Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The high-resolution structure of DNA-binding protein HU from *Bacillus stearothermophilus*

Stephen W. White,[a,b]* Keith S. Wilson,[c] Krzysztof Appelt[d] and Isao Tanaka[e]

[a]Department of Structural Biology, St Jude Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 38105, USA, [b]Department of Biochemistry, University of Tennessee, 858 Madison, Suite G01, Memphis, TN 38163, USA, [c]Department of Chemistry, York University, Heslington, York YO1 5DD, England, [d]Agouron Pharmaceuticals, Inc., San Diego, CA 92121, USA, and [e]Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060, Japan

Correspondence e-mail: stephen.white@stjude.org

Received 31 September 1998
Accepted 11 January 1999

PDB Reference: DNA-binding protein HU, 1huu.

Protein HU is a ubiquitous prokaryotic protein which controls the architecture of genomic DNA. It binds DNA non-specifically and promotes the bending and supercoiling of the double helical structure. HU is involved in many DNA-associated cellular processes, including replication, transcription and the packaging of DNA into chromosome-like structures. Originally determined at medium resolution, the crystal structure of HU has now been refined at 2.0 Å resolution. The high-resolution structure shows that the dimeric molecule is essentially a compact platform for two flexible and basic arms which wrap around the DNA molecule. To maximize the protein's stability, non-secondary structural regions are reduced to a minimum, there is an extensive aromatic hydrophobic core and several salt bridges and hydrogen-bonded water molecules knit together crucial regions. Based on the original medium-resolution structure of HU, several proposals were made concerning the structural basis of HU's ability to bind, bend and supercoil DNA. Each of these proposals is fully supported by the high-resolution structure. Most notably, the surfaces of the molecule which appear to mediate protein–DNA and protein–protein interactions have the ideal shapes and physicochemical properties to perform these functions.

1. Introduction

In the cell, there are many situations in which the linear and rather stiff DNA molecule has to be bent or arranged into higher order structures. For example, the DNA is locally supercoiled at sites of replication and transcription, extensively supercoiled when packaged into nucleosomes and is often bent to bring together specific sites which are widely separated along the linear molecule. A large number of DNA-binding proteins have evolved to facilitate the formation of these structures, most notably the histone proteins in eukaryotic cells, and an important family of such proteins in prokaryotic cells are the HU-like molecules.

HU is a small basic protein which binds DNA non-specifically and introduces negative supercoils within the circular molecule (Rouviere-Yaniv *et al.*, 1979; Broyles & Pettijohn, 1986). HU is important for maintaining and mediating the topological structure of supercoiled DNA and it may also function as a bacterial histone protein to help create the bacterial chromosome structure (Griffith, 1976). It has also been shown that HU has an important role in the initiation of replication (Dixon & Kornberg, 1984), site-specific DNA rearrangements (Craigie *et al.*, 1985; Johnson *et al.*, 1986), DNA-strand transfer (Lavoie & Chaconas, 1995) and gene regulation (Flashner & Gralla, 1988). These diverse processes

© 1999 International Union of Crystallography
Printed in Denmark – all rights reserved

share the common feature that the DNA needs to be locally bent, wrapped and/or supercoiled, and it is this organization which HU appears to promote.

HU was first detected and isolated in *Escherichia coli*, but it has now been found in many bacteria and appears to be ubiquitous in prokaryotes and possibly archaebacteria (Oberto *et al.*, 1994). The molecule has a monomeric molecular weight of 9500, is highly conserved and is usually present as a homodimer (Oberto *et al.*, 1994). In *E. coli*, the protein exists as a heterodimer of α- and β-subunits which are 70% identical (Mende *et al.*, 1978; Rouviere-Yaniv & Kjeldgaard, 1979). The genes for both subunits (*hupA* and *hupB*, respectively) have been cloned (Kano *et al.*, 1986, 1987) and they appear to be regulated to maintain a correct balance of subunits within the cell (Rouviere-Yaniv *et al.*, 1990). However, homodimers of each subunit appear to be functional since *hupA*$^-$ and *hupB*$^-$ mutants are both viable (Wada *et al.*, 1988; Huisman *et al.*, 1989). Thus, the *E. coli* situation is probably an evolutionary anomaly with little functional significance. *hupA*$^-$ and *hupB*$^-$ double mutants are viable but exhibit poor growth, filamentation and cold sensitivity (Wada *et al.*, 1988; Huisman *et al.*, 1989). Therefore, HU is an important but not essential protein.

Two other proteins have been identified which are highly homologous to HU and belong to the same family of molecules. Integration host factor (IHF) from *E. coli* is a heterodimer of α- and β-subunits with molecular weights close to 10000 (Nash *et al.*, 1987). It plays an important role in several cellular processes (Friedman, 1988) and was first identified as an essential host factor for bacteriophage λ integration (Nash & Robertson, 1981). It binds specifically to a well characterized DNA consensus sequence and introduces a sharp bend or kink into the DNA at this location (Stenzel *et al.*, 1987; Prentki *et al.*, 1987; Robertson & Nash, 1988). Transcription factor 1 (TF1) from bacteriophage SPO1 is a homodimeric protein (Greene *et al.*, 1984) which preferentially binds SPO1 DNA (Johnson & Geiduschek, 1977), selectively inhibits the transcription of SPO1 DNA by bacterial RNA polymerases (Wilson & Geiduschek, 1969) and is essential for viral multiplication (Sayre & Geiduschek, 1988). Although a precise function has yet to be demonstrated, TF1 does have the ability to bend DNA (Schneider & Geiduschek, 1990). Clearly, the common property of HU, IHF and TF1 is to bind and bend DNA, and the three proteins have diverged evolutionarily to apply this attribute to various specific functions.

We have previously reported the crystallization (Dijk *et al.*, 1983) and 3.0 Å crystal structure of HU from *Bacillus stearothermophilus* (Tanaka, Appelt *et al.*, 1984; Tanaka, White *et al.*, 1984), and have proposed a general model for how this class of protein binds, bends and supercoils DNA (White *et al.*, 1989). Here, we describe the high-resolution structure of HU which has been refined by simulated annealing to 2.0 Å. The dimeric protein has a compact body which acts as a platform for two arms which interact non-specifically with double-stranded DNA. The 2.0 Å refined structure reveals that the body is extremely stable and has evolved to contain the minimal number of unstructured loop regions and the maximum number of protein-stabilizing features. The latter include an extensive aromatic hydrophobic core, strategically placed salt bridges and a number of integral water molecules. Models for HU–DNA and HU–HU interactions based on the lower resolution structure are fully supported by the 2.0 Å structure.

