

Free-Energy Component Analysis of 40 Protein–DNA Complexes: A Consensus View on the Thermodynamics of Binding at the Molecular Level

B. JAYARAM,¹ K. MCCONNELL,² S. B. DIXIT,^{1,2} A. DAS,¹ D. L. BEVERIDGE²

¹*Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi, 110016, India*

²*Department of Chemistry and Program in Molecular Biophysics, Wesleyan University, Middletown, Connecticut 06459*

Received 23 February 2001; Accepted 2 August 2001

Abstract: Noncovalent association of proteins to specific target sites on DNA—a process central to gene expression and regulation—has thus far proven to be idiosyncratic and elusive to generalizations on the nature of the driving forces. The spate of structural information on protein–DNA complexes sets the stage for theoretical investigations on the molecular thermodynamics of binding aimed at identifying forces responsible for specific macromolecular recognition. Computation of absolute binding free energies for systems of this complexity transiting from structural information is a stupendous task. Adopting some recent progresses in treating atomic level interactions in proteins and nucleic acids including solvent and salt effects, we have put together an energy component methodology cast in a phenomenological mode and amenable to systematic improvements and developed a computational first atlas of the free energy contributors to binding in ~40 protein–DNA complexes representing a variety of structural motifs and functions. Illustrating vividly the compensatory nature of the free energy components contributing to the energetics of recognition for attaining optimal binding, our results highlight unambiguously the roles played by packing, electrostatics including hydrogen bonds, ion and water release (cavitation) in protein–DNA binding. Cavitation and van der Waals contributions without exception favor complexation. The electrostatics is marginally unfavorable in a consensus view. Basic residues on the protein contribute favorably to binding despite the desolvation expense. The electrostatics arising from the acidic and neutral residues proves unfavorable to binding. An enveloping mode of binding to short stretches of DNA makes for a strong unfavorable net electrostatics but a highly favorable van der Waals and cavitation contribution. Thus, noncovalent protein–DNA association is a system-specific fine balancing act of these diverse competing forces. With the advances in computational methods as applied to macromolecular recognition, the challenge now seems to be to correlate the differential (initial vs. final) energetics to substituent effects in drug design and to move from affinity to specificity.

© 2002 John Wiley & Sons, Inc. J Comput Chem 23: 1–14, 2002

Key words: protein–DNA recognition; binding free energy computations; noncovalent association

Introduction

Thermodynamics of noncovalent associations plays a pivotal role in much of molecular biology as in replication, transcription, and translation for instance, wherein the underlying molecular process typically involves protein–nucleic acid recognition, protein–protein association, etc. Developing an energetic perspective of molecular recognition, complementing the structural view has thus become a thrust area to facilitate a better understanding of function as well as to drug discovery. The growing structural database of protein–nucleic acid complexes and advances in theoretical and computational methods for biological molecules sets the stage for obtaining a deeper understanding of structure, functional energetics, and thermodynamics of binding in these systems.^{1–23} In many studies

to date, binding is interpreted in terms of the interfacial contacts observed in the cocrystal structures and treated in the context of electrostatic/hydrogen bond complementarity. However, interpretations of binding in terms of crystallographic contacts are at risk with respect to errors of omission and of commission. The protein–DNA complex is the final state of a thermodynamic binding process, with the uncomplexed forms of protein and DNA in solution forming the initial state (Fig. 1). Thus any successful

Correspondence to: D. L. Beveridge; e-mail: dbeveridge@wesleyan.edu

Contract/grant sponsor: Indo-French Centre for the Promotion of Advanced Research (to B.J.)

Contract/grant sponsor: National Institutes of Health; contract/grant number: GM 37909 (to D.L.B.)

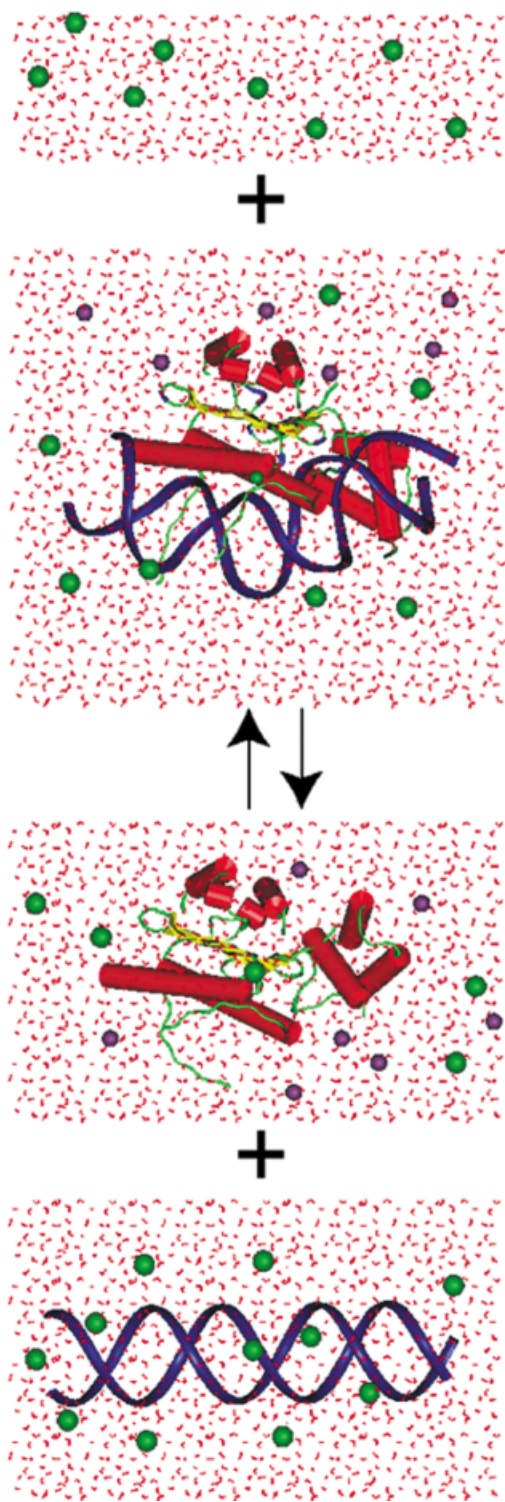


Figure 1. A schematic representation of DNA-protein binding in aqueous medium illustrating the structural variations in the reactants and the release and reorganization of small ions upon binding.

analysis must consider the relative effects on initial and final state species, whereas electrostatic (and hydrogen bond) complementarity is a property only of the final state. Moreover, the binding of proteins to DNA involves a number of features in addition to electrostatics, including shape complementarity and solvent release, both considered to be net favorable to binding, and entropy effects resulting from the loss of translational, rotational, and some internal degrees of freedom upon complex formation, generally considered net unfavorable to binding. Electrostatics per se has the direct contribution due to the protein interacting with DNA, but an indirect contribution due to the relative effects of solvent polarization in the initial and final-state species. This latter makes an electrostatic contribution to the free energy of binding in opposition to direct effects, because charged groups in contact are not as available for solvent polarization. With these factors and the possibilities of others yet to be considered, the interpretation of binding in terms of crystallographic contacts is obviously perilous, and valid only when the energetics associated with electrostatic complementarity in the complex is dominant. Recent theoretical studies of binding in diverse systems indicate that assuming electrostatics as always net favorable to binding, much less the dominant contribution is questionable.^{24, 25}

Further understanding of the nature of binding in complexes in terms of the diverse biophysical features of the process requires a detailed knowledge of the relative magnitudes of the various components of the free energy of formation of initial and final-state species. Although overall binding constants and thus free energies of binding can be measured at a given salt concentration and temperature, no set of experiments can quantify the free energy components essential to a phenomenological interpretation of the results. The only route into further penetration of this problem is by theoretical and computational methods, which thus have a unique purview on this class of problems. Even so, hurdles remain. The diverse components of the binding free energy can be calculated, but for systems of this complexity a considerable degree of approximation is involved. Furthermore, uncertainties in the calculated values may propagate to a significant level in both magnitude and sign of the calculated net binding free energy to the extent that errors in the individual terms are not compensatory. Thus, theoretical studies of any individual system, even at the state of the art, are hardly expected to be unequivocal at the present state of play.

As a step forwards on this problem we tackle the case of protein-DNA binding with what we shall term the “consensus view” approach: consider a large number of systems, calculate the components of the free energy of binding systematically using the best methods we can under well-defined protocols, and determine the average or consensus values of each component and nature of binding for the class of systems. Although it would be unreasonable (for reasons mentioned above) to claim a high degree of quantitative accuracy for any given case, our hypothesis is that an accurate qualitative view of the nature of binding in a set of complexes will emerge, and lead us to an improved theoretical view of the relative contributions viz. electrostatics (intramolecular and intermolecular), energetics of shape complementarity (packing) as reflected in van der Waals energies, solvent release, and reorganization on complex formation including the hydrophobic effect, internal entropies, etc. The necessity to develop a comprehensive molecular thermodynamic view of affinity is all the more pressing to further

understand key issues related to specificity, which depend upon differential energetics.

In this study, we consider a set of over 40 protein–DNA complexes for which crystal structures are available and a variety of binding motifs are represented. Citations to protein–DNA literature have been provided elsewhere.^{25–29} We treat interactions at the interatomic level with well-validated energy functions wherever possible, employ well-calibrated semiempirical approaches to solvation from continuum methods, and otherwise use basic techniques and methods drawn from classical statistical mechanics. We review and discuss the statistical mechanical theory for noncovalent associations and then proceed to apply the theory to a large number of protein–DNA complexes in an attempt to elucidate some principles common to the thermodynamics of macromolecular binding. We address here such questions as, which of these components are net favorable or unfavorable to binding in the consensus view, and which terms are large and small, and what are the relative magnitudes. In the consensus view of this class of systems, we find that packing and hydrophobic effects favor binding and that, surprisingly, electrostatics (including solvent, salt, and hydrogen bond contributions) are net unfavorable to binding. Our results dramatically illustrate the necessity to consider diverse competing effects in constructing a structure-based interpretation of binding free energies.

