***a***, Representative region of the 1.6 Å resolution $2F_o$ - F_c electron density map. ***b***, Same region of the MIR map after convergence of 12-fold NCS averaging at 3 Å shown with the refined model. ***c***, 2.05 Å resolution $2F_o$ - F_c electron density showing Tris molecule capping the acidic pore connecting the outside of the Dps dodecamer to the hollow core.

structure and localizes this putative iron sequestering activity to where it can provide maximum protection of the DNA. Although the crystal structure reveals no obvious DNA-binding site in the Dps dodecamer, the structure is consistent with a novel mode of DNA binding in which nine lysine side chains from three neighboring hexagonally packed Dps dodecamers interact with the DNA backbone.

Structure determination

The structure was solved using a combination of multiple isomorphous replacement (MIR) methods (Table 1) and non-crystallographic symmetry (NCS) averaging. MIR phases were derived from a mercury derivative, a lead derivative and a selenium derivative, and were refined using the NCS. The selenium derivative data were collected from crystals of wild type protein, but the other two derivative data sets and most of the native data were collected from crystals of a mutant protein (S164C) with a cysteine substituted for the serine at residue 164. The structure solution was complicated by interactions between the molecular (non-crystallographic) symmetry and the crystallographic symmetry. This interaction results in pseudo-A centering in the diffraction data which made it impossible to distinguish between two choices of space group ($P2_12_12$ and $P2_12_12_1$) until late in the structure determination.

Despite the ambiguity of the space group it was possible to determine the identity, orientation and position of the molecular point group from inspection of the native Patterson, the self rotation function (which indicated a 23 molecular point group) and packing analysis. Knowledge of the NCS made it possible to determine the location of heavy atom substitution sites in the mercury derivative using a tetrahedrally locked Patterson vector search. The heavy atom sites in the lead derivative were located using low resolution (5 Å) phases from the mercury derivative. These heavy atom positions were used to calculate MIR phases at 3 Å resolution, and the phases were refined by NCS averaging. Analysis of the NCS averaging statistics suggested that the initial choice of space group ($P2_12_12_1$) was incorrect, so the MIR phases were recalculated and NCS refined at 3 Å resolution in the alternative space group ($P2_12_12$). This produced an improvement in the NCS refinement statistics and electron density maps that suggested the space group was correct. Phases from the transform of the 3 Å averaged map were used to locate the selenium sites in the third derivative and a new set of MIR phases from all three derivatives was calculated and refined by NCS. The majority of the molecular model was built into the resulting averaged electron density at 3 Å. The phas-

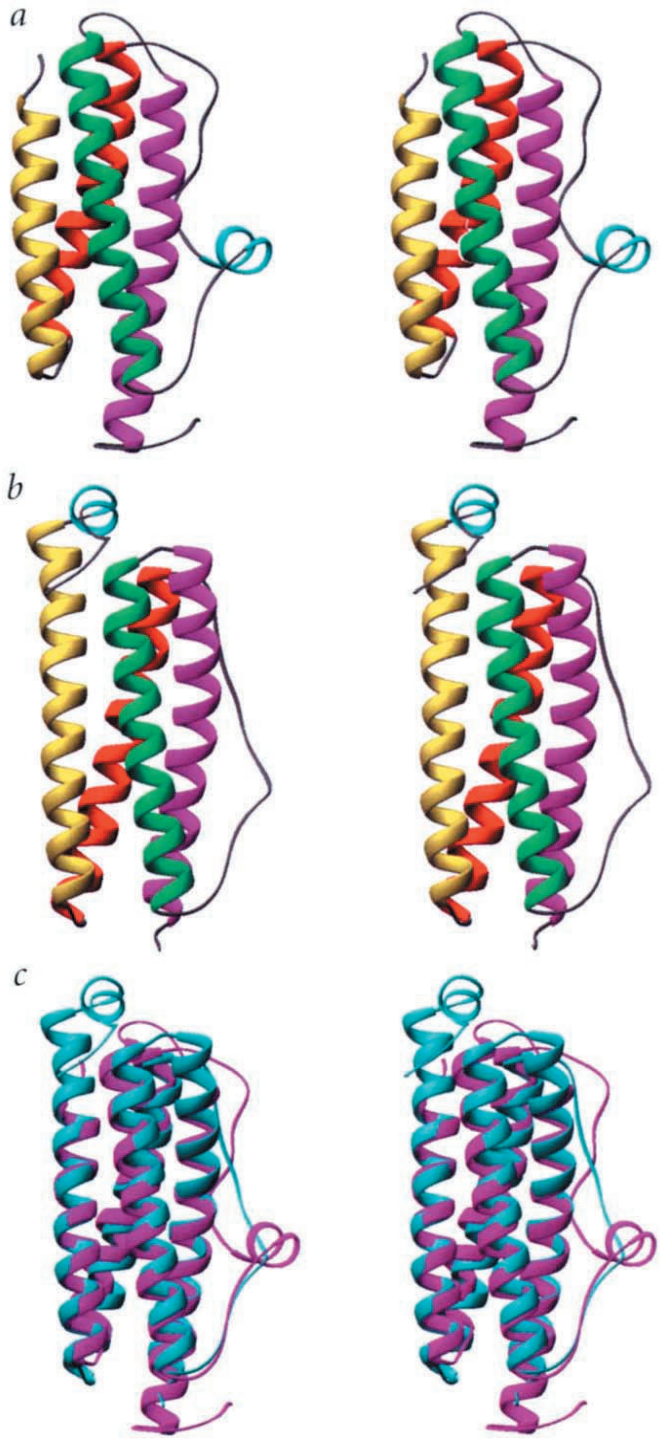

Fig. 2 Comparison of Dps and Bfr monomers. ***a***, Stereo view of Dps monomer. The N-terminus is located at the bottom of the four helix bundle on the right; the C-terminus is at the top on the left. The A helix is colored magenta, the B helix green, the BC helix cyan, the C helix red and the D helix yellow. ***b***, Stereo view of the Bfr monomer in the same orientation as the monomer in panel A. The color coding is the same as in (*a*) except that the E helix is cyan. ***c***, Superposition of the Dps and Bfr monomers based on a least squares minimization of the structurally conserved features identified in Fig. 3. The Dps monomer is colored magenta and the Bfr monomer is cyan.

es were extended to 2.05 Å and refined using strict NCS, then NCS restraints. Solvent atoms and individual temperature factors were incorporated into the model at this stage. Finally, the phases were extended and refined at 1.6 Å using NCS restraints and a bulk solvent correction was applied.

The final model (Table 2) of the Dps dodecamer includes residues 14–167 (all except the N-terminus) of all 12 monomers in the asymmetric unit of the crystal, plus residues 9–13 (with some side chain disorder, notably in Lys 10) of one of the monomers. The missing N-terminal residues appear to be disordered. The final *R*-factor for the model refined at 1.6 Å resolution is 18.8%. The model geometry is consistent with the high quality and resolution of the diffraction data. All of the solvent atoms in the structure are modeled as water, with the exception of 12 ions (modeled as sodium) bound on the outer surface of the dodecamer.