2. Experimental

2.1. Crystallization

The protein was purified from *B. stearothermophilus* cells as described previously (Dijk *et al.*, 1983), although it should be noted that the gene has now been cloned and overexpressed in *Escherichia coli* (Padas *et al.*, 1992). Crystals were grown by vapor diffusion using the hanging-drop technique (Davies & Segal, 1971). Several crystal forms were obtained, but the best in terms of diffraction quality were long needles in space group *P*2.

2.2. Data collection

2.0 Å diffraction data were collected at room temperature on film at beamline 7.2 of the Daresbury synchrotron radiation source. The crystal-to-film distance was 60 mm and the X-ray wavelength was 1.488 Å. Data were collected using standard oscillation geometry with 2° rotation ranges about the long *b* axis of the needle-shaped crystals. A complete 180° data set was collected from a single large needle (3.0 × 0.4 × 0.3 mm) by translating it three times to expose fresh regions to the beam. To fill in the blind region, a needle fragment was mounted across the capillary, which enabled 20° of data to be collected about the *a** axis. The data were processed using the *MOSFLM* package and combined into a unique set using *CCP*4 programs (Collaborative Computational Project, Number 4, 1994).

2.3. Refinement

The previously published 3 Å structure (Tanaka, Appelt *et al.*, 1984) minus water molecules was the starting point of the 2 Å refinement by simulated annealing using the *X-PLOR* program (Brünger, 1988). All data between 6.0 and 2.0 Å with $F > 2\sigma$ were used, and 10% of the data were partitioned into a test set for calculation of the $R_{\rm free}$ value (Brünger, 1992). The parameter set developed by Engh & Huber (1991) was used throughout the refinement. The structure was annealed from an initial temperature of 3000 K using the protocol suggested in the *X-PLOR* manual. Following the annealing run, the atomic temperature factors were refined and $2F_o - F_c$ and $F_o - F_c$ electron-density maps were calculated. The model was carefully inspected and compared with these maps using the *O* program (Jones *et al.*, 1991). Adjustments were made to the model and many water molecules were incorporated at this stage. Several more rounds of refinement were necessary to delineate some surface side-chain orientations and to produce a final set of water molecules. In these later rounds, the simulated annealing was started at 1500 K. Many water molecules were clearly visible in the electron-density maps,

and the marginal ones were decided by the quality of the density, proximity to hydrogen-bond acceptors and donors and temperature-factor values. The quality and geometry of the final model was analyzed using the *PROCHECK* program (Laskowski *et al.*, 1993).

3. Results

3.1. Crystallization, data collection and structure determination

The HU protein from *B. stearothermophilus* readily forms crystals in phosphate buffer around pH 8.0 from either ammonium sulfate or 2,4-methylpentandiol (MPD). A total of five crystal forms were obtained, and the monoclinic *P*2 crystals were eventually chosen for analysis because they diffracted to the highest resolution (2.0 Å). These crystals are needle-shaped with their long axis coincident with the *b* axis and have unit-cell parameters $a = 65.5$, $b = 37.3$, $c = 65.5$ Å, $\alpha = 114.5^\circ$. They typically grow to their maximum size (1–3 × 0.4 × 0.3 mm) within a week. The structure of the protein was initially determined at 3.0 Å by single isomorphous replacement using a uranyl derivative ($K_3UO_2F_5$). The initial electron-density map revealed that the asymmetric unit comprises three HU monomers and the quality of the map was greatly improved by threefold non-symmetry averaging (Tanaka, Appelt *et al.*, 1984; Tanaka, White *et al.*, 1984). This map was of sufficient quality to build a medium-resolution model of the protein, which was partially refined to 3 Å resolution using *PROLSQ*. It was decided not to perform an extensive refinement of the structure until the higher resolution 2.0 Å data described here (Table 1) had been collected.

Table 1
The refinement and the refined structure of protein HU from *B. stearothermophilus*.

X-ray data and refinement	
Resolution (Å)	2.0
Merging *R* (on *F*s) (%)	7.0
Resolution range for refinement (Å)	6.0–2.0
Completeness within the range (%)	89
σ cutoff	2.0
Final *R* factor (%)	19.3
Final R_{free}† (%)	21.3
Model geometry (molecules *A* + *B* + *C*)	
Final *G* factor‡	0.37
R.m.s. deviations from ideal geometry	
Bonds (Å)	0.011
Angles (°)	1.525
Mean temperature factors (*B*)§ (Å^2)	
Main-chain atoms	23.9
Side-chain atoms	29.9
Comparison of molecules in the asymmetric unit¶	
Missing residues	
Molecule *A*	59–68
Molecule *B*	57–72
Molecule *C*	56–74
Number of water molecules common to *A*, *B* and *C*	35
Total number of water molecules	
Molecule *A*	101
Molecule *B*	79
Molecule *C*	86
R.m.s. on C^α atoms§ (Å)	
Molecules *A*/*B*	0.166
Molecules *A*/*C*	0.215
Molecules *B*/*C*	0.238

† Brünger (1992). ‡ Laskowski *et al.* (1993). § Residues 1–52 and 77–90, *i.e.* excluding the arms. The molecules were superimposed with the *O* program (Jones *et al.*, 1991) using C^α atoms 1–52 and 77–90. ¶ The asymmetric unit of the HU crystal contains three monomers *A*, *B* and *C*.

3.2. Refinement

The starting structure had an *R* factor of 27.5% when calculated with the 2.0 Å data; this dropped to 25.9% during the first annealing run from 3000 K. The R_{free} dropped from 35.7 to 31.9%. Following temperature-factor refinement, these values dropped to 22.3 and 28.1%, respectively. At this point, $2F_o - F_c$ and $F_o - F_c$ maps were calculated and compared with the model. The protein itself required very little adjustment, but some 200 water molecules could be inserted unambiguously. The new model was subjected to several further rounds of simulated annealing, temperature-factor refinement and comparison to electron-density maps. These additional rounds were necessary to finalize the water molecules and to build as much of the molecule's arm structure as possible. The arms are very flexible and very poorly defined in the electron-density map (Tanaka, Appelt *et al.*, 1984).

