Accelerated Publications

NMR Evidence for Similarities between the DNA-Binding Regions of *Drosophila melanogaster* Heat Shock Factor and the Helix–Turn–Helix and HNF-3/*forkhead* Families of Transcription Factors†

Geerten W. Vuister,‡ Soon-Jong Kim,§ Carl Wu,§ and Ad Bax*,‡

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, and Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

Received October 8, 1993®

ABSTRACT: Heteronuclear multidimensional NMR experiments of residues 33–163 of the DNA-binding domain of *Drosophila* heat shock factor, dHSF(33–163), were recorded, using only 3 mg of uniformly ^{15}N-labeled or 2 mg of uniformly $^{15}N/^{13}C$-labeled protein. The polypeptide consists of a structured part comprising three helices, a three-stranded antiparallel β-sheet, with the first two strands connected by a four-residue type I tight turn. The second helix is disrupted at its C-terminal end by a proline residue and is followed by an extended turn, leading to the third helix. The dHSF(33–163) protein is unstructured at its N- and C-termini, and a third unstructured region is found from Thr[113] to Arg[124]. Exchange broadening of the ^{15}N-^{1}H correlations upon titration of ^{15}N labeled HSF with a 13-base-pair DNA duplex suggests a DNA-binding motif in which the third helix acts as the recognition helix. Both the secondary structure and DNA-binding pattern of dHSF(33–163) suggest that the overall topology resembles that the helix–turn–helix bacterial activator CAP [Weber, I. T., & Steitz, T. A. (1987) *J. Mol. Biol. 198*, 311–326] and the liver-specific transcription factor HNF-3γ, the prototype of the HNF-3/*forkhead* protein family [Clark, K. L., Halay, E. D., Lai, E., & Burley, S. K. (1993) *Nature 364*, 412–420].

In response to heat or chemical stress, all organisms rapidly induce the expression of heat shock protein genes (Morimoto et al., 1990). Recent studies suggest a central role for heat shock proteins as molecular chaperones, mediating protein–protein interactions, protein folding, and the translocation of proteins across cellular membranes (Gething & Sambrook, 1992). The synthesis of heat shock proteins is regulated at both transcriptional and translational levels. The transcriptional response to heat shock in eukaryotes is mediated by the transcription factor HSF[1] (heat shock factor) (Lis & Wu, 1992). In higher eukaryotes, such as the fruitfly *Drosophila* and human, HSF exists in a latent, monomeric form which trimerizes upon heat stress (Westwood & Wu, 1993). This monomer to trimer transition constitutes the first stage of HSF activation, which results in binding of HSF to its target DNA site with high affinity. The DNA sequence recognized by HSF, the heat shock element (HSE), is composed of three repeats of a conserved, five-base-pair module, nGAAn, arranged in inverted orientation: nGAAnnTTCnnGAAn (Lis & Wu, 1992).

The formation of a HSF trimer is dependent on intermolecular interactions between several hydrophobic heptad repeats or leucine zipper motifs located in the N-terminal region of the protein (Sorger & Nelson, 1989; Clos et al., 1990). The interactions between N-terminal leucine zipper motifs are masked under non-heat-shock conditions, possibly by intramolecular contacts with an additional C-terminal leucine zipper motif. Under heat shock conditions the contacts involving the C-terminal zipper are apparently disrupted, thereby allowing trimer formation (Rabindran et al., 1993). The trimerization of HSF increases its affinity for the HSE by a factor of 10^3–10^4 (Wu et al., 1987; S.-J. Kim et al., manuscript in preparation). In addition to DNA binding, a second stage of HSF activation involves the acquisition of transcriptional activity, which is correlated with an increase in the phosphorylation of a number of serine and threonine residues (Sorger, 1990).

Complementary DNAs encoding HSF have been cloned from a variety of organisms including yeast, tomato, chicken, mouse, human, and *Drosophila* (Lis & Wu, 1993; Scharf et al., 1993; Nakai & Morimoto, 1993). The HSF sequences vary widely in composition with the exception of a highly conserved block of 102 amino acids at the N-terminal end of the protein and the aforementioned leucine zipper motifs. Deletion mapping of HSF domains in the yeast HSF protein identified the region important for specific DNA binding as a 118-residue fragment which contains the highly conserved

† This work was supported by the AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health.

‡ National Institute of Diabetes and Digestive and Kidney Diseases.

§ National Cancer Institute.

® Abstract published in *Advance ACS Abstracts*, December 15, 1993.

[1] Abbreviations: 2D, 3D, and 4D, two, three, and four dimensional, respectively; CAP, catabolite gene activator protein; CBCA(CO)NH, 3D ($^{13}C^{\alpha}$,$^{13}C^{\beta}$)–^{15}N–^{1}H spectrum correlating amide ^{1}H and ^{15}N with C^{α} and C^{β} shifts of the preceding residue; CBCANH, 3D ($^{13}C^{\alpha}$,$^{13}C^{\beta}$)–^{15}N–^{1}H spectrum correlating amide ^{1}H and ^{15}N with sequential and intraresidue C^{α} and C^{β} shifts; C(CO)NH, 3D ^{13}C–^{15}N–^{1}H spectrum correlating amide ^{1}H and ^{15}N with side chain ^{13}C shifts of the preceding residue; CT, constant time; dHSF(33–163) *Drosophila melanogaster* heat shock transcription factor residues 33–163; HBHA(CO)NH, 3D ($^{1}H^{\alpha}$,$^{1}H^{\beta}$)–^{15}N–^{1}H spectrum correlating amide ^{1}H and ^{15}N with H^{α} and H^{β} shifts of the preceding residue; HNCA, 3D $^{13}C^{\alpha}$–^{15}N–^{1}H spectrum correlating amide ^{1}H and ^{15}N with sequential and intraresidue C^{α} shifts; HNCO, 3D $^{13}C'$–^{15}N–^{1}H spectrum correlating amide ^{1}H and ^{15}N with the C′ shift of the preceding residue; HNF-3γ, hepatocyte nuclear factor-3γ; HNHA, 3D ^{1}H–^{15}N–^{1}H spectrum correlating amide ^{1}H and ^{15}N with H^{α}; HSE, heat shock element; HSF, heat shock transcription factor; HSQC, heteronuclear single-quantum correlation; NOE, nuclear Overhauser effect; NOESY, 2D NOE spectroscopy.