With the exception of the first averaged map calculated in the wrong space group, the averaged maps and the NCS restrained maps were of exceptionally high quality (Fig. 1) except in the vicinity of the disordered N termini, where a significant amount of uninterpretable density was present even at the end of the refinement.

The Dps monomer

The Dps monomer has a four helix bundle core that is almost identical to the core of the bacterioferritin (Bfr) monomer (Fig. 2). The r.m.s. deviation between Cα positions in the cores of the aligned bundles (Fig. 3) is 1.1 Å. In both molecules the first two helices (the A and B helices of Bfr) are connected to each other by a short loop, as are the last two helices (the C and D helices of Bfr). The two pairs of helices are connected by a long loop that stretches across the surface of the bundle from the end that contains the N-terminus to the end containing the C-terminus. In Bfr, the BC loop has no secondary structure, but there is a small E helix (residues 146–151) at the C terminus of the molecule. In the Bfr 24-mer, the E helix interacts with symmetry related copies of itself around the four-fold axes (Fig. 4*a*). In Dps, there is no E helix but there is a short helix (residues 95–101) in the middle of the BC loop. In the Dps dodecamer, this small helix interacts with a symmetry related copy of itself across the two-fold axes (Fig. 4*b*). The C-terminal ends of the A and D helices of Bfr are slightly longer than the corresponding helices of Dps. This, along with the presence of the E helix, makes the Bfr molecule slightly longer than Dps at the C-terminal end of the bundle.

The very accurate alignment of the structural elements of Dps and Bfr allows a precise comparison of the structure of Dps in the region homologous to the ferroxidase center of Bfr. The ferroxidase center, located in the middle of the Bfr four helix bundle, was first identified as the location of a bimetallic cluster in the crystal structure of human H ferritin[18] (L ferritin lacks a ferroxidase center). Site-directed mutagenesis experiments have since verified the role of this structure in accelerating the oxidation of ferrous iron by Bfr[20]. In Bfr, four aspartate and two histidine side chains coordinate the two metal ions. The corresponding residues in the Dps structure include one histidine, two phenylalanines, two leucines, and a tyrosine (Fig. 3) whose side chains make predominantly hydrophobic interactions with each other at the core of the bundle.

The Dps homolog from *Listeria innocua* (a functional ferritin) also lacks the ferroxidase center, but appears to have a cluster of acidic side chains similar to the putative iron core nucleation site

a

```
                                  αααααααααα A αααααααααα
Dps  : mstaklvkskatnllytrndvsdseKKATVELLNRQVIEFIDLSLITKQA  50
Bfr  :                        mkgDTKVINYLNKLLGNELVAINQYFLH  28
                                    αααααααααα A αααααααααα

       αα      αααααααααα B αααααααααα        α BC α
Dps  : HWnmrganfIAVHEMLDGFRTALIDHLDTMAERAVQLGGVALgttqvins  100
Bfr  : ARmfknwglKRLNDVEYHESIDEMKHADRYIERILFLEGLPNlqdlgkln  78
       αααααα  αααααααααα B αααααααααα

       α           αααααααααα C αααααααααα      ααααα
Dps  : ktplksypldihnvqdhlKELADRYAIVANDVRKAIGEa---kDDDTADI  147
Bfr  : iged---------veemlRSDLALELDGAKNLREAIGYadsvhDYVSRDM  119
                    αααααααααα C αααααααααα

       ααααα D ααααααα
Dps  : LTAASRDLDKFLWFIESNIe                          167
Bfr  : MIEILRDEEGHIDWLETELdliqkmglqnylqaqireeg        158
       αααααααααα D αααααααααα   α E α
```

b

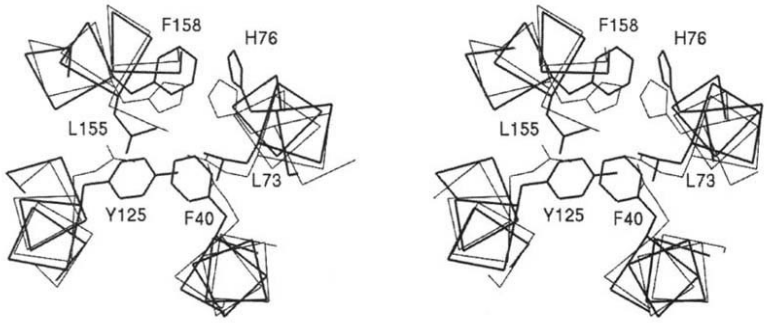

Fig. 3 ***a***, Structural alignment of the Dps and Bfr amino acid sequences α-helices (α) and identities (.) are indicated. The ferroxidase center of Bfr is highlighted by the bold font. The alignment is based on a least squares superposition of the Cαs (1.1 Å r.m.s. displacement) of the residues in uppercase. In a search for other homologous proteins in the PDB using the program Dali[39] the top three hits were rubrerythrin[40], the H chain of human ferrtin, and Bfr, aligning with r.m.s. displacements of 2.8 Å, 2.8 Å and 2.6 Å respectively. At lower scores were a large number of other proteins containing four helix bundles. ***b***, The Bfr ferroxidase center (fine lines) and the corresponding structure in Dps (bold lines) are shown in a stereo view of the two overlapped structures viewed down the center of the four helix bundles. In Bfr four glutamate and two histidine side chains (not labeled) coordinate two metal ions. In Dps two leucines, two phenylalanines, a histidine, and a tyrosine side chain occupy the equivalent positions, forming a typical hydrophobic core of a protein.

of L ferritin. This cluster is located on the face of the helix bundle that forms the inner surface of the hollow core of the dodecamer. In the *Listeria* protein an additional acidic side chain (Asp 48) nearby may confer a ferroxidase activity to the site as well[9]. In Dps, neither of the two critical glutamate residues (52 and 55 in the *Listeria* protein) of the nucleation site nor the additional aspartate is conserved, but an aspartate residue (residue 75 in Dps) is substituted at the Glu 55 position.

The Dps dodecamer

The Dps dodecamer (Fig. 4) measures ~90 Å in diameter. It has a hollow core ~45 Å in diameter. In assembling the dodecamer, the Dps monomers make two-fold and three-fold interactions that closely resemble the packing of the monomers in the ferritin oligomer (Fig. 5). However, there are two obvious differences in the architectures of Dps and ferritin. The first is the absence of any four-fold interactions (mediated in ferritin by the E helix) in the Dps dodecamer. The second is that ferritin's 432 symmetry dictates that only one type of trimeric interaction between monomers is possible. Because of the octahedral symmetry, the chemical environment around the three-fold axes is the same on the 'front' and 'back' sides of the ferritin oligomer. However, the 23 symmetry of the Dps dodecamer leads to two nonequivalent environments around the three-fold axes, one corresponding to the interactions between monomers on the 'front' of the dodecamer, and the other to interactions around the same threefold on the 'back' side of the dodecamer. At the N-terminal end of the Dps four helix bundle, the packing around the three-fold axes is similar to the packing in ferritin. The other three-fold interaction takes place at the C-terminal end of the Dps bundle, where the ferritin monomer makes four-fold interactions with its neighbors.