The *R* and R_{free} values of the final model are 19.3 and 21.3%, respectively. The R_{free} value is particularly low and this may be related to the high degree of secondary structure in the molecule and the lack of extended loop regions. Complete details of the final model are listed in Table 1. Representative electron density from the final $2F_o - F_c$ map is shown in Fig. 1. When analyzed with the *X-PLOR* (Brünger, 1988) and *PROCHECK* (Laskowski *et al.*, 1993) programs, the overall geometry was found to be excellent. The various geometrical criteria analyzed by *PROCHECK* are well within accepted limits, apart from those of the visible residues within the flexible arms. The Ramachandran plot is particularly well clustered into the accepted regions.

3.3. Crystal packing

The *P*2 unit cell contains a total of six HU polypeptide chains arranged as three dimers around three of the four crystallographic twofold rotation axes. This creates an unusual asymmetric unit comprising the three halves of each dimer. HU dimers are stacked in a similar fashion along the twofold axes and the empty fourth axis creates a channel in each unit cell along the *y* direction. Viewed normal to the *y* axis, the dimers form a pseudohexagonal close-packing arrangement with holes, and this is reflected in the unit-cell parameters, in which $a = c$ and β is close to 120°. The three sets of dimers have no preferred azimuthal orientation with respect to each other and generally display little interaction. With regard to their alignment along the *y* direction, dimers *B* and *C* are at the same relative 'height' whereas dimer *A* is some 12 Å 'higher'.

3.4. The compact body of HU

The majority of the HU protein comprises an extremely compact dimeric structure which has two distinct but highly associated halves. A cartoon of the structure is shown in Fig. 2. Each monomer has two N-terminal α-helices $\alpha1$ (residues 4–14) and $\alpha2$ (18–38) which are oriented at a relative 'V' angle of 25° and a crossing angle of 60°. Two such pairs intertwine to form a four-helix bundle at the 'bottom' of the dimer. Asn2 OD1 bonds to the amide N atom of Glu5 and thereby N-caps helix $\alpha1$. Ser17 OG performs a similar role in helix $\alpha2$ by bonding to the amide N atom of Asp20. The C-terminal half of each monomer contains a three-stranded antiparallel β-sheet comprising β-strands $\beta2$ (41–44), $\beta3$ (47–51) and $\beta4$ (78–80). Two of these associate in a side-to-side fashion on 'top' of the dimer to create a saddle-like structure normal to and covering the α-helical base. The distance between the two inner strands of the dyad-related β-sheets is 6.3 Å and the relative orientation of their surfaces is approximately 65°. Thus, the two β-sheets cannot be regarded as a single six-stranded structure. The final secondary-structural element is the short α-helix $\alpha3$ (82–90) at the extreme C-terminus, which is horizontally disposed on opposite sides of the dimer.

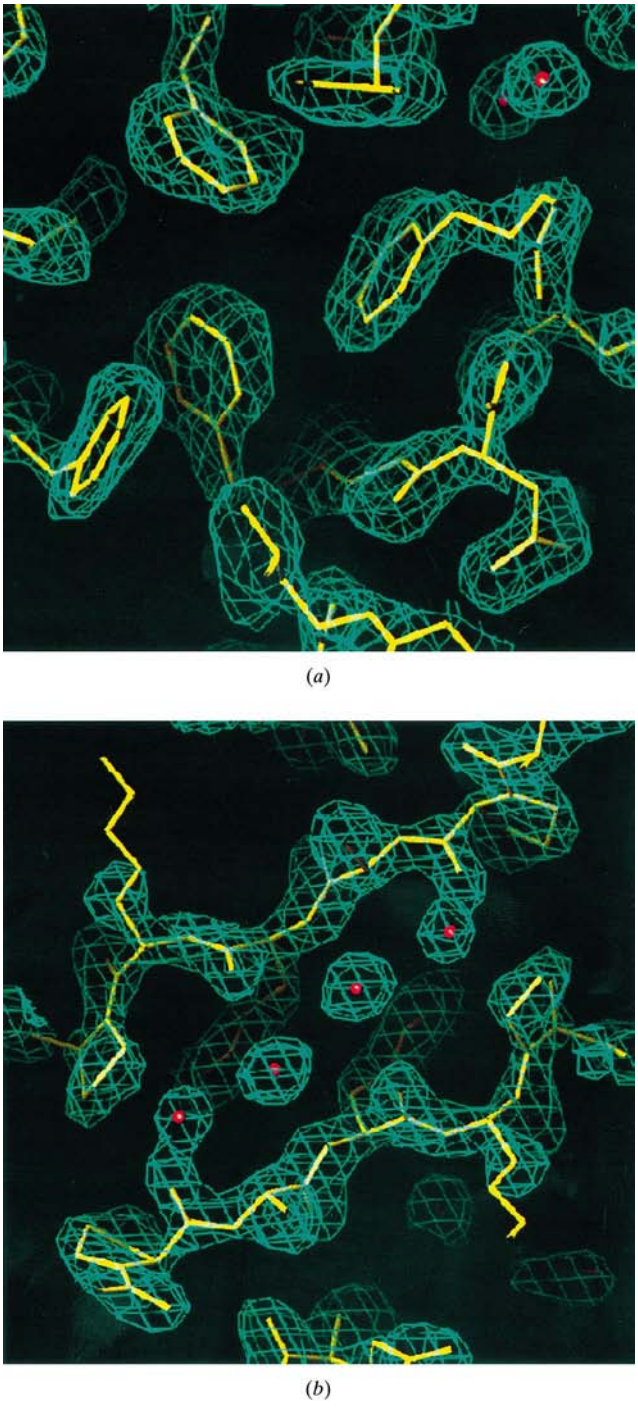

Figure 1
Electron-density maps of protein HU from *B. stearothermophilus*. Both are from the final $2F_o - F_c$ map and have been displayed using the program *O* (Jones *et al.*, 1991) with contouring at 1.5σ. The regions correspond to parts of the HU structure which are described in the text and shown in Fig. 3. Shown on the left is the hydrophobic core in the region of the phenylalanine cluster which corresponds to Fig. 4(*b*). Shown on the right is the junction of the dimer-related β-pleated sheets that contains the four structural water molecules which corresponds to Fig. 4(*c*).

3.5. Monomer–monomer interactions

The HU dimer has five features which optimize its compactness and stability. Firstly, the molecule has a pronounced aromatic hydrophobic core. Secondly, apart from the DNA-binding arms, the protein has no loops or extended turns. Thirdly, both termini are extremely short and anchored to the body of the molecule. Fourthly, strategically located water molecules knit together important elements of the structure. Finally, several salt bridges between secondary-structure elements act as additional stabilizing elements. These features are discussed in greater detail below. Note that residues and secondary-structural elements labeled with a prime (′) refer to the dyad-related monomer within the dimer.

