Protein Science (1994), *3*:1040–1051. Cambridge University Press. Printed in the USA.

Interaction of the DNA-binding domain of *Drosophila* heat shock factor with its cognate DNA site: A thermodynamic analysis using analytical ultracentrifugation

SOON-JONG KIM,[1] TOSHIO TSUKIYAMA,[1] MARC S. LEWIS,[2] AND CARL WU[1]

[1] Laboratory of Biochemistry, National Cancer Institute, and [2] Biomedical Engineering and Instrumentation Program, National Center for Research Resources, National Institutes of Health, Bethesda, Maryland 20892

(RECEIVED January 5, 1994; ACCEPTED April 12, 1994)

Abstract

Heat shock transcription factor (HSF) mediates the activation of heat shock genes by binding to its cognate sites with high affinity and specificity. The high-affinity binding of HSF is dependent on the formation of an HSF homotrimer, which interacts specifically with the heat shock response element (HSE), comprised of 3 inverted repeats of the 5-bp sequence NGAAN. In order to investigate the thermodynamic basis of the interaction between HSF and HSE, we have overexpressed and purified a polypeptide (dHSF(33–163)) encompassing only the DNA-binding domain of HSF from *Drosophila* and analyzed its binding to DNA by equilibrium analytical ultracentrifugation using a multiwavelength scan technique. We demonstrate that dHSF(33–163) can bind as a monomer with 1:1 stoichiometry to a synthetic 13-bp DNA containing a single NGAAN sequence. The values of the thermodynamic parameters obtained from the temperature dependence of the equilibrium binding constants indicate that the changes of free energy for the binding of dHSF(33–163) to the wild-type site and a mutant DNA site are predominantly characterized by substantial negative changes of enthalpy. Binding to the wild-type DNA is characterized by a significant positive change of entropy, whereas binding to the mutant DNA is distinguished by a negative change of entropy of comparable magnitude. The binding to the mutant DNA was also highly sensitive to increasing salt concentrations, indicating a dominance of ionic interactions. The sequence-specific, 1:1 binding of dHSF(33–163) to the NGAAN sequence provides a basis for the analysis of higher order interactions between HSF trimers and the HSE.

Keywords: HSE; HSF; protein-DNA interactions; ultracentrifugal analysis

The transcriptional induction of heat shock genes in eukaryotes is initiated by the binding of the transcription factor heat shock factor (HSF) to its cognate site, the heat shock element (HSE), which is composed of 3 contiguous repeats of the 5-bp sequence NGAAN arranged in alternating orientation: 5′-[NGAAN][NTTCN][NGAAN]-3′. The high-affinity binding of HSF to a complete HSE is dependent on the formation of an HSF homotrimer (for a review, see Lis & Wu, 1993). Although the exact stoichiometry of the interaction between HSF subunits and the HSE has not been determined, each subunit of the HSF trimer is generally assumed to recognize 1 of the 3 NGAAN repeats of the HSE. Thus, the high-affinity binding of HSF is thought to result from the combined interactions of the DNA-binding domains of all 3 HSF subunits.

Sequence analysis of cDNA clones encoding HSFs from a wide variety of species reveals 2 conserved regions in the N-terminal part of the HSF protein that represent the DNA-binding domain and trimerization domain. The overall sequence of the DNA-binding domain does not show extensive similarity to any known category of DNA-binding motif, and it was therefore anticipated that a determination of the structure of HSF should reveal a new motif for specific DNA recognition. However, recent X-ray and NMR studies reveal that, despite the lack of overall sequence similarity, there is a structural relatedness of the DNA-binding domain of HSF to DNA-binding motifs exemplified by the helix-turn-helix and HNF-3 protein families (Harrison et al., 1994; Vuister et al., 1994).

The trimerization domain of HSF is situated next to the DNA-binding domain and is separated by a short linker. The con-

Reprint requests to: Carl Wu, Building 37, Room 4C09, National Institutes of Health, Bethesda, Maryland 20892, or Marc S. Lewis, Building 35, Room B101C, National Institutes of Health, Bethesda, Maryland 20892.

served residues in this region essentially define several arrays of hydrophobic heptad repeats or coiled-coil motifs (Sorger & Nelson, 1989; Clos et al., 1990). Physicochemical studies of HSF trimerization domains synthesized as ~90-residue polypeptides indicate that they form stable homotrimers in solution of predominantly α-helical character, suggesting a triple-stranded α-helical coiled-coil structure (Peteranderl & Nelson, 1993). How the 3 DNA-binding domains of the HSF trimer interact with the linear HSE sequence is unclear at present.

All 3 NGAAN boxes of the HSE are required for high-affinity binding to the HSF trimer, although moderate binding can be detected using an incomplete HSE carrying 2 contiguous repeats (Perisic et al., 1989). Because both the head-to-head (→ ←); ([NGAAN][NTTCN]) configuration and the tail-to-tail (← →); ([NTTCN][NGAAN]) configuration possess similar binding affinities for HSF, it has been suggested that the NGAAN box is the fundamental unit of recognition by the HSF trimer and that a minimum of two 5-bp repeats is required for a stable interaction (Perisic et al., 1989). Unlike these studies of the HSF trimer, which were carried out using footprinting and electrophoretic mobility shift assays, the interaction of the latent, monomeric form of HSF with DNA has not been investigated. This is due in part to the inability to purify the monomeric form of HSF from nonheat-shocked cells and to the tendency of HSF expressed in *Escherichia coli* to assemble as trimers constitutively, even at low growth temperatures (Clos et al., 1990; Rabindran et al., 1991). Therefore, in order to elucidate the interaction of monomeric HSF with its cognate site in the absence of knowledge of a physiological mechanism to convert HSF trimers to monomers, we have utilized recombinant DNA techniques to engineer the expression of only the HSF DNA-binding domain. Accordingly, a polypeptide encompassing residues 33–163 of HSF from *Drosophila* (dHSF(33–163)) was expressed in *E. coli* and purified to homogeneity.

By means of equilibrium analytical ultracentrifugation we show that dHSF(33–163) exists as a monomer in solution. We have determined the equilibrium binding constants for the interaction between dHSF(33–163) and a synthetic 13-bp DNA containing a single, centrally located NGAAN sequence, as well as the same DNA sequence in which the genetically critical G nucleotide at position 2 of the NGAAN sequence was changed to a C nucleotide. By appropriate analysis of data, we have been able to calculate the values of the thermodynamic parameters characterizing the equilibrium between dHSF(33–163) and the wild-type and mutant DNA sites. The effects of monovalent salt (K^+) on the equilibrium constants are also discussed. The recently developed multiwavelength scan method (Lewis et al., 1994) was used to analyze the equilibrium sedimentation data and to obtain the values of the natural logarithms of the association constants for the protein–DNA interactions.

Results

Purification and mass analysis of dHSF(33–163)

The dHSF(33–163) polypeptide is comprised of residues 33–163 of *Drosophila* HSF, with 3 additional amino acids (Ala-Ala-Glu) introduced at the C-terminal end; the addition is caused by cloning the cDNA insert into a bacterial expression vector. Residues 33–163 of *Drosophila* HSF encompass the DNA-binding domain of HSF (residues 47–148) that is highly conserved among a wide variety of animal, plant, and fungal species (Scharf et al., 1993). The structure of dHSF(33–163) is under investigation by multidimensional NMR, and it is notable that the structured part of the polypeptide is coincident with the highly conserved domain (Vuister et al., 1994). As shown in Figure 1, the molecular mass of dHSF(33–163), as analyzed by mass spectrometry, is 15,263 ± 1 Da, in excellent agreement with the molecular mass of 15,259 Da calculated from the predicted amino acid sequence. The homogeneity of the preparation after chromatography on heparin-Sepharose CL-6B and S-Sepharose was found to be in excess of 99% by SDS-PAGE and Coomassie blue staining (Fig. 1 inset).

Functional activity of dHSF(33–163): DNase I footprinting

The functional activity of dHSF(33–163) was analyzed by means of a DNase I footprinting assay. As shown in Figure 2, dHSF(33–163) is capable of binding to a natural hsp70 promoter containing 2 adjacent HSEs (5 complete NGAAN boxes are indicated by arrows and 2 incomplete boxes are indicated by dots). This result clearly demonstrates that the recombinant polypeptide possesses sequence-specific binding activity. The footprint of the full-length *Drosophila* HSF trimer expressed in *E. coli* and purified as described (Clos et al., 1990) is also shown for comparison. It is worthwhile to note that the extent of the footprints of the full-length trimer and of dHSF(33–163) are essentially identical. This result indicates that, despite the absence of the remainder of the protein, the interaction between dHSF(33–163) and the HSE as probed with DNase I remains unchanged. However, other properties of the interaction appear to be different. As the protein concentration of the full-length HSF trimer is increased stepwise, the footprints clearly show cooperative binding to the 2 adjacent HSEs, as previously reported (Topol et al., 1985; Xiao et al., 1991). In contrast, such cooperativity between HSEs is absent from the footprints of dHSF(33–163), although modest cooperation can be observed for the binding

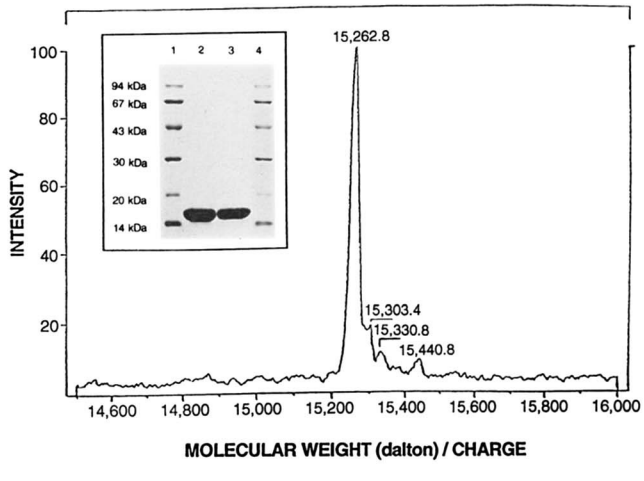

Fig. 1. Mass spectrum of purified dHSF(33–163). **Inset:** SDS-PAGE analysis of purified dHSF(33–163) at 2 different loading concentrations. Lanes 1 and 4, molecular weight markers: 94 kDa, phosphorylase b; 67 kDa, bovine serum albumin; 43 kDa, ovalbumin; 30 kDa, carbonic anhydrase; 20 kDa, soybean trypsin inhibitor; 14 kDa, α-lactalbumin. Lanes 2 and 3, dHSF(33–163).