102-residue sequence block (Wiederrecht et al., 1988). Although some short stretches of HSF showed some similarity with other DNA-binding proteins (Clos et al., 1990; Scharf et al., 1993), an overall sequence comparison between the HSF DNA-binding domain with other sequence-specific factors provided little indication that HSF could be classified with known DNA-binding motifs (Lis & Wu, 1992, 1993). Therefore, it has generally been assumed that HSF would bind to DNA through a novel fold.

Here, we report the results from high-resolution multidimensional heteronuclear NMR studies of the HSF DNA-binding domain from *Drosophila* [dHSF(33–163)]. The polypeptide consists of three α-helices, with the second and third helix separated by an extended turn, and a three-stranded, antiparallel β-sheet. Selective exchange broadening (Dekker et al., 1993) of the ^{15}N–^{1}H correlations indicates that residues in the extended turn and in the third helix are most affected by DNA-binding. Both the secondary structure of the HSF DNA-binding domain and the pattern of residues involved in DNA-binding are reminiscent of the bacterial activator CAP (Weber & Steitz, 1987) and the liver-specific transcription factor HNF-3γ, the prototype of the HNF-3/*forkhead* family of transcription factors (Clark et al., 1993).

EXPERIMENTAL PROCEDURES

Sample Preparation. *Escherichia coli* (BL21/DE3) cells transformed with pHSF33-163 plasmid were grown at 37 °C in M9 minimal media containing 0.3% glucose/0.1% $^{15}NH_4Cl$ or 0.3% [$^{13}C_6$]glucose/0.1% $^{15}NH_4Cl$ in order to obtain ^{15}N- and $^{15}N/^{13}C$-labeled proteins, respectively. pHSF33-163 contains the gene fragment for residues 33–163 of *Drosophila* HSF cloned in the expression vector pJC20 (Clos et al., 1990) and includes three additional C-terminal amino acids, Ala-Ala-Glu, resulting in a total length of 134 amino acids. The total molecular mass for the unlabeled polypeptide is 15 259 Da. Cells were grown in log phase to OD_{600nm} = 0.5, and protein expression was induced for 4 h with 0.6 mM isopropyl β-thiogalactopyranoside. Cells were harvested by centrifugation for 10 min at 6000*g* at 4 °C and resuspended in CB–0.4 M KCl [CB: 10 mM potassium phosphate, pH 6.3, 0.1 mM dithiothreitol, 0.1 mM PMSF (phenylmethanesulfonylfluoride)]. Cells were lysed by sonication at 100 mW for 2 min on ice, repeated twice at 10-min intervals. Debris was removed by centrifugation for 10 min at 6000*g*, 4 °C, and the supernatant was frozen in liquid nitrogen. The crude supernatant (40 mL) was diluted with CB to 0.1 M KCl and applied to a 10-mL heparin-Sepharose CL-6B column (Pharmacia). The column was washed with CB–0.2 M KCl and eluted with CB–0.4 M KCl. Peak fractions were pooled and diluted to 0.05 M KCl with CB and further chromatographed on a 10-mL S-Sepharose Fast Flow column (Pharmacia). After being washed with ca. 10 column volumes of CB–0.1 M KCl, protein was eluted with CB–0.2 M KCl. Sample purity (>95%) was determined by SDS gel electrophoresis and Coomassie Blue staining.

NMR. All NMR experiments were recorded in H_2O solution (5% D_2O), pH 6.3, 10 mM potassium phosphate, and 50 mM KCl, at 27 °C, using a Bruker AMX600 spectrometer equipped with a Bruker triple-resonance probe head and a self-shielded *z*-gradient. Sine-bell-shaped pulsed field gradients (25 G/cm at the center of the sine bell) were generated with an in-house developed pulse shaping unit and amplifier. Experiments were conducted using a 200-μL microcell NMR tube (Shigemi Inc., Allison Park, PA) and a protein concentration in the 0.7–1.1 mM range.

All experiments were recorded without presaturation and were modified to incorporate the WATERGATE solvent suppression scheme described by Piotto et al. (1992) in the last reverse INEPT step. In order to suppress artifacts and reduce the need for phase cycling, pulsed field gradients were incorporated in all experiments in the manner described by Bax and Pochapsky (1992).

The following experiments were recorded on a 1 mM solution of uniformly ^{15}N-labeled (>95%) protein: water-flip-back ^{15}N HSQC (Grzesiek & Bax, 1993a), water-flip-back ^{15}N–{^{1}H}NOE (Grzesiek & Bax, 1993a), water-flip-back ^{15}N-separated NOESY experiment, and HNHA (Vuister & Bax, 1993).