Statistical Mechanical Theory of Protein–DNA Binding in Aqueous Media

Taking off from the discussions on the statistical thermodynamics of binding put forth by Ben Naim,³⁰ Gilson et al.,³¹ Janin,^{32, 33} Ajay and Murcko,³⁴ and Atkins,³⁵ we present here, within the statistical mechanical framework, a hierarchy of methods accessible to theory of varying levels of rigor and computational requirements.

Let P and D be the reactants and P^*D^* , the product of binding in aqueous medium.

$$[P]_{\text{aq}} + [D]_{\text{aq}} = [P^*D^*]_{\text{aq}} \quad (1)$$

At equilibrium

$$\mu_{P,\text{aq}} + \mu_{D,\text{aq}} = \mu_{P^*D^*,\text{aq}} \quad (2)$$

$\mu_{P,\text{aq}}$ is the chemical potential of species P in the solvent medium (partial molar Gibbs free energy) and $\mu_{P,\text{aq}}^\circ$ is its standard chemical potential, i.e., under conditions of 1 bar in the gaseous state and 1 molar (designated as C°) in the liquid state.

$$\begin{aligned} \mu_{P,\text{aq}}^\circ + RT \ln(a_P) + \mu_{D,\text{aq}}^\circ + RT \ln(a_D) \\ = \mu_{P^*D^*,\text{aq}}^\circ + RT \ln(a_{P^*D^*}) \end{aligned} \quad (3)$$

where a_P ($=\gamma_P C_P / C^\circ$), is the activity of P , γ_P is the activity coefficient of species P , and C_P its concentration. The standard molar Gibbs free energy of the reaction (standard absolute molar Gibbs free energy of binding) is

$$\begin{aligned} \Delta G_{\text{aq}}^\circ &= \mu_{P^*D^*,\text{aq}}^\circ - (\mu_{P,\text{aq}}^\circ + \mu_{D,\text{aq}}^\circ) \\ &= -RT \ln[a_{P^*D^*} / (a_P a_D)] \\ &= -RT \ln K_{\text{eq},\text{aq}} \end{aligned} \quad (4)$$

In terms of canonical partition functions (Q)

$$\begin{aligned} \Delta G_{\text{aq}}^\circ &= \Delta A_{\text{aq}}^\circ + P \Delta V_{\text{aq}}^\circ = -RT \ln K_{\text{eq},\text{aq}} \\ &= -RT \ln \left[\frac{Q_{P^*D^*,\text{aq}}}{N_A Q_W} \right] \\ &\quad \left/ \left[\frac{Q_{P,\text{aq}}}{N_A Q_W} \right] \left[\frac{Q_{D,\text{aq}}}{N_A Q_W} \right] \right\} \\ &\quad + P \Delta V_{\text{aq}}^\circ \end{aligned} \quad (5)$$

ΔA° is the standard Helmholtz free energy of the reaction. The Avogadro number N_A in the above equation originates in expressing partition functions Q as molar partition functions (following the notation of Atkins³⁵) and $P \Delta V_{\text{aq}}^\circ$ is the pressure–volume correction to Helmholtz free energy in the solvent medium. Q_W denotes the partition function for pure solvent (water). Equation (5) is an exact expression for noncovalent associations in aqueous medium. The assumption that translations and rotations are separable from intrasolute degrees of freedom as well as those of solvent, a common practice is gas phase statistical mechanics, then leads to

$$\begin{aligned} \Delta G_{\text{aq}}^\circ &= -RT \ln \left[\frac{Q_{P^*D^*}^{\text{tr}} Q_{P^*D^*}^{\text{rot}} Z_{P^*D^*,\text{aq}}^{\text{int}} Q_{P^*D^*}^{\text{el}} N_A Q_W}{\left[\frac{Q_P^{\text{tr}} Q_P^{\text{rot}} Z_{P,\text{aq}}^{\text{int}} Q_P^{\text{el}} \right] \left[\frac{Q_D^{\text{tr}} Q_D^{\text{rot}} Z_{D,\text{aq}}^{\text{int}} Q_D^{\text{el}} \right]} \right]} \\ &\quad + P \Delta V_{\text{aq}}^\circ \end{aligned} \quad (6)$$

Z^{int} is the configurational partition function. It includes contributions from vibrations and internal motions as well as solvation (hydration) effects. The translational and rotational terms have been separated out (i.e., momentum-dependent terms have been integrated out from an integral of the following type).

$$\begin{aligned} Z_{P,\text{aq}}^{\text{int}} &= \int \cdots \int \exp \left\{ -E(X_P^N, X_W^M) / k_B T \right\} dX_P^N dX_W^M \\ &= \langle \exp E(X_P^N, X_W^M) / k_B T \rangle \end{aligned} \quad (7)$$

X_P^N and X_W^M represent the configurational space accessible to the solute P and solvent W , respectively, in the presence of each other. $E(X_P^N, X_W^M)$ denotes the total potential energy of the system describing nonidealities. It includes intramolecular interactions within the solute P and solvent W as well as intermolecular interactions between the solute and the solvent. $k_B T$ is the product of Boltzmann constant and temperature (in Kelvin).

$$Q_P^{\text{el}} \sim 1 \quad (\text{assumed for noncovalent associations}) \quad (8)$$

$$\begin{aligned} \Delta G^\circ &= -RT \ln \left[\frac{Q_{P^*D^*}^{\text{tr}} N_A}{(Q_P^{\text{tr}} Q_D^{\text{tr}})} \right] \\ &\quad - RT \ln \left[\frac{Q_{P^*D^*}^{\text{rot}}}{(Q_P^{\text{rot}} Q_D^{\text{rot}})} \right] \\ &\quad - RT \ln \left[\frac{Z_{P^*D^*,\text{aq}}^{\text{int}} Q_W}{(Z_{P,\text{aq}}^{\text{int}} Z_{D,\text{aq}}^{\text{int}})} \right] + P \Delta V_{\text{aq}}^\circ \end{aligned} \quad (9)$$

Equation (9) is an exact expression for noncovalent associations in aqueous media, and is subject only to the approximation of the separability of translations and rotations from the remaining degrees of freedom in the system including those of solvent. The third term in eq. (9) is accessible to free energy molecular simulations³⁶ configured in the canonical ensemble, albeit they are computationally

expensive. The corresponding expression for associations in the gas phase is given as

$$\begin{aligned} \Delta G_{\text{g}}^{\circ} = & -RT \ln[Q_{P^*D^*}^{\text{tr}} N_A / (Q_P^{\text{tr}} Q_D^{\text{tr}})] \\ & -RT \ln[(Q_{P^*D^*}^{\text{rot}} / Q_P^{\text{rot}} Q_D^{\text{rot}})] \\ & -RT \ln[(Z_{P^*D^*}^{\text{int}}) / (Z_P^{\text{int}} Z_D^{\text{int}})] + P \Delta V_{\text{g}}^{\circ} \quad (10) \end{aligned}$$

In the following, we consider some approximations commonly employed to bring the binding free energy computations via evaluation of the right hand side expression in eq. (9), into feasibility domain.

Approximation-1: Formal separation of external degrees of freedom.

The molecular translational partition function of P is

$$q_P^{\text{tr}} = V / \Lambda_P^3 = V / (h^2 / 2\pi m_P k_B T)^{3/2} \quad (11)$$

The molar partition function of P is $Q_P^{\text{tr}} = (q_P^{\text{tr}})^{N_A}$.

Note that the volume, V , has been included in the translational part consistent with ideal gas statistical mechanics. This requires that the Z^{int} be divided by V to quantify nonidealities (excess free energies). The translational part of the free energy in eq. (9) is now given by the Sackur–Tetrode equivalent as

$$\begin{aligned} \Delta G_{\text{tr}}^{\circ} = & -RT \ln[(N_A/V)(\Lambda_P^3 \Lambda_D^3 / \Lambda_{P^*D^*}^3)] \\ = & -RT \ln[(N_A/V)(h^2/2\pi k_B T)^{3/2} m_{P^*D^*} / (m_P m_D)^{3/2}] \quad (12) \end{aligned}$$

The expression in the square brackets in eq. (12) is dimensionless, (N_A/V) may be replaced by a concentration term. Note that the form of this expression is the same whether in the gas phase or the liquid phase, provided the translational and rotational motions of the solute are unaffected by the solvent. Mathematically, this implies that the integrations in the momentum space can be carried out separately for the solute and solvent. This will be true only in a continuum, frictionless solvent influencing the position-dependent potential energy but not the velocity-dependent kinetic energy of the solute. Hence, in a transfer process (an experiment involving transfer of species P from one phase to another phase such as from the gas phase to the liquid phase or octanol to water, etc.), this term cancels out. In binding processes, however, no such cancellation occurs. Also, if P , D , and P^*D^* are each treated as a collection of nonbonded mono-atomic particles, then, again, the translational partition function for each species is written as a product of the individual partition functions of the constituent atoms, and because the number of atoms is conserved during binding, these terms cancel out. Again, this is not so for polyatomic species where the mass in translational partition function $m_P (= \sum_i m_i)$ is evaluated as a sum of the masses of the constituent atoms.

Similar arguments apply to the rotational partition functions. Separating the rotational part from internal motions implies working under rigid rotor approximation.

$$\begin{aligned} \Delta G_{\text{rot}}^{\circ} = & -RT \ln[(\sigma_P \sigma_D / \sigma_{P^*D^*}) (1 / (8\pi^2)) (h^2 / 2\pi k_B T)^{3/2} \\ & \times \{(I_{P^*D^*}^a I_{P^*D^*}^b I_{P^*D^*}^c) / (I_P^a I_P^b I_P^c I_D^a I_D^b I_D^c)\}^{1/2}] \quad (13) \end{aligned}$$

I_P^a , I_P^b and I_P^c are the components of moments of inertia of species P along the principal axes and σ_P its symmetry number. As an aside, the similarity between eqs. (12) and (13) is worth noting and maintaining. Contributions from external degrees of freedom having been accounted for by eqs. (12) and (13), the net binding free energy is expressed as

$$\begin{aligned} \Delta G^{\circ} = & \Delta G_{\text{tr}}^{\circ} + \Delta G_{\text{rot}}^{\circ} - RT \ln[(Z_{P^*D^*}^{\text{int}} Q_W V) / (Z_{P,\text{aq}}^{\text{int}} Z_{D,\text{aq}}^{\text{int}})] \\ & + P \Delta V_{\text{aq}}^{\circ} \quad (14) \end{aligned}$$

“For the process of bringing P and D from fixed positions and orientations at infinite separation to a final fixed position and orientation of the complex P^*D^* , i.e., freezing the translational and rotational degrees of freedom of the reactants and products,”³⁰ both $\Delta G_{\text{tr}}^{\circ}$ and $\Delta G_{\text{rot}}^{\circ}$ are individually zero.

