Structural and Thermodynamic Strategies for Site-Specific DNA Binding Proteins

Linda Jen-Jacobson,* Lisa E. Engler, and Lewis A. Jacobson
Department of Biological Sciences
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

Summary

Background: Site-specific protein-DNA complexes vary greatly in structural properties and in the thermodynamic strategy for achieving an appropriate binding free energy. A better understanding of the structural and energetic engineering principles might lead to rational methods for modification or design of such proteins.

Results: A novel analysis of ten site-specific protein-DNA complexes reveals a striking correspondence between the degree of imposed DNA distortion and the thermodynamic parameters of each system. For complexes with relatively undistorted DNA, favorable enthalpy change drives unfavorable entropy change, whereas for complexes with highly distorted DNA, unfavorable ∆H is driven by favorable ∆S. We show for the first time that protein-DNA associations have isothermal enthalpy-entropy compensation, distinct from temperature-dependent compensation, so ∆H and ∆S do not vary independently. All complexes have favorable ∆H from direct protein-DNA recognition interactions and favorable ∆S from water release. Systems that strongly distort the DNA nevertheless have net unfavorable ∆H as the result of molecular strain, primarily associated with the base pair destack. These systems have little coupled protein folding and the strained interface suffers less immobilization, so ∆S is net favorable. By contrast, systems with little DNA distortion have net favorable ∆H, which must be counterbalanced by net unfavorable ∆S, derived from loss of vibrational entropy (a result of isothermal enthalpy-entropy compensation) and from coupling between DNA binding and protein folding.

Conclusions: Isothermal enthalpy-entropy compensation implies that a structurally optimal, unstrained fit is achieved only at the cost of entropically unfavorable immobilization, whereas an enthalpically weaker, strained interface entails smaller entropic penalties.

Introduction

The structure and function of site-specific DNA binding proteins are matters of considerable interest because of the crucial roles of these proteins in the expression and regulation of genetic information. From a purely structural viewpoint, it is now evident that nature has devised a wide range of strategies to build proteins that recognize a specific DNA base sequence. Protein recognition of a DNA sequence may involve nonpolar contacts and direct or indirect (i.e., water-mediated) hydrogen bonds with the DNA bases as well as direct and water-mediated contacts with the sugars and phosphates of the DNA backbone. These recognition contacts may take place in the DNA major groove, in the minor groove, or in both grooves simultaneously and may be made by α helices, β sheets, or loops [1]. The protein recognition elements in turn may be presented in the context of a variety of structural motifs (e.g., helix-turn-helix, basic leucine zipper, zinc finger, etc.).

In this paper we examine the thermodynamics of site-specific protein-DNA binding in search of some engineering principles that will help take us beyond the structural perspective on the recognition function. Although fewer site-specific DNA binding proteins have been characterized thermodynamically than structurally, we were able to find reasonably complete data for ten proteins representing a range of biological functions, including transcriptional repressors and activators from both prokaryotes and eukaryotes, and two restriction endonucleases. By looking at structural and thermodynamic properties across many systems, we have uncovered a striking systematic relationship between the degree of DNA distortion in the recognition complexes and the thermodynamic strategy for achieving a favorable binding free energy.

Results and Discussion

Enthalpic and Entropic Contributions to Binding Free Energy

Figure 1 shows the enthalpic (∆H) and entropic (∆S) contributions to binding free energy (∆Gbind) at 298 K for ten different site-specific DNA binding proteins of diverse biological functions. It is evident that the thermodynamic strategies for achieving a favorable ∆Gbind are also diverse. For some proteins, a favorable enthalpy change (∆H < 0) drives an unfavorable entropy change (−T∆S > 0), whereas for others unfavorable ∆H is driven by favorable −T∆S. For the set of proteins, both ∆H and −T∆S have an impressively large range, about 60 kcal/mol.

It may seem remarkable at first glance that a single line in Figure 1 fits all the data so well (r = 0.99), but there is a relatively simple explanation. For this set of proteins, the mean ∆Gbind is −11.7 ± 1.6 kcal/mol, with a range from −9 kcal/mol (lac repressor with operator O3) to −16 kcal/mol (lac repressor with a symmetrized lac operator). The point here is not the numerical values (which depend on experimental conditions of salt, pH, etc.) but that ∆Gbind varies over a much smaller range than either ∆H or T∆S. Thus, the familiar Gibbs relationship is ∆G = ∆H − T∆S may be recast as ∆H = T∆S + ∆G so that in a plot (Figure 1) of ∆H versus −T∆S for systems where ∆Gbind varies in a relatively narrow range (∆G ≤ 8±G), we expect to see a linear relationship with slope = −T and an intercept (∆H at −T∆S = 0) of ∆G. For the data in Figure 1, the slope is −0.96 ± 0.02 and the intercept is −11.7 ± 0.3 kcal/mol. Thus, the strong apparent correlation in Figure 1 reflects nothing more than the clustering of the ∆Gbind values.