All other 3D experiments were recorded on a 0.7 mM solution of uniformly $^{15}N/^{13}C$-labeled (>95%) protein: CT-HNCO (Grzesiek & Bax, 1992a), CBCA(CO)NH (Grzesiek & Bax, 1992b), CBCANH (Grzesiek & Bax, 1992b), CT-HNCA (Grzesiek & Bax, 1992a), HBHA(CO)NH (Grzesiek & Bax, 1993b), C(CO)NH (Grzesiek & Bax, 1993b; Logan et al., 1992), and water-Roe/exchange (Grzesiek & Bax, 1993c). Details regarding the experimental parameters are provided in supplementary Table 1.

After assignments were completed, additional $^{13}C/^{15}N$-enriched protein became available, and NOE data were double-checked with a 4D $^{15}N/^{13}C$-separated NOESY experiment (Muhandiram et al., 1993) using a 1.1 mM protein concentration.

Data were processed on Sun Sparc workstations using in-house written software (F. Delaglio, unpublished). In the acquisition dimension, all data sets were processed identically. A solvent-suppression filter was applied to the time-domain data (Marion et al., 1989), followed by apodization with a 66°-shifted squared-sine-bell window, zero-filling to the next power of 2, Fourier transformation, and phasing. The data were apodized in t_2 by a 72°-shifted sine-bell window prior to zero-filling to 256*, Fourier transformation, and phasing. Except for the ^{15}N-separated NOESY data sets, the lengths of the nondecaying ^{15}N time-domain data were doubled by mirror-image linear prediction (Zhu & Bax, 1990), apodized by a squared cosine-bell window, zero-filled to 128*, and Fourier transformed. Peak positions were determined by independent parabolic interpolation in all three dimensions using the program PIPP (Garrett et al., 1991).

RESULTS AND DISCUSSION

NMR Experiments. Figure 1 shows the most crowded region of the ^{15}N–^{1}H HSQC spectrum of uniformly ^{15}N labeled dHSF(33–163). The resonance dispersion in the ^{1}H dimension is rather poor due to the high degree of flexibility found for many of the residues in dHSF(33–163) (*vide infra*), resulting in ^{1}H chemical shifts close to their random coil values. However, the dispersion in the ^{15}N dimension is sufficient to resolve most resonances, although some regions with considerable overlap remain. Five correlations are actually composites of overlapping pairs of ^{15}N–^{1}H cross peaks [(Asp^{38}, Trp^{68}), (Ala^{98}, Glu^{126}), (Asp^{125}, Ala^{81}), (Arg^{147}, Asn^{151}), (Asn^{152}, Asn^{154})], whereas two are composites of three overlapping ^{15}N–^{1}H correlations [(Met^{106}, Ala^{40}, Ile^{115}), (Arg^{121}, Phe^{122}, Phe^{129})]. The intensities of the individual ^{15}N–^{1}H correlations vary from very high to vanishingly weak. For five residues of dHSF(33–163), three of which are in a consecutive stretch (Arg^{136}–Ser^{138}), the intensity of the ^{15}N-^{1}H correlation falls below the signal-to-noise threshold. The other two residues, Asp^{58} and Asn^{97}, are both located in stretches with very weak ^{15}N–^{1}H correlations. Very low

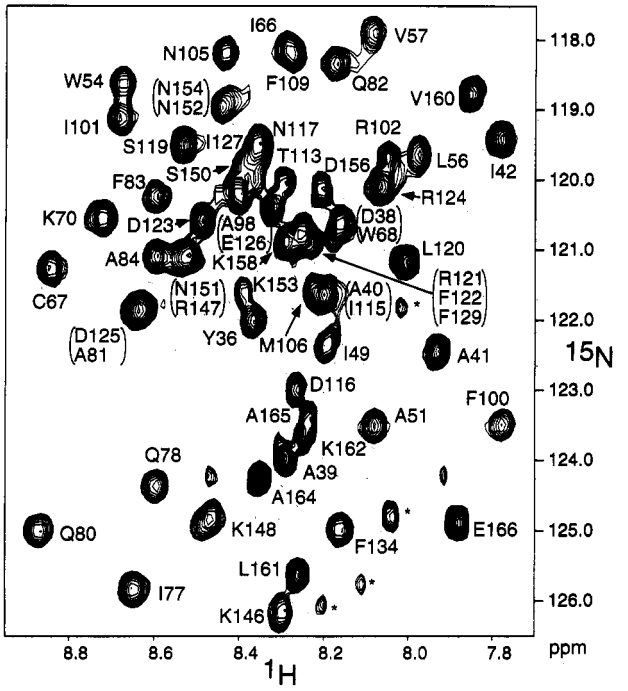

FIGURE 1: Most crowded region of the water-flip-back HSQC spectrum (Grzesiek & Bax, 1993a) of 1 mM solution of uniformly ^{15}N-labeled dHSF(33–163) in ^1H_2O.

intensities are caused by either rapid exchange of the amide proton or slow conformational averaging processes which result in line broadening. Considering that the H_2O magnetization in the water-flip-back HSQC experiment is kept close to its thermal equilibrium value (Grzesiek & Bax, 1993a), hydrogen exchange would affect the ^{15}N–^1H correlation intensity only if the exchange rate is fast compared to the reciprocal of the ^1H line width (i.e., $\gg 30$ s^{-1}). Because the intrinsic amide hydrogen exchange rate for random coil linear peptides at 27 °C, pH 6.3, is slower than 30 s^{-1} (Molday et al., 1972), unusually low intensities in the present HSQC spectrum can be attributed to slow conformational averaging. The weak ^{15}N–^1H correlations and the resonance overlap in the ^{15}N–^1H correlation spectrum are the biggest obstacles in a sequential assignment process which relies primarily on experiments that detect the amide proton.

