© 1994 Nature Publishing Group http://www.nature.com/nsmb

Solution structure of the DNA-binding domain of *Drosophila* heat shock transcription factor

Geerten W. Vuister[2], Soon-Jong Kim[1], András Orosz[1], John Marquardt, Carl Wu[1] and Ad Bax

The solution structure of the DNA-binding domain of the *Drosophila* heat shock transcription factor, as determined by multidimensional multinuclear NMR, resembles that of the helix-turn-helix class of DNA-binding proteins. The domain comprises a four-stranded antiparallel β-sheet, packed against a three-helix bundle. The second helix is significantly distorted and is separated from the third helix by an extended turn which is subject to conformational averaging on an intermediate time scale. Helix 3 forms a classical amphipathic helix with polar and charged residues exposed to the solvent. Upon titration with DNA, resonance shifts in the backbone and Asn and Gln side-chain amides indicate that helix 3 acts as the recognition helix of the heat shock transcription factor.

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases and [1]Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA
[2]Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584-CH Utrecht, The Netherlands

Correspondence should be addressed to A.B.

Organisms respond to elevated temperatures and to a variety of chemical treatments by rapidly increasing the synthesis of heat shock proteins, which protect the native structure and activity of proteins from the denaturing effects of the cellular stress[1–3]. In eukaryotes, the regulation of the heat shock response is mediated by a pre-existing transcription factor, heat shock factor (HSF), which binds to a conserved regulatory DNA sequence, the heat shock element (HSE), present upstream of heat shock genes[4–6]. The HSE is composed of contiguous, alternating repeats of the five-base pair sequence NGAAN; three NGAAN repeats are required for high affinity interaction with HSF[5]. For multicellular eukaryotes such as *Drosophila*, vertebrates, and plants, the initial stage of HSF activation involves a stress-induced conversion of the monomeric form of the protein to a homotrimer[6]. Although an isolated, monomeric HSF DNA-binding domain lacking the trimerization motifs is capable of binding specifically to a synthetic 13 bp sequence containing a single NGAAN repeat, the affinity of this interaction

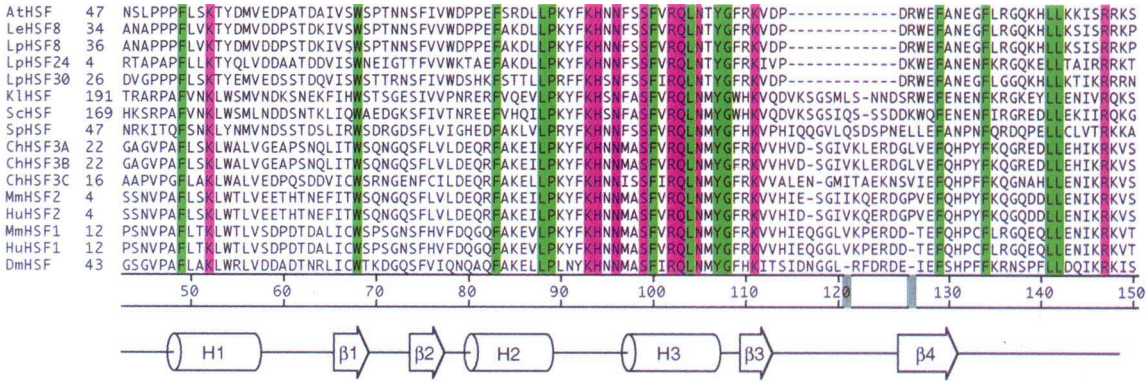

Fig. 1 Alignment of HSF DNA-binding domain sequences from *Arabidopsis thaliana* (AtHSF)[31], *Lycopersicon esculentum* (LeHSF8), *Lycopersicon peruvianum* (LpHSF8, 24, and 30)[32], *Kluyveromyces lactis* (KlHSF)[33], *Saccharomyces cerevisiae* (ScHSF)[34], *Schizosaccharomyces pombe* (SpHSF)[35], *Gallus gallus* (ChHSF)[36], *Mus musculus* (MmHSF1 and 2)[37] and Human (HuHSF1 and 2)[38,39] to the *Drosophila melanogaster* (DmHSF)[22] DNA-binding domain using the CLUSTAL-V algorithm[40]. Absolutely conserved polar and ionic residues are highlighted in red, while absolutely conserved hydrophobic residues are highlighted in green. Secondary structural elements for dHSF residues 43–148 are shown below the sequence alignment.

© 1994 Nature Publishing Group http://www.nature.com/nsmb

is relatively low (K_d ~10^{-7} M) [7]. Trimerization of HSF increases the affinity for the HSE to the picomolar range through the assembly of three DNA binding domains in one complex, thereby allowing the potential for concerted interactions with all three NGAAN boxes of the HSE[4–6].

Sequence analysis of cDNA clones encoding HSFs from a wide selection of animal, plant and fungal species reveals conserved regions in the amino-terminal part of the HSF protein that represent the DNA-binding domain and the trimerization domain. The overall amino acid sequence of the HSF DNA-binding domain (Fig. 1) does not show extensive similarity to any known category of DNA-binding motifs, and it was anticipated that a determination of the structure of HSF would reveal a new motif for specific DNA recognition. Here we report the three-dimensional structure of the DNA-binding domain of *Drosophila melanogaster* HSF (dHSF), determined by heteronuclear nuclear magnetic resonance (NMR) spectroscopy of two bacterially expressed polypeptides comprising residues 33–163 and residues 33–155 of dHSF. We have previously analysed the secondary structure of dHSF[33–163] by NMR spectroscopy, and found the structurally ordered portion of the polypeptide to be essentially coincident with the conserved DNA-binding region (Fig. 1), whereas the amino- and carboxy-terminal residues flanking this conserved region were structurally disordered[8]. To eliminate intense signals from part of the disordered region, we analysed the shorter dHSF[33–155] during the later phase of the structure determination process. The chemical shifts in this shorter polypeptide are indistinguishable from those in dHSF[33–163], confirming that the two polypeptides have identical structures. Residues 43–148 of *Drosophila* HSF form a structurally ordered domain of 106 amino acids, with an overall protein fold surprisingly related to the helix-turn-helix family of DNA-binding proteins[8]. We refer to this region of the polypeptide as the DNA-binding domain of *Drosophila* HSF (dHSF-DBD).

