Energetics of Sequence-Specific Protein−DNA Association: Binding of Integrase Tn916 to Its Target DNA†

Stoyan Milev,‡ Alemayehu A. Gorfe,‡ Andrey Karshikoff,§ Robert T. Clubb,|| Hans Rudolf Bosshard,‡ and Ilian Jelesarov*,‡

Institute of Biochemistry, University of Zürich, Winterthurerstrasse 190m Room 44 L42, CH-8057 Zürich, Switzerland, Center for Structural Biochemistry, Department of Biosciences, NOVUM, Karolinska Institutet, S-141 57 Huddinge, Sweden, and Department of Chemistry and Biochemistry and UCLA-DOE Laboratory of Structural Biology and Genetics, University of California, 405 Hilgard Avenue, Los Angeles, California 90095

Received October 1, 2002; Revised Manuscript Received February 5, 2003

ABSTRACT: The DNA binding domain of the transposon Tn916 integrase (INT-DBD) binds to its DNA target site by positioning the face of a three-stranded antiparallel β-sheet within the major groove. Binding of INT-DBD to a 13 base pair duplex DNA target site was studied by isothermal titration calorimetry, differential scanning calorimetry, thermal melting followed by circular dichroism spectroscopy, and fluorescence spectroscopy. The observed heat capacity change accompanying the association reaction (ΔC_p) is temperature-dependent, decreasing from -1.4 kJ K^{-1} mol^{-1} at 4 °C to -2.9 kJ K^{-1} mol^{-1} at 30 °C. The reason is that the partial molar heat capacities of the free protein, the free DNA duplex, and the protein−DNA complex are not changing in parallel when the temperature increases and that thermal motions of the protein and the DNA are restricted in the complex. After correction for this effect, ΔC_p is -1.8 kJ K^{-1} mol^{-1} and temperature-independent. However, this value is still higher than ΔC_p of -1.2 kJ K^{-1} mol^{-1} estimated by semiempirical methods from dehydration of surface area buried at the complex interface. We propose that the discrepancy between the measured and the structure-based prediction of binding energetics is caused by incomplete dehydration of polar groups in the complex. In support, we identify cavities at the interface that are large enough to accommodate ~10 water molecules. Our results highlight the difficulties of structure-based prediction of ΔC_p (and other thermodynamic parameters) and emphasize how important it is to consider changes of thermal motions and soft vibrational modi in protein−DNA association reactions. This requires not only a detailed investigation of the energetics of the complex but also of the folding thermodynamics of the protein and the DNA alone, which are described in the accompanying paper [Milev et al. (2003) *Biochemistry 42,* 3492−3502].

Noncovalent binding of proteins to DNA is fundamental to the processing of genetic information. Reactions such as transcription and replication depend on specific DNA recognition by enzymes and regulatory proteins. To understand sequence-specific DNA recognition, one has to understand the structural, kinetic and thermodynamic requirements and characteristics of protein−DNA association reactions. Here we focus on the thermodynamic properties of the reaction of a bacterial integrase with its target DNA.

Sequence-specific protein−DNA complexes are held together by van der Waals contacts, oriented hydrogen bonds and salt bridges. Binding often requires conformational adaptation of both the protein and the DNA molecule, and water can be present at the complex interface, contributing to the overall energetic balance of binding. As a consequence, protein−DNA association is a dynamic process between partly flexible molecules and depends on a balance of many different forces. In the search of unifying principles for the structure-based prediction of binding energetics, a challenging goal is to link the energetics of a protein−DNA association reaction with the type and number of noncovalent intermolecular contacts at the binding interface, taking into account also the energetic consequences of structural adaptation and changes in hydration. The starting point is the determination of the changes of enthalpy (ΔH), entropy (ΔS), and heat capacity (ΔC_P) accompanying protein−DNA association. The best estimates of these parameters are obtained by isothermal titration calorimetry (ITC[1]).

The heat capacity change has attracted special attention, and its accurate determination is crucial, as it provides a rigorous description of the temperature dependence of ΔG,

† This work was supported in part by the Swiss National Science Foundation.

* Corresponding author. Telephone: +41 1 655 5547. Fax: +41 1 635 6805. E-mail: iljel@bioc.unizh.ch.

‡ University of Zürich.

§ Karolinska Institutet.

|| University of California, Los Angeles.

[1] Abbreviations: ACES, *N*-(2-acetamido)-2-aminoethanesulphonic acid; ASA, solvent-accessible surface area; bp, base pair; CD, circular dichroism; DSC, differential scanning calorimetry; HEPES, N-2-hydroxypiperazine-*N*′-2-ethanesulphonic acid; ITC, isothermal titration calorimetry. Subscripts and superscripts to thermodynamic parameters have the following meaning: A, value referring to complex formation from folded components; C, value referring to dissociation of complex into unfolded components; cal, value from direct calorimetric measurement; CD, value from CD melting curve; conf, conformational change; D, DNA; m, value at temperature midpoint of thermal transition; P, protein; rb, rigid body association; vH, value from van't Hoff analysis.

10.1021/bi0269355 CCC: $25.00
Published on Web 03/08/2003

ΔH, and ΔS. Values of ΔC_P for protein−DNA association are negative when using the free components as the reference state. This is ascribed to the loss of bound water molecules from interacting surface areas in the complex. The magnitude of ΔC_P should therefore correlate with the energetic cost of dehydration of molecular surface (*1*, *2*). Hence, ΔC_p (and also ΔH and ΔS) can be estimated by semiempirical methods from the structure of the complex and its components (*1*, *2*), although such estimates may be ambiguous (*3*−*5*). Indeed, values of ΔC_p for protein−DNA association tend to be larger than those predicted from the burial of polar and nonpolar surface alone (*6*−*10*). Since the conformation of protein and DNA often change when forming a complex (*11*−*14*), conformational changes are invoked to explain the observed discrepancies between measured and structure-based estimates of ΔC_P. Association may change the vibrational modes of the complex as compared to its free components, and this may also increase ΔC_P (*15*, *16*). Moreover, vibrational effects are not restricted to protein and DNA. The decrease of translational and soft vibrational modes of water molecules near to or at the binding interface, can be substantial (*17*, *18*). A recent theoretical study hints at a large heat capacity effect of DNA dehydration (*19*).

Structural interpretation of heat capacity changes can also be hampered if the conformation of both the protein and the DNA are temperature-sensitive. Many DNA-binding proteins are flexible and sometimes partly or even completely unfolded in isolation and have low thermal stability (*11*). This brings us to an important point: Analysis of the energetics of protein−DNA association requires the knowledge of all the conformational transitions of the complex and its free components in as large a temperature range as possible. In particular, knowing the heat capacities of the complex and its components is necessary to interpret discrepancies between measured and calculated heat capacity changes, as shown by one of us before (*10*). To this end, mixing calorimetry (ITC) has to be supplemented by differential scanning calorimetry (DSC) and other "thermal melting" methods. Only from the combined results of such studies performed on the complex and its components can one reliably correlate energetics with structure.

We have used bacterial integrase Tn916 and its target DNA as a relatively simple system for the comprehensive thermodynamic analysis of a protein−DNA complex. Here we report on the energetics of the association reaction of the 74-residue DNA-binding domain of integrase Tn916 with a 13 bp target duplex DNA. The association reaction is endothermic and entropy-driven at low temperature and becomes exothermic and opposed by entropy decrease above about 15 °C. ΔC_P of association is itself temperature-dependent and is significantly larger than the structure-based estimate of ΔC_p. This peculiar behavior of ΔC_p is due to restricted thermal motions of the protein and the DNA in the complex and, possibly, the presence of water-filled cavities at the complex interface. It should be emphasized that the structural interpretation of the thermodynamic parameters presented here would not have been possible without the independent analysis of the thermodynamics of folding of the free protein and the free DNA described in the accompanying paper (*20*).

MATERIALS AND METHODS

Materials. Buffers. Most experiments were conducted in standard buffer composed of 50 mM Na−phosphate, 100 mM NaCl, pH 6. Some ITC experiments were performed in 64 mM HEPES (or ACES), 100 mM NaCl, pH 6.0. All chemicals were of analytical grade and were used without further purification.

Overexpression and Purification of INT-DBD. The subcloning of the N-terminal domain of Tn916 integrase (INT-DBD) comprising residues 2−74, and the Cys57Ala mutation has been described elsewhere (*21*). *E. coli* cultures were grown for 4 h at 37 °C and induced by addition of 1 mM isopropylthio-β-D-galactoside. Cells were harvested by centrifugation, resuspended in 100 mM Tris-HCl (pH 7.2), 5 mM EDTA, 2.5 mM dithiothreitol (DTT), 5 mM benzamidine, and lysed by sonification. The lysate was centrifuged at 15 000 rpm for 1 h, and the supernatant containing INT-DBD was loaded on a heparin-Sepharose CL-6B column (Pharmacia; 1.6 cm × 14 cm) pre-equilibrated with 100 mM Tris-HCl (pH 7.0), 5 mM EDTA, 2.5 mM DTT. After thorough washing, elution was achieved with 1 M NaCl in the same buffer. The eluate was dialyzed overnight against 50 mM Tris-HCl (pH 7.0), applied to a LKB-SP-5PW column, and eluted with a gradient of 1 M NaCl (0−100%) in 50 mM Tris-HCl (pH 7.0). INT-DBD was purified to homogeneity by reversed-phase HPLC on a Nucleosil 300−5 C8 column (Machery & Nagel) and binary gradients of acetonitrile/H_2O containing 0.1% trifluoroacetic acid. The protein was lyophilized and stored as a powder. Ion spray mass spectrometry showed a single peak of 8555 Da corresponding to residues 2−74 (C57A). Protein concentration was measured by UV absorption at 280 nm in 6M GdmCl using the calculated ϵ_{280} of 10.81 M^{-1} cm^{-1}.