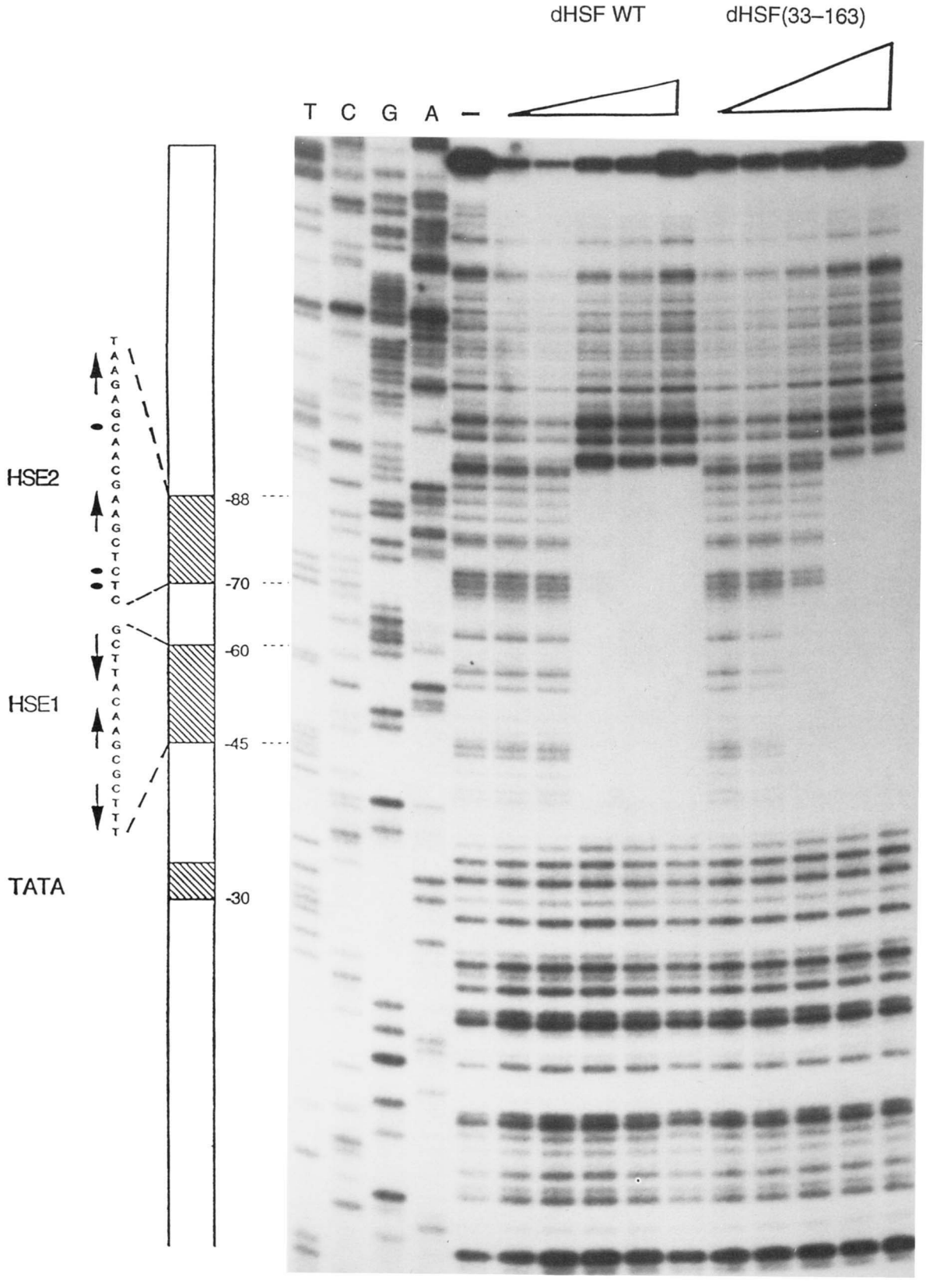

Fig. 2. DNase I footprinting patterns of full-length *Drosophila* HSF (HSF-WT) and of dHSF(33–163). Each lane has ~20 fmol of DNA. Sample volumes of HSF-WT (~0.5 mg/mL) were 0.1, 0.2, 0.4, 1, and 2 μL; sample volumes of the dHSF(33–163) (~15 mg/mL) were 0.1, 0.2, 0.4, 1.0, and 2 μL incubated with DNA in a final volume of 20 μL. The HSE1 and HSE2 positions along with the TATA sequence are shown as shaded boxes. The DNA sequences containing a complete HSE unit (NGAAN) are indicated as arrows, and the incomplete HSE sequences are shown as dots.

to HSE1, which contains 3 complete NGAAN boxes, as compared with HSE2, which has only 2 complete boxes (Fig. 2). Therefore, the sequences responsible for cooperative interaction between HSF trimers are located outside residues 33–163, whereas the slight cooperativity for the binding to HSE1 appears to lie within the DNA-binding domain.

Equilibrium sedimentation analysis of dHSF(33–163) and DNA alone

The associative state of dHSF(33–163) and the homogeneity of the double-stranded 13-bp synthetic DNAs were evaluated by analytical ultracentrifugation. For a homogeneous, thermodynamically ideal solute, the concentration as a function of radial position at equilibrium is given by

$$A_{r,i} = A_{b,i} \exp[\mathrm{A}_i M_i (r^2 - r_b^2)], \quad (1)$$

where $A_{r,i}$ is the concentration, expressed as an absorbency, at the radial position r, and $A_{b,i}$ is the concentration at the cell bottom of either the polypeptide or the DNA; M_i refers to M_P and M_N, the values of the molar mass of the protein and nucleic acid, respectively; $\mathrm{A_P}$ and $\mathrm{A_N}$ (not to be confused with A, the absorbencies) are defined by

$$\mathrm{A}_i = (\partial\rho/\partial c)_\mu \omega^2/2RT, \quad (2)$$

where ω is the rotor angular velocity, R is the gas constant, T is the absolute temperature, and $(\partial\rho/\partial c)_\mu$ is the derivative of solution density with respect to solute concentration at constant chemical potential. Although $(\partial\rho/\partial c)_\mu$ is essentially equal to $(1 - \bar{v}^0\rho)$, where $\bar{v}^0$ is the partial specific volume at the limit as the solute concentration approaches 0 and ρ is the solvent density, we prefer to use $(\partial\rho/\partial c)_\mu$, the thermodynamically more appropriate term (Durchschlag, 1986) in this context. Because Equation 1 can be derived using the principles of reversible thermodynamics, and because the values of M_i are known, we can obtain the experimental values of $(\partial\rho/\partial c)_\mu$ as fitting parameters when Equations 1 and 2 are used as a mathematical model for fitting the concentration distributions of polypeptide and DNA.

Figure 3A and B shows the sedimentation equilibrium concentration distributions of the polypeptide alone and DNA alone, respectively, fit at multiple wavelengths as required for multiwavelength analysis. Figure 3A (inset) also shows the UV spectra of the protein alone, the DNA alone, and a 1:2 DNA:dHSF(33–163) mixture. These spectra show the wavelength region from 230 to 240 nm where the multiple wavelength scans were taken. This range of wavelengths was chosen because it covers a region of protein absorption dominance to a region of DNA absorption dominance in the total observed absorption. Additionally, because the absorption of polypeptide in this range is characteristic of the amino acid backbone, and because this range of wavelengths is well away from the DNA absorption peak at 260 nm, the analysis of the data is less complicated because hypo- or hyperchromic effects are either absent or minimal.