Approximation-2: Formal separation of solvent effects from internal motions of the solute.

$$\begin{aligned} Z_{P,\text{aq}}^{\text{int}} = & Z_P^{\text{vib.conf}} Z_P^{\text{solvn}} \\ = & \int \cdots \int \exp\{-E(X_P^N, X_W^M) / k_B T\} dX_P^N dX_W^M \\ \simeq & \int \cdots \int \exp[-\{E(X_P^N) + E(X_P^{\text{fixed}}, X_W^M)\} / k_B T] \\ & \times dX_P^N dX_W^M \quad (15) \\ \simeq & \int \cdots \int \exp[-E(X_P^N) / k_B T] dX_P^N \\ & \times \int \cdots \int \exp[-E(X_P^{\text{fixed}}, X_W^M) / k_B T] dX_W^M \quad (16) \end{aligned}$$

Equations similar to (15) and (16) can be written for D and P^*D^* and converted to excess free energies. Such a separation allows

$$\Delta G^{\circ} = \Delta G_{\text{tr}}^{\circ} + \Delta G_{\text{rot}}^{\circ} + \Delta G_{\text{intra}}^{\circ} + \Delta G_{\text{solvn}}^{\circ} \quad (17)$$

Equation (17) forms the basis for “master equation”³⁴ methods. Further, if the internal degrees of freedom of the solutes P , D , and P^*D^* are frozen in addition to translational and rotational degrees of freedom, the binding free energy can be written as

$$\Delta G^{\circ} = \Delta E_{\text{g}}^{\circ} + \Delta G_{\text{solvn}}^{\circ} \quad (18)$$

an expression that is commonly employed in some earlier theoretical work. The $P \Delta V_{\text{aq}}^{\circ}$ term in eq. (9) is often neglected in liquid-state work. Equations (9) and (18) constitute two extreme theoretical models for binding with eq. (17) falling in between. Some prescriptions to alleviate approximations in eq. (17) and to make the results correspond as closely as possible to eq. (9) are: (a) compute the translational, rotational contributions in gas phase by forming a suitable thermocycle while adopting static structures such as from modeling, X-ray crystallography or NMR,²⁵ (b) generate an ensemble of structures of the reactants and products separately in the solvent medium with structural inputs from either modeling or experiment and then apply eq. (17).^{37, 38} This corresponds to a post facto analysis of the molecular dynamics trajectories. The former circumvents the objections concerning the

extension of Sackur-tetrode equation for reactions in liquid phase. The latter, (a) helps convert ΔE_g° into ΔH_g° in eq. (18), and (b) avoids decoupling internal motions of the solute from those of solvent since the simulation incorporates solvent explicitly in developing the ensemble of structures. The problem then shifts to solvation energy estimates and whether an ensemble of structures consistent with the solution phase was considered in arriving at the solvation model parameters. This problem may be less severe than it appears if the calibration of solvation parameters is performed against experiment.

Even so, some problems remain. Particularly nagging ones in developing reliable absolute binding free energy estimates pertain to (i) the vibrational and configurational entropies that entropies contribute to the $T\Delta S_{\text{intra}}^\circ$ terms in eq. (17); (ii) the environmental effects such as due to counterion association and release particularly in binding equilibria involving nucleic acids;^{22, 23, 39–43} (iii) the “bound” water/water-mediated interactions;⁴⁴ (iv) complexes involving metal ions exhibiting charge transfer; (v) the temperature sensitivity of the parameter set and whether heat capacity estimates could be generated from free energy estimates to provide an additional link to experiment.

Methods and Calculations

Atomic X-ray coordinates of over 40 protein–DNA complexes were taken from the structural data banks.^{45, 46} Hydrogens were added to the crystal structures and the complexes prepared for analysis via a short energy minimization protocol staying close to the crystal structures.²⁵ The following methodology, based on eq. (17) above, describes the subsequent computational effort to bridge the gap between structure and thermodynamics.

The thermodynamic cycle considered here to obtain the standard free energy of complex formation is shown in Figure 2. Step I involves taking the free DNA, along with its counterion complement to a conformation as in the complexed form. We identify the corresponding energy with the deformation expense of DNA. Step II (deformation of the protein) is similar in spirit to the first step, and involves taking the free protein in solution in its native state to its conformation in the complexed form. The next two steps (III and IV) involve transferring the protein and the DNA from aqueous medium to vacuum, essentially the desolvation steps. In each of these consider the energies to comprise three components

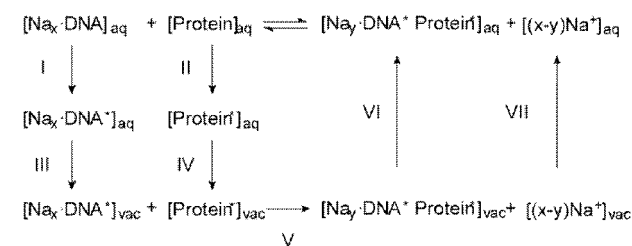


Figure 2. The thermodynamic cycle used for a component-wise analysis of the binding free energies of protein–DNA complexes. The * denotes the structure of the macromolecule in the final state as distinct from the initial state. The vertical arrows denote transfer processes.

viz., the electrostatic, the van der Waals, and the cavity formation expense. Transfer of the solute from aqueous medium to vacuum involves loss of favorable electrostatic and van der Waals interactions with solvent, and a gain from the cavity formation term. In step V, the protein and the DNA are brought together to their bound form as in the final complex. This involves turning on the electrostatic (with the dielectric constant set at unity) and the van der Waals interactions between the protein and the DNA. Some ions are released in this process. We consider the ion effects explicitly. A consideration of the loss in translational and rotational entropies of the free protein and DNA enter this step, as also the loss of vibrational and configurational entropy. In the next step (VI), the complex is transferred from vacuum to water, a solvation step. Also, the released counterions are transferred to the solvent (step VII). These again involve an electrostatic component, a van der Waals component and a cavity formation term, the former two being favorable and the latter unfavorable. The last contribution is the solvation free energy of small ions released during complexation.

The electrostatic contribution to the solvation energy is estimated via the generalized Born methodology.^{47–52} The parameters to be used for GB model solvation have been calibrated recently to be consistent with AMBER⁵³ Cornell et al. force field.⁵⁴ The added salt effects are included via a Debye–Huckel free energy term,²⁵ in a spirit similar to the GB treatment for solvation. The nonelectrostatic contributions due to van der Waals interactions of the solute with solvent and the work done to form a cavity in water are treated as linear functions of the solvent-accessible surface area with an empirical coefficient of 7.2 cal/Å following Still et al.⁴⁷ We treat this 7.2 cal as a composite of +47 cal/Å⁵⁵ as the hydrophobic component and –39.8 cal/Å⁵⁶ as the van der Waals component. The surface areas here refer to those of all atoms and not just the nonpolar atoms. There is considerable debate in the literature on the exact value to be employed for hydrophobic estimates. Whatever the decomposition of this surface area-based free energy, the net value of 7.2 cal/Å for nonelectrostatic contribution to solvation is calibrated on small molecules to be consistent with the GB treatment of solvation and the parameters therein.^{47, 50} All the molecular surface area calculations were performed using the ACCESS program based on the algorithm of Lee and Richards⁵⁷ and parm94 vdW radii.⁵⁴ The direct interactions between the protein and DNA in the complex in vacuum are calculated with the AMBER force field.⁵⁴

The polyelectrolyte effects, the configurational/vibrational entropy losses, and the deformation expense proved refractory to a simple treatment. These, in principle, can be treated rigorously via molecular dynamics simulations on the unbound and bound protein and DNA, followed by a free energy component analysis.^{37, 38, 58, 59} Identifying these as areas for further refinement, here we explore simpler theoretical alternatives to enumerate the diverse phenomenological contributors. The sodium counterions each of an effective charge +0.76, are placed on the bisector of phosphates 7 Å from the phosphorous atom. The interactions of these ions with DNA are computed. This procedure is repeated in the complex in presence of the protein, counterions clashing with the protein are attempted to be relocated in the vicinity, and those that could not be accommodated are treated as having been transferred to the bulk and considered released. Each condensed

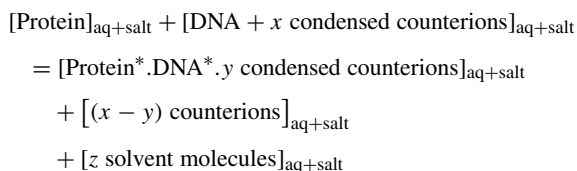
ion relative to the released ion is estimated to have lost an entropic contribution ($-T\Delta S$) of 2 kcal/mol. Further details of the methodology adopted for the calculation of each of the component terms can be accessed in ref. 25, where in the case of the EcoRI endonuclease–DNA complex is discussed elaborately as a prototypical system.

The loss in rotational and translational entropies of the protein and the DNA upon binding, have been calculated using the statistical mechanical framework by forming appropriate partition functions [eqs. (12) and (13) above]. The loss in side-chain conformational entropy has been estimated as 656 calories for each protein side chain contacting DNA ($R \ln 3$ at a temperature of 298 K). This involves a small modification of Janin's^{32, 33} approach, and partially accounts for the loss of conformational entropy of DNA as well. Finally, the deformation expense is derived for now, from the net binding free energy (assumed to be about -15 kcal/mol).