Methods. Preparation of DNA Duplex and Protein−DNA Complex. Single-stranded oligonucleotides were purchased from Metabion GmbH (Martinsried) and were purified by HPLC. For duplex preparation, equimolar amounts of the two complementary strands were mixed and annealed by heating to 70 °C and slowly cooling to room temperature. Concentrations were determined, after complete digestion by phosphodiesterase I, from light absorption at 260 nm (Sigma). The complex was formed by adding the protein to a solution of DNA until the calculated 1:1 ratio was reached.

Circular Dichroism Measurements. All measurements were carried out on a Jasco J-715 instrument equipped with computer-controlled water bath, using jacketed cylindrical cells of 1-mm or 10-mm path length. Spectra shown are the average of three scans recorded at 5 nm min^{-1}. Thermal unfolding curves were measured by following the ellipticity change upon continuous heating or cooling between 3 and 65−85 °C at a scan rate of 0.5 or 1 °C min^{-1} and with data collection every 20 s. Reversibility of unfolding was checked by repeated scans and was always better than 95%. Thermal melting curves were analyzed as described (*22*).

Fluorescence Spectroscopy. Fluorescence measurements were made on a Perkin-Elmer LS 50B luminescence spectrometer. Excitation was at 295 nm, and emission spectra were recorded from 300 to 450 nm at 1 nm intervals with 1.2 s integration time. For binding experiments, 2−5 μM protein placed in the thermostated cuvette was titrated with 2 μL aliquots of concentrated DNA with continuous stirring.

Final dilution was 4% or less. The association constant, K_A, was calculated from

$$F = \frac{F_{max}\{D_t + P_t + 1/K_A - \sqrt{(D_t + P_t + 1/K_A)^2 - 4D_tP_t}\}}{2D_t} \quad (1)$$

describing a 1:1 association reaction with D_t, P_t, and F_{max} representing the total concentrations of DNA duplex and protein, respectively, and the maximal change of fluorescence intensity reached at saturation.

Isothermal Titration Calorimetry. ITC was performed on a MCS ITC instrument (MicroCal Inc., Northampton, MA). The calorimeter was calibrated according to the manufacturer's description. Samples of protein and DNA were prepared in, and thoroughly dialyzed against, the same batch of buffer to minimize artifacts due to minor differences in buffer composition. Concentration was determined after dialysis. The sample cell (1.36 mL) was loaded with 30 μM DNA duplex. A titration experiment typically consisted of 25−27 injections of a 150−350 μM protein stock solution, each of 8 or 10 μL volume and 10 or 12 s duration, with a 5 min interval between additions. Stirring rate was 350 rpm. Nonspecific heat effects were estimated from the magnitude of the peaks appearing after complete saturation. Raw data were integrated, corrected for nonspecific heats, normalized for concentration, and analyzed according to a 1:1 binding model assuming a single set of identical binding sites.

Scanning Calorimetry. DSC experiments were performed on a VP-DSC calorimeter (MicroCal Inc.) equipped with twin coin-shaped cells of 0.52 mL volume. Details on the instrument's performance are given elsewhere (*23*). The heating rate was 1 °C min^{-1}. Protein, DNA, and complex samples were dialyzed for 18−24 h against the same batch of buffer used to establish the baseline. Reversibility was checked by 2−3 cycles of heating and cooling. The raw experimental data were corrected for the instrumental buffer-buffer baseline and transformed to partial molar or partial specific heat capacity using partial specific volumes of 0.715, 0.561, and 0.659 cm^3 g^{-1} for protein, DNA duplex, and protein−DNA complex, respectively (see below). The analysis of heat capacity traces of the protein−DNA complex followed the formalism detailed elsewhere (*24*−*26*). Data handling and analysis were carried out using the program CpCalc 2.1 (Applied Thermodynamics), subroutines for Origin provided by MicroCal, and in-house written scripts for NLREG (Phillip H. Sherrod).

Calculation of Absolute Heat Capacities. The absolute heat capacity C_p was derived from the concentration dependence of the apparent heat capacity, C_{app}, according to (*27*)

$$C_{app} = (C_p - \upsilon_M)m_M + \text{const} \quad (2)$$

where υ_M is the partial specific volume of the macromolecule and m_M is its mass in the calorimetric cell. The partial specific volumes of the protein, 13 bp DNA duplex, and the complex were computed according to Karshikoff and Ladenstein (*28*). This method reproduces with high precision υ_M of globular proteins. Indeed, the calculated υ_M is within 2% of υ_M estimated from the amino acid sequence according to Makhatadze et al. (*29*). For the calculation, 20 NMR conformers of the complex and 25 NMR conformers of the protein were used, and the DNA was modeled in an ideal B-form conformation. Values of υ_M used in eq 2 were 0.561 g cm^{-3} for DNA, 0.715 ± 0.006 g cm^{-3} for protein and 0.659 ± 0.003 g cm^{-3} for the complex.

Binding Simulations. To simulate the temperature dependence of the excess heat capacity function of an equimolar mixture of protein and DNA interacting to a 1:1 complex, the theoretical framework developed by Brandts and Lin was used (*30*). In the following, SS_A and SS_B denote, respectively, the total molar concentrations of DNA single strands A and B. D_N and D_T are, respectively, free DNA duplex and total DNA in duplex equivalents. The total molar concentration of protein is P_T. The concentrations of free protein in the unbound native state and in the unbound unfolded state are designated P_N and P_U, respectively. The following expressions describe all the equilibria of the system. DNA melting with strand dissociation, $D_N \rightleftarrows SS_A + SS_B$, is described by the equilibrium constant $K_D = SS_A \times SS_B/D_N$, the enthalpy of DNA dissociation $\Delta H_D(T_D)$ at the reference temperature (melting temperature) T_D, and the heat capacity increment of DNA unfolding $\Delta C_{p,D}$. The protein conformational equilibrium $P_N \rightleftarrows P_U$ is described by the equilibrium constant $K_P = P_U/P_N$, the protein unfolding enthalpy $\Delta H_P(T_P)$ at the reference temperature T_P, and the protein unfolding heat capacity change $\Delta C_{p,P}$. The Gibbs free energies of DNA unfolding, ΔG_D, and protein unfolding, ΔG_P, are defined by the Gibbs−Helmholtz eqs 3 and 4:

$$\Delta G_D(T) = \Delta H_D(T_D)\cdot\left(1 - \frac{T}{T_D}\right) + \Delta C_{p,D}\left[T - T_D - T\ln\left(\frac{T}{T_D}\right)\right] - RT\ln(0.85D_T) \quad (3)$$

$$\Delta G_P(T) = \Delta H_P(T_P)\cdot\left(1 - \frac{T}{T_P}\right) + \Delta C_{p,P}\left[T - T_P - T\ln\left(\frac{T}{T_P}\right)\right] \quad (4)$$

Native protein and DNA duplex associate to form the 1:1 complex, PD, according to $D_N + P_N \rightleftarrows PD$. The association constant is $K_A = PD/(D_N \times P_N)$ and the change of the enthalpy (ΔH_A), entropy (ΔS_A), and free energy (ΔG_A) of association with temperature is given by

$$\Delta G_A(T) = -RT\ln K_A = \Delta H_A(T_R) + \int_{T_R}^{T}\Delta C_{p,A}\,dT - T[\Delta S(T_R) + \int_{T_R}^{T}\Delta C_{p,A}\,d\ln T] \quad (5)$$

where T_R denotes an appropriate reference temperature.

Since SS_A and SS_B are always present in equimolar amount and since $SS_A = SS_B$ equals the concentration of unfolded DNA in duplex equivalents, mass conservation can be written as

$$D_T = D_N + SS_A + PD = \frac{SS_A^2}{K_D} + SS_A + K_A\cdot P_N\cdot\frac{SS_A^2}{K_D} \quad (6)$$

$$P_T = P_N + P_U + P_D = P_N + K_P\cdot P_N + K_A\cdot P_N\cdot\frac{SS_A^2}{K_D} \quad (7)$$

Substitution of $P_N = P_T/(1 + K_P + SS_A^2 \times (K_A/K_D))$ from eq 7 into eq 6 leads to an equation that can be numerically solved for SS_A. Once SS_A is known, the population of all other species at any temperature can be calculated from the above equations.