3.5.1. The hydrophobic core. The core of HU can be divided into two distinct halves. The first comprises typical aliphatic residues (listed in Tanaka, Appelt *et al.*, 1984) which fill the space between the α-helices in the base. Covering this is the second half, which consists of a remarkable grouping of eight phenylalanines, residues 29, 47, 50 and 79 from each monomer.

Interactions between aromatic side chains within proteins have been extensively analyzed by Burley & Petsko (1985). Three general features are commonly observed and all are seen in the HU structure (Fig. 3*a*). Firstly, interacting groups are between 4.5 and 7.0 Å apart, and the edge-to-face interactions between aromatic rings which are seen in small molecule structures are also observed in proteins. Of the 28 possible phenylalanine–phenylalanine interactions in the HU dimer, 15 are within 4.5–7.0 Å and ten involve edge-to-face interactions. Six are particularly clear, 29–47′, 47–50′, 50–79′ and their three dimeric partners. Secondly, aromatic side chains frequently form networks and this is clearly the case in HU. Finally, the interactions usually link secondary-structural elements. This is exemplified in HU, where the aromatic cluster links the tops of the four α-helices and the hydrophobic inner surface of the β-sheet region. Burley & Petsko (1985)

concluded that aromatic interactions represent a major contribution to protein stability, and HU may be one of the best examples of this in currently known structures.

3.5.2. The turns. The five turns which connect the secondary-structure elements are extremely tight and are mediated by highly conserved glycine residues. Each turn produces minimal disruption to the hydrogen-bonding patterns of the flanking secondary-structure elements.

In turn 1 ($\alpha1$–$\alpha2$), which is centered on Gly15, only the backbone of Leu16 is not formally involved in α-helical hydrogen-bonding interactions. Its amide N atom is hydrogen bonded to Ser14 OG and the its carbonyl O atom interacts with a conserved water molecule (107). The precise details of this turn were not apparent in the original 3 Å structure, where it was thought to be more extensive (residues 14–20). It is now clear that residues 14–15 and 17–20 form the termini of the flanking α-helices, which agrees with more recent NMR results on HU (Vis *et al.*, 1995).

Turn 2 ($\alpha2$–$\beta2$) is centered on Gly39 and is extremely abrupt, with no disruption of the hydrogen-bonding patterns of the flanking α-helix and β-strand. This is accomplished by the carbonyl O atom of Ala35 which makes two hydrogen-bonding interactions. The first is to the amide N atom of Gly39 at the C-terminus of $\alpha2$ and the second is across the turn to the amide N atom of Asp40 at the N-terminus of $\beta1$.

Turn 3 ($\beta2$–$\beta3$) spans five residues (44–48) and can best be described as a type II β-turn with a β-bulge. The turn contains the typical glycine (46) at the third position, but the carbonyl O atom of Leu44 at the first position forms two hydrogen bonds, one to the amide N atom of Phe47 and a second to the amide N atom of Gly48 (Fig. 3*b*).

Turn 4 at the end of the DNA-binding arm is not visible in the electron-density map. We had originally predicted that the arm would form a simple β-ribbon and that Pro63 would occupy position 2 of a β-turn at the end. Recent NMR results on HU (Vis *et al.*, 1995) have shown that our basic idea of a β-ribbon is correct, but the details of the arm structure are not. Pro63 is now known to occupy position 1 of a type I β-turn.

The final turn, turn 5 ($\beta4$–$\alpha3$), is more of a kink centered on Pro81 and Gly82. The carbonyl O atom of Gly82 forms bifurcated hydrogen bonds to the amide N atoms of residues 85 and 86 at the start of $\alpha3$ (Fig. 3*b*). The kink structure also allows a stabilizing tertiary interaction with turn 3. This is discussed below.

3.5.3. The termini. The N-terminus can be regarded as a dimer clamp which attaches to the β-sheet of the other monomer. Specifically, residues 1–3 form a short β-strand ($\beta1$) which hydrogen bonds in a parallel orientation to the outer ($\beta2'$) strand. The N-terminal methionine is further anchored by its side chain, which is buried in the hydrophobic core, and by a salt bridge between its amino group and Asp40′. An additional salt bridge to Glu5 completes this local structure, which is shown in Fig. 3(*b*). The C-terminus has no extended structure and the molecule ends abruptly with the final residue of helix $\alpha3$. The inner surface of this helix is typically hydrophobic, which fastens this end of the molecule onto the hydrophobic core.

3.5.4. Water structure. Each of the three HU monomers in the crystal unit cell has some 90 water molecules bound at the surface, but only 35 are common to all three (these water molecules are labeled 100–135). The majority are simply bound to free hydrogen-bond acceptors and donors such as carbonyl O atoms and amide N atoms in the turns and at the ends of secondary-structure elements. However, six are integral parts of the structure and act to knit together potentially flexible regions of the dimers. The importance of these water molecules is reflected in their low temperature factors, which approach those of the main-chain atoms.

Three water molecules act as pseudo-β-sheet elements. Waters 130 and 132, together with 130′ and 132′, create a hydrogen-bonded pseudo-β-strand which links the two β-sheets on the top of the molecule. 130 is hydrogen bonded to the carbonyl O atom of residue 77, and 132 bridges the amide N atom of residue 79 and the carbonyl O atom of 79′ (Fig. 3*c*). Water 119 spans β-strands $\beta2$ and $\beta3$ at one end of the β-sheet where they start to diverge and are too far apart to form a true antiparallel structure. Specifically, 119 links the carbonyl O atom of residue 40 at the start of $\beta1$ to the amide N atom of residue 52 at the end of $\beta2$ (Fig. 3*b*).