The narrow range of ∆Gbind for these proteins is itself largely the result of how we operationally define “site-specific” proteins. To perform a site-specific biological function, a protein must bind significantly more tightly to its DNA recognition site

Key words: distortion; DNA-protein interactions; enthalpy-entropy compensation; molecular strain; thermodynamics
than to competing non-specific DNA. Equilibrium binding constants (intrinsic $K_d$) for non-specific DNA are generally in the range of $10^{-10} - 10^6$ (M$^+$ sites)$^{-1}$ [2], corresponding to $\Delta G_{\text{bound}} = -4$ to $-7$ kcal/mol$^+$ sites at 298 K. Thus, the proteins in this set bind from 1000-fold to $>10^7$-fold more tightly to specific than to non-specific DNA. At the opposite end of the scale, binding affinity must not be excessive in order that binding may be both physiologically reversible and not hypersensitive to minor fluctuations in the in vivo concentration of site-specific protein [3]. It was argued from first principles more than 20 years ago [3, 4], when few data were available, that these considerations would place a practical upper bound of $K_d = 10^{12}$ M$^{-1}$, and this estimate has proven remarkably prescient.

Molecular Strain and Design Strategy

In compiling data from ten diverse systems rather than focusing narrowly on any one system, we were provoked to ask a novel question: on what basis can we understand why any protein lies at its own particular position along the line of Figure 1? In what follows, we attempt to unite the structural and thermodynamic perspectives by identifying the various factors that contribute to $\Delta H^\circ$ and $\Delta S^\circ$. In no system is it yet possible to apportion these contributions quantitatively. These are first steps, necessarily tentative and incomplete, to building a language for future discussion of these issues.

Consider first the values of $\Delta H^\circ$. The formation of the recognition contacts, that is, hydrogen-bond, ion-pair, and nonpolar contacts between protein and DNA, is expected to be enthalpically favorable for all systems, although the magnitude of this favorable $\Delta H^\circ$ certainly varies somewhat for structurally different interfaces. Opposing this is an unfavorable $\Delta H^\circ$ from desolvation. As the protein-DNA interfaces form, both polar and nonpolar surfaces are brought together. The desolvation of nonpolar surfaces makes only a small contribution to $\Delta H^\circ$ at 298 K [5], although it is a dominant source of favorable $\Delta S^\circ$ (see below). Polar surfaces apposed at protein-DNA interfaces may entrap water molecules [6-9] or may be relatively “dry” [9, 10] (i.e., desolvated); in the latter case, the desolvation of polar groups would make a significant unfavorable contribution to $\Delta H^\circ$. However, we find no persuasive correlation between the observed $\Delta H^\circ$ for each system and the amount of polar surface buried or the degree of residual solvation of such surface in the protein-DNA complex.

Another potential source of unfavorable $\Delta H^\circ$ is “strain” in the complexes. By strain we mean that atoms, functional groups, or entire residues (DNA bases or protein side chains) have adopted positions in the complex that do not coincide with their own positions of minimum potential energy, as the result of bond bending or rotation and/or unfavorable nonbonded interactions (e.g., steric clashes, electrostatic repulsion). Of course, the net result must be to minimize the free energy of the entire complex (see below).

The most obvious source of strain in protein-DNA complexes is distortion of the DNA, and in this the various complexes differ considerably. At one extreme, in the site-specific complexes of GCN4 transcription factor and λcl repressor, the DNA is in nearly ideal B-form conformation, with little or no distortion of the path of the backbone [1]. This pair of proteins has the most favorable $\Delta H^\circ$ in Figure 1. At the other extreme, the CAP protein bends the DNA by 87° [11, 12] and TATA box binding protein (TBP) introduces a three-dimensional writhe of 80-92° (in various complexes) as well as unwinding the bound DNA [12]. This pair of proteins has the most unfavorable $\Delta H^\circ$ in Figure 1. Other proteins have varying degrees of local DNA distortion and intermediate values of $\Delta H^\circ$. For example, the DNA distortion in the EcoRI endonuclease complex [13] is greater than that in the BamHI endonuclease complex [14], and $\Delta H^\circ$ is more unfavorable for EcoRI [6]. Even for one protein, the trend is discernible: GCN4 binds to the AP-1 site without DNA bending [15] and $\Delta H^\circ$ is more favorable [16] than that for binding to the ATF/CREB site where the DNA is bent by about 20° [17]. We are presently unable to evaluate strain in the proteins. Crystal structures of the free proteins are not available in all cases, and in any event the less regular structure of proteins will usually make strain more difficult to discern. In unusual instances where the parent protein structure is relatively regu-
lar and simple [18], strain in the bound protein is obvious from structure, but the energetic consequences have yet to be determined.