Table 1 Structural statistics and atomic r.m.s. differences[a]

***a* Structural statistics**

	<SA>	$(SA)_r$
r.m.s.d. from experimental distance restraints (Å)[b]		
All (1135)	0.038±0.005	0.042
Intraresidue (378)[c]	0.032±0.007	0.039
Interresidue sequential (|i - j| = 1) (158)	0.044±0.008	0.073
Interresidue medium range (1 < |i - j| < 5) (190)	0.033±0.005	0.025
Interresidue long range (|i - j| 5) (377)	0.042±0.010	0.035
H-bond (32)[d]	0.033±0.015	0.037
r.m.s.d. from experimental dihedral restraints (deg) (65)[e]	0.48±0.17	0.35
Deviations from idealized covalent geometry		
Bonds (Å) (1763)	0.004±0.000	0.004
Angles (deg) (3172)	0.586±0.020	0.534
Impropers (deg) (962)[f]	0.365±0.043	0.236
E_{L-J} (kcal mol^{-1})[g]	-378±33	-515

***b* Atomic r.m.s. differences (Å)**

	Backbone atoms[h]	All atoms[h]	Ordered backbone atoms[i]	All ordered atoms[j]
<SA> vs $(SA)_r$	1.07	1.59	0.94	1.04
<SA> vs SA	0.89	1.35	0.82	0.87
$(SA)_r$ vs SA	0.63	0.91	0.61	0.65

[a]Notation of NMR structures: <SA> are the final 28 simulated annealing structures; SA is the mean structure obtained by averaging the coordinates of the individual structures, best fitted to each other (including only residues 48–112 and 126–145). $(SA)_r$ is the restrained minimized structure obtained by restrained regularization of the mean structure[10].
[b]The total set of 28 structures contains six intraresidue, three sequential, no medium-range, and ten long range NOE violations. $(SA)^r$ contains no distance violations.
[c]Excluding restraints between protons three or less bonds apart and other non-constraining NOEs.
[d]For each of the 16 backbone hydrogen bonds, there are two distance restraints: r_{NH-O}, 1.7–2.5 Å; r_{N-O}, 2.3–3.5 Å.
[e]The torsion angle constraints comprise 28 ϕ, 25 χ_1, and 12 χ_2 angles. The minimum ranges employed for these restraints were ±20°, ±10°, and ±20°,
[f]The improper torsion constraints serve to maintain planarity and chirality.
[g]E_{L-J} is the Lennard-Jones van der Waals' energy, calculated with the CHARMM[42] empirical energy function and is not included in the target function for simulated annealing or restrained minimization.
[h]For residues 48–112 and 126–145.
[i]For residues 48–53, 60–92, 97–112, 126–134, and 140–145.
[j]The atoms included comprise all N, C^α, C, O, and C^β atoms of residues 48–53, 60–92, 97–112, 126–134, and 140–145, the complete side chains of residues F49, L50, L53, T62, L65-T69, S74-I77, F83, L87-P89, Y92, M97-I101, L104, M106-H110, I112, I127, F129, H131-F134, F140-L142, and I145, the side chains of R64, Q73, Q78, Q80, Q82, E86, R102, E128, and Q144 up to C^γ, the side chain of K85 up to C^δ, and the side chain of K52 and K111 up to C^ε.

© 1994 Nature Publishing Group http://www.nature.com/nsmb

Structural restraints

Phenylalanine H^{ζ} protons are particularly important for obtaining correct packing of the hydrophobic core of a protein when determining its solution structure by NMR. These protons are usually located in this core, where they exhibit primarily long-range NOE interactions. In contrast to the other ring protons, the H^{ζ} position is not affected by rapid ring flips and distances measured to H^{ζ} are not subject to the additional uncertainty caused by such flips. Unfortunately, technical problems related to large ^{13}C–^{13}C aromatic J couplings and small ^{13}C chemical shift dispersion usually prohibit assignment of these most valuable resonances for all but the smallest of proteins. dHSF-DBD contain ten Phe residues, nine of which are located in the hydrophobic core of the molecule. Complete resonance assignments of these ten residues could only be made using a novel 'reverse labelling' technique, where natural abundance ($^{12}C/^{14}N$) Phe was incorporated in an otherwise fully $^{13}C/^{15}N$ labelled protein[9]. The Phe protons in this reverse labelled protein also have advantageous relaxation properties, and allow the NOE interactions between Phe and the other residues in the protein to be measured with high sensitvity in a ^{13}C-edited, ^{12}C-filtered 3D HMQC-NOESY experiment. Fig. 2 shows a cross-section through such a 3D spectrum and illustrates the long-range NOE interactions between Phe 134 H^{ζ} and aliphatic protons of Leu 53, Trp 54 and Val 57. One hundred and thirty eight long-range distance restraints to phenylalanine aromatic ring protons were obtained from this 3D spectrum.

A total of 1,135 interproton distance restraints could be extracted from the 3D and 4D NOE data. These distance restraints were supplemented by 65 dihedral restraints obtained from ^{1}H–^{1}H, ^{1}H–^{13}C, ^{13}C–^{13}C, and ^{13}C–^{15}N J coupling data, and 32 hydrogen-bond restraints obtained from 16 slowly exchanging amides. The structural statistics are summarized in Table 1.

The N- and C-terminal residues of dHSF(33–163) are subject to fast amide hydrogen exchange and have low ^{15}N-{^{1}H} heteronuclear NOE values and near random coil H^{α}, C^{α} and C^{β} chemical shift values, indicating that these terminal regions do not adopt a defined structure[8]. Therefore, residues 33–42 and 149–163 were not included in the structure calculations. However, the solvent-exposed and highly flexible loop[8], extending from Thr 113 through Arg 124, was included in the calculations. The 28 structures of dHSF–DBD, calculated with the combined distance geometry and simulated annealing protocol[10,11], are shown in Fig. 3 together with the average structure. Of these, 13 structures satisfy all experimental constraints with no violations greater than 0.5 Å and 5° for distance and dihedral constraints, respectively. The remaining 15 structures show several minor NOE or dihedral violations (cf. Table 1). All structures display only small deviations from idealized geometry and exhibit good non-bonded contacts.

As well as the residues mentioned above, which have rapid, large amplitude motions, that give rise to intense NMR signals, dHSF(33–163) also contains regions with motional averaging on an intermediate NMR time scale (μs–ms). This latter type of motion results in very weak or absent resonances for Trp 54-Asp 59, Lys 93–Asn 96 and Lys 135-Pro 139. Consequently, the number of restraints per residue (Fig. 4*a*) is low for these regions, resulting in an increase in the positional r.m.s.d. of both backbone and side chain atoms (Fig. 4*b*). Overall, residues 48–112 and 126–145 of the 28 structures have a positional r.m.s.d. of 1.4 Å for all non-hydrogen atoms, and 0.9 Å for the backbone. When omitting solvent-exposed, disordered side chains and the three regions with intermediate time scale motional averaging, these r.m.s.d. values reduce to 0.87 Å and 0.82 Å, respectively, indicating a well-packed hydrophobic core. This desirably small positional r.m.s.d. is primarily due to the large number of distance restraints obtained for the phenylalanine residues, using the ^{12}C reverse labelling technique, and to the stereospecific assignments for most leucine and valine methyl groups and the χ^{1} and χ^{2} dihedral constraints obtained for leucine and isoleucine residues.

All backbone angles in the well defined portions of the molecule are in the allowed regions of the Ramachandran ϕ, ψ map. For three of the residues in less well defined regions, Asp 59, Lys 93 and Ser 138, positive ϕ angles are found in the average structure. However, the unusually large ϕ r.m.s.d. values, especially for Asp 59 and Ser 138 (76° and 67°), indicate that the backbone angles for these particular residues are poorly determined by the NMR data.