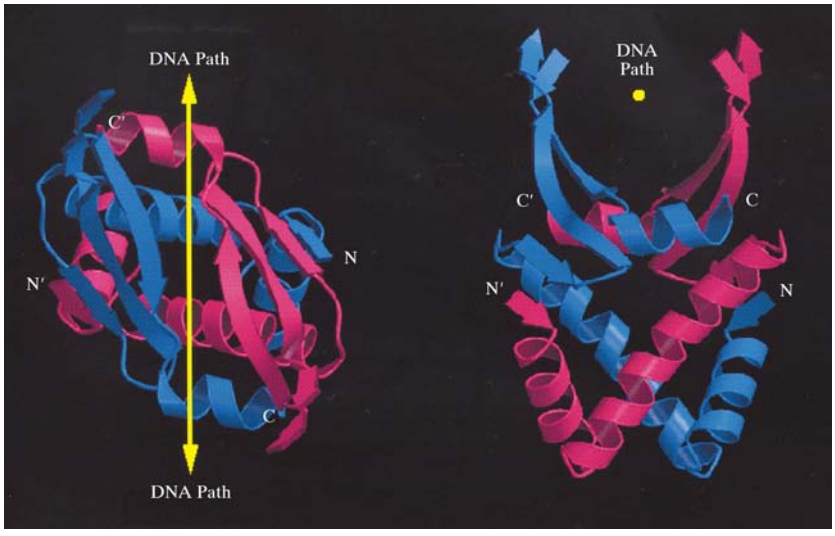

Figure 2
Orthogonal views of the structure of protein HU from *B. stearothermophilus*. The cartoons were made using the program *MOLSCRIPT* (Kraulis, 1991). Each monomer is coloured differently to show how they wrap around each other. Features of the structure mentioned in the text include (i) the short β-strand at the N-terminus which hydrogen bonds to the outside of the β-sheet of the other monomer, (ii) the absence of extended loop regions in the body of the molecule, (iii) the hinge in the arm and (iv) the beginning of the distal β-ribbon in the arm. The ends of the arm are not visible in the crystal structure. In the top view (left), the path of the bound DNA is vertical, and in the side view (right), the path of the bound DNA would be towards the viewer between the arms. The protein surface shown on the right is proposed to mediate protein–protein interactions during DNA supercoiling.

The remaining three structural waters span the two monomers. Waters 118 and 121 help to secure the position of turn 3. 118 links turn 3 to $\alpha2'$ by bridging the carbonyl O atom of Gly46 and Thr33′ OG. Thr33 is highly conserved and this local structure is probably common to many HU molecules. 121 connects turn 3 and the start of $\alpha1'$ *via* the amide N atoms of residues 45 and 3′ (Fig. 3*b*). Finally, water 129 bridges the start of $\beta4$, where the arm re-enters the dimer, and $\alpha3$ *via* the carbonyl O atoms of residues 75 and 86′.

3.5.5. Salt bridges and hydrogen bonds. The molecule contains five salt bridges in addition to those which immobilize the N-terminus described above. Asp26–Lys3 spans helices $\alpha1$ and $\alpha2$ and appears to stabilize their relative orientation. Asp20–Lys23 and Glu34–Arg37/Lys38 are disposed along the outer surface of $\alpha2$ and may stabilize this relatively long α-helix (Marqusee & Baldwin, 1987). Finally, Glu51–Arg53/Lys80 and Glu54–Lys75 are at the start of the DNA-binding arm and may help to maintain the β structure in this region. Arg53 and Lys80 are proposed to interact with DNA (White *et al.*, 1989) and Glu51 may help to align them prior to binding. However, Glu51 is poorly conserved and clearly not crucial to HU structure or function.

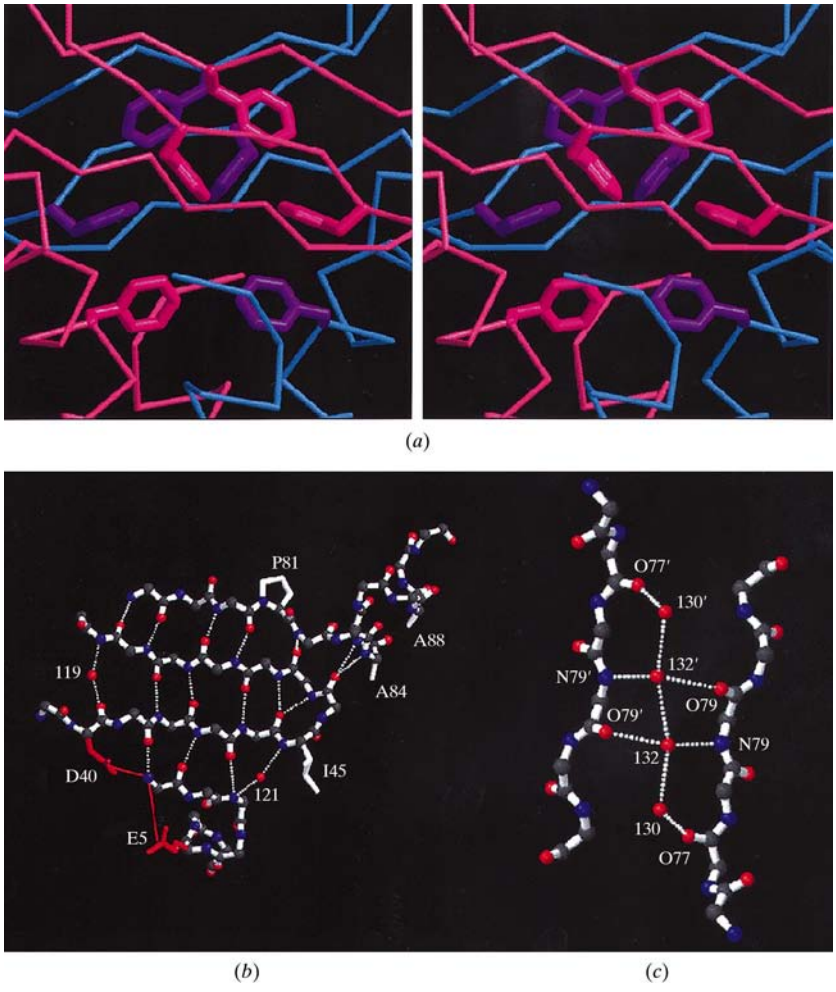

Figure 3
Details of the structure of protein HU from *B. stearothermophilus*. Each figure was made using the *MOLSCRIPT* program (Kraulis, 1991). (*a*) A stereoview of the aromatic cluster within the hydrophobic core. The four phenylalanine residues and the C^{α} backbone from one monomer have similar colours. (*b*) A view centered on the interactions which immobilize the N-terminus. Shown in red are the acidic residues Glu5 and Asp40 which form salt bridges (also in red) to the terminal amine group. Note the structural water 121 (red sphere) and the hydrogen bonding of the N-terminal β-strand to the β-sheet of the dimeric partner. Also shown in the figure is the structural water 119 (red) which connects the diverging β-strands of the β-sheet, the hydrogen bonding within turns 3 and 5, and the conserved alanines 84 and 88 and isoleucine 45 which may mediate protein–protein interactions. (*c*) The four water molecules which mediate the hydrogen-bonding interaction between the inner strands of the dimer-related β-pleated sheets.