Sequential assignments were completed through the combined use of the CBCA(CO)NH and CBCANH, HNHA, and HBHA(CO)NH experiments. Strips along the ^{13}C axes of the CBCA(CO)NH, C(CO)NH, and CBCANH experiments, respectively, taken at the ^{15}N–^1H resonance frequencies of Arg^{102}–Asn^{105}, are shown in Figure 1A–D of the supplementary material. If experiments which detect amide protons during acquisition are exclusively used for the sequential assignment procedure, relay of magnetization from the intraresidual and sequential C^β to the amide proton becomes essential for establishing sequential connectivities, since the dispersion of the C^α resonances alone is insufficient. Specifically, for 75 non-glycine residues the C^α chemical shift differs by less than 0.05 ppm from one or more other C^α resonances. Using only the C^α chemical shift for establishing sequential connectivities would rapidly lead to a combinatorial explosion of possible assignments. In contrast, only two pairs of residues show overlap for both the C^α and C^β resonances. In addition to C^β, H^α resonances were also correlated to both the intraresidual and the sequential amide by means of the HNHA and HBHA(CO)NH experiments, respectively.

This sequential assignment procedure becomes ambiguous whenever correlations to C^β and H^α are missing, or when ^1H–^{15}N correlations are indistinguishable at the spectral resolution of the experiment. For 12 residues with weak ^1H–^{15}N correlations (including the five residues for which no ^1H–^{15}N correlation was observed), neither the intraresidue C^β nor intraresidue H^α resonances were observed. In these cases the CT-HNCA experiment was used for establishing potential sequential connectivities since it yields better resolution and sensitivity than the CBCANH experiment. Ambiguities resulting from overlapping ^1H–^{15}N correlations were resolved by mapping longer fragments, in which the residues with overlapping ^1H–^{15}N correlations constitute the end of the fragment, onto the primary sequence. As the C^α and C^β chemical shifts are highly indicative of residue type, they allow for straightforward positioning of such fragments (Grzesiek & Bax, 1993b). The residue types were further confirmed by the C(CO)NH experiment, which correlates all aliphatic side chain carbons with the amide of the next residue, but which is significantly lower in sensitivity. Once completed, assignments were also checked on the basis of sequential NOE connectivities observed in the water-flip-back ^{15}N-separated NOESY experiment and the 4D $^{15}N/^{13}C$-separated NOESY spectrum. The latter spectrum was recorded using a more concentrated protein solution (1.1 mM). However, it is noteworthy to point out that essentially all triple-resonance NMR experiments on this 15-kDa protein were carried out using only 2 mg (i.e., 0.7 mM) of $^{13}C/^{15}N$-labeled material. Backbone, C^β, and H^β assignments are given in Table 2 of the supplementary material.

Secondary Structure. Figure 2 summarizes the amide exchange and ^{15}N–{^1H} heteronuclear NOE results determined from the water-ROE and water-flip-back ^{15}N-{^1H} NOE spectra, respectively, the short- and medium-range NOE data obtained from the water-flip-back 3D ^{15}N-separated NOESY spectrum, and the $^3J_{HNH\alpha}$ values calculated from the HNHA spectrum. Also shown in Figure 2 are the deviations from random coil values of the H^α, C^α, and C^β chemical shifts, which correlate well with secondary structure (Dalgarno et al., 1983; Szilagyi & Jardetzky, 1989; Wishart et al., 1991; Spera & Bax, 1991). The secondary structural elements, as inferred from the NMR data, are indicated in Figure 2 as well.

The amides of residues Met^{33}–Gly^{45}, Thr^{113}–Arg^{124}, and Arg^{147}–Glu^{166} all exhibit fast amide hydrogen exchange and have low values for the ^{15}N–{^1H} heteronuclear NOE. This is indicative of a high degree of internal mobility and non-hydrogen-bonded amide protons. Moreover, H^α, C^α, and C^β resonances of these residues all resonate close to their random coil values, which is further evidence of extensive conformational averaging.

The structured portion of the dHSF(33–163) protein correlates almost exactly with the evolutionarily conserved region, extending from Pro^{47} through Lys^{148}, and comprises three α-helices, a three-stranded antiparallel β-sheet, and the aforementioned flexible loop, Thr^{113}–Arg^{124}. Helix 1 starts at Ala^{48} as judged from the combined presence of strong d_{NN} connectivities, small values of the $^3J_{HNH\alpha}$, negative secondary H^α shifts, and large positive secondary C^α shifts. The amide resonances of residues Trp^{54}–Asp^{61} are weaker than for the remainder of the structured part of the protein, indicating the presence of conformational averaging that is intermediate on the NMR time scale. NOE intensities observed to these amide protons are correspondingly weak, and for most of these no $d_{\alpha N}(i,i+3)$ is observed, although the chemical shift patterns are not inconsistent with a helical conformation. Thr^{62} shows a large value for $^3J_{HNH\alpha}$ (10 Hz), and its H^α and C^α shifts are