In the ferritin oligomer there are hydrophilic pores, lined by acidic side chains, at the three-fold axes. These pores connect the 80 Å diameter hollow core to the outside of the particle. In the Dps dodecamer, there are similar pores at the N-terminal three-fold interfaces. TRIS molecules were observed capping these pores (Fig. 1*c*) in maps calculated from the 2.05 Å data set (Table 1), which was collected from crystals that were frozen in buffer containing TRIS. Another type of pore is found at the other (C-terminal) three-fold interfaces. This pore is smaller and less acidic and contains a hydrophobic constriction formed by the Cβ of Ala 61.

Heavy atom sites

In the mercury derivative, two mercury atoms are covalently attached to each of the twelve cysteine side chains of the dodecamer. A cysteine was engineered into the protein by site-directed mutagenesis (there are no cysteines in the wild-type protein) to make the mutant S164C with the expectation that its reactive side chain might produce a useful derivative. The cysteine side chains are located on the outer surface of the dodecamer, near the C terminus of each monomer. The lead binding sites are on the inner surface of the hollow core at the interfaces between monomers related by the dodecamer two-fold symmetry. Each lead ion binds in a pocket formed by side chains from the A helix (His 51 and Lys 48) and B helix (Asp 67) from one monomer and from the C-terminal end of the B helix (Glu 82 and Asp 78) of the symmetry related monomer.

Dodecamer electrostatics

In order to investigate how the Dps dodecamer may bind DNA, the electrostatic potential of the molecule (Fig. 6) was calculated using DelPhi[21] and displayed using GRASP[22]. It was expected that surface features with highly positive electrostatic potential that might interact with the phosphate backbone of the DNA would be found. Instead it is clear that, with the exception of some small patches of positive charge, the electrostatics of the dodecamer surface are dominated by the concentration of negative charge

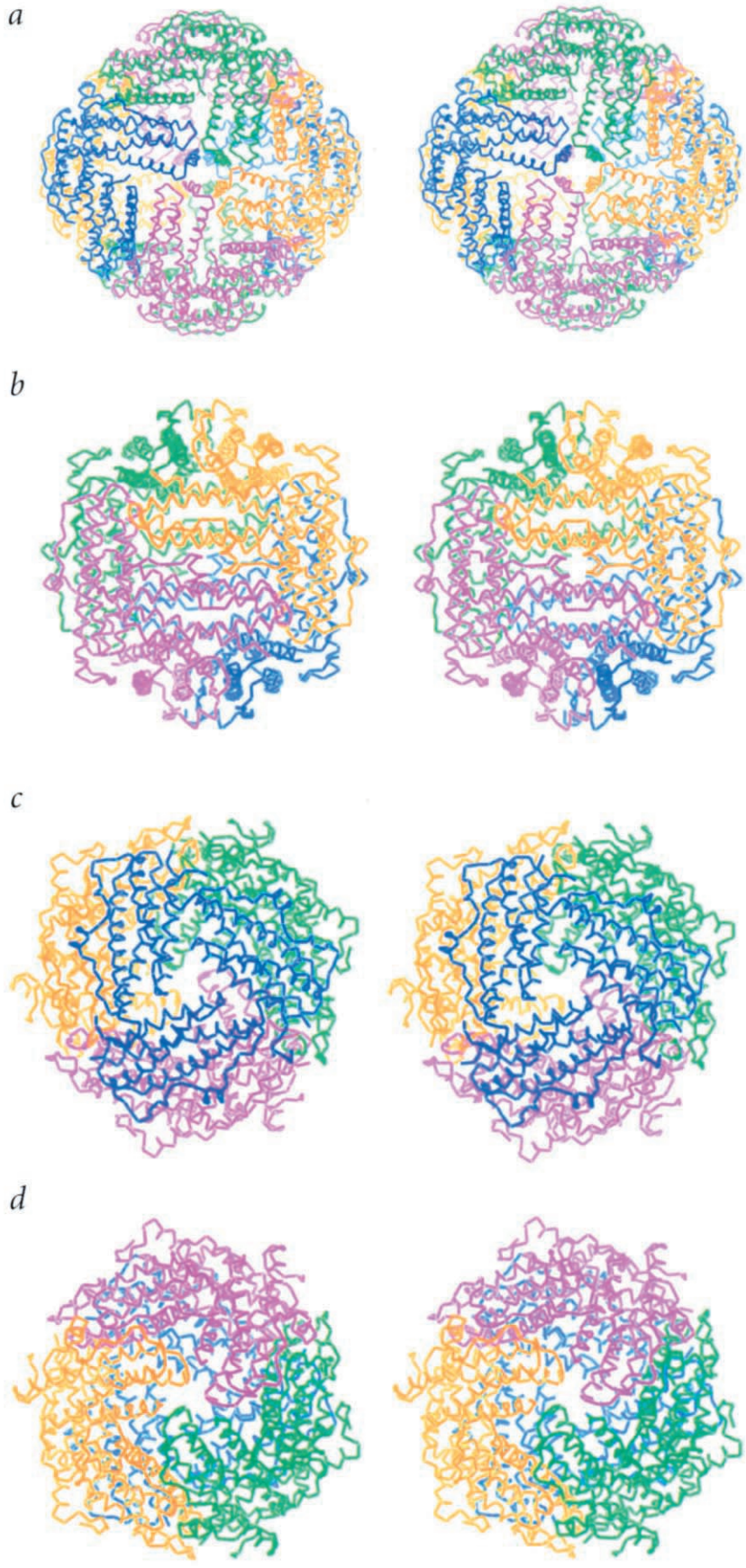

Fig. 4 Comparison of the Bfr 24-mer and the Dps dodecamer. ***a***, Stereo view of the Bfr 24-mer down the four-fold axis. Eight different trimers surrounding the pores at the threefold axes are highlighted by color coding. Trimers related by twofold symmetry are the same color. A more detailed view of the trimer structure is presented in Fig. 5. ***b***, Stereo view of the Dps dodecamer looking down a twofold axis. Each of the four trimers of the type present in the Bfr 24-mer is colored differently. ***c***, Stereo view down a threefold axis of the Dps dodecamer. The trimeric interaction at the front of the dodecamer is similar to the one in ferritin. ***d***, Stereo view of the dodecamer in (*c*) rotated 180° around the horizontal axis. The threefold interaction at the front of this view (between green, magenta and yellow trimers) has no analogous interaction in ferritin.

around the two types of three-fold interfaces, particularly the interfaces that form the ferritin-like pores. The electrostatic potential is least negative in the area surrounding the twofold interfaces, where there are four lysine side chains (two from each monomer) which could interact with the DNA. No other electrostatic interaction with the outside surface seems likely. It is possible that the DNA-binding species is not the dodecamer but a monomer, dimer, trimer or hexamer. The predominant species detected after crosslinking purified Dps with glutaraldehyde are monomers, trimers, and dodecamers, with trace amounts of dimer and hexamer (S.E.F., unpublished data). It has been reported that DpsA binds DNA as a hexamer[7]. The hexameric subassemblies that comprise the Dps dodecamer have large clefts into which DNA would easily fit. However, it is unlikely that DNA-binding requires dodecamer dissociation for two reasons. First, the interactions between monomers in the dodecamer are very tight, involving numerous salt bridges, hydrogen bonds and hydrophobic interactions. Second, the electrostatic potential of the surfaces that line the inside of the dodecamer are highly and exclusively negative (not shown). Dissociation of the dodecamer would expose negatively charged surfaces that would repel DNA.