Tertiary structure

dHSF–DBD forms a three-helix bundle packed against a small four-stranded antiparallel β-sheet. The ribbon diagram of Fig. 5 shows helix 1 on the bottom, helix 2 on the left, helix 3 in the front, and the four-stranded β-

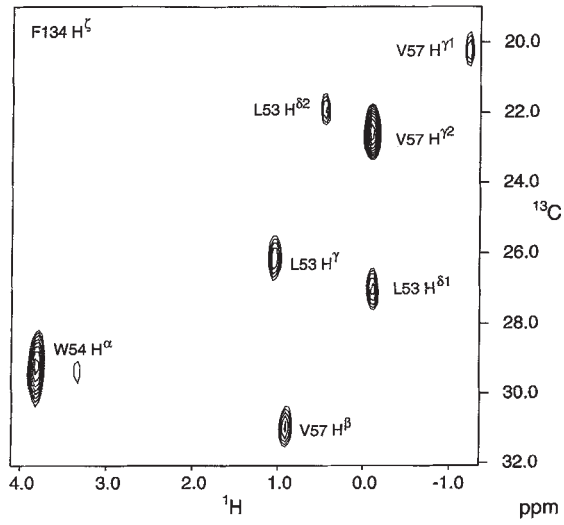

Fig. 2 Small region of a F_1/F_2 cross section, at the F_3 frequency of Phe 134 H^{ζ}, through the 3D ^{13}C-edited, ^{12}C-filtered HMQC-NOESY experiment. The 3D spectrum was recorded in 38 h for a 2 mM solution of {^{12}C-Phe, U-^{13}C/^{15}N} reverse-labelled dHSF(33-155) in D_2O (ref. 9). ^{1}H-^{13}C correlations are seen for all protons that experience an NOE interaction with Phe 134 H^{ζ}. The spectral window in the ^{13}C dimension is 33.1 p.p.m. and the Trp 54 H^{α} resonance is folded once.

© 1994 Nature Publishing Group http://www.nature.com/nsmb

sheet, comprising Ile 66–Thr 69, Ser 74–Ile 77, His 110–Ile 112, and Glu 126–Ser 130 on top. Although the proximity of His 110 and Ser 130 was noted previously[8], our initial NMR data did not conclusively indicate that His^{110}–Ile^{112} formed a β strand. However, upon calculating the structures, this strand (labelled β3 in Fig. 5) is found, oriented at an angle of approximately 30° with respect to β4. Only the distance between the H^{α} protons of Lys 111 and Phe 129 (2.6 Å) is close to the 2.3 Å distance expected in a regular antiparallel β-sheet[12]. The short nature and severe twist of β3 apparently weakens the stable hydrogen-bonding pattern usually found in β-sheets and none of the amide hydrogens of His 110, Ile 112, Glu 128 or Ser 130 are found in very slow exchange with solvent[8]. The β-sheet is anchoured to the hydrophobic core by Ile 66 and Trp 68 in β1, Phe 75 and Ile 77 in β2, and Ile 127 and Phe 129 in β4.

Helix 1 is formed by residues Ala 48-Val 57 as judged from the backbone angles and the hydrogen-bonding pattern. Residues Phe 49, Leu 50, Leu 53, Leu 56 and Val 57 of this helix all show extensive interactions with the hydrophobic core of the protein. Residues at and directly following the C-terminal end of helix 1 are subject to slow conformational exchange and residues Asp 58-Asp 61 therefore display high positional and angular r.m.s.d. values. Residues Thr 62–Leu 65 constitute a non-classical turn leading up to the first strand of the antiparallel β-sheet (Ile 66-Thr 69). Residues Ala 60–Leu 65 are among the least conserved residues of the DNA binding domains of HSF from different HSF species (Fig. 1).

Helix 2, following β-strand 2 and a two-residue turn, comprises residues Gln 80-Pro 89. This helix is highly irregular and its backbone angles deviate substantially from canonical α-helical values. Moreover, the hydrogen-bond pattern is disrupted and only one very slowly exchanging amide is found in this helix. Residues Phe 83 and Leu 88 are tightly packed in the hydrophobic core of the protein and the homologous residues in *Kluyveromyces lactis* HSF have been implicated in causing an α-helical bulge[13]. In addition, the presence of two juxtaposed leucine residues in dHSF, Leu 87 and Leu 88, is likely to contribute to some of the observed unwinding of the helix in this region. The end of helix 2 is marked by Pro 89 which is absolutely conserved among all known HSF species.

Residues Leu 90–Lys 93 form a non-classical turn resulting in short distances between the H^{α} protons of Pro 89 and Lys 93. The aromatic sidechain of Tyr 92 is packed against the hydrophobic core, showing interactions with residues of helix 1, that is Ala 48, Phe 49, Lys 52, and Leu 56, as well as Leu 88 in helix 2. Residues His 94-Asn 96 are in an approximately extended conformation, leading up to helix 3. As mentioned above, conformational averaging in this region results in poor definition of the structure, reflected in higher positional and dihedral r.m.s.d. values for these residues.

Helix 3 comprises residues Met 97–Tyr 107 and is well characterized by its regular hydrogen-bonding pattern

Table 2 Structural comparison of helix-turn-helix segments

a Atomic r.m.s. Differences Å

	KlHSF[c]	CAP[d]	HNF-3γ[e]	Oct-1[f]	λ-rep[g]	434-rep[h]	434-Cro[i]
DmHSF[b]	0.78	1.52	2.10	1.87	2.01	1.95	1.83
KlHSF		1.42	1.89	2.00	2.12	2.06	1.92
CAP			2.32	2.10	2.25	2.20	2.15
HNF-3γ				1.93	2.12	2.04	2.01
Oct-1					0.46	0.52	0.55
λ repressor						0.43	0.50
434 repressor							0.52

b Interhelical angles (deg)[j]

	Helices 1+3	Helices 2+3
DmHSF	117	80
KlHSF	106	81
CAP	78	103
HNF-3γ	84	120
Oct-1	105[k]	115
λ repressor	92[k]	102
434 repressor	99[k]	116
434 Cro	113[k]	126

[a]Helix backbone atoms (N, C^{α}, C, and C^{β}) for each helix-turn-helix motif were aligned and r.m.s. differences (Å) calculated using X-PLOR[11].
[b]Residues 80–85 and 97–105 of the DmHSF NMR structure reported in this work.
[c]Residues 36–41 and 53–61 of the KlHSF X-ray structure[13].
[d]Residues 169–174 and 183–191 of the CAP–DNA X-ray structure[43].
[e]Residues 145–150 and 164–172 of the HNF-3γ–DNA X-ray structure[15].
[f]Residues 27–32 and 44–52 of the Oct-1-POU_{sp}–DNA X-ray structure[44].
[g]Residues 33–38 and 44–52 of the λ repressor–DNA X-ray structure[45].
[h]Residues 17–22 and 28–36 of the 434 repressor–DNA X-ray structure[46].
[i]Residues 17–22 and 28–36 of the 434 Cro-DNA X-ray structure[47].
[j]Calculated using helical vectors generated by Quanta (Molecular Simulations, Inc) Protein Design Module.
[k]The C-terminal helix analogous to the DmHSF helix-1 was used.