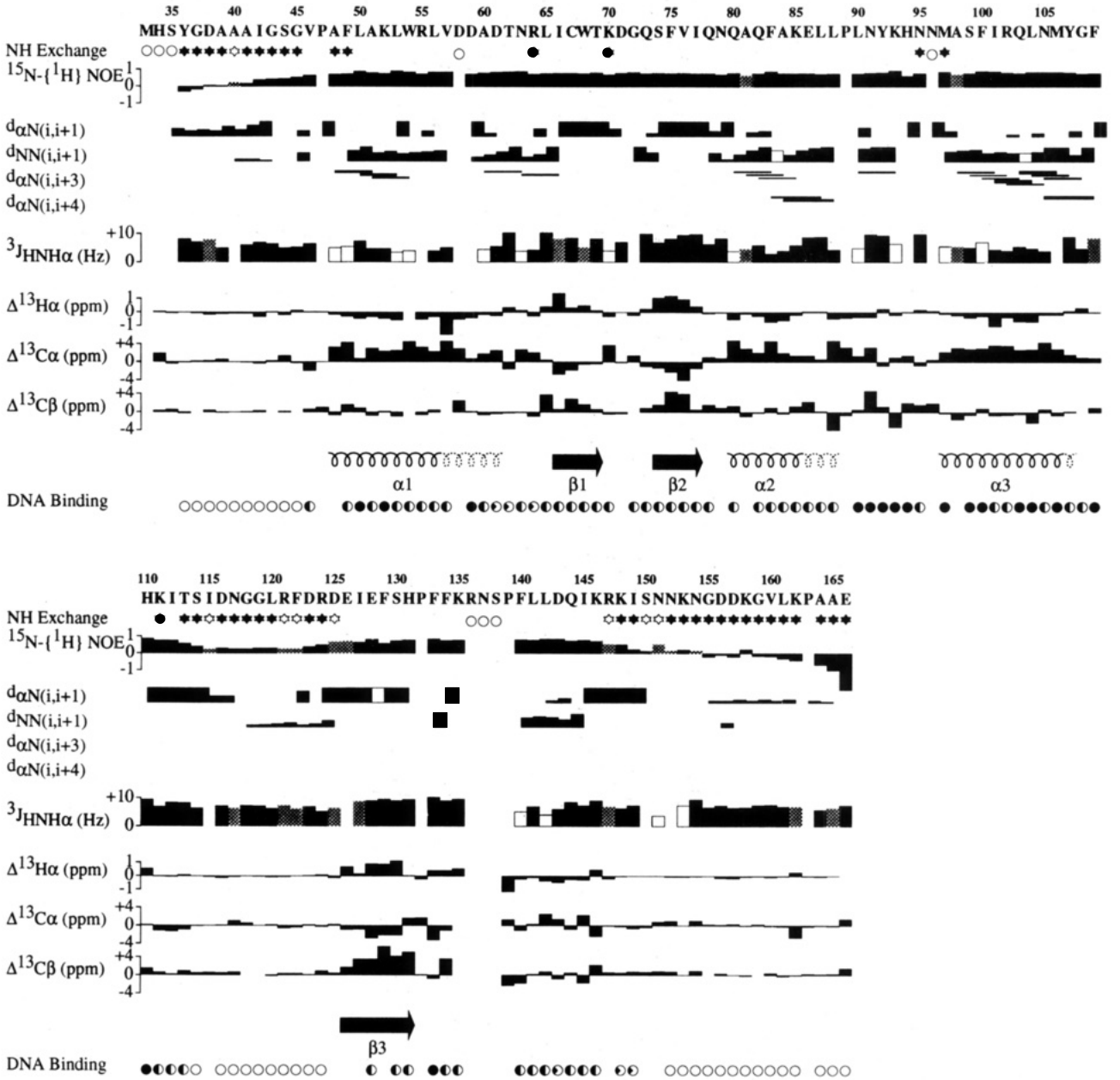

FIGURE 2: Summary of the NMR data obtained for dHSF(33–163) on NH exchange, ^{15}N–{^1H} NOE, sequential and medium-range NOEs involving H^N, $^3J_{HNH\alpha}$ coupling constants, H^α, C^α, and C^β secondary shifts, the secondary structure deduced from these data, and the exchange broadening of ^{15}N–^1H re sonances upon DNA binding. NH exchange: Filled stars indicate ^{15}N–^1H correlations subject to rapid hydrogen exchange ($>\sim 0.5\ s^{-1}$); open stars indicate ambiguity due to overlap; open circles denote absent ^{15}N–^1H correlations, presumably due to slow conformational averaging. ^{15}N–{^1H} NOE: Results from the water-flip-back ^{15}N–{^1H} NOE experiment (Grzesiek & Bax, 1993a). Shaded boxes denote uncertain measurements due to overlap. $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$: Open boxes denote ambiguity due to overlap. $^3J_{HNH\alpha}$: Coupling constants obtained from the HNHA experiment (Vuister & Bax, 1993). Open boxes indicate upper limits based on a cross-peak intensity below the noise threshold. Shaded boxes refer to amides for which the ^{15}N–H^N correlation is (partially) overlapping resulting in a less precise measurement of $^3J_{HNH\alpha}$. DNA binding: Open circles denote residues that do not shift upon DNA binding and which are present at a 0.67:1 DNA:HSF molar ratio. Quarterly filled circles are attenuated at a 0.67:1 DNA:HSF molar ratio. Half-filled circles are attenuated at a 0.25:1 DNA:HSF molar ratio, and filled circles are attenuated beyond detection at a 0.25:1 DNA:HSF molar ratio.

also consistent with a nonhelical arrangement. Two $d_{\alpha N}(i,i+3)$ NOEs between Ala^{60} and Asn^{63} and between Asn^{63} and Ile^{66} suggest a multiple-turn-like structure leading up to the first β-strand.

The presence of strong $d_{\alpha N}(i,i+1)$ NOEs for Ile^{66}–Lys^{70}, together with large values for the $^3J_{HNH\alpha}$ coupling constants and positive H^α and negative C^α secondary shifts are indicative of an extended conformation. A similar pattern is observed for residues Ser^{74}–Ile^{77} and Glu^{126}–His^{131}. Several interstrand $d_{NN}(i,j)$ and $d_{\alpha N}(k,l)$ NOEs could be identified, indicating that these three strands form an antiparallel β-sheet. The topology of the β-sheet and long-range interstrand NOEs are shown in Figure 3.