Packing of dodecamers in the crystal

The dodecamers pack in the unit cell with each of the N-termini in a dodecamer making a close approach to a neighboring dodecamer. Actual crystal contacts are made with ten of the twelve neighbors. Four of these contacts are with dodecamers related by one screw axis, four are with dodecamers related by the other screw axis and the other two contacts are across twofold axes. The neighbors that do not make crystal contacts are related by the twofold

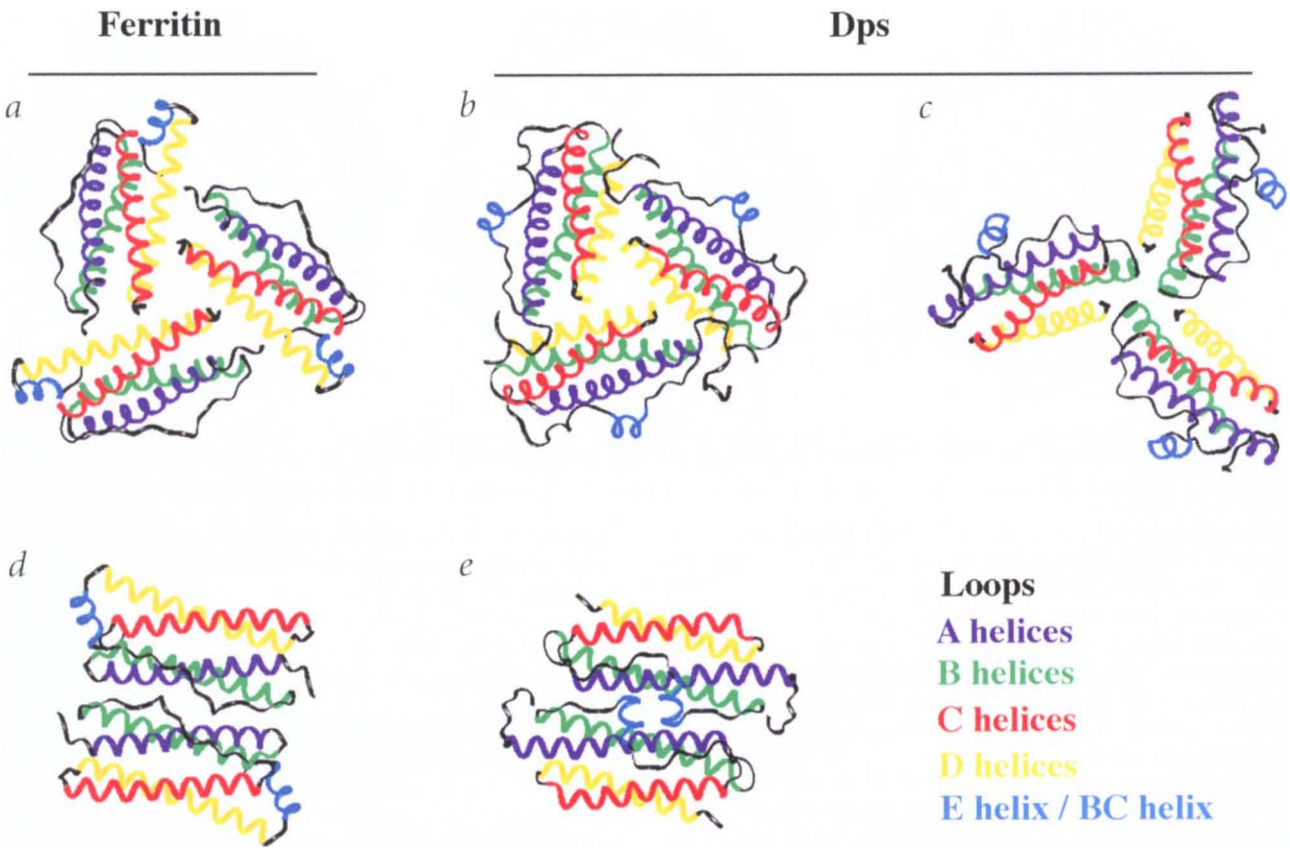

Fig. 5 Comparison of monomer–monomer interactions in Dps and Bfr. ***a***, Ribbon diagram of trimer from the Bfr 24-mer. Color coding of the helices is shown on the lower right of the figure. ***b***, Ribbon diagram displaying one of the two types of trimeric interaction between monomers in the Dps dodecamer. The acidic hole at the center of the trimer connects the exterior of the dodecamer to the hollow core. Note the similarity of this trimer to the Bfr trimer in (a). ***c***, Ribbon diagram displaying the second type of trimeric interaction. The hole at the center of this trimer is smaller than that in the other type of trimer and contains a hydrophobic constriction. ***d***, Ribbon diagram displaying the dimeric interaction between Bfr monomers. ***e***, Ribbon diagram displaying the dimeric interaction between Dps monomers. Note the similarity to the Bfr dimer.

symmetry. Due to the approximate alignment of the dodecamer (non-crystallographic) two-fold axes with the crystallographic axes (Table 2), dodecamers related by the 2_1 screw axes have an approximate local two-fold between the N-terminal sequences making the crystal contacts. When viewed down the three-fold axis of a dodecamer, this packing arrangement resembles multiple layers of dodecamers in two-dimensional sheets one dodecamer thick. The packing of dodecamers within these sheets is approximately hexagonal (Fig. 7) and is remarkably similar to that observed in negatively stained electron microscopy (EM) preparations of Dps-plasmid complexes[2]. It is clear from inspection of the EM images that the sheet is composed of densely packed dodecamers that interact with their six nearest neighbors. The specific interactions seen in pseudo-hexagonal array in the crystal may be similar to those in the Dps-plasmid complex. If so, the crystallization conditions may be promoting some of the same protein–protein interactions produced by DNA binding, but the requirement for order in three dimensions produces a distortion of the two-dimensional lattice.

Model for DNA binding and protection

Since it is clear from the structure that the dodecamer is the probable DNA-binding species, and Dps–DNA complexes have been shown to form hexagonally packed two-dimensional arrays, we propose a hypothetical model for DNA binding by hexagonally packed dodecamers. The pseudo-hexagonal sheets in the crystal are penetrated by three crystallographically distinct kinds of holes where solvent channels separate three neighboring dodecamers. The holes are occupied by the disordered N-terminal residues not present in the model. Each disordered N terminus contains three lysine residues, so each hole contains nine lysine side chains that could interact with DNA. One of the three different kinds of holes is approximately the diameter of a DNA B helix; one of the other two holes is slightly smaller; the third is significantly larger. If a sheet were allowed to adjust so that all of the interactions between neighboring dodecamers were identical, a perfect hexagonal two-dimensional lattice would be created in which all the holes would be similar and sufficiently large to easily accommodate a DNA helix.