A simple postulate is that the low-DNA-distortion systems have a “baseline” value of $\Delta H^\circ$, and $\Delta H^\circ$ becomes progressively less favorable in other systems primarily as the result of increasing strain. The more pronounced DNA distortions generally involve large changes in roll angles between adjacent base pairs [12], so the resulting destacking is likely to be a major factor in how strain contributes unfavorable $\Delta H^\circ$. Estimates of the enthalpic cost of destacking a single DNA base pair are in the neighborhood of $7\pm 11$ kcal/mol [19, 20], so it is not unreasonable that complexes in which the DNA is strongly bent (e.g., TBP destacks $\sim 6$ base pairs) should include an unfavorable enthalpic component ranging up to $+50$ to $+60$ kcal/mol. A calculation [21] on the EcoRI endonuclease–DNA complex, in which the DNA is kinked (but not bent) so as to destack the 3 central base pairs, estimates an enthalpic cost of $+46$ kcal/mol for DNA distortion. This latter case illustrates why the detailed inter–base pair structural parameters, rather than net deviation of the DNA backbone [1], should be considered when evaluating the contribution of strain to $\Delta H^\circ$. Aspects of DNA strain other than destacking (e.g., steric clashes [22]) may contribute to some specific structures.

If DNA distortion is so costly, why should any site-specific binding protein evolve so as to distort the DNA? Previous discussions of this question have often emphasized the structural view that the DNA must be distorted to accommodate some particular protein (e.g., to widen a DNA groove to allow insertion of recognition elements), but this avoids the issue of why the protein has evolved to a form that requires DNA distortion. The alternative view has been that in some particular cases, DNA distortion is an indispensable aspect of protein function, and is, in fact, the primary “purpose” of site-specific protein binding [11]. This seems to us more persuasive. For example, TBP apparently distorts the DNA in order to facilitate the association of other transcription factors and RNA polymerase II [23, 24]. We have argued elsewhere [2] that enzymes such as restriction endonucleases use sequence-dependent DNA distortions closely resembling those in the transition state (but extending far beyond the immediate reaction center) to enhance discrimination between correct and near-correct DNA recognition sites. Conversely, in the low-DNA-distortion systems, bound protein can apparently function simply by establishing itself at a strategic location on the DNA.

What is so striking in these data is how these diverse functional requirements correlate with the choice of thermodynamic strategy for protein-DNA binding. At one extreme, the low-DNA-distortion systems have very favorable $\Delta H^\circ$ and unfavorable $T \Delta S^\circ$ (Figure 1), whereas at the other extreme, the high-DNA-distortion systems use favorable $T \Delta S^\circ$ to pay for the enthalpic cost (unfavorable $\Delta H^\circ$) of DNA strain. Moderate DNA distortions incur moderate enthalpic costs and require only moderate entropic subsidies.

Given any particular value of $\Delta H^\circ$, how is $\Delta S^\circ$ determined so as to give a functionally appropriate value of $\Delta G^\circ$? Favorable $\Delta S^\circ$ in protein-DNA binding is contributed primarily by the release of water from nonpolar surfaces [25] and, in lesser proportion, by water release from polar surfaces and the redistribution of salt ions that were associated with the free protein and DNA [26, 27]. The water-release contribution thus differs among systems because of variation in the amounts of buried nonpolar and polar surfaces and/or the residual retention of solvent at polar regions between protein and DNA. However, we can find no tendency for those systems with a more strongly positive net $\Delta S^\circ$ to have a larger net burial of previously solvated surface. For example TBP (high distortion, positive $\Delta S^\circ$) and GCN4 (low distortion, negative $\Delta S^\circ$) have nearly the same change in total solvent-accessible surface areas, apportioned about the same between nonpolar and polar surfaces.