Other areas for improvement include a more accurate charge description for amino acid residues keeping in view their environment dependent pK shifts, a consideration of bound waters at the interface, the precise role of cofactors, the energetic aspects of folding/oligomerization of proteins upon binding, etc. This contribution thus essentially constitutes a progress report in generating the first computational atlas of the energetics of protein–DNA complexation, amenable to successive improvements, and for the present offers a qualitative energetic perspective of protein–DNA recognition.

Results and Discussion

The protein–DNA binding process, shown schematically in Figure 1, may be fully described by the following equation.



The “*” refers to the structural variations between the native protein/DNA and that in the complex upon binding.^{6, 60} The binding process as depicted above is accompanied by the release of water molecules and counterions.^{39–41, 61–63}

The standard molar free energy of formation for each of the complexes is computed via a thermodynamic cycle consisting of seven steps, as shown in Figure 2. The binding free energy is partitioned into 26 components accounting for the complexation process at a molecular level, and best estimates are made for each of the components. A component-wise analysis of the binding free energies for all the 40 complexes is provided in the appendix. Averages for each component are shown in Table 1. Note that of the 26 factors considered in the thermocycle, 10 of them favor complexation. Table 1 essentially presents a consensus view of the protein–DNA binding energetics.

We find that considering a phenomenon as favorable to complexation infact involves a nonunique combination of a few free energy components provided in Table 1. One such combination

used conventionally is shown in Figure 3 and the corresponding components identified in Table 2. In this view, the net electrostatic interactions, which include direct protein–DNA interactions and desolvation steps, become unfavorable. The small ion effects too are unfavorable.²⁵ Similar behavior of the electrostatic effects in protein–ligand and DNA–ligand systems have been observed by Honig and coworkers,^{24, 42, 64} based on finite difference Poisson–Boltzmann studies. Packing and hydrophobic interactions are the only forces favorable to complexation. Figure 3 further shows a split of van der Waals into direct interactions between the protein and the DNA, which is always favorable, and van der Waals component of desolvation, which is unfavorable. If this desolvation is combined with the component identified in Figure 3 as hydrophobic term, then binding becomes attributable predominantly to van der Waals interactions. If, however, the two van der Waals components are combined, then hydrophobic contributions dominate the factors favorable to binding. Similarly, all the desolvation components (electrostatic, van der Waals and hydrophobic) are combined, then desolvation disfavors complexation and electrostatics, and van der Waals are the only two factors favoring complexation. The hazards of coming to conclusions based on compounded subsets as listed in Table 2 are all too obvious.

Figure 3 further shows that the net desolvation combined with direct interactions and rotational, translational entropy losses leads to an average net binding free energy ($\Delta G'$) of -185.4 kcal. The ion effects, the vibrational entropy losses, and the deformation contribute a value of $+170.8$ kcal, resulting in the observed free energy values in the range of -7 to -15 kcal. These latter three components shown after $\Delta G'$ bar are estimates, and carry a lower level of confidence at the present stage of theory, as already pointed out in the previous section. Qualitative features as reported in Figure 3, however, are not expected to change with a larger data base or a more refined theory.

Steric Complementarity

The direct van der Waals interactions between the protein and the DNA, which quantify the extent of steric complementarity correlate well (correlation coefficient = -0.95) with the accessible surface area lost upon complexation and can be fitted to the following equation (for AMBER force field):

$$\begin{aligned} \text{vdW energy (in kcal/mol)} &= m \cdot (\text{Accessible area lost in } \text{\AA}^2) + c \\ m &= -0.0617 \pm 0.0033; \quad c = 17.60 \pm 12.17 \end{aligned}$$

The average energy coefficient for the surface area thus is about $61.7 \text{ cal}/\text{\AA}^2$. The tightness of packing reflected in the uncertainties of the above linear fit, varies from $-68.4 \text{ cal}/\text{\AA}^2$ for TATA binding protein to $-44.1 \text{ cal}/\text{\AA}^2$ for Erythroid transcription factor. Even after discounting for the loss in van der Waals interactions of the macromolecules with solvent upon binding (which is placed at $39.8 \text{ cal}/\text{\AA}^2$), there is an average net gain of $\sim 22 \text{ cal}/\text{\AA}^2$ upon complexation ascribable to net van der Waals interactions or packing forces at the interface.

For the specific recognition of DNA, the proteins can contact either the bases directly in the grooves, which is called the direct recognition mode or the phosphodiester backbone, whose structure is modulated indirectly by the base sequence. This latter is

Table 1. Calculated Average Primary Numbers in the Thermodynamic Cycle for 40 Protein–DNA Complexes.

Step	Term	Component	Contribution to $\Delta G_{\text{net}}^{\circ}$ (kcal/mol) ^a
Step I			
1	$\Delta H_1^{\text{adpt.d}}$	Structural adaptation enthalpy of DNA	
2	$-T\Delta S_2^{\text{adpt.d}}$	Structural adaptation entropy of DNA	
Step II			+55.4
3	$\Delta H_3^{\text{adpt.d}}$	Deformation enthalpy of protein	
4	$-T\Delta S_4^{\text{adpt.d}}$	Structural adaptation entropy of protein	
Step III			
5	$\Delta G_5^{\text{el.d}}$	El. component of DNA desolvation	+10985.1
6	$\Delta G_6^{\text{el.ci.d}}$	Counterion effect on DNA desolvation	-6280.0
7	$\Delta G_7^{\text{vdW.d}}$	vdW component of DNA desolvation	+266.3
8	$\Delta G_8^{\text{Cav.d}}$	Cavity component of DNA desolvation	-314.4
9	$\Delta G_9^{\text{DH.d}}$	Loss of added salt interactions with NaDNA	+24.5
Step IV			
10	$\Delta G_{10}^{\text{el.p}}$	Ele. component of protein desolvation	+4486.5
11	$\Delta G_{11}^{\text{el.ci.p}}$	Counterion effect on protein desolvation	~0.0
12	$\Delta G_{12}^{\text{dWp}}$	vdW component of protein desolvation	+473.2
13	$\Delta G_{13}^{\text{Cav.p}}$	Cavity component of protein desolvation	-558.8
14	$\Delta G_{14}^{\text{DH.p}}$	Loss of added salt interactions with protein	+30.6
Step V			
15	$\Delta H_{15}^{\text{vdW}}$	vdW interactions between protein and DNA in complex	-201.5
16	$\Delta H_{16}^{\text{el}}$	El. interactions between protein and DNA in complex	-6260.6
17	$\Delta H_{17}^{\text{ci}}$	Counterion (complex–DNA–protein) interactions	+3812.1
18	$-T\Delta S_{18}^{\text{ci}}$	Entropy of (complex–DNA–protein) counterions	-12.9
19	$-T\Delta S_{19}^{\text{tr.rot}}$	Rot and transl. entropy	+28.7
20	$-T\Delta S_{20}^{\text{vib.cnf}}$	Vib. and configurational entropy	+19.9
Step VI			
21	$\Delta G_{21}^{\text{el.c}}$	El. component of complex solvation	-9198.1
22	$\Delta G_{22}^{\text{el.ci.c}}$	Counterion effect on complex solvation	+2880.6
23	$\Delta G_{23}^{\text{vdW.c}}$	vdW component of complex solvation	-598.1
24	$\Delta G_{24}^{\text{Cav.c}}$	Cavity component of complex solvation	+706.3
25	$\Delta G_{25}^{\text{DH.c}}$	Added salt interactions with complex	-37.2
Step VII			
26	$\Delta G_{26}^{\text{fci}}$	Solvation free energy of released counterions	-322.2
	$\Delta G_{\text{net}}^{\circ}$	Net binding free energy	-14.6

^a 1 kcal = 4.18 kJ.

called the indirect mode of recognition. To develop a better understanding of recognition employed by the DNA binding proteins (DBPs), we show in Figure 4 the number of residues contacting either the backbone or the base pair per 100 Å² of the DNA area lost upon complexation for all the 40 systems. The upper triangle

in the figure denotes the domain of indirect code for recognition, and the lower triangle that of the direct code. If the DBPs show no such distinction, then with each point on the graph representing one protein–DNA complex, all points should lie along the diagonal. If, however, the DBPs were to adopt either direct or indirect code