Binding of Dps to DNA could promote two-fold dodecamer–dodecamer interactions between N termini similar to those observed in the crystal, inducing the formation of hexagonal sheets with the DNA backbone interacting with lysine side chains lining the holes and perhaps also with those arranged around the two-fold axes. This would result in a sheet of dodecamers with DNA threaded through the holes in the sheet. Even if the sheet were of very limited extent, and if the loops of DNA between holes were very large, this type of

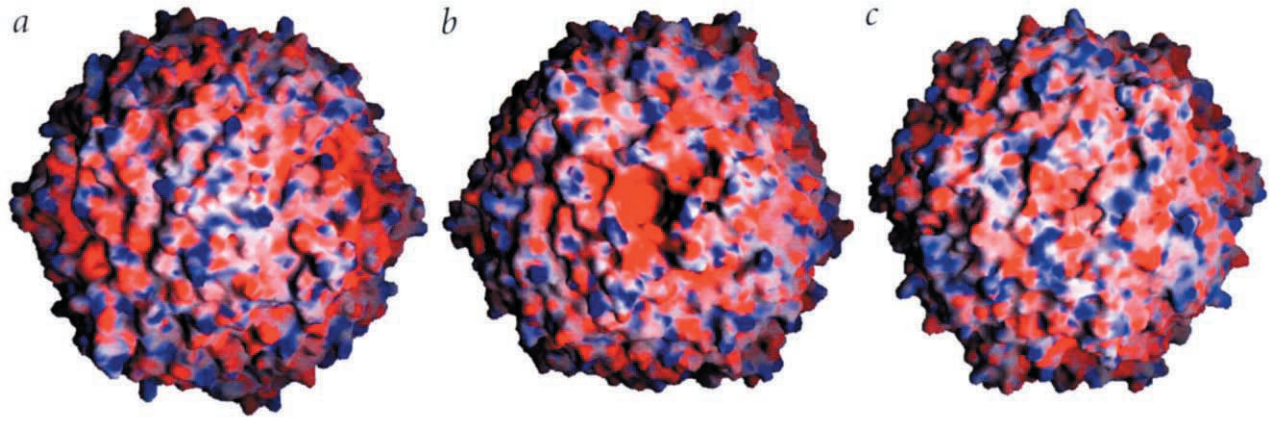

Fig. 6 Electrostatics of the Dps Dodecamer. The molecular surface is colored red where the charge is negative and blue where positive. ***a***, View down a two-fold axis. ***b***, View down the three-fold axis with the acidic pore in front. ***c***, View down the three-fold axis with the acidic pore in back.

interaction could be responsible for the Dps-dependent compaction of DNA observed in stationary phase. Protein–protein interactions between different hexagonal sheets (like those in the crystal) could contribute to the three dimensional organization of the chromosome but would not necessarily be required, since the binding of linearly distant DNA sequences to the same small sheet would result in a significant reduction in the conformational degrees of freedom available to the chromosome.

Gel shift assays have shown Dps binds DNA with a K_d of ~2×10^{-7} M and that DNA fragments smaller than ~90 base pairs do not bind Dps well (S.E.F., unpublished data). If only a single hole in the hexagonal sheet were required for DNA binding, a much smaller fragment should be able to bind. This suggests that a DNA molecule must interact with at least two holes in order to bind well in this model. It also may be significant that a 90 base pair DNA fragment would be approximately the length required to wrap once around a Dps dodecamer. The model also suggests that the lysines in the disordered N-termini of Dps are important if not required for DNA binding. However, in a published sequence alignment[7], similar residues are not found at the corresponding positions in DpsA and MrgA, the other two DNA-binding members of the Dps family. DpsA has no basic residues in its N-terminus. The N-terminus of MrgA is 16 amino acids shorter than that of Dps, so it can have no structural equivalent of the disordered lysines in Dps. MrgA does, however, have two lysines in its truncated N-terminus.

A combination of DNA condensation and masking with protein could account for some of the ability of the Dps dodecamer to protect DNA from oxidative damage, but it is likely that a significant additional protective effect is derived from a ferritin-like sequestration of ferrous (or other) ions that might otherwise generate free radicals. While there is currently no evidence that Dps sequesters iron, the lead derivative used to solve the crystal structure proves that the dodecamer can bind metal ions tightly on the inside surface of its hollow core. Access to the lead binding site was probably through the acidic pore, since any significant dissociation of the dodecamer or large conformational changes in it would have disrupted the crystal during the heavy atom soaking experiment. It seems reasonable to postulate that cations would be attracted to the negatively charged pores leading to the hollow core of Dps. A similar mechanism has been proposed for the import of ions into the core of ferritin[15]. Once inside, even if the iron ions do not actually bind the protein as lead ions do, they could remain trapped by the highly negative electrostatic potential of the core. However, naturally occurring ferritin oligomers have at least some monomers with the ferroxidase center which Dps lacks. The ferroxidase activity is important in the nucleation of mineralization of the iron core of Bfr[20]. The absence of a ferroxidase activity in Dps may preclude significant mineralization of trapped ions.

Relationship between the Dps family and the ferritins

The prediction that the Dps family of proteins is related to the bacterioferritins was based on protein sequence alignments. Evans *et al.*[12] aligned 15 sequences and concluded that many of the Dps homologs, but not Dps, had conserved Bfr ferroxidase centers. While the prediction that Dps lacks a ferroxidase center is correct, it is based on results that fail to correctly predict the alignment of the structurally conserved cores of the four helix bundles of Dps and Bfr. We have used the program ClustalW[23] to align all of the known Dps homologs (not shown). The results are not substantially different from those of Peña and Bullerjahn[7], whose alignment of DpsA, Dps, six other Dps homologs and Bfr is relatively accurate in its prediction of the structural homology of Dps and Bfr. It is very interesting that in both alignments the sequence of the bromoperoxidase from *Streptomyces aureofaciens* has a large N-terminal extension before the A helix of Dps and is truncated before reaching the D helix of Dps. This suggests that if the bromoperoxidase monomer is a four helix bundle like Dps it must have a very different topology, with the N-terminal helix taking the place of the C-terminal helix of Dps (and Bfr).

The striking similarity in secondary and tertiary structure of Dps and the ferritins extends to quaternary structure, but only with respect to the two-fold and N-terminal three-fold interactions between monomers. Dps, which lacks the E helix that mediates the four-fold interactions in ferritin, has no four-fold symmetry. Of the eight examples of the Dps family in the alignment of Peña and Bullerjahn all but one, the protein from *Anabaena variabilis*, terminate before the Bfr E helix sequence. This, along with EM observations of a number of Dps homologs, suggests that the other members of the Dps family form dodecamers, rather than 24-mers like the ferritins. The one likely exception to this prediction is the pilin[10]. While the sequence homology of the pilin to Dps is indisputable, it is not clear how dodecamers ressembling Dps might assemble into a filamentous structure like a pilus, although it is conceivable that quarternary interactions between four helix bundles similar to one or more of those seen in Dps might give rise to a filament.