In Figure 3A, the concentration distributions of the dHSF(33–163) polypeptide alone at different wavelengths are fit by a thermodynamically ideal monomeric species with a molecular mass, calculated from the predicted amino acid sequence, of

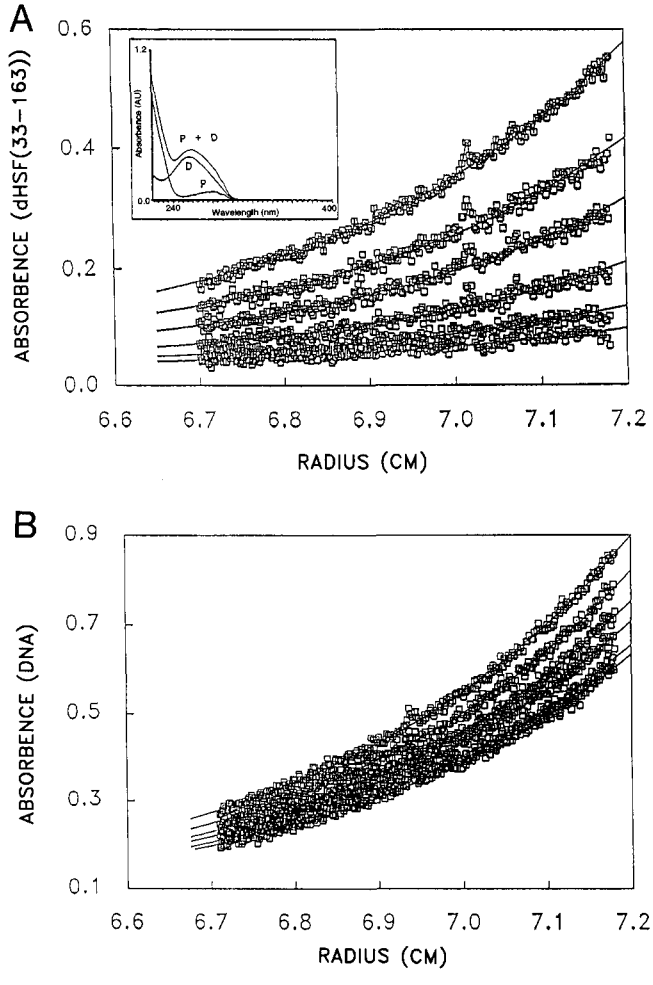

Fig. 3. A: Equilibrium sedimentation distributions of absorbencies of dHSF(33–163) at different wavelengths. The loading concentration was 7.3 μM. The scanned wavelengths are, from top to bottom, 230, 232, 234, 236, 238, and 240 nm. The fitted lines are for a simultaneous fit of the polypeptide as a monomer with a molar mass of 15,259 Da. The reference concentration and the baseline error term were fit as local parameters, whereas the value of $(\partial\rho/\partial c)_\mu$ for the dHSF(33–163) was obtained as a global parameter. **Inset:** Absorption spectra of dHSF(33–163) (P), 13-bp wild-type DNA (D), and a 2:1 molar ratio mixture of both (P + D). **B:** Equilibrium sedimentation distributions of absorbencies of 13-bp wild-type DNA at different wavelengths. The loading concentration was 3.7 μM. The scanned wavelengths are, from bottom to top, 230, 232, 234, 236, 238, and 240 nm. The fitted lines are for a simultaneous fit of the double-stranded DNA with a molar mass of 8,011 Da.

15,259 Da. Figure 3B illustrates the concentration distributions of a 13-bp double-stranded oligonucleotide, 5′-GGGCAGAA CGCCG-3′, in 0.05 M KCl, 10 mM potassium phosphate buffer, pH 6.3, fit with a molecular mass, calculated from the base-pair composition, of 8,011 Da. This wild-type DNA contains a single 5-bp module comprising the fundamental NGAAN box located at the center. In addition, an A nucleotide was chosen for position 1 in the NGAAN box in view of recent evidence for its biological importance (Fernandes et al., 1994).

The results of the multiwavelength fits of the data for both the polypeptide and the DNA show that the chosen models are appropriate. These results confirm that the dHSF(33–163) is a monomer in solution and that the wild-type HSE is a 13-bp

double-stranded DNA with no detectable single-stranded DNA contamination. The data for the mutant 13-bp DNA, 5′-GGG CACAACGCCG-3′, which contains a single nucleotide substitution (NGAAN to NCAAN) were also of similar quality (data not shown). The monomeric state of dHSF(33–163) has been confirmed over the entire salt concentration range studied (0.05–0.2 M KCl, data not shown).

Because determination of the values of the thermodynamic parameters characterizing the interaction of polypeptide and DNAs requires measurements at a number of temperatures, it is necessary to know the values of $(\partial\rho/\partial c)_\mu$ as a function of temperature. Thus, at each temperature, multiwavelength data sets were jointly fit with the values of A_b, the absorbance at the cell bottom, as a parameter local to each data set and the value of $(\partial\rho/\partial c)_\mu$ as a parameter global to all the data sets. The temperature dependence of $(\partial\rho/\partial c)_\mu$ was well fit with a power series having the form

$$(\partial\rho/\partial c)_{\mu,T} = (\partial\rho/\partial c)_{\mu,0} + a_1 T + a_2 T^2, \quad (3)$$

where T is the Celsius temperature and $(\partial\rho/\partial c)_{\mu,0}$, a_1, and a_2 are fitting parameters. As shown in Figure 4A and B, the values of $(\partial\rho/\partial c)_\mu$ for the DNA and the dHSF(33–163) polypeptide exhibit temperature dependence. The base-pair composition shows some effect on the value of $(\partial\rho/\partial c)_\mu$; other HSE DNAs, which have different base-pair compositions, have different values (data not shown). Accordingly, these values must be considered as valid only for this DNA in the buffer used here.

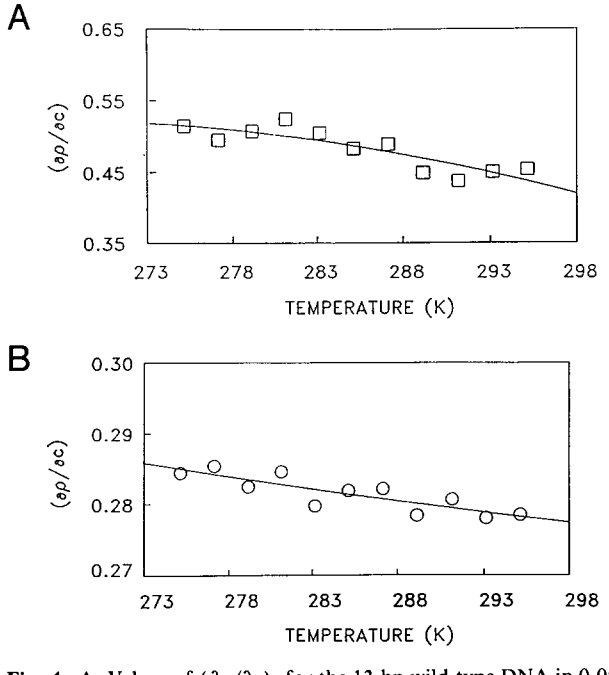

Fig. 4. A: Values of $(\partial\rho/\partial c)_\mu$ for the 13-bp wild-type DNA in 0.05 M KCl with 10 mM potassium phosphate buffer, pH 6.3, and 0.1 mM EDTA as a function of temperature calculated from concentration distributions at equilibrium at a rotor speed of 14,000 rpm. The loading concentration was 3.7 μM. **B:** Values of $(\partial\rho/\partial c)_\mu$ for the dHSF(33–163) polypeptide in 0.05 M KCl with 10 mM potassium phosphate buffer, pH 6.3, and 0.1 mM EDTA as a function of temperature calculated from concentration distributions at equilibrium at a rotor speed of 14,000 rpm. The loading concentration was 7.3 μM.

The composition of this particular 13-bp DNA is 77% G·C and 23% A·T.

The value of A_b, the absorbance at the cell bottom obtained at the wavelength λ, may be used to obtain the values of E_λ, the molar extinction coefficient at that wavelength. Because the molar extinction coefficients are known at 280 nm for the polypeptide and at 260 nm for the DNA, the extinction coefficient at any wavelength λ is $E_\lambda = E_{280} \cdot A_{b,\lambda}/A_{b,280}$ for the polypeptide and $E_\lambda = E_{260} \cdot A_{b,\lambda}/A_{b,260}$ for the DNA. Extinction coefficients obtained in this manner are optimal because they are obtained with the XL-A monochromater at exactly the same wavelength for the polypeptide, the DNA, and the complex. Extinction coefficients obtained with a spectrophotometer must be validated by satisfying these conditions if optimal results are to be obtained. This was the case for data obtained here.

Equilibrium sedimentation analysis of dHSF(33–163):DNA complexes

Figure 5 shows scans of the equilibrium distributions of the complexes of the wild-type (square) and mutant (circle) DNAs and dHSF(33–163) measured at 230 nm and 240 nm at chemical and centrifugal equilibrium. The data sets measured at the other wavelengths (232, 234, 236, and 238 nm) are not shown in order to simplify the figure. On the basis of observable differences between the concentration gradients, the wild-type DNA: dHSF(33–163) complex shows a stronger interaction than that of the mutant DNA:dHSF(33–163) complex. The data in Figure 5 represent the distribution of the absorbances of the mixtures before the conversion to molar concentrations as described below.

Multiwavelength analysis of dHSF(33–163):DNA complexes

The multiwavelength analysis used here is described in detail in a recent publication (Lewis et al., 1994) and will be discussed only briefly here. The method requires constructing 2 matrices. The first matrix, denoted $\boldsymbol{A}$, is an n by m matrix of absorbencies, where n is the number of radial positions and m is the number of wavelengths; this is constructed from the radius and wavelength-dependent absorbency data from the polypeptide–DNA mixture. The second matrix, denoted $\boldsymbol{E}$, is a 2 by m matrix of extinction coefficients; the rows represent the polypeptide and the DNA, respectively; the columns represent the same wavelengths as the columns of the $\boldsymbol{A}$ matrix. These matrices are then used to calculate an n by 2 matrix $\boldsymbol{C}$, using $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{E}^+$, where $\boldsymbol{E}^+$ is the Moore–Penrose pseudoinverse of $\boldsymbol{E}$ (Strang, 1986). The 2 columns of $\boldsymbol{C}$ contain total molar concentrations, uncomplexed and complexed, of polypeptide and nucleic acid, respectively, and each column of $\boldsymbol{C}$ is a function of r. The values of $\boldsymbol{C}$ in a given row represent the linear least-squares fit of the values of $\boldsymbol{A}$ in a corresponding row. The matrix $\boldsymbol{C}$, the dependent variables, is concatenated with a vector of the radial positions, the independent variable, to form a 3-column data matrix suitable for analysis by nonlinear, least-squares curve fitting.