3.6. Dimer flexibility

Since the crystal asymmetric unit contains three monomers, it was possible to compare their conformations and to distinguish features which are structurally important from those which are simply the result of crystal packing. This was particularly useful for identifying the structurally important water molecules, salt bridges and hydrogen bonds discussed above. The comparison also allowed a unique insight into the flexibility of the molecule.

As would be expected for such a compact protein, the three crystallographically independent molecules are extremely similar. Using the least-squares superposition algorithm in the *O* program (Jones *et al.*, 1991), the r.m.s. deviation of the C^{α} atoms in the body of the molecules was calculated (Table 1). When the superimposed molecules are compared pairwise, residue-by-residue, some flexibility is revealed which correlates well with the backbone temperature factors. The turns and the C-terminus are less rigid than the body of the protein, as are the amino acids flanking the arm. Two regions of the dimer are particularly variable, turn 3 and the C-terminal α-helix ($\alpha3$), and this presumably explains the consistently poor electron density for these elements during the initial model building and the subsequent refinement. Turn 3 is adjacent to $\alpha3$ in the molecule, and they interact *via* the carbonyl O atom of residue 46 and the amide N atoms of residues 84 and 85. The variability involves the conformation of Gly46. In molecules *A* and *C*, its carbonyl group is pointing up and makes direct hydrogen bonds (Fig. 3*b*), whereas in molecule *B*, it points sideways and a water molecule (B240) is hydrogen bonded to the two amide N atoms. The apparent flexibility of the turn 3/$\alpha3$ region may have functional reasons related to dimer–dimer interactions (see below). As regards the relatively high flexibility of $\alpha3$, this may reflect that its attachment to the body of the dimer is mediated primarily by a few hydrophobic interactions.

3.7. The saddle and the arms

The most distinctive feature of the HU dimer is a pair of symmetry-related arms which extend out from β3 and β4 of the saddle-like β-sheet structure to create a concave β surface. As discussed previously (Tanaka, Appelt *et al.*, 1984; White *et al.*, 1989) and further elaborated below, this part of HU almost certainly constitutes the DNA-binding region. Unfortunately, the arms are also the most flexible parts of the molecule and the least defined in the electron-density map. However, the refinement at 2.0 Å has revealed important features which were not previously visible.

Interpretable electron density extends to varying degrees from the three dimers in the unit cell, but it is most clear in molecule *A* (Table 1). Our discussion of the arm conformation is therefore based on molecule *A*. The β-ribbon represents a simple extension of β-strands β3 and β4 and formally starts and ends at residues 52 and 77. There is a slight kink at residue 77 which is a highly conserved proline. The hydrogen-bonding pattern from the β-sheet into the β-ribbon proceeds normally from 47–82 to 55–74, but then jumps to 58–71 after which the electron density rapidly deteriorates. The most likely explanation for this is that residues 56–57 and 72–73 constitute a hinge or possibly a bend which allows for a redirection and/or flexibility of the distal part of the arm. This suggestion is supported by the fact that alanines and prolines dominate the hinge/bend region in all known HU sequences. An NMR analysis of HU has revealed the arm conformation, which agrees well with the region which is visible in our X-ray structure (Vis *et al.*, 1995).

4. Discussion

4.1. The dimer structure

It is clear that the body of the HU dimer is primarily a stable platform for the extended DNA-binding arms. Apart from two surfaces, one which binds DNA and a second which may mediate protein–protein interactions in the DNA supercoiling process (White *et al.*, 1989), the molecule has evolved to incorporate the maximum number of protein-stabilizing features. It is possible that, with 85% secondary structure, the body of HU has one of the largest fractions of residues involved in α-helices and β-strands amongst known protein structures. Also, the dimeric conformation of HU depends on highly interlocking monomer–monomer interactions and considerable unfolding of each monomer would be necessary to separate them. When one calculates the difference between the solvent-accessible surfaces of the HU monomer and dimer (Argos, 1988), a total of 3594 Å^2 is buried during dimer formation, which is by far the largest value for comparably sized proteins (Janin *et al.*, 1988). It has been estimated that each Å^2 of buried surface contributes between 117 and 179 J to the free energy of association (Janin *et al.*, 1988) and the total contribution in the HU dimer is therefore between 418 and 627 kJ.

It is interesting to speculate how such an interlocking structure might have evolved. It was shown recently that HU is homologous to, and may have evolved from, prokaryotic ribosomal protein S7 (Wimberly *et al.*, 1997; Hosaka *et al.*, 1997). S7 has an α-helical structure and contains a single β-ribbon extension which is thought to bind double-stranded RNA. Two of the α-helices and the β-ribbon superimpose rather well onto those of the HU monomer. Particularly striking, however, is that the packing of four of the S7 α-helices resembles that of the four-helical bundle in the HU dimer (Hosaka *et al.*, 1997). Thus, deletion of the appropriate S7 α-helices followed by dimerization might have generated the primitive HU dimer, and subsequent mutations could then have optimized the monomer–monomer interactions.

4.2. DNA-binding and DNA-bending

We have proposed that HU interacts with DNA *via* the exposed surface of the β-sheet saddle and its associated β-ribbon arms. We have further suggested that conserved elements in the arms are responsible for bending the DNA around the protein. Although these features have been fully discussed elsewhere (Tanaka, Appelt *et al.*, 1984; White *et al.*, 1989), it is convenient to summarize the main points here.

First, the two arms and the surface of the β-saddle match a DNA molecule both in shape (Fig. 2) and charge (Fig. 4). Conserved arginine and lysine residues are ideally positioned to interact electrostatically with the DNA sugar–phosphate backbone and the β elements with their natural twist combine to form a concave helical structure. Second, the hinge/bend region of each arm located at residues 56–57 and 72–73 allows the distal region to wrap around the 'back' of the DNA and may also appropriately space the arginines and lysines at the extremity of the arm. Third, HU interacts with DNA *via* the minor groove. Based on the structure alone, it was not possible to decide between the minor and major grooves (Tanaka, Appelt *et al.*, 1984). However, subsequent chemical protection experiments with the highly homologous IHF protein (Craig & Nash, 1984) clearly showed that this family of proteins interact predominantly with the minor groove. Finally, a pair of highly conserved and adjacent hydrophobic residues (Met69 and Ile71 in the *B. stearothermophilus* protein) which would point towards the bound DNA are postulated to prize open the minor groove and bend the DNA (Fig. 4). Biochemical, mutagenesis and NMR studies on HU and its complex with DNA fully support our model. (Goshima *et al.*, 1990, 1992; Shindo *et al.*, 1993). Also, DNA affinity cleavage data on the binding of HU within the mu transpososome have enabled a model of the HU–DNA complex to be derived using distance constraints (Lavoie *et al.*, 1996). This low-resolution model confirms the dramatic bending of the DNA molecule, binding within the minor groove and the predominance of electrostatic interactions.