Table 1 Data collection and phasing statistics

Data Set	Resolution	Total # ofobservations	# unique hkl	R_{sym}[1]	R_{iso}[2]	Phasing Power[3]	R_{cullis}[4]
Native_wt[5]	40–2.05 Å	779,516	135,336	0.074			
			(98.3%)	(0.127)	-	-	-
S164C_1[6]	40–2.5 Å	388,458	77,645	0.050			
			(95.6%)	(0.091)	-	-	-
S164C_2	40–2.5 Å	591,460	72,221	0.063			
			(89.3%)	(0.144)	-	-	-
S164C_3	40–1.6 Å	812,348	258,456	0.054			
(CHESS)			(88.9%)	(0.231)	-	-	-
S164C_4[7]	40–1.6 Å	408,332	267,680	0.046			
			(92.0%)	(0.062)	-	-	-
TAMM	40–2.95 Å	278,130	46,405	0.078			
			(99.8%)	(0.147)	0.248	1.64 (a)	0.70 (a)
						1.37 (c)	0.64 (c)
Pb	40–2.95 Å	134,899	41,329	0.077	0.249	1.23 (a)	0.79 (a)
			(87.5%)	(0.219)		0.87 (c)	0.77 (c)
Se	40–2.5 Å	602,015	80,251	0.066			
			(99.7%)	(0.131)	0.190	1.47 (a)	0.73 (a)
						1.08 (c)	0.71 (c)

Space Group $P2_12_12$; a = 134.41 Å, b = 139.65 Å, c = 118.11 Å
Overall figure of merit = 0.502 (for 3 Å MIR phase set)

[1] $R_{sym} = \Sigma_{i,(hkl)} | I_{i(hkl)} - I_{(hkl)} | / \Sigma_{i,(hkl)} | I_{(hkl)} |$; where $I_{(hkl)}$ is the statistically weighted average intensity of symmetry equivalent reflections. The number in parentheses is the value for highest resolution bin. All measurements were included (no cutoff was applied).
[2] $R_{iso} = \Sigma_{i,(hkl)} | I_{h(hkl)} - I_{n(hkl)} | / \Sigma_{(hkl)} | I_{n(hkl)} |$, where I_h is the intensity from the scaled derivative data set and I_n is the intensity from the native data set (S164C_1 was the native reference for all derivatives).
[3] Phasing Power = $\Sigma_{(hkl)} | F_H | / (|F_{PH}| - |F_P + F_H|)$; (a) acentric reflections, (c) centric reflections.
[4] $R_{cullis} = \Sigma_{(hkl)} (| F_{PH} | - |F_P + F_H |) / \Sigma_{(hkl)} | F_{PH} - F_P |$; (a) acentric reflections, (c) centric reflections.
[5] These data are from a crystal of wild-type protein, all others are from crystals of the S164C mutant. They were used for refinement at 2.05 Å.
[6] These data were truncated to 3 Å resolution for use in averaging. They were also used for scaling derivative data and the calculation of isomorphous differences.
[7] This data set was obtained by merging the reduced data sets S164C_1, S164C_2, and S164C_3 with SCALEPACK. The R_{sym} value is has no contributions from reflections beyond 2.5 Å resolution. These data were used for the final, high resolution refinement.

The Dps structure clearly represents a new DNA-binding motif, distinct from those of IHF, HN-S, HU and Fis, the other 'histone-like' proteins whose structures have been determined[24–28]. It has been suggested that the Dps family might have evolved as heme or metal binding proteins that later acquired a DNA-binding activity[7]. Indeed, two of the known Dps homologs are linked to metal metabolism. The expression of the *B. subtilis* protein MrgA has been shown to be repressed by metal ions[8] and the homolog from *L. innocua* has been shown to be a functional ferritin, the first ever indentified in a gram-positive bacterium. Bozzi *et al.* suggested that this protein might be representative of the ancestral, metal-binding enzyme[9]. The remarkable structural homology between Dps and the ferritins leaves little doubt that the two families are related. Given the simpler structure of Dps, and its lack of a ferroxidase center, it is tempting to speculate that both families evolved from a dodecameric protein whose function was simply to trap, but not mineralize, metal ions and that the ability to efficiently oxidize and mineralize iron and to form fourfold interactions, came later. In any case, it is clear that the hollow-cored dodecamer motif exemplified by Dps has been adapted to a variety of functions. It will be interesting to see what structural studies of other members of the Dps family reveal about how this motif is used to accomplish their diverse functions.

Methods

Production, purification and crystallization of Dps. Cells expressing the wild type protein were grown as described[2]. The Se-methionyl protein was produced in cells carrying the pBAD18-dps plasmid[3] using a protocol that down regulates methionine biosynthesis[29]. The mutant protein S164C was produced using a modified pBAD18-dps plasmid carrying the mutant gene.

Dps was precipitated with ammonium sulfate from cell lysates that had been passed through a DE-52 column equilibrated with 50 mM Tris (pH 8), 250 mM NaCl, 0.1 mM EDTA. The precipitate was dissolved in and dialyzed against 50 mM Hepes (pH 7.5), 50 mM NaCl, 0.1 mM EDTA, and loaded onto an SP-Sepharose column. Dps was eluted with a 50 mM to 1.0 M NaCl gradient, dialyzed against 50 mM Tris (pH 8), 100 mM NaCl, 0.1 mM EDTA, 0.01% NaN_3, and concentrated to 15–20 mg ml^{-1} by membrane filtration. All buffers used in the purification of S164C also contained 2 mM DTT.

Crystals were grown in hanging drops equilibrated against 50 mM Tris (pH 8.0), 1.55–1.7 M sodium formate, 13–16% PEG-8000 and were harvested into 1.7 M sodium formate, 15% PEG-8000. Heavy atom soaks used 1–2 mM tetrakis(acetoxymercuri)methane (TAMM) or 5 mM lead acetate solutions in harvest buffer. Crystals were frozen in harvest buffer containing 20% glycerol.

X-ray data collection and processing. Diffraction data were collected at -160 °C using a MAR 30 cm image plate on a rotating anode source and on Fuji image plates at the CHESS F1 beamline. Intensities were integrated with DENZO, then scaled and merged with SCALEPACK[30]. The CCP4[31] package was used for calculation of Pattersons, cross-difference Fourier maps, self-rotation functions, and in the refinement of heavy atom sites. Heavy atom and native data were scaled using MAXSCALE[32]. The 1.6 Å native data set was produced using SCALEPACK to combine three reduced SCALEPACK data sets.