© 1994 Nature Publishing Group http://www.nature.com/nsmb

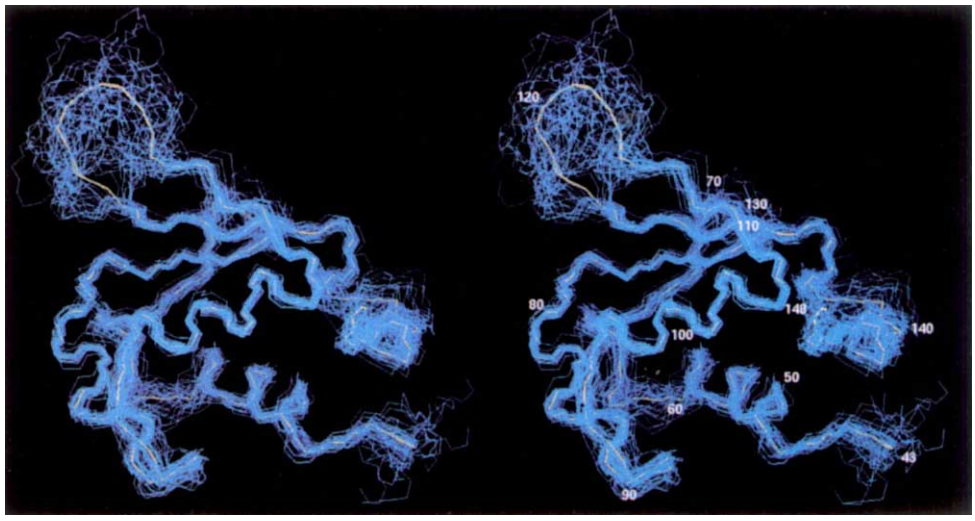

Fig. 3 Best-fit superposition of the backbones of the 28 simulated annealing structures (blue) and the refined average structure of dHSF-DBD (yellow).

and small deviations in its positional and angular r.m.s.d. values. The helix is amphipathic with Met 97, Phe 100, Leu 104, and Tyr 107 packed into the hydrophobic core of the protein. Several hydrophilic residues protrude from the protein, capable of making contacts with the DNA (see below). The end of helix 3 is followed by a two-residue turn, leading up to the third β-strand. The absolutely conserved glycine (Gly 108) in this turn has a positive ϕ angle.

Strands β3 and β4 are connected by a long loop, stretching from Thr 113 through Arg 124. As shown previously[8], this loop is highly flexible, solvent exposed, and is not affected by DNA binding. Its structural significance therefore remains unclear. The corresponding region in the *Lycopersicon* and *Arabidopsis* HSF sequences has a ten-residue deletion (Fig. 1), suggesting that most of this loop is dispensable. Indeed, in our structure of dHSF-DBD the loop can easily be replaced by a tight turn, connecting the third and fourth strands, without disrupting the overall structure. The fourth β-strand is followed by an extended region and a poorly defined short loop, comprising Lys 135–Ser 138, which is subject to slow conformational averaging. This loop is followed by six residues, Pro 139–Gln 144, which form approximately one and a half unwound helical turns. Residues Leu 142 and Ile 145, on one side of these turns, interact with the hydrophobic core and close off one face of the protein. Leu 142 is absolutely conserved among all known HSF species, and Ile 145 is, with one exception, also conserved, suggesting that these interactions are important for correct packing of the molecule. Overall, dHSF-DBD is approximately spherical, with dimensions along the principal axes of 37 Å x 33 Å x 38 Å, excluding the loop containing residues Thr 113–Arg 124 which protrudes extensively from the surface of the molecule into solution.

Homonuclear and heteronuclear coupling constants indicate conformational averaging for several side chains, particularly for residues in the flexible loop, Thr 113–Arg 124. For example, Ile 112, Ile 115, and Leu 120 all show χ^1 and χ^2 rotameric averaging, whereas Phe^{122} and Thr^{113} display χ_1 rotameric averaging. However, several of the side chains in well-defined regions of the protein are also subject to side chain rotameric averaging. For example, the carbonyl and amide nitrogen of Val 76 both show intermediate values for the J couplings to both C^γ methyl carbons.

K. lactis HSF and other HTHS

It is interesting to compare the solution structure of dHSF-DBD with the recently reported crystal structure of the homologous DNA-binding domain of *K. lactis* HSF[13]. Considering the extensive conformational averaging of both backbone and side chain resonances for several regions in dHSF-DBD, and the absence of residues homologous to Phe 134–Lys 148 in the crystallized yeast polypeptide, it is remarkable that the two structures can be superimposed with an r.m.s.d. of only 1.6 Å. The additional residues present in dHSF-DBD, particularly Pro 139–Lys 148, are ordered and several of the absolutely conserved hydrophobic residues in this region interact extensively with the core of the protein, as noted above.

Among the members of the helix-turn-helix family compared in Table 2, the angle between helices 1 and 3 shows considerable variation. This value ranges from 78° for CAP to 117° for DmHSF. The value for KlHSF is 106°, and the 11° difference between the KlHSF and DmHSF structures may be related to slow conformational averaging observed for several of the helix 1 side chains in DmHSF, which reduces the precision at which the position of this helix can be determined. The angles between helices 2 and 3 of the helix-turn-helix motifs are about 80° for both the DmHSF and KlHSF structures (Table 2*b*). This unusually small value distinguishes the DmHSF and KlHSF domains from other members of the helix-turn-helix family which have interhelical angles ranging from 102° to 126°.

Comparing the helix-turn-helix motifs from DmHSF, KlHSF, CAP, HNF-3γ, Oct1-POU$_{sp}$, λ repressor, 434 repressor and 434 Cro structures, the helices in the DmHSF

© 1994 Nature Publishing Group http://www.nature.com/nsmb

NMR structure most closely resemble those of the KlHSF X-ray crystal structure, with a r.m.s.d. in backbone coordinates of 0.78 Å (Table 2*a*). The only other helix-turn-helix proteins exhibiting greater pairwise similarity involve members of the λ and 434 repressor families.

DNA binding

We have previously noted similarities between the topology of dHSF(33–163)[8] and those of the DNA-binding domains of the helix-turn-helix bacterial activator CAP[14], and the liver-specific transcription factor HNF-3γ, the prototype of the HNF/*forkhead* protein family[15]. Harrison *et al.*[13] showed similarities in the relative positions of helices 1 and 3 in the DNA-binding domains of *K. lactis* HSF and CAP, biotin repressor, histone H5, and HNF-3. These proteins use helix 3 for sequence-specific recognition in a classical helix-turn-helix motif or a variation thereof. CAP binds to DNA as a dimer, whereas HNF-3γ binds DNA as a monomer. HNF-3γ makes additional contacts with the DNA through residues distal to helix 3, presumably to increase binding affinity[15]. In contrast, full-length HSF binds the HSE *in vitro* as a trimer, and its high affinity for DNA results from cooperative effects between the three HSF subunits.

The dHSF(33–163) polypeptide lacks the trimerization domain and binds a 13-bp DNA duplex containing one copy of the NGAAN recognition site with submicromolar affinity and 1:1 stoichiometry[7]. Earlier, we used exchange broadening of the ^{1}H-^{15}N correlations upon titration of ^{15}N labelled HSF with the 13 bp DNA duplex to probe the residues most affected by DNA binding. These residues are colour coded blue in the backbone trace of dHSF (Fig. 6). It is interesting to note that several residues with side chains that are packed into the hydrophobic core (Leu 50, Tyr 92, His 94, Met 97, Phe 100, Leu 104 and Phe 109) are among the residues for which the chemical shifts of the backbone amides are strongly affected by DNA binding. As the hydrophobic core would not be expected to change its structure upon DNA binding, this finding might initially seem to be counterintuitive. However, it should be noted that the backbone amide is displaced by about 35° on the helical wheel relative to C^{α}, and most of the corresponding amides are not buried in the hydrophobic core. Moreover, as the chemical shift is exquisitely sensitive to any type of structural variation, even very small rearrangements upon DNA binding can suffice to affect the amide chemical shift significantly.