The high-resolution structure of the HU–DNA complex has not been determined, but the crystal structure of the related IHF–DNA complex was recently solved (Rice *et al.*, 1996). The structure confirms all of the major features of our model, but there are two features which were not predicted. Firstly, the highly conserved proline residue at the very tip of each arm does not simply mediate the the tight turn but actually

creates the 'wedge' which bends the DNA. The proline on each arm intercalates between the bases, and the pair create two kinks which are nine base pairs apart and which constitute the major sites of DNA bending. Our prediction that the conserved adjacent hydrophobic residues in the arm (69 and 71 in HU) would contribute to the DNA bending is correct; specifically, they contact sugar residues in the DNA backbone immediately adjacent to the kink. The second feature concerns the four water molecules which bridge the two β-sheets at the top of the HU dimer. In the complex, these are hydrogen bonded to the water molecules within the DNA minor groove (the so-called spine of hydration). Thus, the structural complementarity between the β-sheet saddle and the DNA minor groove is even better than we originally suggested.

4.3. DNA supercoiling

Our DNA-binding scheme for HU immediately suggested a model for how the protein might supercoil DNA (Tanaka, Appelt *et al.*, 1984). HU has an overall wedge shape (Fig. 2) and upon DNA bending, tandemly bound molecules would interact to form a circle or a helix with the DNA around the periphery. We have previously noted that the putative interacting surfaces have complementary shapes and are also conserved (Tanaka, Appelt *et al.*, 1984; Wilson *et al.*, 1990). The high-resolution structure now reveals that this surface is also predominantly apolar and ideal for hydrophobic interactions with itself (Fig. 4). As discussed above, turn 3 at the center of the surface is also somewhat flexible, which might facilitate the docking process. In IHF, the corresponding surface is hydrophilic and acts as a clamp which binds the periphery of the DNA and forces it to bend by nearly 180°. If our supercoiling model is correct, the DNA must be bent less severely in the HU complex, both to allow for protein–protein interactions and to avoid contact between the charged DNA backbone and this hydrophobic surface. It should also be noted that the overall conformation of the DNA in the IHF–DNA complex is consistent with negative supercoils, which agrees with biochemical observations of HU-induced supercoiling.

Based on observed DNA–HU stoichiometries, the extent of DNA supercoiling (Rouviere-Yaniv *et al.*, 1979; Broyles & Pettijohn, 1986) and the HU–DNA model structure, we proposed that this circular complex would comprise 8–10 HU dimers and 80–100 base pairs per turn. We also noted that such an object would be close in size and DNA supercoiling to the eukaryotic histone core particle. Several other proposals have been made (Oberto *et al.*, 1994; Searcy & Stein, 1980), but recent experiments have largely supported our model. Most notably, it has been demonstrated that tandemly bound HU molecules each occupy some nine base pairs (Bonnefoy & Rouviere-Yaniv, 1991) and also that the HU-stimulated cyclization of DNA requires a minimum of 98 base pairs (Hodges-Garcia *et al.*, 1989).

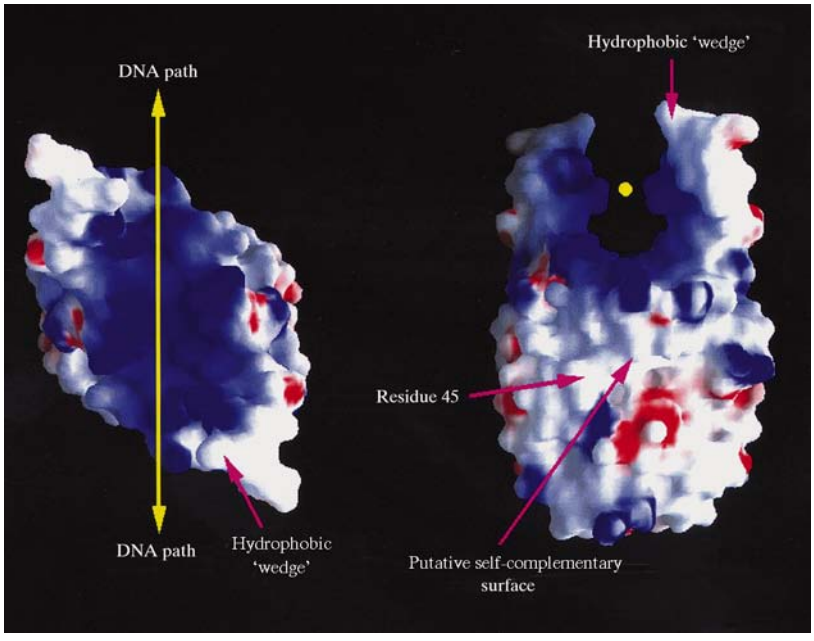

Figure 4
The surface electrostatic potential of protein HU from *B. stearothermophilus*. The two views correspond to the orientations of the dimer shown in Fig. 2. Red and blue represent negative and positive potentials and white denotes non-polar regions. The putative DNA-binding surface is extremely electropositive and shows the path of the DNA. In the view on the left the path is vertical and in the view on the right it is towards the viewer. Note the two conserved hydrophobic patches, which are proposed to mediate DNA bending and dimer–dimer interactions. The first is a 'wedge' on the inside surface of each arm projecting into the putative path of the DNA and the second is in the body of the protein centered on residue 45. The calculations and pictures were made using the *GRASP* program (Nicholls *et al.*, 1991).

SW would like to thank Christopher Davies for his invaluable assistance in this project and for critically reading the manuscript. He would also like to acknowledge the support of CORE grant CA-21765, the American Lebanese Syrian Associated Charities (ALSAC) and the Rippel Foundation.