In opposition to these favorable $\Delta S^\circ$ terms, there are several entropically unfavorable factors whose identities we know but whose relative contributions to different protein-DNA systems are not well characterized. There is an unfavorable $\Delta S^\circ$ from the loss of translational and rotational freedoms of protein and DNA (essentially constant for all systems) and an unfavorable $\Delta S^\circ$ from the restriction of configurational freedoms of bound protein and DNA and of vibrations of entrapped water molecules. (We adopt the nomenclature of Karplus et al. [28] in which the “configurational” entropy is comprised of a “vibrational” term representing local fluctuations in a well-defined structure and a “conformational” term representing the existence of more than one such structure.) Duan et al. [29] point out that vibrational and conformational entropies, which can be considered as distinct for protein folding [28], are difficult or impossible to deconvolute for protein-DNA associations.

Induced folding of the protein upon binding, resulting in an unfavorable reduction in conformational entropy, was proposed as a widespread feature of site-specific protein-DNA binding [25] but appears to us to be most prevalent (and/or greater in scope) for systems in which the protein-bound DNA remains undistorted (lower right sector of Figure 1). The basic DNA binding domains of GCN4 protein, for example, are relatively disordered in the free protein and become folded upon site-specific DNA binding [16, 30]. On the basis of 46 residues converted from “nascent helices” to fully folded $\alpha$ helices upon GCN4 binding to the AP-1 site [16], we estimate that this alone could make an adverse entropic contribution to $\Delta G^\circ$ of $+40$ to $+60$ kcal/mol at 298 K; a similar estimate of $\Delta S^\circ$ has been calculated from NMR backbone dynamics of the bZIP domain [31]. The unusually unfavorable $\Delta S^\circ$ for $\lambda$ repressor-operator binding may relate in part to the induced folding of the N-terminal arms [32], however, cautious interpretation is warranted, because thermodynamic measurements [33] were made with the complete repressor, but the crystal structure was determined on only the N-terminal DNA binding domains [34]. Induced protein folding has been proposed as an explanation for the unfavorable $\Delta S^\circ$ of association [35] of the matz2 homeodomain with undistorted [36] DNA. By contrast, induced protein folding appears to be relatively minor in the high-DNA-distortion systems [11, 37] where net $\Delta S^\circ$ is strongly favorable, although protein conformational transitions that are too small to have major effects on the thermodynamics may nevertheless be important to function.

It is thus apparent that for protein-DNA association, the intermolecular interactions, dynamic restrictions, and reorganizations of ions and solvent molecules that contribute to $\Delta S^\circ$ also contribute to $\Delta H^\circ$ in most instances. In fact, it is not possible to design (or evolve) a site-specific DNA binding protein in such a way as to vary $\Delta H^\circ$ and $\Delta S^\circ$ independently. This is not only because the same aspects of the association affect both parameters, but also because of a phenomenon called “isothermal enthalpy-entropy compensation.”

**Isothermal Enthalpy-Entropy Compensation**

Previous discussions of entropy-enthalpy compensation in protein-DNA association reactions have focused exclusively
on the tendency of $\Delta H^\circ$ and $T\Delta S^\circ$ to oppose and compensate each other as the temperature varies [16, 33, 38]. Figure 2 shows plots of $\Delta H^\circ$ versus $T\Delta S^\circ$ for several protein-DNA binding systems; in each plot the data points are obtained at various temperatures. The linearity of the relationship between $\Delta H^\circ$ and $T\Delta S^\circ$ is evident and in all cases the slope is near 1. This “temperature-dependent” enthalpy-entropy compensation is required by basic thermodynamics [38, 39] for any process in which the change in heat capacity $\Delta C_p^\circ$ is nonzero, albeit only in the restricted range of temperatures where $|\Delta S^\circ| << |\Delta C_p^\circ|$.

If the heat capacity change $\Delta C_p^\circ$ for any process is nonzero and independent of temperature in the range of the experiments, then

$$\Delta H^\circ = \Delta C_p^\circ (T - T_s)$$

(1)

$$T\Delta S^\circ = \Delta C_p^\circ T \ln \frac{T}{T_s}$$

(2)

where $\Delta H^\circ = 0$ at $T = T_s$, and $\Delta S^\circ = 0$ at $T = T_s$. The values $T_s$ and $T_0$ are characteristic of each system. A Taylor expansion of (2), neglecting third-order and higher terms, gives [39]

$$T\Delta S^\circ = \Delta C_p^\circ (T - T_s) + \frac{1}{2} (T - T_s)^2 \frac{T - T_s}{T_s}$$

(3)

Since protein-DNA binding experiments are generally conducted in the range 0–40°C (273–313 K) and $T_s$ is generally in the same range, the second term in (3) is small. Thus,

$$T\Delta S^\circ \approx \Delta C_p^\circ (T - T_s)$$

(4)

From (1) and (4) it follows that

$$\Delta H^\circ \approx \Delta C_p^\circ (T - T_s) - \Delta C_p^\circ (T - T_0)$$

(5)

where the second term contains only constants characteristic of any given protein-DNA pair.