Locating heavy atom sites. Twelve TAMM sites were located at 5 Å resolution using a vector search program written specifically for this problem. The algorithm used exploited the NCS to search real space for heavy atom positions on a 1 Å grid, over a volume

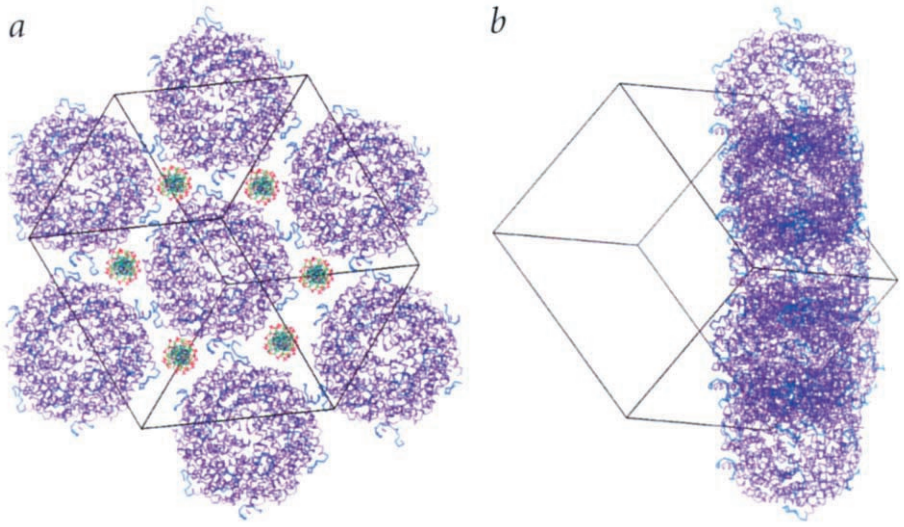

Fig. 7 Pseudo-hexagonal sheets in the Dps crystal and a model for DNA binding. ***a***, Cα traces of a central dodecamer and six of its nearest neighbors are shown to demonstrate the pseudo-hexagonal packing present in the crystal. Most of the trace is colored purple, but the N-terminal portions of the ordered structure (residues 14–25) are shown in blue. The outline of the crystallographic unit cell is shown in black. The holes between three adjacent dodecamers in the hexagonal lattice could accomodate a DNA helix (not present in the crystal but modeled in green with phosphates colored red). Three N-termini line each hole. Each N-terminus includes three lysine residues (disordered in the crystal and therefore not present in the molecular model) which could interact with DNA. ***b***, Edge on view of the dodecamers in (*a*) with the same unit cell outline but no DNA.

corresponding to one twelfth of a sphere of radius 50 Å centered on the NCS origin. Each grid point represented a possible heavy atom site in the monomer. For each grid position, 47 symmetry related sites were calculated from all possible combinations of NCS and crystallographic symmetry. The unique vectors between the 48 sites were calculated and looked up in the Patterson function. The values found were filtered (by discarding the weakest 25% of the values and the highest 10%) and summed to construct a score for the potential heavy atom site. Twelve lead sites were located by a cross-difference Fourier calculation using SIR phases from the TAMM derivative.

Table 2 Model statistics and refinement[1]

R-values	
R_{crys}[2]	0.188
R_{free}[3]	0.220
R_{nonx}[4]	0.140
R.m.s. deviations from target values	
Bond lengths (Å)	0.016
Bond angles (°)	1.425
Average B-factor	16.8

The parameters that describe the position and orientation of the NCS point group with respect to the crystallographic axes were refined extensively and repeatedly during the NCS averaging that preceeded X-PLOR refinement. The final refined values for T, the origin of the NCS point group (the center of the dodecamer), and M, the rotation of its twofold axes away from perfect alignment with the crystallographic axes, were:

$$T^{5} = (0.4591, 0.5003, 0.4999)$$

$$M = \begin{pmatrix} 0.99942 & 0.03394 & -0.00121 \\ -0.03394 & 0.99942 & -0.00227 \\ 0.00113 & 0.00227 & 1.00000 \end{pmatrix}$$

[1]The molecular model includes residues 14–167 of all twelve monomers in the asymmetric unit and residues 9–13 of chain A, with 14,968 protein atoms, 1,268 solvent atoms, and 12 sodium ions.
[2]$R_{crys} = \Sigma_{(hkl)} | F_o | - | F_c | / \Sigma_{(hkl)} | F_o |$; value reported by X-PLOR for 1.6 Å data set S164C_4.
[3]R_{free}, same as R_{crys} except that the sum is over a test set containing a randomly selected 10% of the F_o that were omitted from the X-PLOR refinement process. This statistic is essentially meaningless because of the 12-fold non-crystallographic symmetry.
[4]R_{nonx}, same as R_{crys} except that F_c is obtained from the transform of the 2.05 Å averaged map at the convergence of the XX12 refinement.
[5]T is in fractional coordinates.

Phase and model refinement. MIR phases from the TAMM and lead acetate derivatives were calculated in $P2_12_12_1$ using MLPHARE[31] and refined to convergence by NCS averaging[33] using locally developed programs[34]. The parameters describing the NCS were optimized as described[34]. The resulting averaged map was interpretable but noisy. The NCS averaging residual R_{nonx} (see Table 2 for definition) was reasonably good (29.9%) for reflections with even index parity (k + l = 2n) but much worse (44.5%) for reflections with odd parity, suggesting that the space group was actually $P2_12_12$. The heavy atom positions were re-refined in $P2_12_12$ and MIR phases were re-calculated. These were refined by NCS averaging in $P2_12_12$ at 3 Å then used to determine more accurate heavy atom positions from the TAMM and lead acetate derivatives and to find three Se sites per Dps monomer in the Dps–Se data. With the improved phases the TAMM peaks, initially modeled as a single high occupancy Hg atom per monomer, were each resolved as a pair of distinct peaks 3 8 Å apart. A third set of MIR phases was calculated after refining 12 lead sites, 24 Hg sites and 36 Se sites together in MLPHARE at 3 Å, and the resulting phases were refined to convergence at 3 Å by NCS averaging. The atomic model was built using FRODO[35] and O[36] and refined against the averaged density using XX12[37]. After the refinement converged, a 2.05 Å data set was substituted for the 3 Å data and the refinement was cycled to convergence at 2.05 Å. X-PLOR[38] was used to further refine the model, first with strict NCS constraints, then with NCS restraints that were relaxed as the refinement proceeded. Finally, a 1.6 Å data set was substituted for the 2.05 Å data, and the final steps of the model building and refinement were carried out. During the refinement with X-PLOR all data with $F/\sigma(F) \geq 2$ were included. The final model (Table 1) was refined with NCS restraints (250 kcal mol^{-1} $Å^{-2}$) on the backbone atoms for residues 14–167 of the 12 monomers and no restraints on the other atoms.

Coordinates. The refined model has been deposited in the Brookhaven Protein Data Bank (accession number 1DPS).

Acknowledgments
This work was supported by grants from the NIH (to J.M.H) and the American Cancer Society (to R.K), post-doctoral fellowships from The Medical Foundation, Boston (to R.A.G.) and from the Helen Hay Whitney Foundation (to S.E.F.) & a grant for High Performance Computing and Communication from the NSF (to G. Wagner).