The excess enthalpy function of the system, using the complex as the reference state is

$$\langle H \rangle = \Delta H_P \cdot \left(\frac{P_U}{P_T}\right) + \Delta H_D \cdot \left(\frac{D_U}{D_T}\right) - \left[\Delta H_A \cdot \left(\frac{P_N + P_U}{P_T}\right)\right] \quad (8)$$

and after differentiation, the excess heat capacity function is obtained.

Calculation of Solvent-Accessible Surface Area. ASA was calculated using the program NACCES with the default set of atomic radii and parameters (*31*). The structures of free protein and free DNA in their binding competent conformation (protein* and DNA* in Scheme 1) were generated by removing the protein or the DNA from the coordinate file of the complex (1B69). The NMR structure of the free protein (2BB8) and of the 13 bp duplex DNA in a canonical B-DNA conformation were used to calculate the surface area changes associated with the structural rearrangement induced upon binding. Residues Ser 2, His 72, Asp 73, and Gly 74 missing in the NMR structure of the complex, and residues Ser 2 and Gly 74 missing in the NMR structure of the free protein were added manually and the resulting structures were minimized by standard protocols using CHARMM. ΔC_p ascribed to dehydration of the complex interface was calculated from

$$\Delta C_p = a_{aliph}\Delta ASA_{aliph} + a_{ar}\Delta ASA_{ar} + a_{pol}\Delta ASA_{pol} \quad (9)$$

where the terms ΔASA (in units of Å^2) are the binding-induced changes in aliphatic, aromatic, and polar surface, respectively, and the coefficients a_i are the elementary contributions (in units of kJ K^{-1} mol^{-1} Å^{-2}) to the heat capacity of hydration of the corresponding type of surface (*32*−*34*). Likewise, the hydration enthalpy can be scaled from the amount and type of surface becoming inaccessible to bulk solvent when the complex is formed (*1*, *2*). Numerical values of the per-Å^2-contributions to ΔH were taken from Table 13 in Makhatadze and Privalov (*34*). Other parametrization schemes that do not distinguish between aliphatic and aromatic surface contributions (*2*, *32*) gave very similar estimates of ΔC_p.

RESULTS AND DISCUSSION

Structure of the Protein−*DNA Complex.* Tn916 is a transposon carrying tetracycline resistance. Excision of the Tn916 transposon requires the formation of a nucleoprotein complex in which the N-terminal domain of the transposon-encoded integrase specifically recognizes a DNA target sequence. The N-terminal 75 residues of the Tn916 integrase, called INT-DBD, suffice for DNA recognition.

The solution structure of the free protein (with a Cys57Ala mutation) and of the protein bound to the 13 bp target DNA duplex (sequences 5′-GAGTAGTAAATTC-3′ and 5′-GAATTTACTACTC-3′) have been solved by NMR spectroscopy (*21*, *35*).[2] Figure 1 shows the structures of the protein−DNA complex and the free protein. The protein binds to the DNA through a three-stranded antiparallel β-sheet and contacts bases and backbone groups of the major groove. This is a rare DNA-binding motif found so far in only three DNA-binding proteins (*36*, *37*). The structure of the protein becomes more disordered in the complex, as indicated by the loss of several NOEs in the spectrum of the bound form (*21*). The DNA is bent by ~35° toward the protein but retains an overall B-conformation with a modest compression of the major groove in the center of the recognition motif.

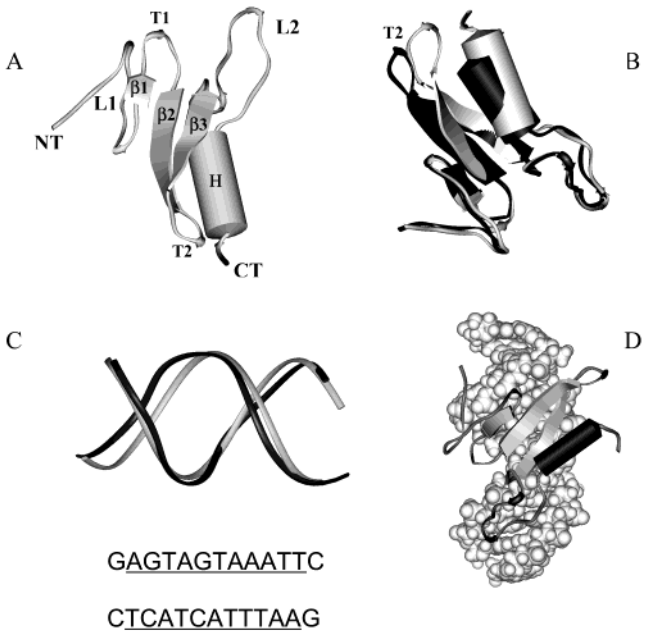


FIGURE 1: NMR structure of the DNA-binding domain of integrase Tn916 and its complex with a 13 bp duplex DNA (*21*, *35*). (A) Free form of INT-DBD. The secondary structure elements are labeled. L, loop; T, turn; β, beta strand; H, α-helix. (B) Comparison of the free (in gray) and bound (in black) forms of the protein. Superposition of the backbone atoms from residues Arg 6 to Tyr 27 and Gln 37 to Glu 66 emphasizes the different positioning of turn T2, the major structural rearrangement induced upon binding. (C) Comparison of the bound form of the INT−DNA duplex (in black) with the cannonical B-DNA conformation (in gray). The oligonucleotide sequence used in the NMR structure and all experiments is listed. The consensus binding site for the integrase protein is underlined. (D) Schematic representation of the structure of the INT-DBD−DNA complex. A ribbon diagram of the protein is shown positioned on the van der Waals surface of DNA.

The protein used in this study contains residues 2−74. The missing N-terminal Met of the original protein used to solve the NMR structure does not participate in binding and has no effect on the stability of the complex. Experiments were performed in the buffer used for NMR structure determination.

Spectroscopic Characterization of Complex Formation. CD Spectroscopy Reveals Conformational Changes. The CD spectrum of the complex differs from the sum of the CD spectra of the protein and the DNA (Figure 2A). Subtraction of the CD spectrum at 85 °C from the spectrum at 3 °C yields the difference spectrum between the folded and the

[2] To facilitate reading, the N-terminal fragment 2−74 of integrase Tn916 and the 13 bp duplex target DNA are called "protein" and "DNA", respectively.

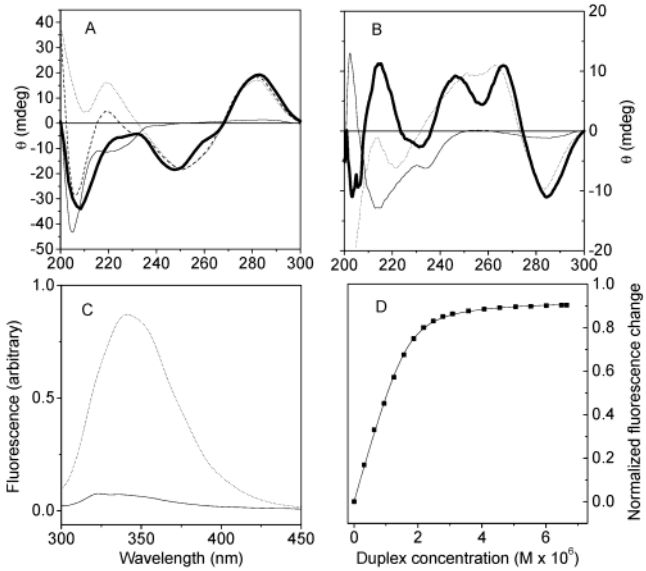


FIGURE 2: Spectroscopic characterization of protein, DNA and protein−DNA complex. **(**A) CD spectra of protein (thin line), DNA (dotted line), 1:1 protein−DNA complex (heavy line), and calculated sum of spectra of free protein and free DNA (dashed line) recorded at 3 °C in standard buffer with 60 μM protein, 60 μM duplex DNA, and 60 μM complex, respectively. (B) Difference spectra calculated by arithmetic subtraction of the spectra at 3 °C from the spectra at 85 °C (lines as in panel A). Spectral changes induced by heating are >95% reversible. (C) Fluorescence emission spectrum (excitation 295 nm) of free protein (dotted line) and of protein saturated with DNA (solid line). (D) Binding isotherm recorded at 20 °C. The data are plotted as the normalized change of the emission intensity. DNA from a 70 μM stock solution was added stepwise to 2 μM of protein. The continuous line is a best fit according to eq 1 for K_A of 8×10^6 M^{-1}.

unfolded species (Figure 2B). Thermal unfolding of the complex is silent at 232 nm for the DNA and at 260 nm for the protein. Ellipticities at these wavelength can be used to separately monitor the unfolding of the protein and the DNA, respectively, during thermal unfolding of the complex (Figure 5).

Dissociation Constant of the Complex Obtained by Fluorescence Titration of Protein with DNA. The fluorescence emission of Trp 42, located at the DNA binding site (*21*), is quenched when this residue is buried at the complex interface (Figure 2C). From the change of fluorescence, the association constant of the complex can be determined (Figure 2D). The solid line in Figure 2D is a best fit for a 1:1 binding reaction (eq 3) with equilibrium constant K_A of $(8 \pm 3) \times 10^6$ M^{-1} at 25 °C.