Thus, “temperature-dependent” enthalpy-entropy compensation (Figure 2) will be nearly exact [$\Delta H^\circ = \Delta (T\Delta S^\circ)$] for $T$ not far from $T_s$ with the consequence that $\Delta G^\circ_{\text{bind}}$ is nearly temperature independent [38]. Protein-DNA associations show nearly exact compensation because $T_s$ happens to fall in a temperature range that is physiologically relevant and experimentally accessible, not because of some peculiar evolutionary adaptation of site-specific DNA binding proteins. We have discussed elsewhere [6] the molecular basis for the strongly negative $\Delta C_p^\circ$ in protein-DNA association processes. It is crucial to emphasize that while the numerical value of $\Delta G^\circ_{\text{bind}}$ depends on molecular properties of each system, the existence of temperature-dependent enthalpy-entropy compensation does not.

It has not been recognized until now that protein-DNA interactions are also profoundly influenced by a completely distinct phenomenon, which we call “isothermal” enthalpy-entropy compensation. It is well-known in small-molecule “host-guest” chemistry [40] that if the binding of a homologous series of ligands is compared under otherwise constant conditions (including constant temperature, hence “isothermal”), $\Delta H^\circ$ and $T\Delta S^\circ$ often show compensating changes. This phenomenon has been called “extra-thermodynamic” enthalpy-entropy compensation [41], but that name seems infelicitous and potentially confusing. We urge that “enthalpy-entropy compensation” should always carry a prefix, either “temperature-dependent” or “isothermal,” to distinguish between the two distinct compensation phenomena.

It turns out there is a close parallel in protein-DNA interactions when a DNA recognition site is held constant while variation is introduced in the sequence context, that is, in the non-contacted base pairs surrounding the recognition site. Figure 3a shows a plot of $\Delta H^\circ$ versus $T\Delta S^\circ$ (at constant 298 K) for EcoRI endonuclease binding to a series of 24 bp oligonucleotides, all of which contain the canonical GAATTC recognition site but have different flanking sequences. This sequence “context” variation produces a surprisingly large effect on binding free energy; the range for the four sequences shown is $\Delta \Delta G^\circ_{\text{bind}} = 3.5$ kcal/mol, or about a 400-fold range of binding affinity $K_a$. It is evident that the variation in $\Delta H^\circ$ and $\Delta S^\circ$ is significantly greater than this and that there is a pronounced linear relationship between $\Delta H^\circ$ and $T\Delta S^\circ$. The calculated slope is 1.2, indicating that $T\Delta S^\circ$ “undercompensates” for $\Delta H^\circ$; that is, as sequence context improves, the benefit in enthalpy ($\Delta \Delta H^\circ < 0$) is ~20% greater than the penalty in entropy ($T\Delta \Delta S^\circ < 0$), so that $\Delta G^\circ_{\text{bind}}$ becomes more negative. A similar plot for BamHI endonuclease binding to its GGATCC site in various contexts is shown in Figure 3b (slope = 1.1). In this case the van’t Hoff enthalpies have been confirmed directly by isothermal titration microcalorimetry [6].

Such a linear relationship may represent true isothermal enthalpy-entropy compensation or may instead be simply an artifact of correlated errors [42] in measuring $\Delta H^\circ$ and $\Delta S^\circ$ by the van’t Hoff method, where both parameters are extracted from the dependence of ln$K_a$ on temperature. (Compensation detected in calorimetric studies does not arise from such an
Thermodynamic Strategies for DNA Binding Proteins

Figure 3. Isothermal Enthalpy-Entropy Compensation

Plots of $\Delta H^\circ$ versus $\Delta S^\circ$ at a constant temperature of 298 K are shown for (a) EcoRI endonuclease binding to the GAATTC site embedded in four different sequence contexts: (1) GGGCGGCGCggaattcGCGGGG, (2) GGGCGGCGGGTgaa ttcACAGGCGG, (3) GGGCGGCGGTgggaattcCAC CGCGG, and (4) GGGCGGAAAGggaattcTTTTGGG; and (b) BamHI endonuclease binding to the GGATCC site embedded in eight different sequence contexts: (1) GGGATGGGTgaaatccCACCCAC, (2) CGCGGGTTATggatccATAAGGGC, (3) TGGGTGggatccCACCCAC, (4) GGGATGGGTGggatccCCACCAC, (5) GGGATGGCAAggatccTTGCCAC, (6) CGCGGGCGGCGggatccGGGCGGGC, (7) GGGATGGGGGggatccCCACCAC, and (8) GGGATGGTGGggatccCCACCAC. Values of $\Delta H^\circ$ and $\Delta S^\circ$ were estimated [6] by van't Hoff analysis.