The mathematical models for fitting the molar concentration data matrix are the equations

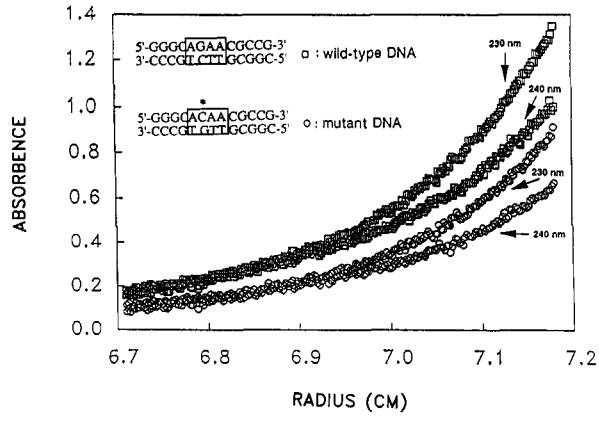

Fig. 5. Equilibrium distributions of the absorbencies of mixtures of wild-type DNA and dHSF(33-163) (squares) and mutant DNA and dHSF(33-163) (circles) in a 1:2 molar ratio at equilibrium at 14,000 rpm at 10 °C in pH 6.3, 10 mM potassium phosphate buffer containing 0.05 M KCl and 0.1 mM EDTA. The loading concentrations of DNA and dHSF(33-163) were 3.7 and 7.3 μM, respectively. The sequences of wild-type and mutant 13-bp DNAs are also shown. The data were collected at wavelengths from 230 to 240 nm in 2-nm increments, but, in order to avoid confusion, only the data at 230 and 240 nm are shown. The corresponding wavelengths are indicated with arrows.

$$C_{r,P} = C_{b,P}\exp[A_P M_P(r^2 - r_b^2)] + C_{b,P}C_{b,N} \times \exp[\ln K_a + (A_P M_P + A_N M_N)(r^2 - r_b^2)] + \epsilon_P \quad (4)$$

$$C_{r,N} = C_{b,N}\exp[A_N M_N(r^2 - r_b^2)] + C_{b,P}C_{b,N} \times \exp[\ln K_a + (A_P M_P + A_N M_N)(r^2 - r_b^2)] + \epsilon_N. \quad (5)$$

The global analysis then involves jointly fitting columns 1 and 2 with Equation 4 and columns 1 and 3 with Equation 5. The value of $\ln K_a$, the natural logarithm of the molar equilibrium constant, and the values of $C_{b,P}$ and $C_{b,N}$, the concentrations of uncomplexed polypeptide and nucleic acid at the radius of the cell bottom, r_b, are fitting parameters global to both equations, while the values of ϵ_P and ϵ_N, the small baseline error terms, are fitting parameters local to Equations 4 and 5, respectively. The form of these equations is such that the system is very highly constrained and the fitting is a well-conditioned problem. In contrast, attempting to obtain a value of $\ln K_a$ by analysis of data obtained at a single wavelength is an ill-conditioned problem and is usually not successful unless the method of implicit constraints is used (Lewis, 1991). However, this method has the experimentally stringent requirement of conservation of mass of the solute in the ultracentrifuge cell, and, from a computational point of view, is as complex as the method described here. The $\boldsymbol{A}$ and $\boldsymbol{E}$ matrices for 2 wavelengths represent the minimum set that can be solved for $\boldsymbol{C}$ and in this case only, gives an exact solution for that set of values. However, the effect of relatively small errors in the extinction coefficients on the concentration values can be magnified if only 2 wavelengths are used; the use of multiple wavelengths, even with normally distributed error as great as 10% in the extinction coefficients, significantly reduces this problem. Simulation studies and practical experience indicate that data from 5 or more wavelengths are needed to attain good reliability for the values in $\boldsymbol{C}$ (Lewis et al., 1994). Thus, although it is possible to do the analysis with scans at only 2 wavelengths, doing so gives up the advantages of obtaining molar concentrations that are the result of linear least-squares fits of multiwavelength scans. There are additional advantages in using the multiwavelength method: because the pseudoinverse is a linear transform, normally distributed error in the absorbencies becomes normally distributed error in the molar concentrations; because the pseudoinverse is fitting the absorbencies for the molar concentrations, even non-normally distributed error in the absorbencies tends to become more normally distributed error in the molar concentrations.

Figure 6A shows the simultaneous fit of Equations 4 and 5 to molar concentration distributions of dHSF(33-163) plus DNA:dHSF(33-163) complex and DNA plus DNA:dHSF(33-163) complex at 10 °C. For this wild-type DNA:dHSF(33-163) interaction, a value of 17.63 ± 0.98 was obtained for $\ln K_a$ at that temperature. Figure 6B shows the distribution of the residuals about the fitting lines. The same procedure was applied to data for the mutant DNA:dHSF(33-163) interaction, and the results are shown in Figure 6C. A value of 14.20 ± 0.11 for $\ln K_a$ was obtained at the same temperature. Figure 6D shows the distribution of the residuals about the fitted lines. The distributions are quite adequately random with no significant systematic deviations, demonstrating not only the quality of the fit, but also that a 1:1 stoichiometry appears to be an appropriate model for the interaction. The value of $\ln K_a$ for the wild-type species agrees well with those obtained using a fluorescence titration assay (Kim et al., in prep.). This agreement is a cross-validation of the value and also serves to demonstrate the reliability of the multiwavelength scan technique for the study of protein-DNA interactions. The reliability of this method was also confirmed by Monte Carlo simulation studies (Lewis et al., 1994). All of the data measured at other temperatures (see below) are comparable in quality to the data shown in Figure 6. It should be noted that this method of analysis gives very low standard errors for $\ln K_a$ for values of $\ln K_a$ up to 16 for this case. Simulation studies show that the standard errors rise significantly at values of $\ln K_a$ above this, but are acceptable up to a value of 18. At values of $\ln K_a$ of 18 and above, it is necessary to perform Monte Carlo analyses in order to obtain the standard errors. When $\ln K_a$ exceeds a value of 21, the concentrations of free reactants are so low that the obtained values of $\ln K_a$ are questionable. Under such circumstances it would be advisable to work in the presence of higher salt concentrations, which would have the desirable effect of reducing the strength of the association.

Analysis of thermodynamic parameters characterizing the interaction of wild-type and mutant DNA binding to dHSF(33-163)

Analytical ultracentrifugation has the advantage that it is rigorously based upon reversible thermodynamics; the reactants and the product of an interaction each have a uniquely defined concentration gradient that can be resolved by appropriate mathematical analyses to give the desired value of $\ln K_a$, and hence ΔG^0, with a minimum of assumptions. In order to determine the values of $\ln K_a$ for the polypeptide-DNA interactions at different temperatures, data sets were collected at 2° intervals from 2 to 18 °C and converted to molar concentration distributions. The data were then analyzed using Equations 4 and 5 as mathematical models to obtain the values of $\ln K_a$ as be-