Received 12 January, 1998; accepted 3 March, 1998.

1. Farr, S.B. & Kogoma, T. Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium*. *Microbiol. Rev.* **55** 561–585 (1991).
2. Almirón, M, Link, A.J., Furlong, D. & Kolter, R. A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev.* **6** 2646–2654 (1992).
3. Lomovskaya, O.L., Kidwell, J.P. & Matin, A. Characterization of the σ^{38}-dependent expression of a core *Escherichia coli* starvation gene, *pexB*. *J. Bacteriol.* **176** 3928–3935 (1994).
4. Martinez, A. & Kolter, R. Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps. *J. Bacteriol.* **179** 5188–5194 (1997).
5. Altuvia, S., Almirón, M., Huisman, G., Kolter, R. & Storz, G. The *dps* promoter is activated by OxyR during growth and by IHF and σ^{S} during stationary phase. *Mol. Microbiol.* **13** 265–272 (1994).
6. Schmid M.B. More than just 'histone-like' proteins. *Cell* **63** 451–453 (1990).
7. Peña, M.O. & Bullerjahn, G.S. The DpsA protein of *Synechococcus sp.* strain PCC7942 is a DNA-binding hemoprotein. *J. Biol. Chem.* **270** 22478–22482 (1995).
8. Chen, L. and Helmann, J.D. *Bacillus subtilis* MrgA is a Dps (PexB) homologue: evidence for metalloregulation of an oxidative stress gene. *Mol. Microbiol.* **18** 295–300 (1995).
9. Bozzi, M. *et al.* A novel non-heme iron-binding ferritin related to the DNA-binding proteins of the Dps family in *Lysteria innocua*. *J. Biol. Chem.* **272** 3259–3265 (1997).
10. Brentiens, R.J., Ketterer, M., Apicella, M.A. & Spinola, S.M. Fine tangled pili expressed by *Haemophilus ducreyi* are a novel class of pili. *J. Bacterol.* **178** 808–816 (1996).
11. Pfiefer, O., Pelletier, I, Altenbuchner, J. & Van Pee, K.-H. Molecular cloning and sequencing of a non-haem bromoperoxidase gene from *Streptomyces aureofaciens*. *J. Gen. Microbiol.* **138** 1123–1131 (1992).
12. Evans, D.J., Evans, D.G., Lampert, H.C. & Nahano, H. Identification of four new prokaryotic bacterioferritins, from *Helicobacter pylori, Anabaena variabilis, Bacillus subtilis* and *Treponema pallidum*, by analysis of gene sequences. *Gene* **153** 123–127 (1995).
13. Fehniger, T.E., Radolf, J.D. & Lovett, M.A. Properties of an ordered ring structure formed by recombinant *Treponema pallidum* surface antigen 4D. *J. Bacteriol.* **165** 732–739 (1986).
14. Sato, N. Hypothetical 20.2K protein from *Anabeana variabilis*. PIR accession No. JUO384 (1991).
15. Harrison, P.M. & Arosio, P. The ferritins: molecular properties, iron storage function and cellular regulation. *Biochim. Biophys. Act.* **1275** 161–203 (1996).
16. Banyard, S.H., Stammer, D.K. & Harrison, P.M. Electron density map of apoferritin at 2.8 Å resolution. *Nature* **271** 282–284 (1978).
17. Trikha, J. *et al.* Crystallization and structural analysis of bullfrog red-cell L-subunit ferritins. *Proteins* **18** 107–118 (1994).
18. Lawson, D.M. *et al.* Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* **349** 541–544 (1991).
19. Frolow, F., Kalb (Gilboa) A.J. & Yariv, J. Structure of a unique twofold symmetric haem-binding site. *Nature Struct. Biol.* 453–460 (1994).
20. Le Brun, N.E. *et al.* Identification of the feroxidase center of *Escherichia coli* bacterioferritin. *Biochem. J.* **312** 385–392 (1995).
21. Nichols, A. and Honig, B. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.* **12** 435–445 (1991).
22. Nichols, A., Sharp, K.A. & Honig, B. Protein folding and association: insights from interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11** 281–296 (1991).
23. Thompson, J.D., Higgins, D.G. & Gibson, T.J. ClustalW: Improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22** 4673–4680 (1994).
24. Rice, P.A., Yang, S.-W., Mizuuchi, K. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87** 1295–1306 (1996).
25. Shindo, H. *et al.* Solution structure of the DNA-binding domain of a nucleoid-associated protein, HN-S, from *Escherichia coli*. *FEBS Lett.* **360** 125–131 (1995).
26. Tanaka, I., Appelt, K., Dijk, J., White, S.W. & Wilson, K.S. 3 Å resolution structure of a protein with histone-like properties in prokaryotes. *Nature* **310** 376–381 (1984).
27. Kostrewa, D. *et al.* Three-dimensional structure of the *E. coli* DNA-binding protein FIS. *Nature* **349** 178–180 (1992).
28. Yuan, H.S. *et al.* The molecular structure of wild-type Fis protein, relationship between mutational changes and recombinatorial enhancer function or DNA binding. *Proc. Nat. Acad. Sci. USA* **88** 9558–9562 (1992).
29. Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. & Clardy, J. Atomic structures of of the human immunophilin FKBP-12 complexes with FK-506 and rapamycin. *J. Mol. Biol.* **229** 105–124 (1993).
30. Ottwinowski, Z. and Minor, W. Processing of X-ray diffraction data collected in ocsillation mode. *Meth. Enz.* **276** 307–326 (1997).
31. Collaborative computational project no. 4 The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr.* D **50** 760–763 (1994).
32. Rould, M.A. Screening for heavy-atom derivatives and obtaining accurate isomorphous differences. *Meth. Enz.* **276** 461–472 (1997).
33. Bricogne, G. Geometric sources of redundancy in intensity data and their use in phase determination. *Acta Crystallogr.* **A30** 395–495 (1974).
34. Filman, D.J.,Wien, M.W., Cunningham, J.A., Bergelson, J.M. & Hogle, J.M. The structure determination of echovirus I. *Acta Crystallogr.*, **in the press** (1998).
35. Jones, A.T. Interactive computer graphics: FRODO. *Meth. Enz.* **115** 157–171 (1985).
36. Jones, A.T., Zou, Y-J.,Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A42** 140–149 (1994).
37. Jacobson D.H., Hogle, J.M. & Filman, D.J. A pseudo-cell based approach to efficient crystallographic refinement of viruses. *Acta Crystallogr.* D **52** 693–711 (1996).
38. Brunger, A.T. X-PLOR, Version 3.1 A system for X-ray crystallography and NMR. (Yale University Press, New Haven, Connecticut; 1992).
39. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233** 123–138 (1993).
40. deMare, F., Kurtz, D.M. Jr., Norland, P. The structure of *Desulfovibrio vulgaris* rubrerythrin reveals a unique combinations of rubredoxin-like FeS_4 and ferritin-like diiron domains. *Nature Struct. Biol.* **3** 539–546 (1996).