Direct calorimetric measurements of heat of reaction at four different temperatures [for BamHI and sequences (1), (2), (5) and (6); others not tested] completely confirmed the predicted van’t Hoff values of $\Delta H^\circ$. The indicated slopes $\frac{d\Delta H^\circ}{dT}$ are 1.2 and 1.1, respectively. Panels (c) and (d) show the same data plotted as $\Delta G^\circ$ versus $\Delta S^\circ$. In each case the slope (the so-called “compensation temperature”) differs significantly from the experimental temperature (298 K), indicating that the $\Delta H^\circ-\Delta S^\circ$ correlation arises from isothermal compensation rather than from correlated errors [42] in estimating the parameters by van’t Hoff analysis. Panel (e) shows $\Delta G^\circ$ plotted against $\Delta H^\circ$ for both proteins (Figures 3c and 3d). A second test [43] is that true isothermal enthalpy-entropy compensation is indicated at true isothermal enthalpy-entropy compensation almost certainly occurs within each system [see below]; however, the various systems with pronouncedly different structures are not related to each other by a compensation relationship.)

How does sequence context affect binding free energy? This is discussed in detail elsewhere [2, 6], so we only summarize here. For the EcoRI context variants, both crystallographic and biochemical data indicate that there is no difference in the identity or number of direct protein-base or protein-phosphate contacts or in the degree of DNA distortion achieved in the complexes [2, 6]. The fact that the dependence of binding on cosolvent (osmolyte) concentrations does not differ between sequence variants further implies that sequence context does not alter the number of water molecules released upon binding (D. F. Cao and L. J.-J., unpublished data). If data on the temperature-dependence of $\Delta S^\circ$ are extrapolated to 386 K (Figure 4), where $\Delta S^\circ$ (hydrophobic) $\approx 0$ [44], it becomes clear that there is a large residual negative $\Delta S^\circ$ that varies with sequence context. The most plausible source of this variable negative $\Delta S^\circ$ is stiffening of internal vibrations [45] in the complexes.

Sequence context variation also affects heat capacity: $\Delta C^\circ_P$ becomes progressively and dramatically more negative (from $-1.2$ to $-2.5$ kcal/mol for EcoRI endonuclease) as $\Delta G^\circ_{\text{bind}}$ becomes more negative [6]. The negative $\Delta C^\circ_P$ of association has a major contribution from the hydrophobic effect [38], which also contributes positive $\Delta S^\circ$, but other factors (Table 1) can increase $\Delta C^\circ_P$ and increase $\Delta S^\circ$ (dehydration of polar surfaces) or decrease $\Delta C^\circ_P$ and decrease $\Delta S^\circ$ (vibrational tightening). A more negative $\Delta C^\circ_P$ for sequence-context variation cannot be
attributed to an increase in the hydrophobic contribution to $\Delta C^r$, because $\Delta S^r$ becomes more negative as $\Delta C^r$ becomes more negative. A similar trend has been observed for sequence differences within a recognition site [16]. Thus, we [6] have inferred that the sequence-dependent differences in $\Delta C^r$, reflect primarily a change in the vibrational contribution [45]; that is, as sequence context improves, elements of both protein and DNA become more immobilized at the interface, so that heat capacity decreases.

**Molecular Interpretations of Isothermal Enthalpy-Entropy Compensation**

Ideas about the source(s) of isothermal enthalpy-entropy compensation have developed from an initial emphasis on solvent reorganization and the unique properties of water [46] to a more general view that compensation is an intrinsic consequence of weak interactions [47, 48]. The latter view may be summarized as follows: as enthalpy becomes more negative (stronger bonding, deeper potential-energy well), the spacing between vibrational energy levels increases and entropy thus decreases; conversely, weaker bonding (less negative enthalpy) is associated with access to more vibrational levels and thus greater entropy. In simplest terms, weaker interactions are sloppier.

For the better sequence contexts in Figure 3, $\Delta H$ becomes more favorable because the energy required to strain the DNA and protein in the complexes becomes less unfavorable. That is, the better complexes have a more perfectly complementary interface, the consequences of which can be viewed in two ways: (1) the geometry of many nonpolar contacts and hydrogen bonds is more optimized, perhaps very subtly, so the overall enthalpy of forming the set of direct protein-DNA interactions is more favorable; and (2) differences in DNA sequence context may change the enthalpic difference between free and bound DNA by reducing strain energy in the complex or (synonymously) reducing the enthalpic cost of DNA distortion.