The residues most affected by DNA binding are clustered in the loop region just before helix 3 and throughout helix 3 itself, identifying helix 3 as the main site for interaction with the DNA. Methylation interference experiments indicate that HSF binds the HSE primarily through major groove interactions[16,17]. Therefore, we propose that helix 3 acts as the recognition helix of dHSF in a variation of the helix-turn-helix motif[18]. Helix 3 of dHSF is amphipathic with the side chains of residues Ser 99, Arg 102, Gln 103, Asn 105 and Met 106 protruding from the surface, capable of making direct or water-mediated contacts with DNA. Side chain amide resonances of asparagine and glutamine residues are sensitive monitors for studying interaction with DNA[19]. In the absence of such an interaction, the solvent-exposed side chain NH_2 moiety frequently yields a motionally averaged, intense set of resonances. These resonances broaden and shift significantly once immobilized by contacts with DNA. The asparagine and glutamine sidechains affected in this manner by DNA binding are highlighted in Fig.

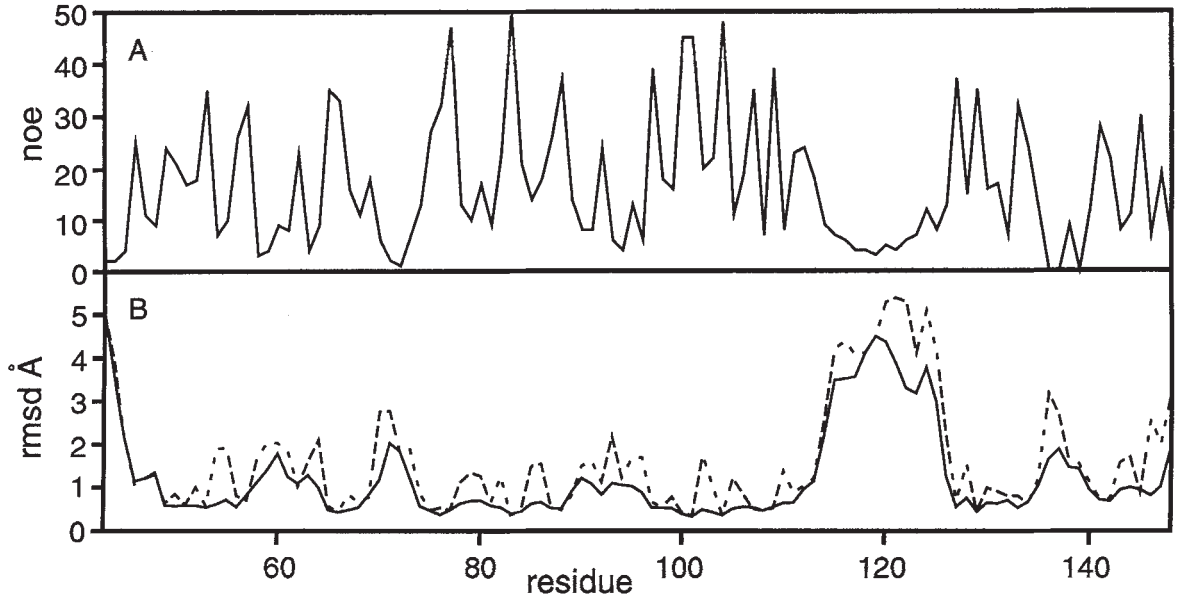

Fig. 4 *a*, Total number of sequential, medium-range, and long-range restraints per residue and *b*, positional r.m.s.d. values per residue for the backbone (solid lines) and all heavy atoms (dashed lines). Restraints are counted for both donor and acceptor atoms. R.m.s.d. values are calculated from the 28 refined structures.

© 1994 Nature Publishing Group http://www.nature.com/nsmb

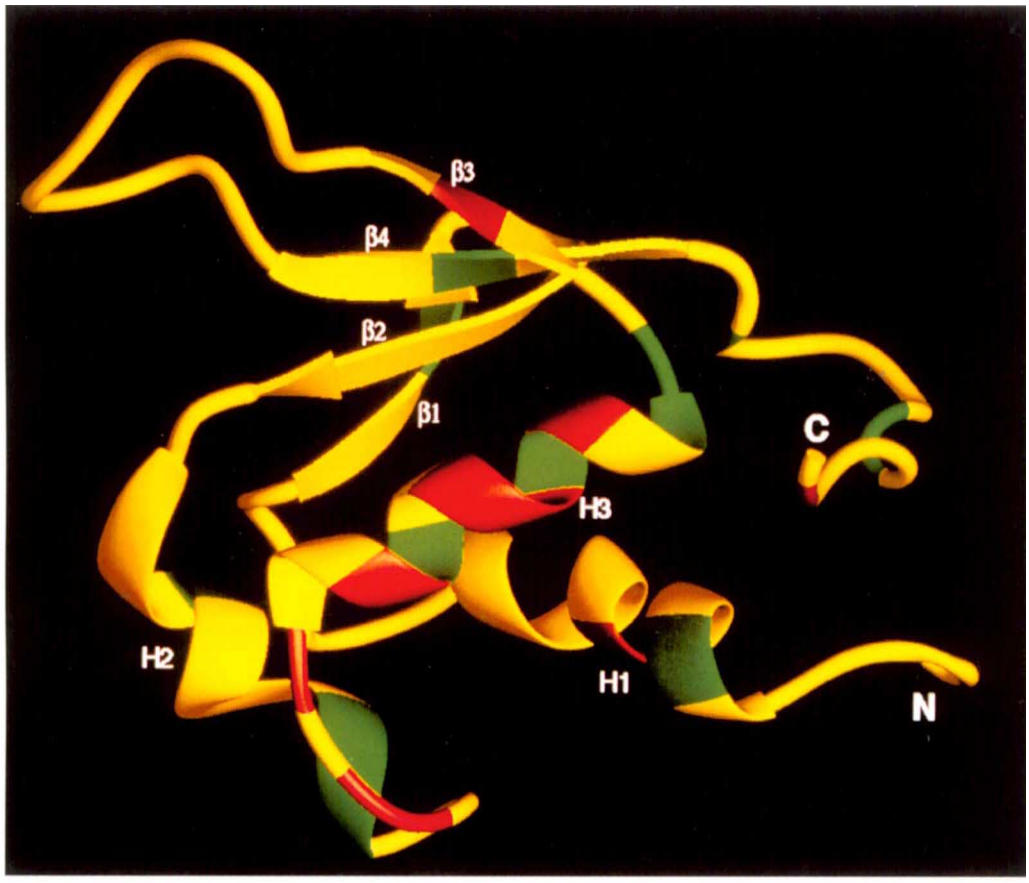

Fig. 5 Ribbon diagram of the refined average structure of dHSF-DBD showing helix one (H1), helix two (H2), and helix 3 (H3), and the four stranded antiparallel β-sheet (β1–4). Absolutely conserved polar and ionic residues are highlighted in red, while absolulely conserved hydrophobic residues are highlighted in green. The plot was generated using the program Ribbons[41].