References

Argos, P. (1988). *Protein Eng.* **2**, 101–113.
Bonnefoy, E. & Rouviere-Yaniv, J. (1991). *EMBO J.* **10**, 687–696.
Broyles, S. & Pettijohn, D. E. (1986). *J. Mol. Biol.* **187**, 47–60.
Brünger, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Burley, S. K. & Petsko, G. A. (1985). *Science*, **229**, 23–28.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Craig, N. L. & Nash, H. A. (1984). *Cell*, **39**, 707–716.
Craigie, R., Arndt-Jovin, D. & Mizuuchi, K. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 7570–7574.
Davies, D. R. & Segal, D. M. (1971). *Methods Enzymol.* **22**, 266–269.
Dijk, J., White, S. W., Wilson, K. S. & Appelt, K. (1983). *J. Biol. Chem.* **258**, 4003–4006.

Dixon, N. & Kornberg, A. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 424–428.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Flashner, Y. & Gralla, J. D. (1988). *Cell*, **54**, 713–721.
Friedman, D. I. (1988). *Cell*, **55**, 545–554.
Goshima, N., Kano, Y., Tanaka, H., Tanaka, H., Kohno, K., Yasuzawa, K. & Imamoto F. (1992). *Gene*, **121**, 121–126.
Goshima, N., Kohno, K., Imamoto, F. & Kano, Y. (1990). *Gene*, **96**, 141–145.
Greene, J. R., Brennan, S. M., Andrew, D. J., Thompson, C. C., Richards, S. H., Heinrikson, R. L. & Geiduschek, E. P. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 7031–7035.
Griffith, J. D. (1976). *Proc. Natl Acad. Sci. USA*, **73**, 563–567.
Hodges-Garcia, Y., Hagerman, P. J. & Pettijohn, D. E. (1989). *J. Biol. Chem.* **264**, 14621–14623.
Hosaka, H., Nakagawa, A., Tanaka, I., Harada, N., Sano, K., Kimura, M., Yao, M. & Wakatsuki, S. (1997). *Structure*, **5**, 1199–1208.
Huisman, O., Faelen, M., Girard, D., Jaffe, A., Toussaint, A. & Rouviere-Yaniv, J. J. (1989). *J. Bacteriol.* **171**, 3704 -3712.
Janin, J., Miller, S. & Chothia, C. (1988). *J. Mol. Biol.* **204**, 155–164.
Johnson, G. G. & Geiduschek, E. P. (1977). *Biochemistry*, **16**, 1473–1485.
Johnson, R. C., Bruist, M. F. & Simon, M. I. (1986). *Cell*, **46**, 531–539.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110 -119.
Kano, Y., Osato, K., Wada, M. & Imamoto, F. (1987). *Mol. Gen. Genet.* **209**, 408–410.
Kano, Y., Wada, M., Nagase, T. & Imamoto, F. (1986). *Gene*, **45**, 37–44.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lavoie, B. D. & Chaconas, G. (1995). *Curr. Top. Microbiol. Immunol.* **204**, 83–99.
Lavoie, B. D., Shaw, G. S., Millner, A. & Chaconas, G. (1996). *Cell*, **85**, 761–771.
Marqusee, S. & Baldwin, R. L. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 8898–8902.
Mende, L., Timm, B. & Subramanian, A. R. (1978). *FEBS Lett.* **96**, 395–398.
Nash, H. A. & Robertson, C. A. (1981). *J. Biol. Chem.* **256**, 9246–9253.
Nash, H. A., Robertson, C. A., Flamm, E., Weisberg, R. A. & Miller, H. I. (1987) *J. Bacteriol.* **169**, 4124–4127.
Nicholls, A., Sharp, K. A. & Honig, B. H. (1991). *Proteins*, **11**, 281–296.
Oberto J., Drlica K. & Rouviere-Yaniv J. (1994). *Biochimie*, **76**, 901–908.
Padas, P. M., Wilson, K. S. & Vorgias, C. E. (1992). *Gene*, **117**, 39–44.
Prentki, P., Chandler, M. & Galas, D. J. (1987). *EMBO J.* **6**, 2479–2487.
Rice, P. A., Yang, S., Mizuuchi, K. & Nash, H. A. (1996). *Cell*, **87**, 1295–1306.
Robertson, C. A. & Nash, H. A. (1988). *J. Biol. Chem.* **263**, 3554–3557.
Rouviere-Yaniv, J., Bonnefoy, E., Huisman, O. & Almeida, A. (1990). *The Bacterial Chromosome*, edited by K. Drlica & M. Riley, pp. 247–257. Washington, DC: American Society of Microbiology.
Rouviere-Yaniv, J., Germond, J. & Yaniv, M. (1979). *Cell*, **17**, 265–274.
Rouviere-Yaniv, J. & Kjeldgaard, N. (1979). *FEBS Lett.* **106**, 297–300.
Sayre, M. H. & Geiduschek, E. P. (1988). *J. Virol.* **62**, 3455–3462.
Schneider, G. J. & Geiduschek, P. E. (1990). *J. Biol. Chem.* **265**, 10198–10200.
Searcy, D. G. & Stein, D. B. (1980). *Biochim. Biophys. Acta*, **609**, 180–195.
Shindo, H., Kurumizaka, H., Furubayashi, A., Sakuma, C., Matsumoto, U., Yanagida, A., Goshima, N., Kano, Y. & Imamoto, F. (1993). *Biol. Pharm. Bull.* **16**, 437–443.
Stenzel, T. T., Patel, P. & Bastia, D. (1987). *Cell*, **49**, 709–717.
Tanaka, I., Appelt, K., Dijk, J., White, S. W. & Wilson, K. S. (1984). *Nature (London)*, **310**, 376–381.
Tanaka, I., White, S. W., Appelt, K., Wilson, K. S. & Dijk, J. (1984). *FEBS Lett.* **165**, 39–42.
Vis, H., Mariani, M., Vorgias, C. E., Wilson, K. S., Kaptein, R. & Boelens, R. (1995) *J. Mol. Biol.* **254**, 692–703.
Wada, M., Kano, Y., Ogawa, T., Okazaki, T. & Imamoto, F. (1988). *J. Mol. Biol.* **204**, 581–591.
White, S. W., Appelt, K., Wilson, K. S. & Tanaka, I. (1989). *Proteins Struct. Funct. Genet.* **5**, 281–288.
Wilson, D. L. & Geiduschek, E. P. (1969). *Proc. Natl Acad. Sci. USA*, **62**, 514–520.
Wilson, K. S., Vorgias, C. E., Tanaka, I., White, S. W. & Kimura, M. (1990). *Protein Eng.* **4**, 11–22.
Wimberly, B., White, S. W. & Ramakrishnan, V. (1997). *Structure*, **5**, 1187–1198.