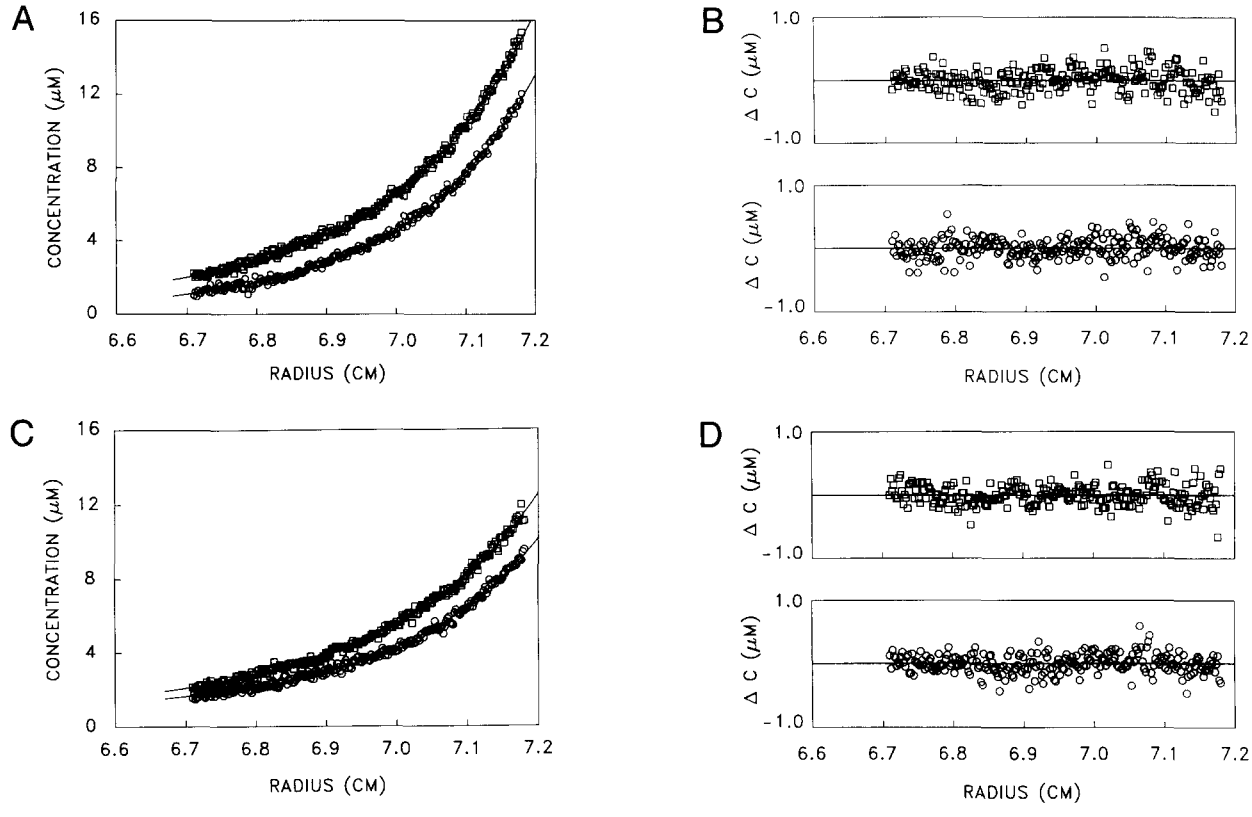

Fig. 6. A: Equilibrium sedimentation distributions of the molar concentrations for the interaction of wild-type 13-bp DNA and dHSF(33-163) loaded in a 1:2 molar ratio at equilibrium at 14,000 rpm and 10.0 °C. The squares are for dHSF(33-163) plus wild-type DNA:dHSF(33-163) complex; the circles are for wild-type DNA plus wild-type DNA:dHSF(33-163) complex. **B:** Distributions of the residuals for the wild-type DNA and dHSF(33-163) interaction shown in A. The data symbols are the same as those in A. The joint fit RMS error is 3.54×10^{-7} M. **C:** Equilibrium sedimentation distributions of the molar concentrations for the interaction of mutant 13-bp DNA and dHSF(33-163) loaded in a 1:2 molar ratio at equilibrium at 14,000 rpm and 10.0 °C. The squares are for dHSF(33-163) plus mutant DNA:dHSF(33-163) complex; the circles are for mutant DNA plus mutant DNA:dHSF(33-163) complex. **D:** Distributions of the residuals for the mutant DNA and dHSF(33-163) interaction shown in C. The data symbols are the same as those in C. The joint fit RMS error is 3.37×10^{-7} M.

fore. The temperature dependence of the binding constants for the mutant and the wild-type DNA interaction with dHSF(33-163) are summarized in Table 1. Throughout the temperature range studied, the binding of wild-type DNA to dHSF(33-163) is ~20-45-fold greater than the binding to mutant DNA.

In order to obtain the thermodynamic parameters for the interaction of the mutant and the wild-type DNA to dHSF(33-163), the temperature dependence of the values of $\ln K_a$ for both DNAs were analyzed by the method of Clarke and Glew (1966) using a Taylor series expansion about the reference temperature θ:

$$R \ln K_a = -\Delta G^0_\theta/\theta + \Delta H^0_\theta(1/\theta - 1/T) + \Delta C^0_{p,\theta}[\theta/T - 1 + \ln(T/\theta)], \quad (6)$$

where ΔG^0_θ is the standard free-energy change, ΔH^0_θ is the standard enthalpy change, and $\Delta C^0_{p,\theta}$ (defined as $(\partial \Delta H^0_\theta/\partial T)_p$) is the standard heat capacity change of binding at constant pressure, all at the reference temperature θ. This form of the expression describing the temperature dependence of equilibrium constants has the desirable property that the fitting parameters ΔG^0_θ, ΔH^0_θ, $\Delta C^0_{p,\theta}$, and $(d\Delta C^0_{p,\theta}/dT)_\theta$, if used, are independent of each other. Fitting for these parameters is easier with the Clarke and Glew equation than with the conventional van't Hoff plot, particularly if $\Delta C^0_{p,\theta}$ and $(d\Delta C^0_{p,\theta}/dT)_\theta$ are not 0. If they are 0, then either plot is linear.

The reference temperature θ was taken to be 283.15 K (10 °C), which is the midpoint of the temperature range studied. The values of ΔS^0_θ, the standard entropy change, were calculated from the values of ΔG^0_θ and ΔH^0_θ by using the relationship $\Delta G^0_\theta = \Delta H^0_\theta - T\Delta S^0_\theta$. The Clarke and Glew plots of $R \ln K_a$ vs. T for the wild-type (square) and mutant DNA (circle) interaction with dHSF(33-163) are shown in Figure 7. In the nonlinear least-squares curve-fitting procedure, each $\ln K_a$ datum was weighted with the normalized reciprocal of its variance, which was calculated from the estimated standard error of $\ln K_a$ obtained in the analysis of the concentration distribution data. The validity of using these values for this purpose has been demonstrated (Lewis et al., 1994). As shown in Figure 7, the differences in the relative affinity of dHSF(33-163) for the 2 DNAs are obvious throughout the temperature range studied. The mutant DNA: dHSF(33-163) interaction shows a greater temperature dependency of $\ln K_a$ under the conditions studied.

Table 1. *Temperature dependence of the association constant of dHSF(33-163) binding to wild-type and mutant DNA*

Temperature (°C)	Wild-type DNA ($\ln K_a$)	Mutant DNA ($\ln K_a$)	Ratio[a]
2	17.84 ± 1.91	14.43 ± 0.13	30.3
4	17.81 ± 1.05	14.33 ± 0.14	32.5
6	17.52 ± 1.44	14.53 ± 0.11	19.9
8	17.25 ± 0.60	14.20 ± 0.11	21.1
10	17.63 ± 0.98	14.20 ± 0.11	30.9
12	17.47 ± 0.70	14.14 ± 0.27	27.9
14	17.32 ± 1.21	13.84 ± 0.10	32.5
16	17.00 ± 0.43	13.71 ± 0.11	26.8
18	17.13 ± 0.57	13.31 ± 0.10	45.6

[a] Ratio = $(K_{a,\text{wild-type}}/K_{a,\text{mutant}}) = \exp(\ln K_{a,\text{wild-type}} - \ln K_{a,\text{mutant}})$. Because $K_d = \exp(-\ln K_a)$, $K_d = 3.63 \times 10^{-8}$ M for wild-type DNA and $K_d = 1.66 \times 10^{-6}$ M for mutant DNA at 18 °C.

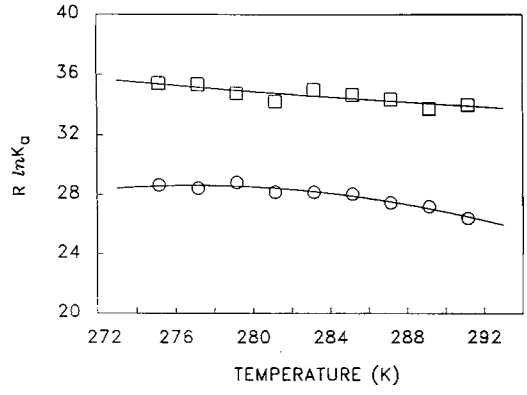

Fig. 7. Temperature dependence of the values of $R \ln K_a$ for the interactions of the 13-bp wild-type (square) and mutant (circle) DNAs with dHSF(33-163) in 0.05 M KCl with 10 mM potassium phosphate buffer, pH 6.3, and 0.1 mM EDTA at equilibrium at a rotor speed of 14,000 rpm. The data are fit with the Clarke and Glew equation with a reference temperature, θ, of 283.15 K. The estimated standard errors obtained when fitting the equilibrium concentration distributions are summarized in Table 1.

Table 2 summarizes the thermodynamic parameters characterizing the interaction of dHSF(33-163) binding to wild-type and mutant DNA. It is clear that the characteristics of wild-type and mutant DNA binding to dHSF(33-163) are quite different. Because the value of $\Delta C^0_{p,\theta}$ for the wild-type DNA binding is small in comparison to its standard error, it might be argued that it is not really possible to distinguish this value from 0 ($\Delta C^0_{p,\theta} = 0.31 \pm 1.11$). On the other hand, the value for the mutant DNA binding is significantly different from 0 ($\Delta C^0_{p,\theta} = -1.67 \pm 0.45$). The changes in values of ΔC^0_p for some proteins have been correlated with changes in protein conformation (Privalov & Khechinaschvili, 1974; Spolar et al., 1989; Livingstone et al., 1991). Therefore, although the near-0 value of $\Delta C^0_{p,\theta}$ for the wild-type DNA binding to dHSF(33-163) suggests that minimal changes in polypeptide conformation occur upon DNA binding, this is probably not the case for the binding of the mutant DNA. The near-0 value of ΔC^0_p for the binding of the wild-type DNA is in contrast to recent studies on sequence-specific protein-DNA interactions, which do show such temperature dependence (Ha et al., 1989; Jin et al., 1993; Lundbäck et al., 1993; Spolar & Record, 1994). It should be noted, however, that reactants used in these studies were of different size and had different stoichiometries than those studied here.

The nondetectability of conformational changes of the dHSF(33-163) upon binding to wild-type DNA is further supported by CD spectra measured in the presence and absence of DNA (Fig. 8). The CD spectra show almost identical spectral profiles both in the absence and presence of DNA. Therefore, based on the results of both ultracentrifugation and CD analysis, the binding of the HSF DNA-binding domain to its cognate site seems to occur without causing significant structural changes in the protein. The effect of mutant DNA on the CD spectra was not studied because it was felt that the much weaker binding of this DNA at the same molar stoichiometry used for the wild-type DNA would preclude obtaining meaningful results.