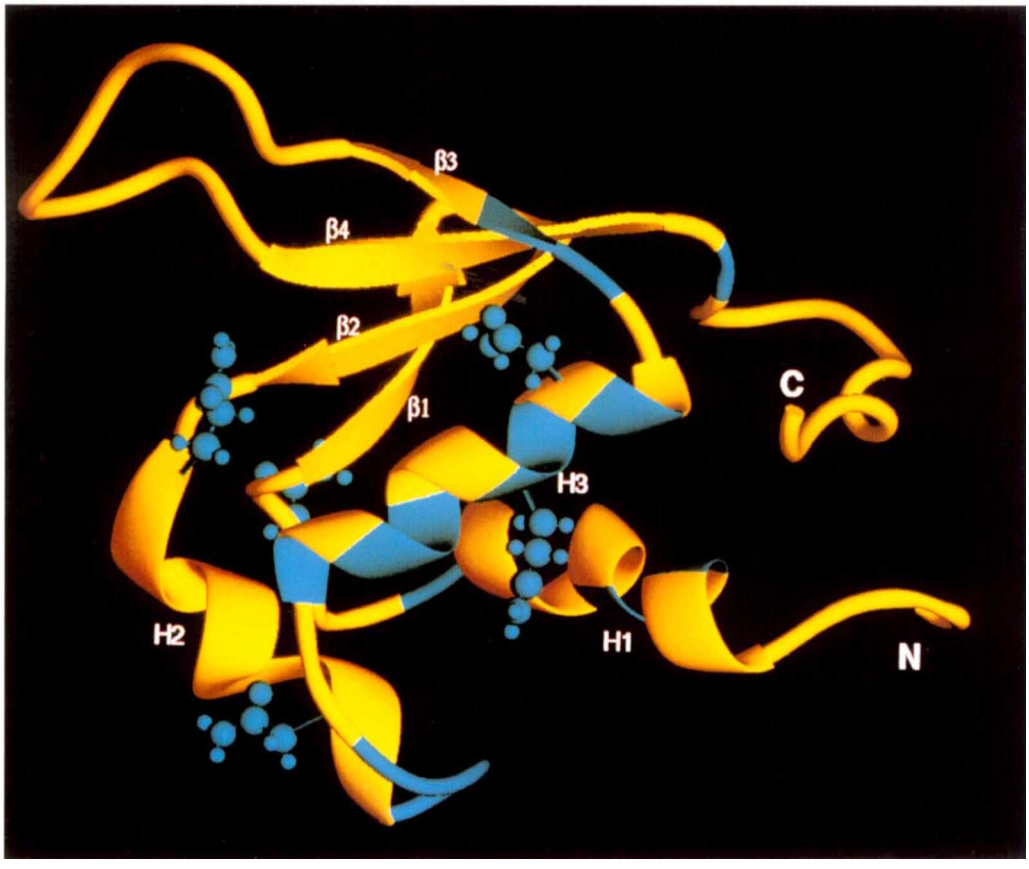

Fig. 6 Ribbon trace of the backbone residues of the refined average structure of dHSF-DBD, (yellow ribbon) illustrating the residues with backbone amides most affected by DNA titration (coloured blue) and asparagine and glutamine residues whose amide sidechain resonances are affected upon DNA binding (blue side-chains). The plot was generated using the program Ribbons[41].

© 1994 Nature Publishing Group http://www.nature.com/nsmb

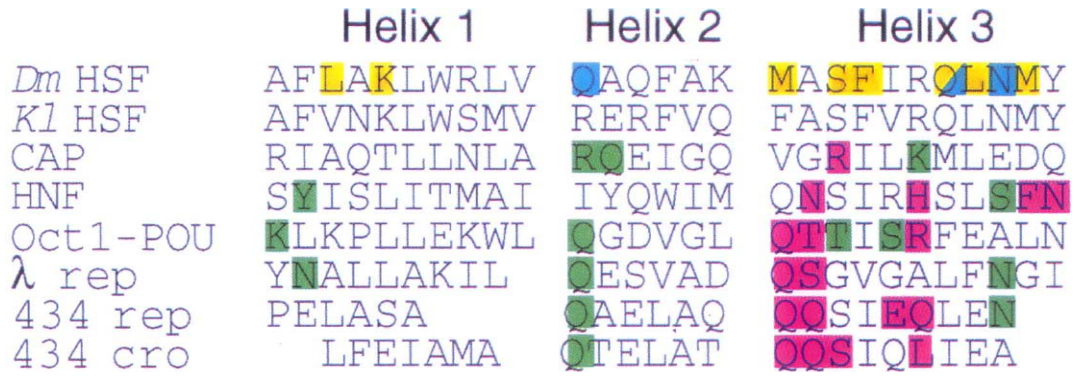

Fig. 7 Alignment of amino acid sequences of the helix-turn-helix proteins listed in Table 2. Residues reported to be involved in contacting the DNA backbone are highlighted in green and residues involved in contacting the DNA bases are highlighted in red. Residues of the DmHSF sequence whose side-chain resonances are affected by DNA binding are highlighted in blue and residues with affected backbone resonances are are highlighted in yellow.

6, including the amides of Gln 80 in helix 2, Gln 103 and Asn 105 in helix 3, and Asn 95 in the loop before helix 3. In addition, the backbone amide resonances for residues Met 97, Ser 99, Phe 100, Gln 103, Leu 104 and Met 106 in helix 3 broaden beyond detection upon titrating DNA into the NMR sample. The helical residues of DmHSF which are affected most by the addition of DNA are indicated in the sequence alignment of Fig. 7 which also highlights residues reported to be involved in base and backbone contacts for the CAP, HNF-3γ, Oct1-POU^{sp}, λ repressor, 434 repressor and 434 Cro proteins. Although each protein exhibits a distinct pattern for the third helix, the DmHSF side chain amide of Gln 80 at the N-terminal end of helix 2 has a clear analogue in many of the other DNA-binding domains, that is in CAP, Oct1, λ repressor, 434 repressor and 434 Cro. In the crystal structures of the DNA complexes of these proteins, both the backbone and the side chain amides of this residue make hydrogen bonds to phosphates in the DNA backbone. At the present stage of our investigation, it cannot be determined whether any of the DmHSF-DNA contacts are direct or water-mediated.

Role of conserved residues

Many of the absolutely conserved residues of dHSF-DBD (cf. Fig. 1) are important for the correct packing of the hydrophobic core of the molecule and the interactions of Phe 49, Trp 68, Phe 83, Leu 88, Pro 89, Phe 100, Leu 104, Tyr 107, Gly 108, Phe 129, Phe 134 and Leu 142 have been discussed. Sequence conservation in conjunction with structural analysis provides additional clues about the possible mode of interaction with DNA. The conserved residues in amphipathic helix 3, Ser 99 and Arg 102 are likely to interact with the DNA, whereas titration studies indicate that Gln 103 and Asn 105, also located in helix 3, are involved in DNA binding. The extended turn connecting helix 2 and helix 3 includes the absolutely conserved residues Lys 93, His 94 and Asn 96, in good positions for potential interaction with the DNA. This extended turn also shows a concentration of residues with backbone amide chemical shifts that are affected by DNA binding, although backbone amides tend to be less specific reporters for intermolecular interactions than side chain amides. Two more absolutely conserved residues can be found in positions where they may potentially interact with DNA. Lys 111 in β-strand 3 has a side chain pointing toward the face of the protein that interacts with the DNA, as does Arg 147, which makes long-range contacts with the C-terminal end of helix 3. In this respect, it is interesting that Arg 147 is followed by a second highly conserved basic residue.