Thermodynamics of Complex Formation Measured by Titration Calorimetry. Dissociation Constant, Enthalpy of Association, and Complex Stoichiometry Deduced from ITC. ITC is the most direct method to measure the change of enthalpy (ΔH_A) and heat capacity ($\Delta C_{p,A}$) of the association reaction and to confirm the 1:1 stoichiometry of the complex. Addition of small aliquots of protein to DNA in the calorimetric cell produces measurable heat effects, which saturate as the molar ratio of protein to DNA increases (Figure 3) (*38*). From the shape of the titration curve the binding stoichiometry, *n*, the association constant, K_A, and the apparent calorimetric enthalpy of association, ΔH_A^{cal}, is calculated (*38*). The stoichiometry is 1.02 ± 0.15 (mean ± SD of 20 experiments), in agreement with a 1:1 complex.

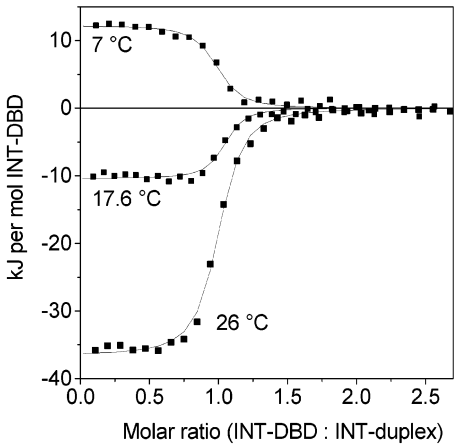


FIGURE 3: Formation of protein−DNA complex followed by ITC. Binding isotherms from titration of 30 μM DNA placed in the calorimetric cell with protein from a 300 μM stock solution in standard buffer. Symbols represent the integrated heats after correction for nonspecific heat effects and normalization for the molar concentration. Continuous lines are nonlinear fits for a 1:1 binding model. Experimental temperatures are indicated.

Within experimental error, K_A is temperature-independent between 4 and 30 °C. The average K_A is $(7 \pm 5) \times 10^6$ M^{-1}, in agreement with K_A from fluorescence titration (Figure 2D).

The apparent enthalpy and entropy changes of association vary with temperature (Figure 4A). Complex formation is endothermic at low temperature and exothermic at high temperature, ΔH_A^{cal} changing signs at around 13 °C. The same ΔH_A^{cal} is measured in buffers of different heats of protonation indicating that there is no change in the protonation state of the protein or the DNA upon binding (*1*). The entropy of complex formation is positive below about 29 °C and negative at higher temperature (Figure 4A). The thermodynamic binding parameters from 20 ITC experiments performed in the range of 4 to 30 °C are summarized in Table 1 of the Supporting Information.

The Heat Capacity Change of Association Determined by ITC. The temperature variation of the binding enthalpy, $\delta\Delta H_A^{cal}/\delta T$, represents the heat capacity change of association, $\Delta C_{p,A}$. Linear regression of the data shown in Figure 4B yields $\Delta C_{p,A}$ of -2.3 ± 0.2 kJ K^{-1} mol^{-1}. However, the data are not well described by a straight line. Fitting statistics improve significantly if a second-order polynomial is used (heavy line in Figure 4B). The curvature of the $\delta\Delta H_A^{cal}/\delta T$ function indicates a temperature variation from -1.4 kJ K^{-1} mol^{-1} at 5 °C to -2.9 kJ K^{-1} mol^{-1} at 30 °C.

Thermal Dissociation and Unfolding of the Protein−DNA Complex. Thermal Unfolding Followed by CD. Unfolding was monitored at 232 and 260 nm, taking advantage of the observation that the two wavelengths reflect the thermal transition of protein alone and DNA alone (Figure 2B). The melting curves are shown in Figure 5. From the normalized curves (inset of Figure 5), it is seen that the transitions monitored at the two wavelengths are superimposable. This hints at a highly cooperative transition from folded complex to unfolded protein and DNA. The data of Figure 5 were subjected to van't Hoff analysis according to eq 2 of the Supporting Information. This analysis requires that the reaction order n of the thermal unfolding reaction is known. An order of 2 means that the complex unfolds into its folded

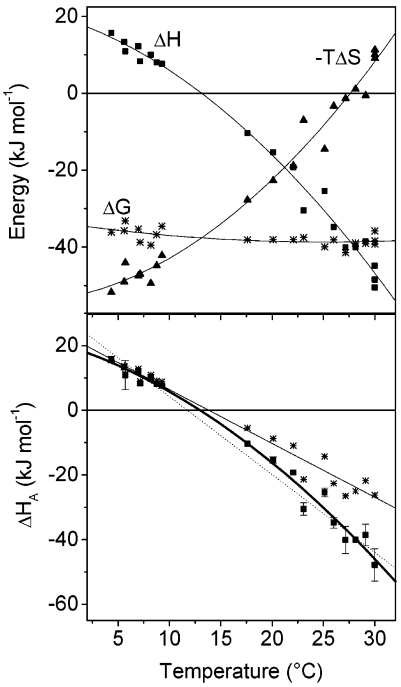


FIGURE 4: Energetics of protein−DNA association measured by ITC. (A) Changes of enthalpy, entropy and free energy in the range of 4−30 °C. (B) Determination of the heat capacity change $\Delta C_{p,A}$ from $\delta\Delta H_A/\delta T$ (Kirchoff plot). Filled squares, ΔH_A^{cal} (data from Table 1); asterisks, $\Delta H_{A,corr}$ (data from Table 1) for a hypothetical rigid body association reaction; see the text for detailed discussion. Dotted line: linear fit of $\delta\Delta H_A^{cal}/\delta T$ yields $\Delta C_{p,A}$ of -2.3 ± 0.2 kJ K^{-1} mol^{-1}. Heavy line: second-order polynomial fit of $\delta\Delta H_A^{cal}/\delta T$ yields increasing $\Delta C_{p,A}$ from -1.4 kJ K^{-1} mol^{-1} at 4 °C to -2.9 kJ K^{-1} mol^{-1} at 30 °C. Thin solid line: linear fit of $\delta\Delta H_A^{cal}/\delta T$ yields $\Delta C_{p,A,corr}$ of -1.8 ± 0.2 kJ K^{-1} mol^{-1} for a hypothetical rigid-body association reaction.

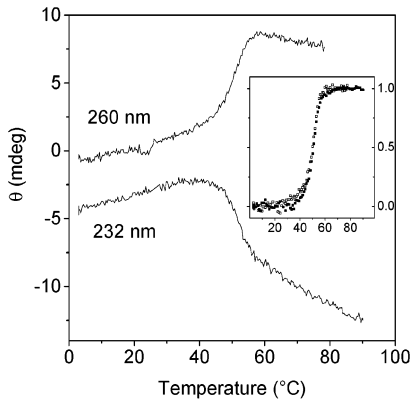


FIGURE 5: Thermal melting of the protein−DNA complex followed by CD. The change of ellipticity at 260 nm (melting of DNA) and 232 nm (melting of protein) was measured with 60 μM complex in standard buffer at a heating rate of 1 deg min^{-1}. Inset: progress of reaction expressed as fraction of melted complex. Closed circles, melting of protein followed at 232 nm; open circles, melting of DNA followed at 260 nm.

components. An order of 3 means simultaneous dissociation and unfolding of the components. The enthalpy changes calculated from the CD melting curves, ΔH_m^{CD}, are 470 kJ mol^{-1} for $n = 2$ and 610 kJ mol^{-1} for $n = 3$. As will be shown below, the analysis of the DSC melting trace indicates that n has a value between 2 and 3.

Thermal Unfolding Followed by DSC. The temperature-induced conformational transitions of the complex are highly

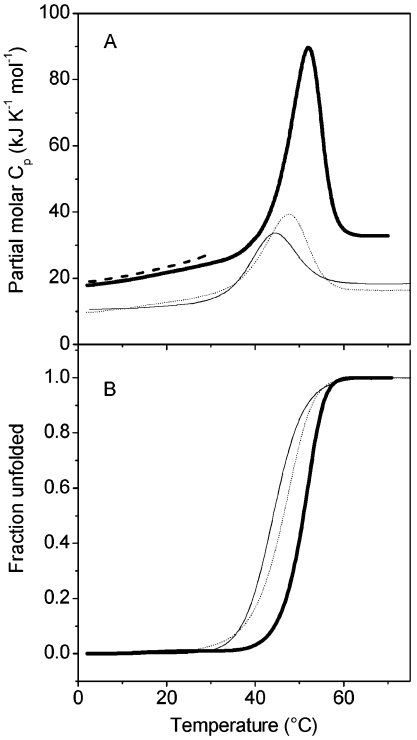


FIGURE 6: Heat capacity curves of the complex and the isolated components obtained by DSC. (A) Thermograms recorded in standard buffer of pH 6.0 at a heating rate of 1 deg min^{-1} with sample concentrations of 60 μM. Heavy line, complex; thin line, free protein; dotted line, DNA; dashed line, calculated sum of partial molar heat capacities of free protein and free DNA. Data for free protein and free DNA are from (*20*). (B) Progress of melting calculated from the changes of heat absorption. Line type is the same as that in panel A.

reversible at pH 6 in 0.1 M NaCl. The partial molar heat capacity function is shown in Figure 6A together with the traces recorded for the isolated components described in the accompanying paper (*20*). Melting of the complex produces a single sharp heat absorption peak whose temperature of maximum heat absorption (which can be regarded as the apparent transition temperature or melting temperature T_m) is higher than the melting temperatures of protein and DNA alone. The shape of the DSC trace of the complex is asymmetric, which indicates a cooperative unfolding process accompanied by subunit dissociation (*25*). The DSC trace demonstrates that the components are thermally stabilized in the complex (Figure 6B). The dissociation of the complex and the concurrent unfolding of the protein and the DNA are taking place within a narrow temperature interval.