If we compare the association constants measured at 0.05 M KCl (Table 1) at the reference temperature (10 °C), the wild-type DNA shows a 31-fold greater affinity for dHSF(33-163) than does the mutant DNA. This difference is reflected in a $\Delta\Delta G^0$ value of -1.77 kcal mol^{-1} from the thermodynamic analysis and is caused by a single base-pair substitution in the mutant DNA (Table 2). The value of ΔG^0 (-9.76 kcal mol^{-1}) for the binding of wild-type DNA and dHSF(33-163) is characterized by a negative ΔH^0 (-7.48 kcal mol^{-1}) augmented by a positive ΔS^0 (8.06 cal mol^{-1} K^{-1}). In contrast, the value of ΔG^0 (-7.99 kcal mol^{-1}) for the binding of the mutant DNA is characterized by a larger ΔH^0 (-10.78 kcal mol^{-1}) opposed by a negative ΔS^0 (-9.49 cal mol^{-1} K^{-1}). On the basis of the negative entropic term, the binding of mutant DNA to dHSF(33-163) may be accompanied by immobilization of water molecules or

Table 2. *Summary of the thermodynamic parameters characterizing the interaction of dHSF(33-163) and wild-type and mutant DNAs*[a]

DNA	$\Delta C^0_{p,\theta}$ (kcal mol^{-1} K^{-1})	$-\Delta G^0_\theta$ (kcal mol^{-1})	$-\Delta H^0_\theta$ (kcal mol^{-1})	ΔS^0_θ (cal mol^{-1} K^{-1})	RMS[b]
Wild type	0.31 ± 1.11	9.76 ± 0.06	7.48 ± 2.73	8.06 ± 9.63	0.38
Mutant	−1.67 ± 0.45	7.99 ± 0.03	10.78 ± 1.00	−9.49 ± 3.52	0.19

[a] Buffer: 10 mM potassium phosphate, pH 6.3, containing 0.05 M KCl and 0.1 mM EDTA.
[b] Weighted RMS error obtained when fitting the values of $R \ln K_a$ as a function of absolute temperature to obtain the values of the thermodynamic parameters by using the Clarke and Glew equation.

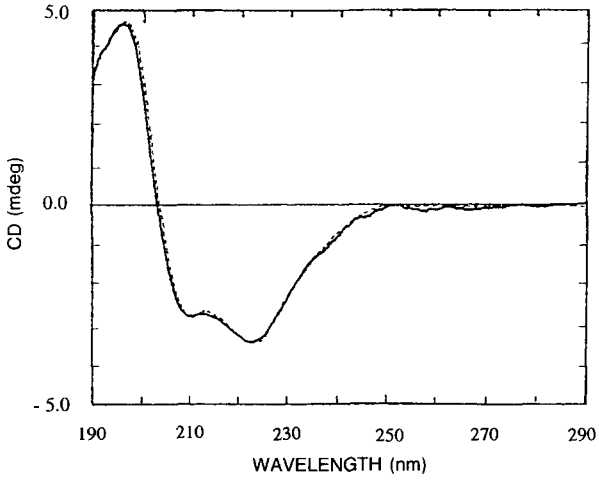

Fig. 8. CD spectra of dHSF(33–163) in the presence (dotted line) and absence (solid line) of 13-bp wild-type DNA in 0.05 M KCl with 10 mM potassium phosphate buffer, pH 6.3, and 0.1 mM EDTA. The concentration of dHSF(33–163) was 33 μM. The protein–DNA complex had a 1:1 molar ratio. These spectra were recorded using a Jasco J-700 spectropolarimeter using a 0.05-cm pathlength cell at 20 °C.

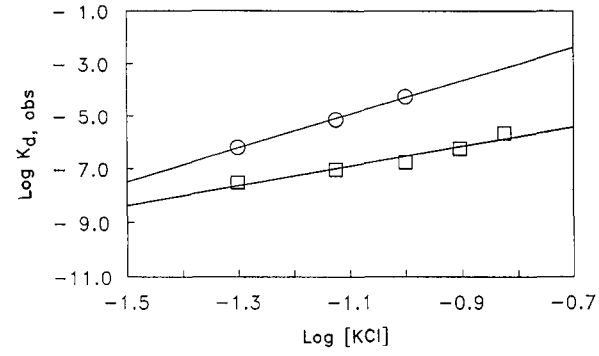

Fig. 9. Salt dependence of the values of log $K_{d,obs}$ for the interactions of the 13-bp wild-type (square) and mutant (circle) DNAs with dHSF(33–163) in 10 mM potassium phosphate buffer, pH 6.3, and 0.1 mM EDTA at equilibrium at a rotor speed of 14,000 rpm at 4 °C.

a reduction of the degrees of freedom of the side chains, or a combination of both, resulting in a net increase of order in the system.

Effects of monovalent salt: Specific vs. mutant site DNA binding

The effects of monovalent salt (KCl) concentration on the association of dHSF(33–163) with wild-type and mutant DNA were studied at 4 °C. The low temperature and the concentration range of KCl (0.05–0.15 M) were chosen because of the low binding affinity of the mutant DNA to dHSF(33–163) (see Table 1). Under these conditions, it is clear that the binding of the dHSF(33–163) to mutant DNA shows greater salt dependence than does its binding to the wild-type DNA (Fig. 9; Table 3). The interactions of dHSF(33–163) with mutant site DNA are more dependent on the salt concentration than on the temperature. In the concentration range studied for the mutant DNA (0.05–0.1 M), the ratio of the association constants of the wild-type and mutant DNA:dHSF(33–163) binding radically increased from 33-fold at 0.05 M KCl to 308-fold at 0.1 M KCl (Table 3), a 9.3-fold increase in these ratios. This difference is also reflected in the increase of the $\Delta\Delta G^0$ value from 1.92 kcal mol^{-1} at 0.05 M KCl to 3.15 kcal mol^{-1} at 0.1 M KCl. This salt dependency is in marked contrast to the temperature dependency if we compare the ratio of association constants measured as a function of salt concentration to those measured as a function of temperature for the protein–DNA interaction (see Table 1).

The data were then analyzed using the theory of protein–polyelectrolyte interactions (Record et al., 1978, 1991; Mascotti & Lohman, 1992) in order to evaluate the effects of salt concentration on the binding to dHSF(33–163). Accordingly, the association of an oligocation L^{z+}, in this case (dHSF(33–163)), with helical B-DNA can be described by the cation-exchange equation with the monovalent ion,

$$L^{z+} + \text{B-DNA} \leftrightarrow \text{complex} + z\psi\ K^+,$$

where z is the net positive charge on the polypeptide and ψ is the thermodynamic extent of cation binding per phosphate. The values of ψ are predicted to range from 0.88 for double-stranded B-form DNA to ~0.7 for single-stranded polynucleotides (Record et al., 1991; Mascotti & Lohman, 1992). The depen-

Table 3. *Salt dependence of the binding of dHSF(33–163) to DNA*

	Wild-type DNA		Mutant DNA			
[KCl]	ln K_a	$-\Delta G^0$ (kcal mol^{-1})	ln K_a	$-\Delta G^0$ (kcal mol^{-1})	$\Delta\Delta G^0$ (kcal mol^{-1})	Ratio[a]
0.05	17.81 ± 1.91	9.81	14.33 ± 0.13	7.89	1.92	32
0.075	16.22 ± 0.63	8.93	11.86 ± 0.12	6.53	2.40	78
0.10	15.57 ± 0.55	8.57	9.84 ± 0.12	5.42	3.15	308
0.125	14.38 ± 0.32	7.92	N.D.[b]			
0.15	13.01 ± 0.14	7.16	N.D.			

[a] Ratio = $K_{a,\text{wild-type}}/K_{a,\text{mutant}} = \exp(\ln K_{a,\text{wild-type}} - \ln K_{a,\text{mutant}})$.
[b] N.D. = not determined because of the low affinity of mutant DNA for dHSF(33–163) at these salt concentrations.

dence of K_{obs} on monovalent cation concentration for the stability of an L^{z+}-B-DNA complex can be described by

$$\log K_{d,obs} = \log K_{d,nonionic} + 0.88z \log [K^+], \quad (7)$$

where $\log K_{d,obs} = -\ln K_a/2.303$. Figure 9 shows that the plot of $\log K_{d,obs}$ vs. $\log [K^+]$ is linear, as predicted by this equation. The value of z calculated from the slope for wild-type DNA binding is 4.2, implying that there are 4–5 ionic interactions involved in this binding. The theory assumes that $\log K_{d,nonionic} = \log K_d$ when $[K^+] = 1.0$ M; the value of $K_d = \sim 3$ mM indicates a 33% ionic contribution to the equilibrium constant at 0.05 M KCl. In contrast, the mutant DNA:dHSF(33–163) association appears to be characterized primarily by ionic interactions. The calculated z value of 7.5 for this binding, implying 7–8 ionic interactions, suggests that the binding of dHSF(33–163) to the mutant site involves almost twice the number of ionic interactions as are involved in site-specific binding and that ionic interactions represent a dominant force in mutant site binding.

Discussion

The sequence-specific binding of the heat shock transcription factor HSF to DNA initiates the activation of the heat shock response in multicellular eukaryotes. In *Drosophila* and other eukaryotes, the latent HSF protein acquires high DNA-binding affinity by conversion from a monomeric form to a homotrimer (Westwood & Wu, 1993). The trimeric form of HSF binds specifically to the HSE in vitro and is redistributed upon heat stress from nonspecific sites in chromatin to the heat shock chromosomal loci and other specific sites in vivo (Westwood et al., 1991). Because the occurrence of a trimeric DNA-binding protein is unusual in nature, how the HSF trimer specifically recognizes its cognate DNA site may provide useful insights in protein–DNA interactions.