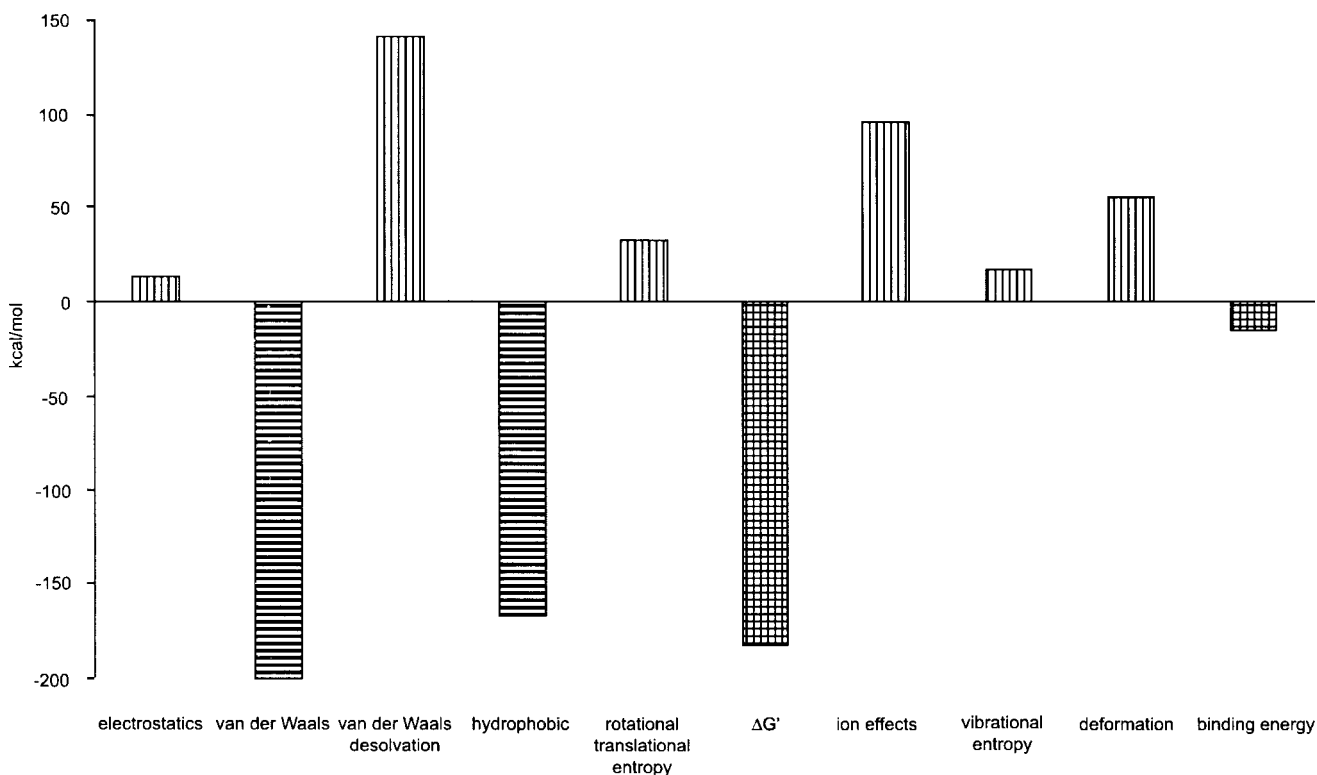


Figure 3. A histogram of the calculated primary contributions to the binding free energy of 43 protein–DNA complexes presented as composite subsets formed to be consistent with a traditional view of binding. The reference state is separated protein and DNA in aqueous medium at 25°C. Negative values (horizontal shade) are favorable and positive values (vertical shade) unfavorable to binding. The $\Delta G'$ (crosshatched) represents an intermediate sum of the first five bars viz., the net electrostatic, the van der Waals, the cavitation, and the rotational–translational entropy contributions. To this are added the next three bars viz. estimates of the ion effects, the vibrational entropy losses, and the deformation expense to obtain the net binding free energy (last green bar) (1 kcal = 4.18 kJ).

exclusively, then all points should lie closer to the axes either the abscissa or the ordinate. Figure 4 indicates that none of these extremes is true. Instead, a definite clustering of points in the interval of one to two residues, above and below the diagonal emerges as an interesting result. As to why the number of residues contacting DNA does not exceed three per 100 Å² of DNA area may be

a matter of packing density in the biomolecules and at the interface. It is clear that the DBPs can be said to interact with DNA predominantly but not exclusively via either direct code or indirect code. A further normalization of the DNA area contacted in terms of the accessible areas of phosphates and bases (the former being more accessible) can be contemplated with the goal of seek-

Table 2. A Traditional Combination of the Computed Thermodynamic Components for Protein–DNA-Specific Complex Formation (in kcal/mol^a).

Combination	Contribution	Components
1 van der Waals (direct + desolvation)	−60.1	7 + 12 + 15 + 23
2 Electrostatics (direct + desolvation)	+12.9	5 + 10 + 16 + 21
3 Hydrophobic (cavity term)	−166.9	8 + 13 + 24
4 Entropy loss (rot., trans., vib.) ($-T\Delta S$)	+48.6	19 + 20
5 Deformation enthalpy	+55.4	1 + 2 + 3 + 4
6 Small ion (counterion + added salt) effects	+95.5	6 + 9 + 11 + 14 + 17 + 18 + 22 + 25 + 26
Net binding free energy	−14.6	

^a 1 kcal = 4.18 kJ.

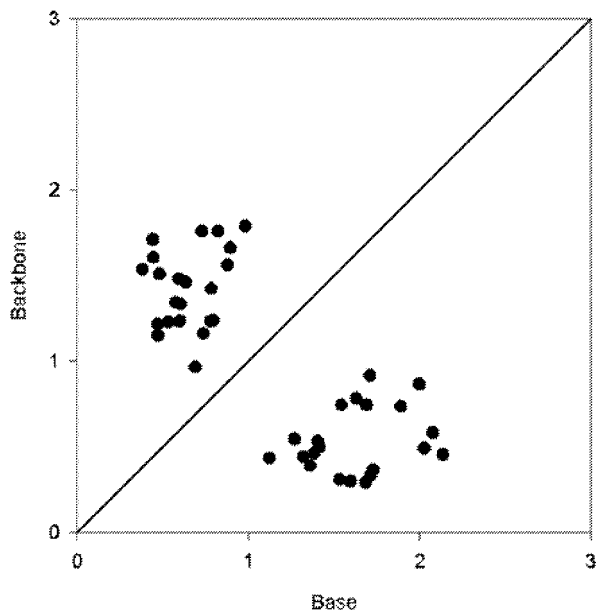


Figure 4. Number of protein residues per 100 Å² of the DNA area lost contacting primarily either the sugar-phosphate backbone or a base, averaged over all the complexes studied.

ing class or motif-dependent trends in DNA binding. This work is in progress.

Electrostatic Complementarity

Irrespective of the net charge on the protein, all the DNA binding proteins examined here carry a net positive charge at the interface. Including the direct Coulomb interactions between the protein and the DNA, the electrostatics of desolvation of protein and DNA and the small ion effects, the overall electrostatic contribution is net unfavorable to binding in all cases. With the ion effects separated out, net electrostatics (i.e., direct interactions + desolvation) becomes case specific—favorable in about half of the cases considered and unfavorable in the rest. Compensations between direct electrostatic interactions and desolvation energetics occur in all cases. To pin down the precise role of electrostatics in binding, we performed a computational experiment of switching off all the charged residues (formal charges) on the proteins. This led to an unfavorable net electrostatics in all cases with an average of +52 kcal/mol in contrast to the consensus value of 12.9 kcal/mol with the charges on. A neutral protein—effectively a collection of dipoles resulting from the partial charges on neutral residues and backbone—interacts favorably with DNA but incurs a heavier desolvation penalty and, hence, does not prefer to bind to DNA electrostatically. The effect of formal charges on the protein appears to be to decrease this unfavorable electrostatics to promote binding. The inference to be drawn here is that formal charges embedded on a protein gain more via direct interactions with DNA and lose less via desolvation, and that the electrostatics of dipole–dipole interactions alone is insufficient to cause binding of macromolecules in aqueous media.

A case-by-case perusal of the results suggests that endonucleases (EcoRI, EcoRV, PvuII), binding in an enveloping mode to short stretches of DNA (six base pairs), make a strong unfavorable electrostatic contribution to the overall binding free energy (biasing in fact the computed averages) due mainly to desolvation expense. The average net electrostatic contribution comes down from +12.9 to +2 kcal without the endonucleases. The length and extent of DNA coverage as well as the charge distribution appear to dictate the net electrostatics of complexation.

The spatial distribution of charge on the DNA binding proteins reveals an interesting pattern in the consensus view. In Figure 5, we show the number of charged residues (normalized with respect to the number of DNA base pairs contacted by the protein and averaged over all the 40 systems) as a function of distance from the DNA helical axis. Basic residues dominate the acidic residues in the vicinity of DNA (i.e., up to 12 Å), a primary requirement for binding to polyanionic DNA. That the negative charge population in the grooves and near the phosphates is nonzero, and that the acidic residues occur in larger fraction away from DNA are communicative of structural principles associated with specific recognition. Negative charges in the grooves can interact with the NH₂ groups on adenine and cytosine in the major groove and guanine in the minor groove. How can an Asp or Glu be tolerated near a phosphate remains a question. In a continuum solvent description such as the GB model employed here, the direct Coulomb repulsions are partially offset by favorable desolvation. A structural explanation requires the consideration of interfacial waters (work in progress). Preliminary results on the structural and energetic roles of interfacial waters indicates that water bridges are most likely to occur whenever an Asp or Glu is in the vicinity of a phosphate, thus buffering the repulsions. Preponderance of negative charges away from the DNA axis is clearly suggestive of an orientational role of facilitating the alignment of the face of the protein with excess positive charge towards the DNA.

Ion release

The small ion effects in the consensus view tend to be unfavorable to complexation in the following manner. The counterion–uncomplexed DNA oligonucleotide interactions (including solvent effects and loss in entropy of bound counterions) averaged over all the complexes is about −54.8 kcal/mol/CI, and varies from −51.9 kcal/mol/CI for met repressor (eight base pairs) to −57.1 kcal/mol/CI for γ - δ resolvase (33 base pairs). In the presence of the protein the average interaction energy (including entropy contribution) of the counterion with the complex is \sim −51.1 kcal/mol/CI, and varies from −44.6 kcal/mol/CI for EcoRV to −55.6 kcal/mol/CI for retinoic acid-receptor. This differential of 3.7 kcal/bound CI, between the free DNA and the complex accounts for the unfavorable contribution of the counterions. The net excess positive charge on the protein at the interface accounts for this decrease in interaction strength of the counterions with the complex relative to free DNA. This unfavorable magnitude of the counterion contribution, however, is partly reduced by the released ions, which in our model contribute −56.8 kcal/mol/CI.²⁵

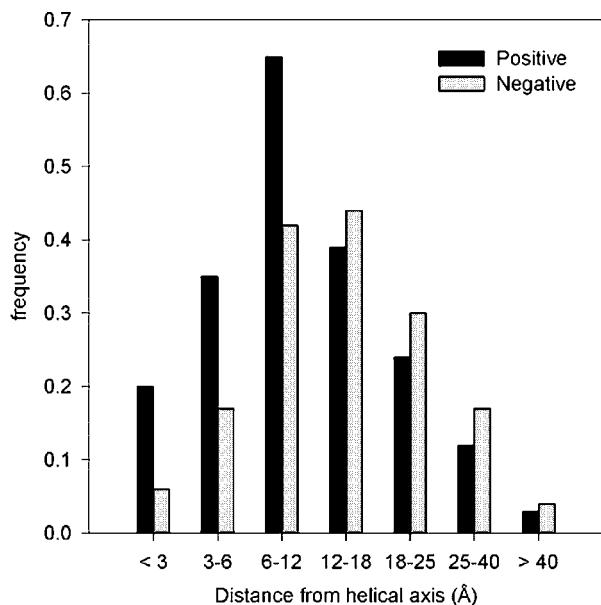


Figure 5. Frequency of occurrence of protein charges, averaged over all the complexes studied and normalized with respect to the number of base pairs contacted, shown as a function of distance from helical axis.