The first and second strand of the β-sheet are connected by a classic four residue tight turn, with Gly^{72} constituting the third residue. A sequential $d_{NN}(71,72)$ connectivity, expected for type I turns (Wüthrich, 1986), could not be observed in the 3D ^{15}N-separated NOESY spectrum due to the small difference in chemical shifts between the amides of Asp^{71} and Gly^{72}. However, neither a $d_{\alpha N}(71,72)$ connectivity, indicative of a type II turn, nor strong $d_{\alpha N}$ NOE for Asp^{71}, expected for type I′ or II′ turns, were observed, and therefore residues Lys^{70}–Gln^{73} most likely constitute a type I turn.

The second helix starts after the second β-strand at residue Gln^{80} and continues until Pro^{89}. The last three residues of helix 2 appear to deviate from α-helical geometry, based on

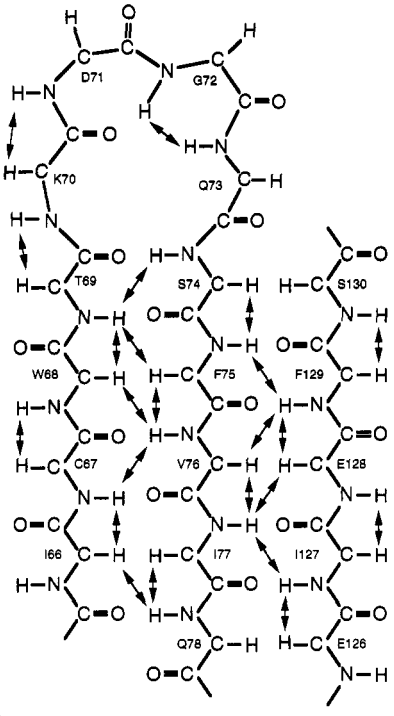

FIGURE 3: Topology of the three-stranded antiparallel β-sheet of dHSF(33–163). Sequential and interstrand NOEs are indicated by arrows.

the absence of $d_{\alpha N}(i,i+3)$ NOEs, the presence of two $d_{\alpha N}(i,i+4)$ NOEs, $^3J_{HNH\alpha}$ values larger than expected for an α-helix, and near-zero H^α and C^α secondary shifts. The presence of strong $d_{NN}(i,i+1)$ NOEs from Leu[90] through Lys[93] and a $d_{NN}(i,i+2)$ NOE between Asn[91] and Lys[93] all support the presence of a turn following Pro[89]. A $d_{\alpha N}(i,i+5)$ NOE between Pro[89] and His[94] and a $d_{\alpha N}(i,i+3)$ NOE between Leu[90] and Lys[93] also point to the presence of an extended turn-like structure. Residues Tyr[92]–Met[97] have very weak ^1H–^{15}N correlations, indicative of conformational averaging on the chemical shift time scale, and a corresponding scarcity of backbone NOEs. The H^α and C^α chemical shifts of His[94]–Asn[96] and the large $^3J_{HNH\alpha}$ value (9 Hz) observed for Asn[95] suggest that the backbone angles of these residues are not in the α-helical region of ϕ/ψ space. The amides of Asn[95] and Met[97] both rapidly exchange with solvent, indicating the absence of hydrogen bonding in this region, just prior to helix 3.

Helix 3 begins at Met[97] and is characterized by a regular pattern of $d_{NN}(i,i+1)$ and $d_{\alpha N}(i,i+3)$ NOEs, small $^3J_{HNH\alpha}$ values, and large positive secondary C^α shifts. Its end is less well-defined: no $d_{\alpha N}(i,i+3)$ NOE is observed for Phe[109], but a relatively intense $d_{\alpha N}(i,i+4)$ NOE is observed between Asn[105] and Phe[109]. A relatively large $^3J_{HNH\alpha}$ value (9 Hz) for Tyr[107] and small positive H^α secondary chemical shifts for Tyr[107] and Gly[108] mark the end of helix 3. Helix 3 is followed by the flexible loop, Thr[113]–Arg[124], which separates it from the last β-strand. Residues Pro[132]–Lys[146] following the last β-strand appear ordered, but secondary shift and NOE data indicate that this region is neither in helical nor sheet conformation.

DNA Binding. As no information is currently available on which residues of dHSF are involved in contacting DNA, we used NMR to probe the residues most affected by DNA binding. In the presence of substoichiometric quantities of DNA, the dHSF(33–163) polypeptide rapidly exchanges between the free and DNA-bound forms. Resonances of amides that experience the largest chemical shift change upon DNA binding experience the most extensive line broadening in the ^1H–^{15}N HSQC spectrum. A similar approach was used for identifying residues involved in the DNA recognition of the POU-specific DNA-binding domain of Oct-1 (Dekker et al., 1993). Changes in the chemical shift have also been used for identifying the site of interaction between IIA^{Glc} and Hpr (Chen et al., 1993).

dHSF(33–163) was titrated with a 13-mer DNA duplex, GGGCAGAACGCCG, containing a single copy of the HSF recognition site, nGAAn. Equilibrium sedimentation and DNase I footprinting studies indicate that dHSF(33–163) binds specifically to this DNA 13-mer with submicromolar affinity (S.-J. Kim, manuscript in preparation). The DNA-binding region is probed by the attenuation of resonances in the ^{15}N HSQC spectrum upon a gradual increase of the DNA/protein molar ratio. The results of this experiment are summarized in Figure 2, where filled circles indicate the largest effect upon DNA binding. The results clearly indicate that the unstructured N- and C-terminal regions of the dHSF-(33–163) protein are not involved in DNA binding. Also, the third unstructured region, Thr[113]–Arg[124], does not interact with the DNA and remains flexible upon DNA binding. In contrast, residues most affected by DNA-binding are concentrated in the extended turn preceding helix 3 and in helix 3 itself. This finding is consistent with spectroscopic studies which indicate fluorescence quenching of Tyr[92] and Tyr[107] upon DNA binding (S.-J. Kim, unpublished results).