These two ways of describing a situation that produces a more negative $\Delta H^r$ are not meaningfully different because of the way we defined strain as the occupation of energetically suboptimal positions. They are certainly indistinguishable experimentally. In either case, the consequence of better fit/less strain energy is a greater immobilization of both DNA and protein at the interface, so $\Delta S^r$ becomes less favorable [49]. There is isothermal enthalpy-entropy compensation.

Recall, however, that sequence-dependent variation in this vibration/strain component shows imperfect compensation (Figure 3) such that as $\Delta H$ becomes more negative, $-\Delta T \Delta S$ does not become more positive by an equal increment. This “undercompensation” is a general property of macromolecular interfaces where binding results from the combined effect of many weak interactions and can be understood as a consequence of the “chelate effect” [48, 50], as follows: the various interacting elements in a crowded DNA-protein interface do not make fully independent entropic contributions [51], so if the interface is “loosened” (that is, some enthalpically favorable interactions are removed or weakened), there is not a full measure of corresponding entropic benefit. Conversely, “tightening” an interface (adding or strengthening enthalpically favorable interactions) does not exact full entropic cost, because some of that cost has already been paid in bringing the macromolecular surfaces into contact.

Isothermal entropy-enthalpy compensation in protein-DNA interactions has been revealed using sequence context variation as a probe, but it should not be considered as a peculiar property of context variation. As a property of all weak intermolecular interactions [47, 48], it should apply to all aspects of direct protein-DNA interaction as well as to the solvent-reorganization processes that accompany the association.

**Enthalpy-Entropy Compensation and Design Strategies**

If the various site-specific binding systems in Figure 1 have their positions along the enthalpy axis largely determined by the degree of strain in the complexes, as required by the func-

<table>
<thead>
<tr>
<th>Factor</th>
<th>Effect on Heat Capacity</th>
<th>Effect on Entropy</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dehydration of nonpolar surface</td>
<td>decrease</td>
<td>increase</td>
<td>[38]</td>
</tr>
<tr>
<td>Dehydration of polar surface</td>
<td>increase</td>
<td>increase</td>
<td>[25]</td>
</tr>
<tr>
<td>Vibrational tightening</td>
<td>decrease</td>
<td>decrease</td>
<td>[6, 7, 45]</td>
</tr>
<tr>
<td>Induced folding (conformational)</td>
<td>little or none</td>
<td>decrease</td>
<td>[25]</td>
</tr>
</tbody>
</table>
Thermodynamic Strategies for DNA Binding Proteins

1021

We can discern several factors that allow adjustment of $\Delta S^\circ$. (1) Low-strain exothermic systems will tend to incorporate coupling between protein folding and DNA binding (e.g., GCN4, λcl repressor, mat2α), whereas high-strain endothermic systems will tend not to couple protein folding to binding, so as to realize the maximum entropic drive from water release. (2) Low-strain exothermic systems, having deeper potential-energy wells in the complexes, will necessarily achieve greater immobilization of the recognition elements at the protein-DNA interface. That is, the residual configurational entropy in the complexes will be less. This factor is likely larger than is generally appreciated, as even vibrational fine-tuning of a single system (Figure 3) can produce a range of $(\Delta T\Delta S) \approx 15$ kcal/mol. Enthalpy-entropy compensation thus helps solve the free-energy problem by generating more adverse $\Delta S^\circ$ in the low-strain systems and less adverse $\Delta S^\circ$ in high-strain systems.

The relative contributions to $\Delta S^\circ$ of protein-folding transitions and of vibrational tightening are not yet clear and should be investigated experimentally in a larger number of systems. It would also be extremely useful to have better estimates of the overall contribution of “strain energy” in any system. Nevertheless, our overall hypotheses lead to predictions for systems for which there are presently structural but not thermodynamic information: high-DNA-distortion systems (e.g., integration Host Factor [52], λC-resolvase [53]) should have strongly unfavorable $\Delta H^\circ$ and be entropy-driven, whereas low-DNA-distortion systems (e.g., NF-κB transcription factor [54], Hin recombinase [55]) should be enthalpy-driven and have unfavorable $\Delta S^\circ$.