Deformation

The deformation expense is deduced here in this study from other components as mentioned in the Methods section and only detailed molecular dynamics simulations can fully resolve theoretically the role of structural adaptation. It is likely that the deformation contributions are overestimated, and the vibrational/configurational entropy losses are underestimated here.⁵⁸

Water Release

The cavitation contribution in our model captures the water release, and is found to favor binding in all cases. Both polar and nonpolar atoms contribute to this phenomenon.

Over all, the length and base sequence of the DNA binding site, the surface area covered, the charge distribution on the protein, the number of contacts to phosphates made by the protein all appear to act in a synergistic manner to accomplish specific binding.

How is the Protein–DNA Specific Binding Accomplished?

Any two macromolecules—for instance, two aliphatic (polymeric) chains of the size of a protein and its DNA binding site, involved in a noncovalent association without steric clashes—would show favorable van der Waals and cavitation contributions and unfavorable entropies, as for a DNA binding protein and its cognate DNA (Fig. 3). One-way specificity (base sequence-specific recognition of DNA) can enter this scenario is via the tightness of the fit. This would improve the van der Waals energy, but would require counterion displacement as well as DNA/protein deformation. Another feature may be the intrinsic propensity of a cognate DNA to adopt a structure complementary to the DNA binding protein, implying a smaller deformation penalty for the right

sequence.⁶⁵ The electrostatics via hydrogen bonds and interactions with backbone introduces an additional handle on specificity. A positively charged protein interface with DNA could facilitate ion displacement and deformation.^{66, 67} Thus, in contrast to a non-specific association of two macromolecules, both van der Waals and electrostatic interactions between the protein and the DNA could act in concert to accomplish specific binding while keeping the net binding free energies in a narrow interval (-9 to -17 kcal¹⁹).

Finally, owing to the nature of the additivity approximation in arriving at the net binding free energies and the large magnitude of some of the components, apprehensions arise regarding the uncertainties in the final free energies computed. Although it is difficult to estimate the level of uncertainty, our previous studies on EcoRI endonuclease–DNA complex,²⁵ λ repressor–operator complex,³⁸ U1A–RNA complex,⁶⁸ and some enzyme–inhibitor complexes^{69, 70} and their correspondence with experiment provide a validation of the methodology. Note that no system-specific parameterization is involved in the procedure. Also, the specific and the nonspecific complexes of EcoRV–DNA, where crystal structures are available for both systems, are well resolved by the protocol. Additionally, the λ -mutant is predicted to bind with a lower affinity than the native protein, as found experimentally. All these indicate to us a certain internal consistency of the methodology and a balance of terms, and generate the optimism that this can be turned into a predictive tool to address issues related to specificity and molecular design.

Conclusions

A statistical mechanical theory of binding based on free energy components is presented and applied to 40 protein–DNA complexes to develop a consensus view on the thermodynamics of macromolecular association. The key factors favoring protein–DNA binding are the direct electrostatic, van der Waals interactions and the cavitation contributions. The electrostatic and van der Waals components of desolvation disfavor complexation and so do the deformation, the small ion effects, and the rotational, translational, and vibrational entropies. Large proteins enveloping short stretches of DNA such as in enzyme–DNA complexes lead to large unfavorable net electrostatics, but highly favorable van der Waals and cavitation contributions. In contrast, DBPs contacting longer stretches of DNA such as in repressor–operator complexes exhibit negligible net electrostatics but large unfavorable ion effects in addition to favorable van der Waals and cavitation. Such compensations may be necessary for optimal binding. Proteins use a mixed mode of recognition with neither direct nor indirect code dominating exclusively. DBPs on an average use about two strategically positioned formal charges per base pair for recognition. Overall, the present study reveals in molecular detail the nature of the thermodynamic balance achieved between diverse opposing forces, which lead to the observed binding free energies in the range of -9 to -17 kcal in all specific protein–DNA complexes.

Appendix

Table 3. Electrostatic Contribution to Solvation and Interaction Energies of Some Protein–DNA Complexes (in kcal/mol)^a.

NDB/PDB Code	System	DNA Protein Charge	DNA Desolvn. (5) ^b	Protein Desolvn. (10)	Complex Solvn. (21)	Desolvn. Energy (5 + 10 + 21)	Pr-DNA el. int. eng. (16)	Net elec. Energy (5 + 10 + 16 + 21)
1. pde001	EcoRI	−24; −4	6892.6	5832.1	−11045.0	1679.7	−1538.0	141.7
2. pde003	EcoRV-sp	−18; +4	4668.1	7175.5	−8167.4	3676.2	−3530.4	145.8
3. pde017	PvuII	−24; −6	6937.6	4191.4	−11389.4	−260.4	+446.7	186.3
4. pdr008	Met rep	−16; −8	3862.2	4004.7	−7216.3	650.6	−623.5	27.1
5. pdr009	Trp rep	−36; −8	12907.8	4209.3	−17891.7	−774.6	+821.9	47.3
6. pdr010	λ Rep	−38; +2	13762.2	3154.7	−13535.9	3381.0	−3358.8	22.2
7. pdr012	Arc-rep	−42; +16	15820.3	5226.6	−9919.5	11127.4	−11151.9	−24.5
8. pdr001	434Cro/r1	−36; +13	12923.7	3050.9	−8695.4	7279.2	−7299.5	−20.3
9. pdr004	434Rep/r1	−36; +8	12930.5	2365.2	−10028.8	5266.9	−5240.0	26.9
10. pdr011	434Rep/r2	−36; +8	12921.1	2211.0	−9841.7	5290.4	−5276.1	14.3
11. pdr015	434Rep/r3	−36; +8	12996.3	2232.5	−9884.8	5344.0	−5337.7	6.3
12. pdr020	Pur Rep	−30; −8	9904.5	8671.7	−14601.5	3974.7	−3886.1	88.6
13. pdt016	Myod bhlh	−26; +18	8029.2	5824.8	−4530.4	9323.6	−9402.7	−79.1
14. pdt023	Max bhlh	−42; +12	16077.6	4872.4	−12426.0	8524.0	−8560.1	−36.1
15. pdt004	Engrl HD	−40; +17	14979.0	4981.9	−9786.4	10174.5	−10256.1	−81.6
16. pdr018	Prd dmn.	−28; +9	8867.3	2815.1	−6558.4	5124.0	−5122.9	1.1
17. pdt019	Oct-1POU	−26; +6	8028.9	3313.7	−6263.3	5079.3	−5038.0	41.3
18. pdr035	Inter. RF	−48; +12	19572.0	5479.4	−15158.4	9893.0	−9828.5	64.5
19. 1FJL	Paired HD	−26; +11	8089.9	3892.9	−4482.0	7500.8	−7461.7	39.1
20. pdt028	Mata2HD	−40; +14	15146.5	3960.2	−9264.6	9842.1	−9856.4	−14.3
21. pdt005	Mata2HD	−40; +16	14996.9	4512.2	−8736.4	10772.7	−10784.0	−11.3
22. pdt031	Evn sk HD	−18; +16	4707.4	4653.2	−3003.2	6357.4	−6342.8	14.6
23a. pdt017	Pyr-Pw Inh	−26; +7	8113.4	4038.1	−7398.0	4753.5	−4765.3	−11.8
23b. pdt017	Pyr-Pw Inh	−26; +19	8142.0	9133.9	−7983.1	9292.8	−9337.7	−44.9
24. pdt011	TT-Tr.reg	−36; +9	12976.7	2024.0	−9359.4	5641.3	−5673.1	−31.8
25. pdtb41	SynZn fng	−24; +8	7099.8	2424.9	−4722.9	4801.8	−4813.2	−11.4
26. pdt008	GLI	−38; +18	14003.2	6727.8	−8348.0	12383.0	−12421.8	−38.8
27. pdt039	Zif268	−20; +11	5452.6	2836.5	−3203.8	5085.3	−5092.0	−6.7
28. pdt003	GAL4	−36; +14	13068.5	4365.7	−9085.7	8348.5	−8401.3	−52.8
29. 1UBD	Ying Yang 1	−38; +11	14023.7	3196.4	−9403.1	7817.0	−7848.0	−31.0
30. 1GAT	Erythroid TF	−16; +7	3783.0	1629.5	−2921.8	2490.7	−2476.4	14.3
31. 2NLL	Reti-Recp	−33; +9	11876.2	4114.0	−9051.1	6939.1	−6962.2	−23.1
32. pdrc01	Gluc-Recp	−34; +14	11835.1	4723.0	−9219.1	7339.0	−7338.8	0.2
33. pdrc03	Estr-Recp	−34; +6	11971.9	3339.6	−10732.5	4579.0	−4580.0	−1.0
34. pdr021	Nuc Recp As	−34; +25	11893.8	9740.8	−7719.0	13915.6	−13999.3	−83.7
35. pdt022	NF κβp50	−36; +4	12511.9	9687.6	−15309.8	6889.7	−6869.2	20.5
36. pde0115	γδ resolv.	−66; +14	30845.7	7552.9	−25219.6	13179.0	−13189.6	−10.6
37. pdt009	TBP	−26; +13	8175.6	3650.3	−6351.2	5474.7	−5448.2	26.5
38. pde009	Hin Rec.	−25; +7	7634.4	1561.1	−5280.6	3914.9	−3888.2	26.7
39. pdeb08	HHAI metr	−22; +1	6206.7	4192.2	−7294.2	3104.7	−3042.7	62.0
40. pde022	Pyri-repair	−24; +6	6991.4	2120.2	−5234.4	3877.2	−3840.2	37.0
41. pdt035	Rapl-telm	−36; +2	12907.5	4429.9	−12655.7	4681.7	−4662.9	18.8
42. pdte01	Repl-Term	−26; +9	7823.8	4798.0	−6599.1	6022.7	−5927.3	95.4
Average	Consensus	−33; +9	10985.1	4486.5	−9198.1	6273.5	−6260.6	+12.9

^a 1 kcal = 4.18 kJ.^b Numbers in parentheses correspond to step numbers in Table 1.