The Heat Capacity Change at T_m Appears Negligible. The heat capacity of the complex increases linearly up to about 40 °C and remains constant above 60 °C, indicating that no further temperature-induced changes occur after the main transition (Figure 6A). Extrapolation of the pre-transitional and post-transitional heat capacities into the transition zone shows that the net heat capacity change of dissociation and unfolding of the complex, $\Delta C_{p,C}$, is almost negligible around T_m. In seven experiments it scatters form -0.9 to $+0.9$ kJ K^{-1} mol^{-1} (dotted line in Figure 7). This is unexpected since $\Delta C_{p,C}$ is composed of the heat capacity change of protein unfolding ($\Delta C_{p,P}$), DNA unfolding ($\Delta C_{p,D}$), and the dissociation reaction (equivalent to $-\Delta C_{p,A}$) according to $\Delta C_{p,C} = \Delta C_{p,P} + \Delta C_{p,D} - \Delta C_{p,A}$. From experiments described in the

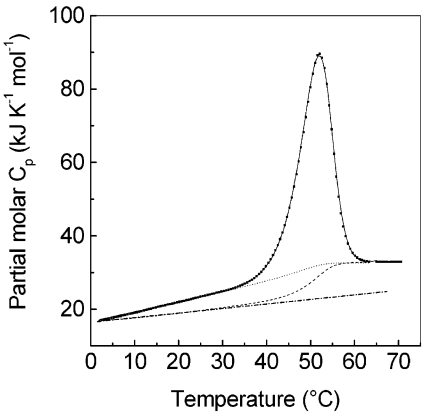


FIGURE 7: Partial molar heat capacity of the protein−DNA complex. Experimental conditions as in Figure 6. Symbols, experimental data. Continuous line, van't Hoff analysis according to eqs 4−6 in Supporting Information with the following best fit parameters: $n = 2.3$, $T_{m0.5} = 51.1$ °C, $\Delta H = 572$ kJ mol^{-1}, and $\Delta C_p = -0.15$ kJ K^{-1} mol^{-1}. (Note that these parameters describe the experimental trace shown and differ from the mean of all experiments given in the text.) Dotted line, apparent intrinsic heat capacity function connecting the pre-transitional and post-transitional parts of the experimental trace and indicating negligible total heat capacity increment. Dashed line, intrinsic heat capacity of a hypothetical "nonfluctuating" complex undergoing thermal melting with a total heat capacity increment of 10 kJ K^{-1} mol^{-1}; see the text for detailed discussion. Dash−dotted line, heat capacity function of a hypothetical "nonfluctuating" complex.

accompanying paper (*20*), $\Delta C_{p,P}$ of protein unfolding is 5 kJ K^{-1} mol^{-1}, and $\Delta C_{p,D}$ of DNA unfolding is 1−2 kJ K^{-1} mol^{-1}. $\Delta C_{p,A}$ of complex association from ITC is −2 to −3 kJ K^{-1} mol^{-1} (Figure 4). Hence, the expected value of $\Delta C_{p,C}$ is 8−11 kJ K^{-1} mol^{-1}, contra to experiment.

Why is the apparent $\Delta C_{p,C}$ deduced from the DSC melting curve so much smaller than expected? A possible explanation is that the complex dissociates into partly folded components so that the heat capacity is smaller than the sum of the heat capacities of the fully unfolded protein and the dissociated unstructured DNA strands. This interpretation can be excluded since the sum of the heat capacities of the protein and the DNA above 60 °C is within ±2 kJ K^{-1} mol^{-1} of the heat capacity of the melted complex (Figure 6A). An alternative explanation may be hidden in the steep heat capacity increase of the native complex below 40 °C.

To further analyze this problem, we calculate the apparent calorimetric enthalpy of complex melting and dissociation, ΔH_C^{cal}. The value corresponds to the area above a line smoothly connecting the pre-transitional and post-transitional portions of the heat capacity trace (dotted line in Figure 7) and equals 540±20 kJ mol^{-1} (mean ± SD of 7 experiments). ΔH_C^{cal} is composed of the enthalpy changes of protein unfolding, DNA unfolding and complex dissociation (equivalent to $-\Delta H_A$): $\Delta H_C^{cal} = \Delta H_P + \Delta H_D - \Delta H_A$. Using the values determined for protein and DNA unfolding reported in the accompanying paper (*20*) and ΔH_A from the ITC experiment extrapolated[3] to the melting region, one obtains ΔH_C^{cal} of about 750 kJ mol^{-1}. This corresponds to the expected total heat change for the disruption of complex-stabilizing contacts and unfolding of the protein and the DNA. The value is significantly higher than ΔH_C^{cal} of 540 kJ mol^{-1} calculated from the peak area of Figure 7. It follows that the DSC melting trace generates neither the expected enthalpy change nor the expected heat capacity change of complex dissociation and unfolding.

Thermal Fluctuations are Gradually Accumulating in the Native Complex before the Start of the Main Melting Transition. We propose that the discrepant values of measured and expected values of ΔH_C^{cal} and $\Delta C_{p,C}$ can be explained if one assumes that thermal fluctuations are gradually accumulating in the complex upon heating. This effect is illustrated in Figure 7. If, as discussed above, one assumes an expected total heat capacity change ($\Delta C_{p,C}$) of complex dissociation plus protein and DNA melting of 8−11 kJ K^{-1} mol^{-1}, the heat capacity of a hypothetical "nonfluctuating" native complex at 65 °C can be obtained by subtracting 8−11 kJ mol^{-1} from the high-temperature end of the trace of Figure 7. This is shown by the dash−dotted line of Figure 7, which represents the heat capacity function of a hypothetical "nonfluctuating" protein−DNA complex lacking a steep increase of thermal fluctuation in the native state. The enthalpy change ΔH_C^{cal} of unfolding and dissociation of such a "nonfluctuating" complex corresponds to the area above the dashed line shown in Figure 7 and is 730 kJ mol^{-1}, in agreement with ΔH_C^{cal} of 750 kJ mol^{-1} calculated from $\Delta H_C^{cal} = \Delta H_P + \Delta H_D - \Delta H_A$. The slope of the heat capacity of the "nonfluctuating" complex is 125 J K^{-2} mol^{-1}, which is almost exactly the sum of the heat capacity temperature slopes of the DNA (69 J K^{-2} mol^{-1}) and the protein (55 J K^{-2} mol^{-1}) reported in the accompanying paper (*20*).

Our analysis of the DSC melting trace indicates that the heat capacity of the complex increases rapidly with temperature. The enthalpic content of the system is significantly reduced at temperatures preceding the main transition. This indicates gradual destabilization of the complex. However, thermal fluctuations are linked to gain in entropy. Obviously, gradual enthalpic and entropic effects are effectively compensating each other, so that the Gibbs energy of association, ΔG_A, remains almost constant (Figure 4A).

Van't Hoff Analysis of DSC Trace Confirms High Cooperativity of Complex Dissociation and Thermal Unfolding. The melting profile of the complex shown in Figure 6A can be subjected to van't Hoff analysis. The reaction order *n* of eqs 3−6 in Supporting Information is not defined. As discussed for CD melting above, limiting values of *n* are 2 for the dissociation to the native components and 3 for the dissociation to unfolded protein and DNA strands. Simulation of the DSC trace according to eqs 4−6 in Supporting Information yields $n = 2.3 \pm 0.1$. The corresponding enthalpy change, ΔH_C^{vH}, is 560 ± 20 kJ mol^{-1}, very similar to the measured ΔH_C^{cal} of 540 ± 20 kJ mol^{-1}. Thus, the ratio $\Delta H_C^{cal}/\Delta H_C^{vH}$ is 1, in support of a highly cooperative unfolding reaction. The enthalpies calculated for fixed n are 530 kJ mol^{-1} ($n = 2$) and 610 kJ mol^{-1} ($n = 3$). These limiting values are almost identical to those obtained from the CD melting curves (Figure 5): $\Delta H_m^{CD} = 470$ kJ mol^{-1} ($n = 2$) and 610 kJ mol^{-1} ($n = 3$). Equally, fitting of the CD melting curves shown in Figure 5 with *n* fixed at 2.3 yields ΔH_m^{CD} of 525 kJ mol^{-1}, in good agreement with ΔH_C^{vH} of 560 ± 20 kJ mol^{-1} and ΔH_C^{cal} of 540 ± 20 kJ mol^{-1}.