The secondary structure of dHSF(33–163) and the pattern of residues involved in DNA binding are similar to those

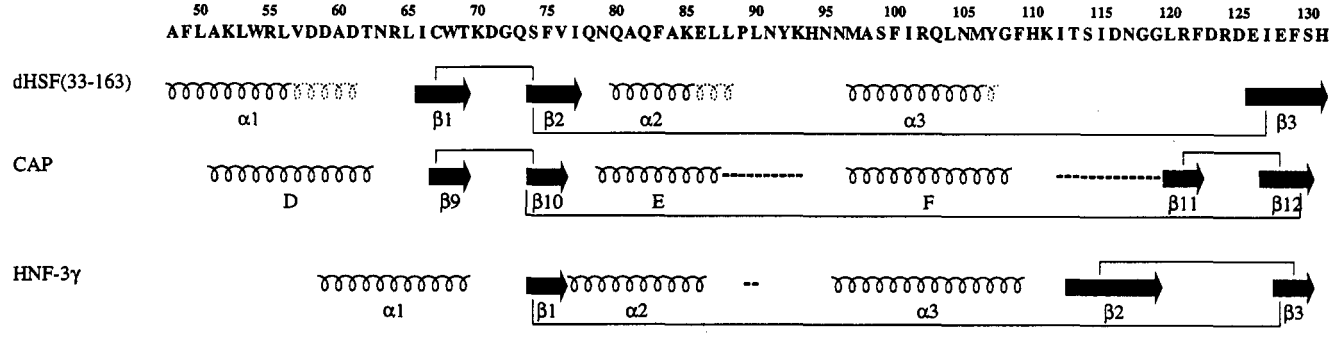

FIGURE 4: Alignment of the secondary structures of the DNA-binding domains of dHSF(33–163), catabolite gene activator protein (CAP) (Weber & Steitz, 1987), and liver-specific transcription factor-3γ (HNF-3γ) (Clark et al., 1993). Insertions for CAP are indicated by dotted lines and were chosen in such a way as to maximize alignment of the secondary structural elements. The two-residue tight turns between β-strands 9 and 10, and β-strands 11 and 12 in CAP (Weber & Steitz, 1987), are commonly referred to as four-residue tight turns in NMR (Wüthrich, 1986). The turns have been extended by two residues, and each strand has been shortened by one residue, accordingly. Dashes in the HNF-3γ sequence refer to residues that are missing when optimizing sequence alignment. The pairing of the β-strands is indicated by connecting lines.

observed for the complexes between DNA and a superclass of sequence-specific DNA-binding proteins exemplified by the *E. coli* catabolite gene activator protein (CAP) (Schultz et al., 1991) and the liver-specific transcription factor HNF-3γ(107–223) (Clark et al., 1993). Figure 4 compares the secondary structure and β-strand pairing for dHSF(33–163) and the analogous regions in CAP and HNF-3γ. The DNA-binding domain of CAP forms a compact three-helix bundle in which the third helix acts as the recognition helix in a helix–turn–helix type DNA-binding motif. The DNA-binding domain of HNF-3γ also consists of a three-helix bundle, binding DNA in a so-called winged-helix motif in which helix3 acts as the recognition helix, with two loops (wings) interacting with more distal parts of the target site. In addition, both CAP and HNF-3γ contain a small antiparallel β-sheet, albeit with a different number of strands. The DNA-binding domains of HNF-3γ and CAP show a similar overall fold and equivalent α-carbons can be superimposed with 1.9-Å rms deviation (Clark et al., 1993).

Our NMR results indicate that the β-strand, helix, extended turn, helix topology of dHSF(33–163) is similar to the topologies of the DNA-binding regions of CAP and HNF-3γ. Sequence alignment shows a 25% identity and 11% similarity for residues 70–113 of dHSF(33–163) and residues 135–181 of HNF-3γ. As in dHSF(33–163), helix 2 of HNF-3γ(107–223) and helix E of CAP are each preceded by a short stretch of β-strand. The end of helix 2 is marked by a proline residue which is absolutely conserved among all known members of the HNF and HSF protein families. In HNF-3γ(107–223) an extended turn also leads up to the recognition helix 3. In contrast, CAP shows a regular three-residue turn separating helix E from the recognition helix F. β-Strands 2 and 3 in dHSF(33–163) appear to be analogous to β-strands 1 and 3 in HNF-3γ(107–223) and β-strands 10 and 12 in CAP. β-Strand 1 of dHSF(33–163) does not have its analog in HNF-3γ(107–223) but is present in CAP (β-strand 9) (Weber & Steitz, 1987), whereas β-strand 11 of CAP and β-strand 2 of HNF-3γ(107–223) are not found in dHSF(33–163). As in CAP, there are two nonhelical residues between the β-sheet and helix 2. Finally, for the fold of dHSF(33–163) to be analogous to the fold of HNF-3γ(107–223) and CAP, β-strand 3 must be in proximity of helix 3. Long-range NOEs between the amides of Ser^{130}, at the end of strand 3, and His^{110}, at the end of helix 3, were indeed observed in the 3D ^{15}N-separated NOESY spectrum of dHSF(33–163). In addition, an independent X-ray study of the DNA-binding domain of yeast HSF has found structural similarities with CAP and HNF-3 (C. Harrison, A. Bohm, and H. Nelson, submitted for publication).