The particular protein-DNA systems cited involve a single protein binding to a single DNA recognition site and as such probably have no significant contribution to $\Delta S^\circ$ from changes in quaternary protein-protein interactions. However, some of these proteins in vivo also participate in more complex interactions (e.g., TBP with other proteins when DNA-bound, λcl repressor binding cooperatively to the tandem λ operator sites) and changes in quaternary interactions are likely to affect $\Delta S^\circ$ significantly in such circumstances and in other multiprotein systems. The coupling of protein-DNA binding to the assembly of multiprotein complexes or to the configurational “tightening” of pre-existing multiprotein complexes could provide a source of unfavorable $\Delta S^\circ$ analogous to that described above for coupled protein folding [56]. We expect that this will be most significant in systems of the low-DNA-distortion class, which require a source of unfavorable $\Delta S^\circ$ to counterbalance the favorable $\Delta H^\circ$.

Biological Implications

The structure and function of site-specific DNA binding proteins are matters of considerable interest because of the crucial roles of these proteins in the expression and regulation of genetic information. A better understanding of both the structural and energetic engineering principles used by nature might lead to rational methods for the modification or design of such proteins. This paper considers the relationship between the structural diversity of site-specific protein-DNA complexes and the diverse thermodynamic strategies by which these systems achieve an appropriate range of binding free energy by analyzing ten diverse complexes for which both crystal structures and thermodynamic data are found in the literature. A key element of the problem is the demonstration for the first time of isothermal enthalpy-entropy compensation, a phenomenon well known in small-molecule chemistry but previously undetected for protein-DNA interactions.

For formation of protein-DNA complexes in which the DNA is relatively undistorted, a favorable enthalpy change ($\Delta H^\circ < 0$) drives an unfavorable entropy change ($-T\Delta S^\circ > 0$), whereas for complexes with highly distorted DNA, unfavorable $\Delta H^\circ$ is driven by favorable $-T\Delta S^\circ$. All such complexes have favorable $\Delta H^\circ$ contributed by direct protein-DNA recognition interactions and favorable $\Delta S^\circ$ contributed by release of water molecules (hydrophobic effect). Molecular strain contributes unfavorable $\Delta H^\circ$ when the DNA is distorted, but there is little or no unfavorable “strain energy” when the DNA is undistorted. In complexes that strongly distort the DNA, molecular strain makes $\Delta H^\circ$ net unfavorable, but the strained interface suffers less immobilization so binding is driven by favorable $\Delta S^\circ$. Conversely, those complexes with little or no DNA distortion have net favorable $\Delta H^\circ$, which must be counterbalanced by net unfavorable $\Delta S^\circ$, derived from greater immobilization of the interface (isothermal enthalpy-entropy compensation) and from coupling between DNA binding and partial protein folding.

Experimental Procedures

Values of $\Delta H^\circ$ and $\Delta S^\circ$ were estimated [6] by van’t Hoff analysis. Values of the equilibrium association constant $K_{ob}$ as a function of temperature were obtained by the membrane filter binding technique. These data were used in nonlinear least squares fits of the equation $\ln K_{obs} = (\Delta C^\circ/R)(T_s/T - 1)$ assuming a temperature-independent $\Delta C^\circ$, and where $\Delta H^\circ = 0$ at $T_0$ and $\Delta S^\circ = 0$ at $T_0$. The predicted values of $\Delta H^\circ$ and $\Delta S^\circ$ at 298 K were calculated from $\Delta H^\circ = -\Delta C^\circ (298 - T_0)$ and $\Delta S^\circ = -\Delta C^\circ \ln (298/T_0)$, with errors derived by propagation of errors in estimating $\Delta C^\circ$, $T_0$, and $T_s$ in some cases (see legend to Figure 3) values of $\Delta H^\circ$ were also obtained by isothermal titration calorimetry, as described [6]. The calorimetric and van’t Hoff methods showed excellent agreement in values of $\Delta H^\circ$ and its temperature-dependence $\Delta C^\circ$.

Structures of protein-DNA complexes were drawn using MIDAS Plus software (UCSF Computer Graphics Lab) and atomic coordinates from the RCSB Protein Databank (http://www.rcsb.org/pdb/) are as follows: TBP [57] (λyb); CAP [58] (Iber); λcl repressor [34] (I1mb); and GCN4 (complex with AP-1 site [15], 1yza).

Acknowledgments

This work was supported in part by grant GM-29207 from the National Institutes of Health. We thank D. F. Cao for contributing some of the unpublished data discussed here.

Received: June 2, 2000
Revised: August 16, 2000
Accepted: August 22, 2000

References

vation of structure and binding energy in the transition state. Biopolymers 4, 153–180.