[3] $\Delta C_{p,A}$ of −2.3 kJ K^{-1} mol^{-1} from the linear regression of the data in Figure 4 was used for this extrapolation.

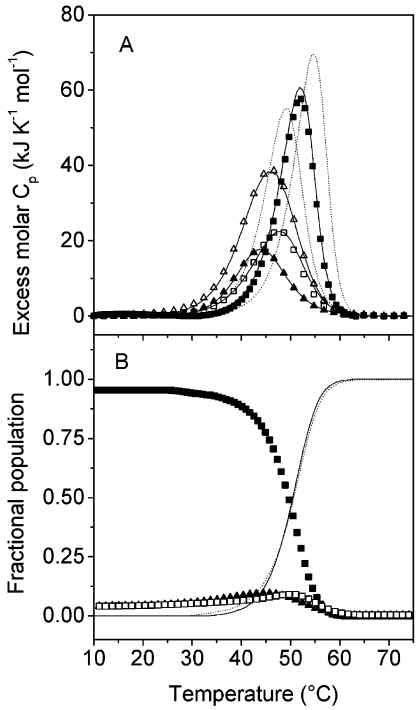
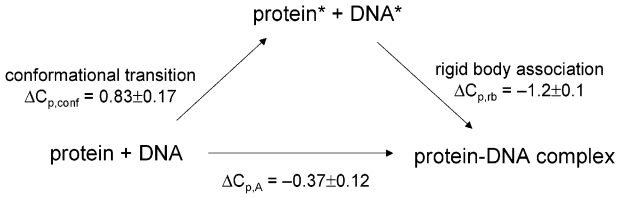


FIGURE 8: Simulations of protein−DNA association according to the procedure of Brandts and Lin (*30*). (A) Symbols are the experimental excess molar heat capacity from DSC for the complex (filled squares), the free protein (open triangles), the free DNA (open squares), and the arithmetic sum of free protein and free DNA (open triangles). Solid lines are simulated excess heat capacity functions calculated with the help of the combined eqs 3−8 and the parameters $T_{m,protein} = 44.2$ °C, $\Delta H_{m,protein} = 245$ kJ mol^{-1}, $\Delta C_{p,protein} = 4.5$ kJ K^{-1} mol^{-1}, $T_{m,DNA} = 46.1$ °C, $\Delta H_{m,DNA} = 320$ kJ mol^{-1}, $\Delta C_{p,DNA} = 1.5$ kJ K^{-1} mol^{-1}, $\Delta H_A(25\ °C) = -30$ kJ K^{-1}, $\Delta C_{p,A} = -2.3$ kJ K^{-1} mol^{-1}. The experimental melting temperature of the complex is reproduced best with $K_A(25\ °C) = 6.4 \times 10^6$ M^{-1}. Dotted lines are simulations performed with K_A 1 order of magnitude lower or higher than 6.4×10^6 M^{-1}. (B) Calculated temperature-induced changes of the fractional population of free native protein, free DNA duplex, and protein−DNA complex (symbols as in panel A). Solid line, free unfolded protein; dotted line, free single-stranded DNA.

The striking similarity of the enthalpy changes deduced from the CD melting curves, from van't Hoff analysis of the DSC trace and from direct calorimetric analysis strongly supports high cooperativity between complex dissociation and unfolding of the components. The reaction order of 2.3, on the other hand, may indicate that the complex dissociates into components that may not yet be completely unfolded.

Simulation of the Heat Capacity Trace by Linked-Equilibria Analysis. The excess heat capacity function can be analyzed further by a model which takes into account the temperature-induced changes in the concentration of all the five molecular species populated in the temperature range of the DSC experiment: complex, native and unfolded protein, duplex DNA, and unfolded DNA strands. This analysis combines the energetics of folding of the components described in the accompanying paper (*20*) with the energetics of complex formation described here. Although developed more than 10 years ago (*30*), the deconvolution technique was applied to protein−DNA interaction only once before (*39*). Figure 8A presents the results of simulations according to the combined eqs 3−8. The experimental excess heat capacity profile of the 1:1 complex is very well simulated. The position of the heat absorption peak, reflecting the mutual stabilization of the individual components due to binding interactions, is reproduced with $K_A(25\ °C)$ of 6.4×10^6 M^{-1}, which is the same within error as K_A of $(7 \pm 5) \times 10^6$ M^{-1} from ITC and K_A of $(8 \pm 3) \times 10^6$ M^{-1} from fluorescence quenching. $\Delta H_A^{cal}(25\ °C)$ determined by ITC is predicted within ±15 kJ mol^{-1}. Figure 8B shows the calculated population of each molecular species. In the transition region, melting of the complex releases small amounts of free folded protein and free DNA duplex in equilibrium with large amounts of unfolded protein and DNA. This is in accord with a reaction order between 2 and 3 obtained from van't Hoff analysis of the DSC trace.[4] The success of the modeling indicates that the thermodynamic parameters obtained by ITC and DSC describe the thermodynamic behavior in a consistent way, bridging the entire temperature and concentration range covered by the ITC and DSC experiments.

Scheme 1



The Heat Capacity Change of Protein−DNA Association. Analysis of the Partial Molar Heat Capacities Explains the Temperature Dependence of $\Delta C_{p,A}$. A peculiar feature of the protein−DNA association reaction studied here is the nonlinear temperature dependence of the association enthalpy seen in Figure 4B. We now wish to explain this behavior in terms of temperature-induced heat capacity changes occurring in the temperature interval where the native form of the complex and its components dominate. Such changes are clearly seen between 4 and 30 °C in Figure 6A. The calculated sum of the heat capacitiy slopes of protein and DNA (dashed line in Figure 6A) increases nonlinearly from 0.24 kJ K^{-2} mol^{-1} near 4 °C to about 0.40 kJ K^{-2} mol^{-1} near 30 °C, the average over the ITC temperature range being 0.32 kJ K^{-2} mol^{-1}. At the same time, the heat capacity of the complex displays a nearly constant temperature slope of 0.24 kJ K^{-2} mol^{-1}. This behavior indicates the following: (i) In the temperature range before the main thermal transition, the free protein, the free DNA and the complex exhibit minor changes of conformation, of thermal fluctuations and of vibrational content, including changes in the water shell around the molecules. (ii) In the complex and its components the changes are not occurring in parallel. (iii) The enthalpy fluctuations of free protein and free DNA duplex are attenuated when association has taken place. These observations explain in a qualitative way $\delta\Delta H_A^{cal}/\delta T$ is curved (heavy line in Figure 4B).

If we were able to account in a quantitative way for the different thermal properties of the system in its associated and dissociated states, we could predict the enthalpic behavior of a hypothetical rigid body association reaction between the protein and the DNA in their binding competent conformations (see Scheme 1 below). In other words, if we

[4] Populations of different molecular species depend on the total concentrations of protein and DNA.

could subtract all the temperature-dependent (nonlinear) changes of C_p of the free components from the main trace of complex melting of Figure 6A, we would obtain the enthalpy change of the hypothetical rigid body association reaction. The necessary procedure has been developed by Privalov and colleagues (*10*). The enthalpy of association at any temperature can be written as

$$\Delta H_A(T) = \Delta H_A(T_R) + \Delta C_{p,A}(T_R) \times (T - T_R) + \lambda \quad (10)$$

where λ is defined as

$$\int_{T_R}^{T} \{[C_p(T) - C_p(T_R)]^c - \sum_i [C_p(T) - C_p(T_R)]^i\}\, dT \quad (10a)$$

Superscript i denotes free protein and free DNA, respectively, and superscript c denotes the complex. $\Delta C_{p,A}(T_R)$ is defined as

$$\Delta C_{p,A}(T_R) = C_p(T_R)^c - \sum_i C_p(T_R)^i \quad (10b)$$

Equation 10 without the integral λ describes the ideal situation of a rigid body reaction for which $\Delta C_{p,A}$ is temperature-independent in the first approximation (in a sufficiently narrow interval of temperatures) and corresponds to the difference between C_p of the rigid complex and C_p of its rigid components. The integral λ represents the difference between the temperature dependence of C_p of the complex and the summed temperature slopes of C_p of free protein and DNA. Thus, we can define a "corrected" enthalpy change of association as

$$\Delta H_{A,corr} = \Delta H_A^{cal} - \lambda \quad (11)$$

Equation 11 defines the association enthalpy change of the rigid body reaction. Values of $\Delta H_{A,corr}$ were calculated from the combined eqs 10b and 11.[5] They are added as asterisks to Figure 4B and are shown in Table 1 in Supporting Information. Most satisfying, the change of $\Delta H_{A,corr}$ with temperature is linear, as expected for rigid body association. The corrected heat capacity change, $\Delta C_{p,A,corr} = \delta \Delta H_{A,corr}/\delta T$, is -1.8 ± 0.08 kJ K^{-1} mol^{-1} (slope of thin straight line in Figure 4B).