In this report, we have initiated a biophysical analysis of the sequence-specific interaction of the *Drosophila* HSF DNA-binding domain using analytical ultracentrifugation. The purified, bacterially expressed polypeptide dHSF(33–163) exists as a monomer in solution and binds to a 13-bp DNA duplex containing a single NGAAN sequence with 1:1 stoichiometry and a standard free energy change of -9.81 kcal mol^{-1}. The binding to a mutant NCAAN sequence was found to decrease by 20–308-fold, depending on temperature and ionic conditions. These findings provide solid experimental support for a model for the high-affinity binding of HSF trimers to DNA, in which 1 HSF subunit of the trimer contacts 1 of 3 NGAAN modules comprising the 15-bp HSE (Perisic et al., 1989; Sorger & Nelson, 1989). Thus, high-affinity binding to the HSE would be achieved by the concerted interactions of all 3 DNA-binding domains.

Throughout the 2–18 °C temperature range studied, the affinity of dHSF(33–163) for wild-type DNA in 0.05 M KCl is 20–45-fold higher than the affinity for mutant DNA. Thus the $\Delta\Delta G^0$ value of 1.77 kcal mol^{-1} is caused by a single base-pair substitution in the mutant DNA. This G to C substitution has been previously reported to neutralize the biological activity of an HSE (Xiao & Lis, 1988), a finding consistent with the studies of other sequence-specific DNA-binding proteins, where a single base pair may simultaneously interact with several different amino acids, e.g., the complexes between DNA and the 434 repressor (Aggarwal et al., 1988), the HNF-3 (hepatocyte nuclear factor-3) DNA-binding domain (Clark et al., 1993). The interaction of dHSF(33–163) with the mutant DNA site shows somewhat greater dependence on temperature and much greater salt dependence than the sequence-specific interaction. Our data suggest that the mutant site binding of dHSF(33–163) to DNA involves about twice as many ionic interactions as to the wild-type sequence.

The differences in thermodynamic parameters for wild-type and mutant site interactions are quite marked. The thermodynamic parameters summarized in Table 2 suggest that the reaction mechanisms for the binding of wild-type and mutant DNA differ significantly. Note that the interaction of mutant DNA with dHSF(33–163) is essentially enthalpically driven, because $\Delta H^0_\theta = -10.78$ kcal mol^{-1}, which is opposed by $T\Delta S^0_\theta = -2.69$ kcal mol^{-1}, giving $\Delta G^0_\theta = -7.99$ kcal mol^{-1}. This is in contrast to the wild-type DNA:dHSF(33–163) interaction, where the change in enthalpy is still dominant ($\Delta H^0_\theta = -7.48$ kcal mol^{-1}) but now $T\Delta S^0_\theta = +2.28$ kcal mol^{-1}, giving $\Delta G^0_\theta = -9.76$ kcal mol^{-1}. These marked differences and the differences in ΔC^0_p suggest that there are very different driving forces in the recognition processes for dHSF(33–163) and the wild-type and mutant DNAs.

As shown in Table 3, the difference in free energy of binding for the monomeric dHSF(33–163):DNA interaction is 3.15 kcal mol^{-1} at 0.1 M KCl. This experimental result for the values of $\Delta\Delta G^0$ for a G to C substitution is close to the theoretical estimate of $\Delta\Delta G^0$ of 5.0 kcal mol^{-1} for a G to T substitution calculated by Lis et al. (1990) using the theory of Berg and von Hippel (1988). However, considering that our estimation is calculated based on the data measured at 0.1 M KCl, the $\Delta\Delta G^0$ of mutant DNA binding and sequence-specific binding could be higher at physiological salt concentrations.

Although the molecular details of how transcription factors scan hundreds of thousands of base pairs and bind to their cognate sites are largely unclear, it appears that the activated form of HSF (assuming homogeneous trimerization) could recognize the HSE with up to 10^6–10^7 times greater affinity at 0.1 M KCl than its ability to recognize nonspecific sites. The change in free energy for site-specific binding would be -26 kcal mol^{-1} (3×-8.57 kcal mol^{-1}) as compared to -16 kcal mol^{-1} for mutant site binding (3×5.42 kcal mol^{-1}). This gives a ~ 10 kcal mol^{-1} difference in the change of free energy of binding, corresponding to a 5.2×10^7-fold increase in the association constant for binding to the 15-bp heat shock element. Considering the differences in analytical techniques and reactants, this value is not inconsistent with an earlier estimate of a $\sim 10^6$-fold difference between specific and nonspecific binding affinities for the full-length HSF trimer obtained using a filter binding assay (Wu et al., 1987). It will be of interest to extend our present findings to an analysis of the binding of full-length HSF trimers to the HSE.

Materials and methods

Sample preparation

BL21(DE3) cells transformed with pHSF(33–163) were grown in 2 L of *E. coli* culture at 37 °C in LB medium containing 100 μM ampicillin. Isopropyl β-thiogalactopyranoside (IPTG) was added to attain a concentration of 0.6 mM when $OD_{600nm} = 0.6$, and the cells were harvested 2 h after IPTG induction. The

bacteria were pelleted by centrifugation (6,000 × *g*, 10 min, 4 °C) and resuspended in 0.4 M KCl solution in buffer A (buffer A: 10 mM potassium phosphate, pH 6.3, with 0.1 mM dithiothreitol [DTT] and 0.1 mM phenylmethanesulfonyl fluoride [PMSF]). After disruption by sonication at 100 mW for 2 min, the lysate was chilled on ice for 10 min. The bacterial debris was removed by centrifugation (6,000 × *g*, 10 min, 4 °C), and the supernatant was frozen in liquid nitrogen and stored at −80 °C until further purification.

For purification of the recombinant dHSF(33–163), 40 mL of the crude supernatant was diluted with buffer A to a KCl concentration of 0.1 M and chromatographed on a 20-mL heparin-Sepharose CL-6B column (Pharmacia). The column was washed with buffer A containing 0.2 M KCl and eluted with 0.4 M KCl in buffer A. The heparin column fractions were pooled and diluted to 0.05 M KCl in buffer A and were further chromatographed on a 10-mL fast-flow S-Sepharose column (Pharmacia). After washing with ca. 10 column volumes of buffer A containing 0.1 M KCl, the sample was eluted with buffer A containing 0.2 M KCl. The extent of purification of the protein was evaluated by SDS–polyacrylamide gel electrophoresis.

The extinction coefficient of dHSF(33–163) was determined from absorbance measurements made using a Hewlett Packard 8452A diode array spectrophotometer. Protein concentration was determined using a Bio-Rad (Richmond, California) protein assay kit. A value of $\epsilon_{280nm} = 9{,}079\ M^{-1}\ cm^{-1}$ was calculated based on a molecular mass of dHSF(33–163) of 15,259 Da from the amino acid sequence.

The DNA was synthesized with a dimethoxytrityl (DMT) group on the 5′-terminus and purified on a Poly-Pak cartridge (Glen Research, Virginia). The concentrations of single-stranded and double-stranded DNAs were determined spectrophotometrically using extinction coefficients that were calculated from a summation of coefficients of individual bases corrected for nearest neighbor contributions (Allen et al., 1972; Puglisi & Tinoco, 1989). The 2 DNAs have the same molar mass of 8,011 Da and the same extinction coefficient of $\epsilon_{260nm} = 90{,}489\ M^{-1}\ cm^{-1}$.

The duplex DNA was formed by mixing equimolar concentrations of 2 single-stranded DNAs and incubating at 80 °C for 30 min, followed by slow cooling to room temperature; the DNA solution was then further cooled to 4 °C. The duplex DNA was isolated by applying the solution to a Centricon-30 concentrator (Amicon, Massachusetts). The DNA retained on the Centricon-30 tube contained the 13-bp double-stranded DNAs. The purity of the double-stranded DNA was confirmed by analytical ultracentrifugation and showed no indication of the presence of single-stranded DNA. These synthetically prepared 13-bp double-stranded DNAs had the sequences 5′-GGGCA GAACGCCG-3′ (wild-type) and 5′-GGGCACAACGCCG-3′ (mutant).

Mass spectroscopy

Mass spectra were obtained at the Harvard Microchemistry Facility (Cambridge, Massachusetts) using a Finnigan-MAT (San Jose, California) TSQ 700 triple quadruple mass spectrometer equipped with an electrospray ion source (Chait & Kent, 1992).

CD spectra

CD spectroscopy was carried out on a JASCO J-700 instrument with a 0.05-cm cell at 33 μM protein concentration at 20 °C. Five accumulated scans were taken for each sample at a scan speed of 50 nm/min. Measurements were made in 10 mM potassium phosphate buffer at pH 6.3 containing 0.05 M KCl and 0.1 mM EDTA. The spectra of the protein–DNA complex were taken from an equimolar mixture of protein and DNA.

Analytical ultracentrifugation

Solutions of polypeptide and DNA, and a 2:1 molar ratio mixture of polypeptide and DNA were simultaneously centrifuged to equilibrium at a rotor speed of 14,000 RPM at temperatures ranging from 2 to 18 °C. Except for experiments evaluating the salt dependence, the buffer used was 50 mM KCl, 10 mM potassium phosphate, pH 6.3, and 0.1 mM EDTA. Equilibrium was considered to have been attained when the scans at 280 nm and 260 nm had been invariant for 24 h. Equilibrium at 2 °C was attained by 72 h, and scans were then taken at 280 nm, at 260 nm, and from 240 nm through 230 nm at 2-nm increments. Similar scans were taken at 2° temperature increments; 12 h were allowed for reequilibration following each increase in temperature.