The solution structure of dHSF-DBD, comprising the evolutionarily conserved DNA-binding domain of HSF, forms a three-helix bundle packed against a four-stranded antiparallel β-sheet. Our results clearly identify the surface of the polypeptide that contacts DNA, with helix 3 acting as the recognition helix in a variation of the helix-turn-helix motif, and implicates several of the residues involved in this interaction. Although study of the structure of the dHSF-DBD complex with DNA by either NMR or crystallography is hampered by the relatively low affinity of the monomeric binding domain, this work provides a first step towards an understanding of the structure of the protein–DNA complex, which, in turn may address the intriguing question of how the three monomeric binding domains of the HSF trimer are naturally aligned on the 15 bp heat shock element.

Methods

The dHSF(33–163) and dHSF(33–155) constructs contain the gene fragments for residues 33–163 and 33–155, respectively, of *Drosophila* HSF and were cloned into expression vector pJC20 as described elsewhere[22]. The dHSF(33–163) construct contains three additional C-terminal residues, resulting in a molecular mass of 15,200. The molecular weight of the dHSF(33-155) peptide is 14,000. DNase I footprinting studies show that both constructs bind specifically to a 13-mer DNA duplex, GCGCAGAACGCCG, containing a single copy of the HSF recognition site[16].

Uniformly ^{15}N and $^{15}N/^{13}C$ labelled dHSF(33–63) was obtained as described before[8]. Reverse-labelled [^{12}C]Phe dHSF(33–155) was obtained using an M9 minimal medium containing 0.3% [$^{13}C_6$] glucose/0.1% $^{15}NH_4Cl$, supplemented by 20 mg l^{-1} ^{12}C L-Phe.

© 1994 Nature Publishing Group http://www.nature.com/nsmb

Sample concentrations were 1–2 mM in 220μl in 5-mm NMR microcells (Shigemi Inc., Allison Park, PA). All NMR experiments were conducted at pH 6.3, 27 ° C on a Bruker AMX-600 spectrometer, equipped with a triple resonance probehead and pulsed field gradients.

Resonance assignments were obtained using an array of heteronuclear multidimensional experiments. Most aliphatic side chains could be correlated with the previously assigned[8] backbone atoms by means of $_1H'$-^{13}C-^{13}C-$_1H$ (HCCH) type correlation spectra[23], recorded in D_2O solution, and from several of the experiments aimed at the measurement of J couplings[24]. In addition, isotope-filtered experiments on reverse-labelled [^{12}C-Phe] dHSF (33–155) provided complete assignments for the Phe aromatic ring protons[9].

Proton–proton distance restraints were obtained from the water-flip-back 3D ^{15}N-separated NOESY[25], the 4D gradient-enhanced ^{15}N/^{13}C separated NOESY[26] and 4D ^{13}C/^{13}C separated NOESY[27] spectra. All distance restraints were classified as strong, (1.8–2.7 Å (1.8–2.9 Å for NOEs involving NH protons), medium, 1.8–3.3 Å (1.8–3.5 Å for NOEs involving NH protons), and weak, 1.8–5.0 Å, in the manner described by Clore *et al.*[28]. Upper distance limits of restraints involving non-stereo specifically assigned methylene protons or methyl groups were corrected appropriately for centre averaging[29]. In addition, upper distance limits of restraints involving methyl groups were increased by 0.5 Å to account for the apparent higher intensity of the methyl resonances[30].

Backbone amides involved in hydrogen bonds were identified from slowly exchanging resonances in the ^{15}N-^{1}H correlation map. A total of 32 restraints, that is r(H^NO) 1.7–2.5Å and r(NO) 2.3–3.5 Å per hydrogen bond, were added for slowly exchanging amides in either α-helical or β-sheet conformation[8].

Quantitative J-correlation experiments[24] yielded $^3J(H_NH_\alpha)$, $^3J(NH_\beta)$, $^3J(COH_\beta)$, ^3J(CC), ^3J(CH), ^3J(CN), and ^3J(CCO) coupling constants. Combined use of these coupling constants yielded stereospecific assignments of the C^γ methyl groups of V46, V57, the C^β methylene protons and C^δ methyl groups of L53, L65, L87, L88, L104, L141, L142, the C_γ methylene protons of I77, I101, I145, and the β-methylene groups of W68, F75, F83, F100, R102, Q103, Y107, F109, K111, F129, F133, F134, and F140. In addition, 65 dihedral restraints, 28 ϕ, 25 χ^1, and 12 χ^2, were obtained from the coupling data.

Forty structures were calculated using the hybrid distance geometry and simulated annealing protocol[10] implemented using the program X-*PLOR* 3.1 (ref. 11). A target function comprising quadratic harmonic potential terms for covalent geometry, 'soft' square-well quadratic potentials for experimental distance restraints, square-well quadratic potentials for experimental dihedral restraints, and a quartic van der Waals' repulsion term for the non-bonded contacts, was minimized during the simulated annealing phase of the protocol. Of these, the 28 lowest energy structures were chosen and subjected to an additional simulated annealing and refinement phase. Coordinates have been deposited with the Brookhaven Protein Data Bank (ident code 1HKS).

Received 25 May; accepted 14 July, 1994

© 1994 Nature Publishing Group http://www.nature.com/nsmb

Acknowledgements
We thank M. Clore for many useful comments and suggestions, J. Kuszewski for expert assistance with the X-*PLOR* program, H. Nelson for an exchange of coordinates, and C. Pabo and S. K. Burley for providing coordinates of the Oct1-POU^{so} and HNF-3γ complexes with DNA. This work was supported by the AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health.