In conclusion, analysis of the heat capacities of free protein and DNA and their complex measured by DSC confirms the curvature of the plot of ΔH_A^{cal} versus T measured by ITC. The temperature dependence of $\Delta C_{p,A}$ is the consequence of nonparallel changes in the thermal fluctuations of the system in its dissociated and associated states. It is important to recognize that the "correction" of the ITC-measured enthalpy does not imply that $\Delta H_{A,corr}$ is the "true" binding enthalpy while the apparent experimental ΔH_A^{cal} is "wrong". Experimental $\Delta H_A{}^{cal}$ is the association enthalpy change of the molecules in their real conformational state, including thermal fluctuations, conformational changes, and perhaps even partial unfolding. Our analysis merely highlights the thermodynamic complexity of the association reaction and provides a rationale for the nonlinear change of ΔH_A^{cal} with temperature. Furthermore, $\Delta C_{p,A,corr}$ of the hypothetical rigid-body association reaction can now be judged against the structure of the complex, that means with regard to the molecular surface buried at the complex interface.

Correlation of the Heat Capacity Change of Association with Structural Features. ΔC_p provides a semiempirical link between structure and energetics and has attracted special attention as a key parameter in estimating the contribution of dehydration to binding (*1, 2*). Experimental ΔC_p often correlates well with the amount and type of surface buried at the interface of a complex. However, in the case of protein−DNA association, significant discrepancies between measured and calculated ΔC_p have been reported (*6*−*10*). The present system is a particularly telling example.

The conformational changes accompanying complex formation are reflected by the different solvent accessible surface area calculated for the complex and its free and bound components. The association reaction can be formally divided into two steps (Scheme 1): (i) the transition of protein and DNA from the free to the binding-competent conformations (marked by asterisks in Scheme 1) and (ii) the association of the binding-competent molecules to the complex.

The heat capacity changes can be estimated from the dehydration of nonpolar and polar surface area according to eq 15. Estimates are based on empirical correlations established for many protein−protein and protein−DNA interactions (*1, 2, 32*−*34*). Interestingly, the protein and the DNA unfold slightly to expose more binding surface in their binding competent conformation (+450 Å^2 of nonpolar surface and −80 Å^2 of polar surface). Therefore, $\Delta C_{p,conf}$ is positive with an estimated value of 0.83 kJ K^{-1} mol^{-1}. Association of the binding competent molecules buries 790 Å^2 of aliphatic, 540 Å^2 of aromatic, and 1040 Å^2 of polar surface, leading to a negative value of $\Delta C_{p,rb}$ of −1.2 kJ K^{-1} mol^{-1}. The estimated overall heat capacity change of association is $\Delta C_{p,A} = \Delta C_{p,conf} + \Delta C_{p,rb}$ and is very much smaller than any of the measured values. Our estimates are based on elementary contributions per Å^2 of polar and nonpolar area (eq 9) taken from (*34*). Estimated values of ΔC_p differ by only 10−20% when other parametrizations are used (*1, 2, 32*−*34*) (not shown).

It should be emphasized that the calculated ΔC_p's are based on the simple assumption that the bulk contribution to ΔC_p originates from dehydration of interacting surface in the complex. Alteration of thermal motions and vibrational modi are not considered. Nevertheless, the structural estimates of ΔC_p provide us with a guideline for assessing the measured thermodynamic parameters.

Circumstantial Evidence for Residual Water at the Complex Interface. Our analysis of the DSC data leading to $\Delta C_{p,A,corr}$ eliminates the contributions arising from conformational adaptation and thermal fluctuations. Therefore, $\Delta H_{A,corr}$ represents the enthalpy of intermolecular contacts plus the changes of hydration upon association of rigid molecular surfaces. Hence, $\Delta C_{p,A,corr}$ should be equal to the structure-based estimate of $\Delta C_{p,rb}$. This is not the case: $\Delta C_{p,A,corr}$ is −1.8 kJ K^{-1} mol^{-1}, and $\Delta C_{p,rb}$ is −1.2 kJ K^{-1} mol^{-1}. How can we explain the "extra" −0.6 kJ K^{-1} mol^{-1}? We propose that they arise from incomplete dehydration of the complex interface.

It has long been recognized that water molecules play an important role in protein−DNA association (*40*). Structural

[5] For integration, T_R = 4 °C was used since the difference $(\delta C_p/\delta T)^c - \Sigma(\delta C_p/\delta T)^i$ is near zero below 10 °C and the experimental value of $\Delta C_{p,A}(T_R)$ obtained from DSC is close to $\Delta C_{p,A}$ obtained from ITC data below 15 °C.

and computational analysis have identified water molecules bridging protein and DNA groups and often participating in hydrogen bonding networks (*41*). Strong evidence for the proposed presence of water comes from packing density calculations. Using the NMR structures of complex, protein and DNA, we identify a total of 140 ± 40 Å^3 of "empty space" distributed over 6−7 cavities at the complex interface. These cavities are large enough to contain together about 10 water molecules that are inaccessible to the bulk solvent. Since the "extra" heat capacity change is negative, it can be attributed to partial dehydration of polar surface (the heat capacity of polar hydration is negative). Thus, the −0.6 kJ K^{-1} mol^{-1} can be accounted for by assuming that the buried polar surface is only ∼60% dehydrated. This estimation is based on an elementary contribution of hydration of −1.27 J K^{-1} mol^{-1} Å^{-2} (*34*). Dehydration is even less than 60% if one takes a value of −1.09 J K^{-1} mol^{-1} Å^{-2} from the parametrization advanced by Freire and colleagues, (*2*).

Incomplete Dehydration Can be Reconciled with the Enthalpy of Association. The association enthalpy measured by ITC is zero near 12−14 °C (Figure 4). The change from endothermic to exothermic heat effect close to room temperature indicates the importance of hydrophobic interactions since the enthalpy of apolar contacts is almost exactly canceled by the corresponding dehydration enthalpy, hence the total enthalpic effect of hydrophobic interaction vanishes at 25 °C (*34*, *42*). The relatively small binding enthalpy in this temperature range results from weak intermolecular interactions, i.e., hydrogen bonds and other polar contacts, or from mutually compensating enthalpic effects. The enthalpy of an association process approximating rigid-body binding can be formally decomposed into four contributions:

$$\Delta H_A = \Delta H_{int}^{apol} + \Delta H_{int}^{pol} + \Delta H_{dehydr}^{apol} + \Delta H_{dehydr}^{pol} \qquad (12)$$

where the terms ΔH_{int} represent all the intermolecular contacts in a vacuum and the terms ΔH_{dehydr} all the enthalpy effects of surface dehydration; superscripts apol and pol refer to contacts between nonpolar and polar surfaces, respectively. For energy parsing we apply the parametrization of Privalov and colleagues (*34*) because it clearly separates between vacuum energies and energy contributions from hydration based on calorimetric results. Using the total amount of buried surface to predict the total enthalpy effect of dehydration, we obtain about 1460 kJ mol^{-1} for ΔH_{dehydr}^{apol} + ΔH_{dehydr}^{pol}, of which about 90% arises from the very unfavorable enthalpy of burying polar groups at the complex interface. The enthalpy of van der Waals contacts involving aliphatic and aromatic groups, ΔH_{int}^{apol}, is only about −195 kJ mol^{-1}. Since the total association enthalpy of the rigid body association reaction at 25 °C, ΔH_A^{corr}, is −19 kJ mol^{-1} (Figure 4B), one obtains ΔH_{int}^{pol} of −1270 kJ mol^{-1} from insertion in eq 12. It is reasonable to assume that the bulk of ΔH_{int}^{pol} originates from hydrogen bonds at the complex interface. Again with the parametrization from (*34*), we take a value of −45 to −60 kJ mol^{-1} for the enthalpic content of a single hydrogen bond and predict about 21−28 hydrogen bonds at the complex interface. There are only 13 residues capable of forming H-bonds between the protein and the DNA. Hence, the predicted number of H-bonds is too large. However, if only 60% of the buried polar surface were dehydrated upon binding, ΔH_{int}^{pol} would be only about −760 kJ mol^{-1} and would agree with 13 to 17 H-bonds. This estimate is corroborated by NMR structural data (*21*) and MD simulations (A. G. and I. J., manuscript in preparation). Thirteen H-bonds can be observed in at least 2−3 conformers of the NMR ensemble. Up to 16 hydrogen bonds can be identified in some conformers after optimization of the hydrogen positions. Ten residues are hydrogen-bonded to DNA bases and backbone groups throughout a 2 ns MD simulation. Furthermore, if water molecules are present at the interface, they also may serve as donors or acceptors of additional H-bonds. Indeed, we observe 7−8 water molecules inaccessible to bulk water in the simulations. Altogether, the enthalpy parsing analysis reconciles structural features of the complex with the observed heat capacity changes if one assumes that only 60% of the complex interface is dehydrated.

In concluding this discussion of enthalpy parsing, we note that polar interactions contribute only little to the enthalpy of binding around room temperature. Incidentally, the calculated enthalpy of hydrophobic bonding (ΔH_{int}^{apol} + ΔH_{dehydr}^{apol}), which includes van der Waals contacts between aliphatic and aromatic groups plus the corresponding dehydration of these groups, is −18 kJ mol^{-1} at 25 °C, which matches perfectly with the association enthalpy of the rigid-body reaction at that temperature (thin line in Figure 4B). Thus, the analysis according to eq 12 seems reasonable.