DNase I footprinting assay

The DNA fragment for the DNase I footprinting assay was prepared by a polymerase chain reaction using as template a plasmid carrying the *Drosophila* hsp promotor (pdhspXX3.2; Tsukiyama et al., 1994). A primer corresponding to −184 to −165 of the *Drosophila* hsp 70 promoter was terminally labeled with [γ-^{32}P]ATP and T_4 polynucleotide kinase, and the PCR was performed with another primer corresponding to sequences from +19 to +36. The PCR fragment was purified by gel electrophoresis.

For the footprinting reaction, approximately 20 fmol of the PCR fragment was incubated with different amounts of recombinant HSF protein in 20 μL of reaction mixture (25 mM Tris HCl, pH 7.4, 50 mM KCl, 0.1 mM EGTA, 0.5 mM DTT, 5% glycerol, 0.5 μg of *E. coli* DNA, 20 μg bovine serum albumin). After incubation at room temperature for 30 min, 2 μL of 10 U/mL DNase I (Boehringer) was added and incubated for 1 min at room temperature. DNase I digestion was terminated by the addition of 80 μL of stop buffer (20 mM Tris HCl, pH 8.0, 20 mM EDTA, 100 mM NaCl, 0.8% SDS, 0.1 μg/μL *E. coli* DNA). The DNA was then purified by organic extraction and ethanol precipitation and separated on a 6% sequencing gel.

Note added in proof

Because of the potential for error in protein extinction coefficients measured using the Bio-Rad method with bovine serum albumin (BSA) as a standard, we have redetermined the extinction coefficient of dHSF(33–163) using the method of Gill and von Hippel (1989). This resulted in an increase of ϵ_{280nm} from $9{,}079\ M^{-1}\ cm^{-1}$ to $15{,}154\ M^{-1}\ cm^{-1}$, giving a decrease in the values of ln K_a of approximately 0.4 (2–4% decrease), with a concomitant increase in the values of ΔG^0 (~0.2 kcal mol^{-1}). The experimentally determined extinction coefficient ($15{,}154\ M^{-1}\ cm^{-1}$) is close to the one calculated from the amino acid composition ($15{,}340\ M^{-1}\ cm^{-1}$). The other thermodynamic parameters given in Tables 2 and 3 (including $\Delta\Delta G^0$) are essentially unaffected.

Acknowledgments

We thank Mrs. Juanita Eldridge for the DNA syntheses, Dr. Jose R. Casas-Finet for help with the CD measurements, Dr. Joachim Clos for constructing the plasmid encoding dHSF33-163, and Dr. B.K. Lee for helpful discussions.

References

Aggarwal AK, Rodgers DW, Drottar M, Ptashne M, Harrison SC. 1988. Recognition of a DNA operator by the repressor of phage 434: A view at high resolution. *Science 242*:899–907.

Allen FS, Gary DM, Roberts GP, Tinoco I. 1972. The ultraviolet circular dichroism of some natural DNAs and an analysis of the spectra for sequence information. *Biopolymers 11*:853–879.

Berg OG, von Hippel PH. 1988. Selection of DNA binding sites by regulatory proteins. *Trends Biochem Sci 13*:207–211.

Chait BT, Kent SBH. 1992. Weighing naked proteins: Practical, high-accuracy mass measurement of peptides and proteins. *Science 257*:1885–1894.

Clark KC, Halay ED, Lai E, Burley SK. 1993. Co-crystal structure of the HNF3/fork head DNA-recognition motif resembles histone H5. *Nature 364*:412–420.

Clarke CW, Glew DN. 1966. Evaluation of thermodynamic functions from equilibrium constants. *Trans Faraday Soc 62*:539–547.

Clos J, Westwood JT, Becker PB, Wilson S, Lambert K, Wu C. 1990. Molecular cloning and expression of a hexameric heat shock factor subject to negative regulation. *Cell 63*:1085–1097.

Durchschlag H. 1986. Specific volumes of biological macromolecules and some other molecules of biological interest. In: Hinz HJ, ed. *Thermodynamic data for biochemistry and biotechnology*. New York: Springer-Verlag. pp 46–50.

Fernandes M, Xiao H, Lis JT. 1994. Fine structure analysis of the *Drosophila* and *Saccharomyces* heat shock factor–heat shock element interactions. *Nucleic Acids Res 22*:167–173.

Gill SC, von Hippel PH. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem 182*:319–326.

Ha JH, Spolar RS, Record MT Jr. 1989. Role of the hydrophobic effect in stability of site-specific protein–DNA complexes. *J Mol Biol 209*:801–816.

Harrison CJ, Bohm AA, Nelson HCM. 1994. Crystal structure of the DNA binding domain of the heat shock transcription factor. *Science 263*: 224–227.

Jin L, Yang J, Carey J. 1993. Thermodynamics of ligand binding to trp repressor. *Biochemistry 32*:7302–7309.

Lewis MS. 1991. Ultracentrifugal analysis of a mixed association. *Biochemistry 30*:11716–11719.

Lewis MS, Shrager RI, Kim SJ. 1994. Analysis of protein–nucleic acid and protein–protein interactions using multi-wavelength scans from the XL-A analytical ultracentrifuge. In: Schuster TM, Laue TM, eds. *Modern analytical ultracentrifugation: Acquisition and interpretation of data for biological and synthetic polymer systems*. Cambridge, Massachusetts: Birkhäuser Boston, Inc. Forthcoming.

Lis JT, Wu C. 1993. Protein traffic on the heat shock promoter: Parking, stalling, and trucking along. *Cell 74*:1–20.

Lis JT, Xiao H, Perisic O. 1990. Modular units of heat shock regulatory region: Structure and function. In: Morimoto R, Tissieres A, Georgopoulos C, eds. *Stress proteins in biology and medicine*. Cold Spring Harbor, New York: Cold Spring Harbor Press. pp 411–428.

Livingstone JR, Spolar RS, Record MT. 1991. Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. *Biochemistry 30*:4237–4244.

Lundbäck T, Cairns C, Gustafsson JA, Carlstedt-Duke J, Hard T. 1993. Thermodynamics of the glucocorticoid receptor–DNA interaction: Binding of wild-type GR DBD to different response elements. *Biochemistry 32*:5074–5082.

Mascotti DP, Lohman TM. 1992. Thermodynamics of single-stranded RNA binding to oligolysines containing tryptophan. *Biochemistry 31*:8932–8946.

Peteranderl RT, Nelson HCM. 1993. Trimerization of the yeast heat shock transcription factor by a triple-stranded α-helical coiled-coil. *Biochemistry 31*:12272–12276.

Perisic O, Xiao H, Lis J. 1989. 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.

Privalov PL, Khechinaschvili NN. 1974. A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study. *J Mol Biol 86*:665–684.

Puglisi J, Tinoco I Jr. 1990. Absorbance melting curves of RNA. *Methods Enzymol 180*:304–325.

Rabindran SK, Giorgi G, Clos J, Wu C. 1991. Molecular cloning and expression of a human heat shock factor, HSF1. *Proc Natl Acad Sci USA 88*:6906–6910.

Record MT, Anderson CF, Lohman TM. 1978. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: The roles of ion association or release, screening, and ion effects on water activity. *Q Rev Biophys 11*:103–178.

Record MT, Ha JH, Fisher MA. 1991. Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. *Methods Enzymol 208*:291–343.

Scharf KD, Materna T, Treuter E, Nover L. 1993. Heat stress promoters and transcription factors. In: Dennis E, Nover L, eds. *Plant promoters and transcription factors*. Berlin: Springer-Verlag. pp 121–158.

Sorger PK, Nelson HCM. 1989. Trimerization of a yeast transcriptional activator via a coiled-coil motif. *Cell 59*:807–813.

Spolar RS, Ha JH, Record MT. 1989. Hydrophobic effect in protein folding and other noncovalent processes involving proteins. *Proc Natl Acad Sci USA 86*:8382–8385.

Spolar RS, Record MT Jr. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science 263*:777–784.

Strang G. 1986. *Introduction to applied mathematics*. Wellesley, Massachusetts: Wellesley–Cambridge Press. pp 138–139.

Topol J, Ruden DM, Parker CS. 1985. Sequences required for in vitro transcriptional activation of a *Drosophila* hsp70 gene. *Cell 42*:527–537.

Tsukiyama T, Becker PB, Wu C. 1994. ATP-dependent nucleosome disruption of a heat-shock promotor mediated by binding of a GAGA transcription factor. *Nature 367*:525–532.

Vuister GW, Kim SJ, Wu C, Bax A. 1994. NMR evidence for similarities between the DNA-binding regions of *Drosophila melanogaster* heat shock factor and the helix-turn-helix and HNF/forkhead families of transcription factors. *Biochemistry 33*:10–16.

Westwood JT, Clos J, Wu C. 1991. Stress-induced oligomerization and chromosomal relocalization of heat-shock factor. *Nature 353*:822–827.

Westwood JT, Wu C. 1993. Activation of *Drosophila* heat shock factor: Conformational change associated with a monomer-to-trimer transition. *Mol Cell Biol 13*:3481–3486.

Wu C, Wilson S, Walker B, Dawid I, Paisley T, Zimarino V, Udea H. 1987. Purification and properties of *Drosophila* heat shock activator protein. *Science 238*:1247–1253.

Xiao H, Lis JT. 1988. Germline transformation used to define key feature of heat shock response elements. *Science 239*:1139–1142.

Xiao H, Perisic O, Lis J. 1991. Cooperative binding of *Drosophila* heat shock factor to an array of a conserved 5-bp unit. *Cell 64*:585–593.