Table 4. Nonelectrostatic (van der Waals and Cavitation) Contributions to Solvation Energies of Some Protein–DNA Complexes (in kcal/mol)^a.

NDB/PDB Code	System	Surface Area Lost (Å ²)	DNA Desol. vdW Eng (7) ^b	DNA Desol. Cav Eng (8)	Protein Desol. vdW Eng (12)	Protein Desol. Cav Eng (13)	Complex Sol. vdW Eng (23)	Complex Sol. Cav Eng (24)	Net ne Desolv. (7 + 8 + 12 + 13 + 23 + 24)
1. pde001	EcoRI	4306	217.7	−257.1	899.3	−1061.9	−945.1	+1116.1	−31.0
2. pde003	EcoRV-sp	4819	163.8	−193.4	901.3	−1064.3	−872.4	+1030.3	−34.7
3. pde017	PvuII	4764	216.0	−255.1	650.9	−768.7	−677.6	+800.2	−34.3
4. pdr008	Met Rep	1861	150.4	−177.6	462.3	−546.0	−539.0	+636.5	−13.4
5. pdr009	Trp Rep	3375	301.0	−355.4	460.8	−544.1	−626.8	+740.2	−24.3
6. pdr010	λ Rep	3847	311.2	−367.5	421.0	−497.2	−579.4	+684.2	−27.7
7. pdr012	Arc-Rep	4583	346.0	−408.5	488.8	−577.2	−651.4	+769.3	−33.0
8. pdr001	434Cro/r1	3194	297.3	−351.1	318.9	−376.6	−488.7	+577.2	−23.0
9. pdr004	434Rep/r1	3028	293.2	−346.3	282.6	−333.7	−455.7	+538.1	−21.8
10. pdr011	434Rep/r2	3153	299.4	−353.5	283.5	−334.8	−457.1	+539.8	−22.7
11. pdr015	434Rep/r3	3083	294.7	−348.0	281.8	−332.8	−454.0	+536.1	−22.2
12. pdr020	Pur Rep	4028	253.3	−299.1	1115.3	−1317.0	−1207.4	+1425.9	−29.0
13. pdt016	Myod bhlh	3042	222.2	−262.4	435.0	−513.7	−535.8	+632.8	−21.9
14. pdt023	Max bhlh	2944	342.1	−404.0	544.7	−643.2	−769.7	+908.9	−21.2
15. pdt004	Engrl HD	2903	324.3	−382.9	367.0	−433.4	−575.6	+679.7	−20.9
16. pdr018	Prd domn.	2847	238.1	−281.2	371.6	−438.8	−496.6	+586.4	−20.5
17. pdt019	Oct-1POU	3750	223.4	−263.8	383.0	−452.3	−457.0	+539.7	−27.0
18. pdr035	Inter. RF	4903	378.2	−446.6	556.4	−657.1	−739.3	+873.1	−35.3
19. 1FJL	Paired HD	4056	222.0	−262.1	378.4	−446.8	−438.1	+517.4	−29.2
20. pdt028	Mata2HD	4069	331.3	−391.3	386.7	−456.7	−556.8	+657.5	−29.3
21. pdt005	Mata2HD	4250	330.1	−389.8	400.8	−473.3	−561.6	+663.2	−30.6
22. pdt031	Evn sk HD	3514	163.9	−193.6	376.5	−444.6	−400.8	+473.3	−25.3
23a. pdt017	Pyr-Pwinh	2292	224.0	−264.5	398.6	−470.7	−530.8	+626.9	−16.5
23b. pdt017	Pyr-Pwinh	2194	221.1	−261.1	393.7	−465.0	−528.3	+623.8	−15.8
24. pdt011	TT-TF	2042	298.4	−352.4	205.0	−242.1	−422.0	+498.4	−14.7
25. pdt041	SynZn fng	3028	209.7	−247.6	252.5	−298.2	−341.4	+403.2	−21.8
26. pdt008	GLI	3819	304.4	−359.4	441.4	−521.3	−593.3	+700.7	−27.5
27. pdt039	Zif-268	3208	180.6	−213.3	277.6	−327.8	−331.0	+390.8	−23.1
28. pdt003	GAL4	2875	294.0	−347.2	387.1	−457.2	−566.8	+669.4	−20.7
29. 1UBD	Ying Yang 1	3500	308.9	−364.7	356.7	−421.3	−526.1	+621.3	−25.2
30. 1GAT	Erythroid TF	2319	139.2	−164.3	230.5	−272.1	−276.6	+326.6	−16.7
31. 2NLL	Reti-Recp	3653	280.1	−330.7	476.6	−562.8	−611.1	+721.6	−26.3
32. pdrc01	Gluc-Recp	3069	283.8	−335.1	429.3	−506.9	−590.3	+697.1	−22.1
33. pdrc03	Estr-Recp	3000	279.9	−330.6	359.2	−424.2	−519.8	+613.9	−21.6
34. pdr021	Nuc Recp As	3583	279.2	−329.8	478.1	−564.6	−615.5	+726.8	−25.8
35. pdt022	NF κβp50	4569	318.6	−376.2	1273.0	−1503.3	−1409.4	+1664.4	−32.9
36. pde0115	γδ resolv.	6306	516.1	−609.5	868.2	−1025.3	−1133.9	+1339.0	−45.4
37. pdt009	TBP	3194	224.4	−265.0	424.1	−500.9	−521.5	+615.9	−23.0
38. pde009	Hin Rec.	3042	211.1	−249.3	184.4	−217.7	−274.1	+323.7	−21.9
39. pdeb08	HHAImetr	3417	205.7	−242.9	604.5	−713.8	−673.7	+795.6	−24.6
40. pde022	Pyr-repair	3000	226.2	−267.1	310.8	−367.0	−417.5	+493.0	−21.6
41. pdt035	Rapl-telm	4639	298.2	−352.1	515.2	−608.4	−628.9	+742.6	−33.4
42. pdte01	Repl-Term	5542	226.0	−266.8	715.3	−844.7	−720.6	+850.9	−39.9
Average	Consensus	3556	266.3	−314.4	473.2	−555.8	−598.1	+706.3	−25.5

^a 1 kcal = 4.18 kJ.^b Numbers in parentheses correspond to step numbers in Table 1.

Table 5. Net Estimated Contributions to the Binding Free Energies of Some Protein–DNA complexes (kcal/mol)^a.

NDB/PDB Code	System	Net el Energy Table 3	Net Non-el Desolv. Table 4	Pro-DNA vdW (15) ^b	Entropy Losses tr,rot,vb (19 + 20)	Ion Effects (6 + 9 + 11 + 14 + 17 + 18 + 22 + 25 + 26)	Deform. (1 + 2 + 3 + 4)	Net Binding Free Energy
1. pde001	EcoRI	141.7	-31.0	-271.0	59.1	36.2	50.0	-15.0
2. pde003	EcoRV-sp	145.8	-34.7	-274.9	59.6	26.5	62.7	-15.0
3. pde017	PvuII	186.3	-34.3	-270.9	56.3	13.6	34.0	-15.0
4. pdr008	Met rep	27.1	-13.4	-97.5	40.0	-26.1	54.9	-15.0
5. pdr009	Trp rep	47.3	-24.3	-166.4	49.4	86.4	0.0	-7.6
6. pdr010	λ Rep	22.2	-27.7	-204.4	50.8	132.2	11.9	-15.0
7. pdr012	Arc-rep	-24.5	-33.0	-311.7	62.4	150.6	141.2	-15.0
8. pdr001	434Cro/r1	-20.3	-23.0	-194.2	47.1	85.0	90.4	-15.0
9. pdr004	434Rep/r1	26.9	-21.8	-187.5	47.0	112.8	7.6	-15.0
10. pdr011	434Rep/r2	14.3	-22.7	-196.6	48.3	95.4	46.3	-15.0
11. pdr015	434Rep/r3	6.3	-22.2	-185.3	46.3	74.6	65.3	-15.0
12. pdr020	Pur Rep	88.6	-29.0	-261.2	54.2	66.2	66.2	-15.0
13. pdt016	Myod bhlh	-79.1	-21.9	-143.8	40.9	183.0	5.9	-15.0
14. pdt023	Max bhlh	-36.1	-21.2	-151.9	41.1	138.6	14.5	-15.0
15. pdt004	Engrl HD	-81.6	-20.9	-131.8	44.1	118.8	56.4	-15.0
16. pdr018	Prd domn.	1.1	-20.5	-158.0	44.2	66.6	51.6	-15.0
17. pdt019	Oct-1POU	41.3	-27.0	-198.6	48.2	48.4	72.7	-15.0
18. pdr035	Inter. RF	64.5	-35.3	-273.4	42.2	162.8	24.2	-15.0
19. 1FJL	Paired HD	39.1	-29.2	-247.1	46.8	140.9	34.5	-15.0
20. pdt028	Matα2HD	-14.3	-29.3	-236.8	46.8	185.6	33.0	-15.0
21. pdt005	Matα2HD	-11.3	-30.6	-240.5	46.9	146.3	74.2	-15.0
22. pdt031	Evn sk HD	14.6	-25.3	-185.9	41.5	84.3	55.8	-15.0
23a. pdt017	Pyr-Pwinh	-11.8	-16.5	-132.5	41.3	54.7	49.8	-15.0
23b. pdt017	Pyr-Pwinh	-44.9	-15.8	-120.2	41.3	94.6	30.0	-15.0
24. pdt011	TT-TF	-31.8	-14.7	-114.7	38.5	88.3	19.4	-15.0
25. pdtb41	SynZn fng	-11.4	-21.8	-143.6	47.0	76.9	37.9	-15.0
26. pdt008	GLI	-38.8	-27.5	-190.0	50.3	129.4	61.6	-15.0
27. pdt039	Zif-268	-6.7	-23.1	-161.8	45.5	74.0	57.1	-15.0
28. pdt003	GAL4	-52.8	-20.7	-169.3	44.2	101.9	81.7	-15.0
29. 1UBD	YingYang 1	-31.0	-25.2	-166.2	48.1	118.1	41.2	-15.0
30. 1GAT	Erythr-TF	14.3	-16.7	-102.2	35.2	13.7	40.7	-15.0
31. 2NLL	Reti-Recp	-23.1	-26.3	-209.7	54.8	50.5	138.8	-15.0
32. pdrc01	Gluc-Recp	0.2	-22.1	-167.9	45.3	46.9	82.6	-15.0
33. pdrc03	Estr-Recp	-1.0	-21.6	-180.6	48.5	30.2	109.5	-15.0
34. pdr021	Nuc Recp As	-83.7	-25.8	-189.4	50.8	123.3	109.8	-15.0
35. pdt022	NF κβp50	20.5	-32.9	-243.3	56.2	110.7	73.8	-15.0
36. pde0115	γδ resolv.	-10.6	-45.4	-379.9	65.7	251.2	104.0	-15.0
37. pdt009	TBP	26.5	-23.0	-218.4	49.6	103.2	47.1	-15.0
38. pde009	Hin Rec.	26.7	-21.9	-196.0	40.9	90.0	45.3	-15.0
39. pdeb08	HHAI metr	62.0	-24.6	-213.5	54.7	41.3	65.1	-15.0
40. pde022	Pyr-repair	37.0	-21.6	-193.1	47.2	111.3	4.2	-15.0
41. pdt035	Rapl-telm	18.8	-33.4	-280.7	60.8	87.8	131.7	-15.0
42. pdte01	Repl-Term	95.4	-39.9	-303.2	59.1	181.3	0.0	-7.3
Average	Consensus	+12.9	-25.5	-201.5	48.6	95.5	55.4	-14.6