It is interesting that the sequence of helix 3 was previously noted to be similar to the putative recognition helix of bacterial sigma transcription factors (Clos et al., 1990) and that limited sequence similarity to the HNF-3/*forkhead* domain was also recognized (Scarf et al., 1993). Despite these limited sequence similarities our results suggest that the overall topology of the DNA-binding domain of dHSF(33–163) is similar to both the helix–turn–helix proteins and the HNF-3 protein families, with helix 3 of dHSF(33–163) acting as the recognition helix.

ACKNOWLEDGMENT

We thank Joachim Clos for constructing pHSF (33–163), Stephan Grzesiek for assistance with the water-flip-back ^{15}N-{^{1}H}NOE experiment, Dan Garrett for help with Figures 2 and 4, Hillary Nelson for discussion of results prior to publication, and Andy Wang for careful reading of the manuscript.

SUPPLEMENTARY MATERIAL AVAILABLE

One figure showing strips along the ^{13}C axes of the CBCA(CO)NH, C(CO)NH, and CBCANH experiments for residues Arg^{102}–Asn^{105}, one table listing parameters of the NMR experiments used, and one table containing H^N, H^α, H^β, ^{15}N, $^{13}C'$, $^{13}C^\alpha$, and $^{13}C^\beta$ chemical shifts and $^3J_{HNH\alpha}$ coupling constant data (6 pages). Ordering information is give on any current masthead page.

REFERENCES

Bax, A., & Pochapsky, S. S. (1992) *J. Magn. Reson. 99*, 638–643.
Chen, Y., Reizer, J., Saier, M. H., Jr., Fairbrother, W. J., & Wright, P. (1993) *Biochemistry 32*, 32–37.
Clark, K. L., Halay, E. D., Lai, E., & Burley, S. K. (1993) *Nature 364*, 412–420.
Clos, J., Westwood, J. T., Becker, P. B., Wilson, S., Lambert, K., & Wu, C. (1990) *Cell 63*, 1085–1097.
Dalgarno, D. C., Levine, B. A., & Williams, R. J. P. (1983) *Biosci. Rep. 3*, 443–452.
Dekker, N., Cox, M., Boelens, R., Verrijzer, C. P., van der Vliet, P. C., & Kaptein, R. (1993) *Nature 362*, 852–855.
Gething, M.-J., & Sambrook, J. (1992) *Nature 355*, 33–45.
Garrett, D. S., Powers, R., Gronenborn, A. M., & Clore, G. M. (1991) *J. Magn. Reson. 95*, 214–220.
Grzesiek, S., & Bax, A. (1992a) *J. Magn. Reson. 99*, 201–207.
Grzesiek, S., & Bax, A. (1992b) *J. Am. Chem. Soc. 114*, 6291–6293.
Grzesiek, S., & Bax, A. (1993a) *J. Am. Chem. Soc.* (in press).
Grzesiek, S., & Bax, A. (1993b) *J. Biomol. NMR 3*, 185–204.
Grzesiek, S., & Bax, A. (1993c) *J. Biomol. NMR* (in press).
Lis, J. & Wu, C. (1992) *Transcriptional Regulation*, pp 907–930, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Lis, J., & Wu, C. (1993) *Cell 74*, 1–4.
Logan, T. M., Olejniczak, E. T., Xu, R. X., & Fesik, S. W. (1992) *FEBS Lett. 314*, 413–418.
Marion, D., Ikura, M., & Bax, A. (1989) *J. Magn. Reson. 84*, 425–430.
Molday, R. S., Englander, S. W., & Kallen, R. G. (1972) *Biochemistry 11*, 150–158.
Morimoto, R. I., Tissières, A., & Georgopoulos, C., Eds. (1990) *Stress Proteins in Biology and Medicine*, pp 1–36, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Muhandiram, D. R., Xu, G. Y., & Kay, L. E. (1993) *J. Biomol. NMR 3*, 463–470.
Nakai, A., & Morimoto, R. I. (1993) *Mol. Cell. Biol. 13*, 1983–1997.
Piotto, M., Saudek, V., & Sklenar, V. (1992) *J. Biomol. NMR 2*, 661–666.
Rabindran, S. K., Haroun, R. I., Clos, J., Wisniewski, J., & Wu, C. (1993) *Science 259*, 230–234.
Scharf, K.-D., Materna, T., Treuter, E., & Nover, L. (1993) *Plant Promoters and Transcription Factors*, pp 121–158, Springer Verlag, Heidelberg.
Schultz, S. C., Shield, G. C, & Steitz, T. A. (1991) *Science 253*, 1001–1007.
Spera, S., & Bax, A. (1991) *J. Am. Chem. Soc. 113*, 5490–5492.
Sorger, P. K. (1990) *Cell 62*, 793–805.
Sorger, P. K., & Nelson, H. C. M. (1989) *Cell 59*, 807–813.
Szilagyi, L., & Jardetzky (1989) *J. Magn. Reson. 83*, 441–449.

Vuister, G. W., & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772–7777.
Weber, I. T., & Steitz, T. A. (1987) *J. Mol. Biol. 198*, 311–326.
Westwood, J. T., & Wu, C. (1993) *Mol. Cell. Biol. 13*, 3481–3486.
Wiederrecht, G., Seto, D., & Parker, C. S. (1988) *Cell 54*, 841–853.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311–333.
Wu, C., Wilson, S., Walker, B., Dawid, I., Paisly, T., Zimarino, V., & Ueda, H. (1987) *Science 238*, 1247–1253.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Zhu, G., & Bax, A. (1990) *J. Magn. Reson. 90*, 405–410.