GENERAL DISCUSSION

This is the first thermodynamic analysis of DNA recognition by a three-stranded β-sheet. It is instructive to compare the presented thermodynamic parameters with the energetics of association reactions promoted by other more common recognition motifs. Sequence-specific protein binding to DNA has been investigated by calorimetry in several systems and representative data have been discussed (*10*, *14*). It appears that significant negative heat capacity changes are the only common feature of several protein−DNA association reactions. This is reasonable since protein−DNA complexes bury a large amount of apolar molecular surface. As for other systems, $\Delta C_{p,A}$ of the integrase−DNA complex is large and exceeds by far the expected heat capacity effect of surface dehydration. It seems that values of $\Delta C_{p,A}$ that are too large to originate from hydration effects are a hallmark of specific protein−DNA binding, although there may be exceptions (*43*). Interestingly, the Kirchoff plot of ΔH_A versus T is curved. We propose that a nonlinear change of association enthalpy with temperature is caused by the nonparallel and nonlinear change of the heat capacity C_p of the free components and the complex before the main thermal transition.

Unlike heat capacity changes, the enthalpies and entropies of protein−DNA association are highly variable. For example, DNA binding of the γ cI repressor has a favorable ΔH_A of −100 kJ mol^{-1} and an unfavorable $T\Delta S_A$ of −60 kJ mol^{-1} at 25 °C (*44*). Just the opposite is seen for DNA recognition by the TATA-box binding protein, which at 25 °C is opposed by ΔH_A of 125 kJ mol^{-1} and favored by $T\Delta S_A$ of 140 kJ mol^{-1} (*9*). As a rule, enthalpic and entropic contributions cancel to a large extent so that the binding free energy is typically in the range of 30−45 kJ mol^{-1} (*14*).

However, because $\Delta C_{p,A}$ is usually large, the enthalpic and entropic contributions to binding are temperature dependent and my change sign at physiological conditions of temperature. At room temperature, ΔH_A and ΔS_A of the integrase−DNA complex is similar to the enthalpy and entropy change of three other major groove binders (*10*): the *Antennapedia* C39S homeodomain, the three zinc finger fragment of TFIIIA, and the methionine repressor dimer MetJ. Only MetJ, a member of the tetrameric ribbon-helix-helix protein family, recognizes the target DNA by β-sheets, similar to integrase Tn916. However, there is little structural homology between MetJ and INT-DBD except that two (MetJ) or three (INT-DBD) β-sheet strands are inserted into the major groove with every second side chain pointing toward the DNA. The energetic similarity of the MetJ−DNA and the integrase−DNA complex could be accidental.

The search for structure-energy relationships in protein−DNA association is greatly hampered by the fact that binding is accompanied by structural rearrangements which are system-specific and difficult to account for. Conformational adaptation is typical for macromolecular recognition. The present study illustrates that adaptation occurs at the cost of introducing conformational disorder in parts of a binding domain. The results emphasize that structure-based methods for the prediction of heat capacity changes (and other thermodynamic parameters) fail when thermal motions and soft vibrational modi of the components are severely altered in the complex. However, we present a useful and promising way to deal with this complication, namely to link ITC measurements of protein−DNA association with the independent DSC analysis of the complex and its free components.

ACKNOWLEDGMENT

We thank Dr. Peter Gehrig for mass spectroscopy analysis.

SUPPORTING INFORMATION AVAILABLE

Equations used for van't Hoff analysis of thermal melting of the protein−DNA complex measured by CD and DSC. Table with thermodynamic parameters of the association of the integrase DNA binding domain with the 13 bp duplex DNA observed by ITC. Next to the directly measured ΔH_A^{cal}, ΔG_A, and $T\Delta S_A$, the Table lists the values of $\Delta H_{A,corr}$ and $T\Delta S_{A,corr}$. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Baker, B. M., and Murphy, K. P. (1998) *Methods Enzymol. 295*, 294−315.
2. Luque, I., and Freire, E. (1998) *Methods Enzymol. 295*, 100−127.
3. Sturtevant, J. M. (1994) *Curr. Opin. Struct. Biol. 4*, 69−78.
4. Cooper, A. (2000) *Biophys. Chem. 85*, 25−39.
5. Cooper, A., Johnson, C. M., Lakey, J. H., and Nollmann, M. (2001) *Biophys. Chem. 93*, 215−230.
6. Berger, C., Jelesarov, I., and Bosshard, H. R. (1996) *Biochemistry 35*, 14984−14991.
7. Ladbury, J. E., Wright, J. G., Sturtevant, J. M., and Sigler, P. B. (1994) *J. Mol. Biol. 238*, 669−681.
8. Lundback, T., Chang, J. F., Phillips, K., Luisi, B., and Ladbury, J. E. (2000) *Biochemistry 39*, 7570−7579.
9. O'Brien, R., DeDecker, B., Fleming, K. G., Sigler, P. B., and Ladbury, J. E. (1998) *J. Mol. Biol. 279*, 117−125.
10. Privalov, P. L., Jelesarov, I., Read, C. M., Dragan, A. I., and Crane-Robinson, C. (1999) *J. Mol. Biol. 294*, 997−1013.
11. Dyson, H. J., and Wright, P. E. (2002) *Curr. Opin. Struct. Biol. 12*, 54−60.
12. Dickerson, R. E., and Chiu, T. K. (1997) *Biopolymers 44*, 361−403.
13. Spolar, R. S., and Record, M. T., Jr. (1994) *Science 263*, 777−784.
14. Jen-Jacobson, L., Engler, L. E., and Jacobson, L. A. (2000) *Structure 8*, 1915−1923.
15. Sturtevant, J. M. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2236−2240.
16. Tidor, B., and Karplus, M. (1994) *J. Mol. Biol. 238*, 405−414.
17. Morton, C. J., and Ladbury, J. E. (1996) *Protein Sci. 5*, 2115−2118.
18. Dunitz, J. D. (1994) *Science 264*, 670.
19. Madan, B., and Sharp, K. A. (1999) *Biophys. J. 81*, 1881−1887.
20. Milev, S., Gorfe, A. A., Karshikoff, A., Clubb, R. T., Bosshard, H. R. and Jelesarov, I. (2003) *Biochemistry 42*, 3492−3502.
21. Wojciak, J. M., Connolly, K. M., and Clubb, R. T. (1999) *Nat. Struct. Biol. 6*, 366−373.
22. Marky, L. A., and Breslauer, K. J. (1987) *Biopolymers 26*, 1601−1620.
23. Plotnikov, V. V., Brandts, J. M., Lin, L. N., and Brandts, J. F. (1997) *Anal. Biochem. 250*, 237−244.
24. Freire, E., and Biltonen, R. L. (1978) *Biopolymers 17*, 463−479.
25. Freire, E. (1989) *Comments Mol. Cell. Biophys. 6*, 123−140.
26. Privalov, P. L., and Potekhin, S. A. (1986) *Methods Enzymol. 131*, 4−51.
27. Kholodenko, V., and Freire, E. (1999) *Anal. Biochem. 270*, 336−338.
28. Karshikoff, A., and Ladenstein, R. (1998) *Protein Eng. 11*, 867−872.
29. Makhatadze, G. I., Medvedkin, V. N., and Privalov, P. L. (1990) *Biopolymers 30*, 1001−1010.
30. Brandts, J. F., and Lin, L. N. (1990) *Biochemistry 29*, 6927−6940.
31. Hubbard, S. J., and Thornton, J. M. (1993). *NACCESS*, Computer Program, Department of Biochemistry and Molecular Biology, University College, London.
32. Spolar, R. S., Livingstone, J. R., and Record, M. T., Jr. (1992) *Biochemistry 31*, 3947−3955.
33. Murphy, K. P., and Freire, E. (1992) *Adv. Prot. Chem. 43*, 313−361.
34. Makhatadze, G. I., and Privalov, P. L. (1995) *Adv. Prot. Chem. 47*, 307−425.
35. Connolly, K. M., Wojciak, J. M., and Clubb, R. T. (1998) *Nat. Struct. Biol. 5*, 546−550.
36. Allen, M. D., Yamasaki, K., Ohme-Takagi, M., Tateno, M., and Suzuki, M. (1998) *EMBO J. 17*, 5484−5496.
37. Flick, K. E., Jurica, M. S., Monnat, R. J., Jr., and Stoddard, B. L. (1998) *Nature 394*, 96−101.
38. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) *Anal. Biochem. 179*, 131−137.
39. Carra, J. H., and Privalov, P. L. (1997) *Biochemistry 36*, 526−535.
40. Schwabe, J. W. (1997) *Curr. Opin. Struct. Biol. 7*, 126−134.
41. Reddy, C. K., Das, A., and Jayaram, B. (2001) *J. Mol. Biol. 314*, 619−632.
42. Baldwin, R. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8069−8072.
43. Gonzalez, M., Weiler, S., Ferretti, J. A., and Ginsburg, A. (2001) *Biochemistry 40*, 4923−4931.
44. Merabet, E., and Ackers, G. K. (1995) *Biochemistry 34*, 8554−8563.

BI0269355