^a 1 kcal = 4.18 kJ.^b Numbers in parentheses correspond to step numbers in Table 1.

Acknowledgments

The authors are grateful to Professors R. Lavery, K. Zakrzewska, I. Russu, L. Jen-Jacobson, J. Rosenberg, and B. Honig for many helpful suggestions and comments on the project.

References

1. Matthews, B. W. *Nature* 1988, 335, 294.
2. Steitz, T. A. *Q Rev Biophys* 1990, 23, 205.
3. Sigler, P. B. *Curr Opin Struct Biol* 1991, 1, 61.
4. Harrison, S. C. *Nature* 1991, 253, 715.
5. Pabo, C. O.; Sauer, R. T. *Annu Rev Biochem* 1992, 61, 1053.
6. Travers, A. *DNA-Protein Interactions*; Chapman and Hall: London, 1993.
7. von Hippel, P. H. *Science* 1994, 263, 769.
8. Harrington, R. E.; Winicov, I. *Prog Nucleic Acid Res Mol Biol* 1994, 47, 195.
9. Lilley, D. M. J. *DNA-Protein: Structural Interactions*; Oxford University Press: Oxford, 1995.
10. Suzuki, M.; Brenner, S. E.; Gerstein, M.; Yagi, N. *Protein Eng* 1995, 8, 319.
11. Burley, S. K.; Roeder, R. G. *Annu Rev Biophys Biomol Struct* 1997, 26, 289.
12. Schwabe, J. W. *Curr Opin Struct Biol* 1997, 7, 126.
13. Choo, Y.; Klug, A. *Curr Opin Struct Biol* 1997, 7, 117.
14. Crothers, D. M. *Proc Natl Acad Sci USA* 1998, 95, 15163.
15. Olson, W. K.; Gorin, A. A.; Lu, X.-J.; Hock, L. M.; Zhurkin, V. B. *Proc Natl Acad Sci USA* 1998, 95, 11163.
16. Clore, G. M.; Gronenborn, A. M. *Prog Biophys Mol Biol* 1994, 62, 153.
17. von Hippel, P. H.; Berg, O. G. *Proc Natl Acad Sci USA* 1986, 83, 1608.
18. Lesser, D. R.; Kupriewski, M. R.; Jen-Jacobson, L. *Science* 1990, 250, 776.
19. Jen-Jacobson, L. *Biopolymers* 1997, 44, 153.
20. Nadassy, K.; Wodak, S. J.; Janin, J. *Biochemistry* 1999, 38, 1999.
21. Pichierrri, F.; Aida, M.; Gromiha, M. M.; Sarai, A. *J Am Chem Soc* 1999, 121, 6152.
22. Jayaram, B.; DiCapua, F. M.; Beveridge, D. L. *J Am Chem Soc* 1991, 113, 5211.
23. Marky, N. L.; Manning, G. S. *J Am Chem Soc* 2000, 122, 6057.
24. Froloff, N.; Windemuth, A.; Honig, B. *Protein Sci* 1997, 6, 1293.
25. Jayaram, B.; McConnell, K. J.; Dixit, S. B.; Beveridge, D. L. *J Comput Phys* 1999, 151, 333.
26. Jayaram, B.; Das, A. Aneja, N. *J Mol Struct (Theochem)* 1996, 361, 249.
27. Jayaram, B. In *Proceedings of the Ninth Conversation in Biomolecular Stereodynamics*; Sarma, R. H.; Sarma, M. H., Eds.; Adenine Press: New York, 1996; vol. 1, p. 109.
28. Dixit, S. B.; Jayaram, B. *J Biomol Struct Dynam* 1998, 16, 237.
29. Dixit, S. B.; Arora, N.; Jayaram, B. *J Biomol. Struct. Dynam.*; *Proceedings of the Eleventh Conversation in Biomolecular stereodynamics*; Sarma, R. H.; Sarma, M. H., Eds.; Adenine Press: New York, 2000; p. 109.
30. Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*; Plenum: New York, 1992.
31. Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys J* 1997, 72, 1047.
32. Janin, J. *Prog Biophys Mol Biol* 1995, 64, 145.
33. Janin, J. *Proteins* 1995, 21, 30.
34. Ajay; Murcko, M. A. *J Med Chem* 1995, 38, 4953.
35. Atkins, P. W. *Physical Chemistry*; W. H. Freeman and Company: New York, 1990.
36. Beveridge, D. L.; DiCapua, F. M. *Annu Rev Biophys Biophys Chem* 1989, 18, 431.
37. Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. *J Am Chem Soc* 1998, 120, 10629.
38. Kombo, D. C.; Jayaram, B.; Beveridge, D. L. *J Mol Sim*, to appear.
39. Manning, G. S. Q. *Rev Biophys* 1978, 11, 179.
40. Jayaram, B.; Beveridge, D. L. *Annu Rev Biophys Biomol Struct* 1996, 25, 367.
41. Record, M. T., Jr.; Anderson, C. F.; Lohman, T. Q. *Rev Biophys* 1979, 11, 103.
42. Misra, V. K.; Hecht, J. L.; Sharp, K. A.; Freidman, R. A.; Honig, B. *J Mol Biol* 1994, 238, 264.
43. Zacharias, M.; Luty, B. A.; Davis, M. E.; McCammon, J. A. *Biophys J* 1992, 63, 1280.
44. Otwinowski, Z.; Schevitz, R. W.; Zhang, R. G.; et al. *Nature* 1988, 335, 321.
45. Bernstein, F. C.; et al. *J Mol Biol* 1977, 112, 535.
46. Berman, H. M.; et al. *Biophys J* 1992, 63, 751.
47. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J. *J Am Chem Soc* 1990, 112, 6127.
48. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J Phys Chem* 1996, 100, 19824.
49. Jayaram, B.; Liu, Y.; Beveridge, D. L. *J Chem Phys* 1998, 109, 1465.
50. Jayaram, B.; Sprous, D.; Beveridge, D. L. *J Phys Chem* 1998, 102, 9571.
51. Cheatham, T. E., III; Srinivasan, J.; Case, D. A.; Kollman, P. A. *J Biomol Struct Dyn* 1998, 16, 265.
52. Bashford, D.; Case, D. A. *Annu Rev Phys Chem* 2000, 51, 129.
53. Pearlman, D. A.; et al. *Comput Phys Commun* 1995, 91, 1.
54. Cornell, W. D.; et al. *J Am Chem Soc* 1995, 117, 5179.
55. Sharp, K. A.; Nicholls, A.; Fine, R. A.; Honig, B. *Science* 1991, 252, 106.
56. Ohtaki, H. *J. Solution Chem* 1992, 21, 39.
57. Lee, B. K.; Richards, F. M. *J Mol Biol* 1971, 55, 379.
58. Duan, Y.; Wilkosz, P.; Rosenberg, J. M. *J Mol Biol* 1996, 264, 546.
59. Sen, S.; Nilsson, L. *Biophys J* 1999, 77, (a) 1782; (b) 1801.
60. Koudelka, G. B.; Harrison, S. C.; Ptashne, M. *Nature* 1987, 326, 886.
61. Spolar, R. S.; Record, M. *Science* 1994, 263, 777.
62. Garner, M. M.; Rau, D. C. *EMBO J* 1995, 14, 1257.
63. Lunback, T.; Hard, T. *Proc Natl Acad Sci USA* 1996, 93, 4754.
64. Misra, V. K.; Honig, B. *Proc Natl Acad Sci USA* 1995, 92, 4691.
65. Flatters, D.; Lavery, R. *Biophys J* 1998, 75, 372.
66. Lebrun, A.; Lavery, R. *Biopolymers* 1999, 49, 341.
67. Gurlie, R.; Zakrzewska, K. *J Biomol Struct Dynam* 1999, 16, 605.
68. Blakaj, D.; Baranger, A.; Beveridge, D. L. *J Am Chem Soc* 2001, 123, 2548.
69. Kalra, P.; Das, A.; Dixit, S. B.; Jayaram, B. *Indian J Chem* 2000, 39A, 262.
70. Kalra, P.; Das, A.; Jayaram, B. *Appl Biochem Biotechnol*, to appear.