1. Lindquist, S. & Craig, E.A. The heat-shock proteins. *A. Rev. Genet.* **22**, 631–677 (1988).
2. Morimoto, R.I., Tissieres, A. &. Georgopoulos, C. The stress response, function of the proteins, and perspectives. In *Stress proteins in Biology and Medicine* (Eds, Morimoto, R.I., Tissieres, A. & Georgopoulos, C.,). 1–36 (Cold Spring Harbor Laboratory Press, New York, 1990).
3. Hendrick, J.P. & Hartl, F.U. Molecular chaperone functions of heat-shock proteins. *A. Rev. Biochem.* **62**, 349–384 (1993).
4. Sorger, P.K. Heat shock factor and the heat shock response. *Cell* **65**, 363–366 (1991).
5. Lis, J.T. & Wu, C. Heat Shock Factor. In *Transcriptional Regulation*, (Eds, McKnight, S.L. & Yamamoto, K.R.,) 907–930 (Cold Spring Harbor Laboratory Press, New York, 1992).
6. Lis, J.T. & Wu, C. Protein traffic on the heat shock promoter: parking, stalling and trucking along. *Cell* **74**, 1–20 (1993).
7. Kim, S.-J., Tsukiyama, T., Lewis, M.S. & Wu, C. The interaction of the DNA binding domain of *Drosophila* heat shock factor with its cognate DNA site: a thermodynamic analysis using analytical ultracentrifugation. *Prot. Sci.* **3**, 1040–1051 (1994).
8. Vuister, G.W., Kim, S.-J., Wu, C. & Bax, A. NMR evidence for similarities between the DNA-binding regions of Drosophila melanogaster heat shock factor and the helix-turn-helix and HNF^3/fork head families of transcription factors. *Biochemistry* **33**, 10–16 (1994).
9. Vuister, G.W., Kim, S.-J., Wu, C. & Bax, A. 2D and 3D NMR study of phenylalanine residues in proteins by reverse isotopic labeling. *J. Am. chem. Soc.* (in the press).
10. Nilges, M., Clore, G. M. & Gronenborn, A. M. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* **229**, 317–324 (1988).
11. Brunger, A.T. *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*, Yale University, New Haven, CT, USA (1992).
12. Wüthrich, K. *NMR of Proteins and Nucleic Acids*, (John Wiley, New York, 1986).
13. Harrison, C.J., Bohm, A.A. & Nelson, H.C.M. Crystral structure of the DNA binding domain of the heat shock transcription factor. *Science* **263**, 224–227 (1994).
14. Weber, I. T. & Steitz, T. A. Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution. *J. molec. Biol.* **198**, 311–326 (1987).
15. Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. Co-crystal structure of the HNF-3/*fork head* DNA-recognition motif resembles histone H5. *Nature* **364**, 412–420 (1993).
16. Perisic, O., Xiao, H. & Lis, J. T. Stable binding of *Drosophila* heat shock factor to head-to-head and tail-to-tail repeats of a conserved 5 bp recognition unit. *Cell* **59**, 797–806 (1989).
17. Fernandes, M., Xiao, H. & Lis, J. T. Fine structure analyses of the *Drosophila* and *Saccharomyces* heat shock factor – heat shock element interactions. *Nucleic Acids Res.* **22**, 167–173 (1994).
18. Harrison, S. C. A structural taxonomy of DNA-binding domains. *Nature* **353**, 715–719 (1991).
19. Dekker, N., Cox, M., Boelens, R., Verrijzer, C. P., var der Vliet, P. C. & Kaptein, R. Solution structure of the POU-specific DNA-binding domain of Oct-1. *Nature* **362**, 852–855 (1993).
20. Pabo, C. O., Aggarwal, A. K., Jordan, S. R., Beamer, L. J., Obeysekare, U. R. & Harrison, S. C. Conserved residues make similar contacts in two repressor-operator complexes. *Science* **247**, 1210–1213 (1990).
21. Assa-Munt, N., Mortishire-Smith, R. J., Aurora, R., Herr, W. & Wright, P. E. The solution structure of the Oct-1 POU-specific domain reveals a striking similarity to the bacteriophage l repressor DNA-binding domain. *Cell* **73**, 193–205 (1993).
22. Clos, J. *et al.* Molecular cloning and expression of a hexameric *Drosophila* heat shock factor subject to negative regulation. *Cell* **63**, 1085–1097 (1990).
23. Clore, G. M. & Gronenborn, A. M. Applications of three- and four-dimensional heteronuclear NMR spectroscopy to protein structure determination. *Progr. NMR Spectr.* **23**, 43–92 (1991).
24. Bax, A., *et al.* Measurement of homo- and heteronuclear *J* couplings from quantitative *J* correlation. *Meth. Enzymol.*, (Eds, James, T. L. & Oppenheimer, N.) **239** 79–125 (Academic Press, San Diego, 1994).
25. Grzesiek, S. & Bax, A. The importance of not saturating H_2O in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. chem. Soc.* **115**, 12593–12594 (1993).
26. Muhandiram, D. R., Xu, G. Y. & Kay, L. E. An enhanced-sensitivity pure absorption gradient 4D ^{13}C-edited NOESY experiment. *J. biomol. NMR* **3**, 463–470 (1993).
27. Vuister, G. W. *et al.* Increased resolution and improved spectral quality in four-dimensional $^{13}C/^{13}C$ separated HMQC-NOESY-HMQC spectra using pulsed field gradients. *J. magn. Reson.* **B 101**, 210–213 (1993).
28. Clore, G. M. *et al.* The three-dimensional structure of α1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *EMBO J.* **5**, 2729–2735 (1986).
29. Wüthrich, K., Billeter, M. & Braun, W. Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. *J. molec. Biol.* **169**, 949–961 (1983).
30. Clore, G. M., Gronenborn, A. M., Nilges, M. & Ryan, C. A. Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. *Biochemistry* **26**, 8012–8023 (1987).
31. Huebel, A. & Schoeffl, F. *Arabidopsis thaliana* HSF gene sequence (submitted to the EMBL Data Library, November 1993).
32. Scharf, K.D., Rose, S., Zott, W., Schoef, F. & Nover, L. Three tomato genes code for heat stress transcription factors with a region of remarkable homology to the DNA-binding domain of the yeast HSF. *EMBO J.* **9**, 4495–4501 (1990).
33. Jakobsen, B. K. & Pelham, H. R. B. A conserved heptapeptide restrains the activity of the yeast heat shock transcription factor. *EMBO J.* **10**, 369–375 (1991).
34. Sorger, P. K. & Pelham, H. R. Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. *Cell* **54**, 855–864 (1988).
35. Gallo, G. J., Prentice, H. & Kingston, R. E. Heat shock factor is required for growth at normal temperatures in the fission yeast *Schizosaccharomyces pombe*. *Molec. cell. Biol.* **13**, 749–761 (1993).
36. Nakai, A. & Morimoto, R. I. Characterization of a novel chicken heat shock transcription factor 3, suggests a new regulatory pathway. *Molec. cell. Biol.* **13**, 1983–1997 (1993).
37. Sarge, K. D., Zimarino, V., Holm, K., Wu, C. & Morimoto, R. I. Cloning and characterization of two mouse heat shock factors with distinct inducible and constitutive DNA binding ability. *Genes Dev.* **5**, 1902–1911 (1991).
38. Rabindran, S. K., Giorgi, G., Clos, J. & Wu, C. Molecular cloning and expression of a human heat shock factor, HSF1. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6906–6910 (1991).
39. Schuetz, T. J., Sheldon, L., Gallo, G. J., Tempst, P. & Kingston, R. E. Isolation of a cDNA for HSF2: evidence for two heat shock factor genes in humans. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6911–6915 (1991).
40. Higgins, D. G. & Sharp, P. M. Fast and sensitive multiple sequence alignments on a microcomputer. *Cabios Comm.* **5**, 151–153 (1989).
41. Carson, M. Ribbon models for macromolecules. *J. molec. Graphics* **5**, 103–106 (1987).
42. Brooks, B.R. *et al.* CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. comput. Chem.* **74**, 187–217 (1983).
43. Schultz, S.C, Shields, G.C. & Steitz, T.A. Crystal structure of CAP-DNA is bent by 90°. *Science* **253**, 1001–1007 (1991).
44. Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. & Pabo, C.O. Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* **77**, 21–32 (1994).
45. Beamer, L.J. & Pabo, C.O. Refined 1.8 Å crystal structure of the λ repressor-operator complex. *J. molec. Biol.* **227**, 177–196 (1992).
46. Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. & Harrison, S.C. Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science* **242**, 899–907 (1988).
47. Mondragon, A. & Harrison, S.C. The phage 434 Cro/O_R1 complex at 2.5 Å resolution. *J. molec. Biol.* **219**, 321–334 (1991).