Journal of Computational Physics **151,** 333–357 (1999)
Article ID jcph.1998.6173, available online at http://www.idealibrary.com on IDEAL®

Free Energy Analysis of Protein–DNA Binding: The EcoRI Endonuclease–DNA Complex

B. Jayaram,* K. J. McConnell, Surjit B. Dixit,* and D. L. Beveridge[1]

*Department of Chemistry and Program in Molecular Biophysics, Wesleyan University, Middletown, Connecticut-06459; and *Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India*

Received August 4, 1998; revised December 7, 1998

A detailed theoretical analysis of the thermodynamics and functional energetics of protein–DNA binding in the EcoRI endonuclease–DNA complex is presented. The standard free energy of complexation is considered in terms of a thermodynamic cycle of seven distinct steps decomposed into a total of 24 well-defined components. The model we employ involves explicit all-atom accounts of the energetics of structural adaptation of the protein and the DNA upon complex formation; the van der Waals and electrostatic interactions between the protein and the DNA; and the electrostatic polarization and screening effects, van der Waals components, and cavitation effects of solvation. The ion atmosphere of the DNA is described in terms of a counterion condensation model combined with a Debye–Huckel treatment of added salt effects. Estimates of entropy loss due to decreased translational and rotational degrees of freedom in the complex relative to the unbound species based on classical statistical mechanics are included, as well as corresponding changes in the vibrational and configurational entropy. The magnitudes and signs of the various components are estimated from the AMBER parm94 force field, generalized Born theory, solvent accessibility measures, and empirical estimates of quantities related to ion release.

The calculated standard free energy of formation, -11.5 kcal/mol, agrees with experiment to within 5 kcal/mol. This net binding free energy is discerned to be the resultant of a balance of several competing contributions associated with chemical forces as conventionally defined, with 10 out of 24 terms favoring complexation. Contributions to binding compounded from subsets of the 24 components provide a basis for advancing a molecular perspective of binding in terms of structural adaptation, electrostatics, van der Waals interactions, hydrophobic effects, and small ion reorganization and release upon complexation. The van der Waals interactions and water release favor complexation, while electrostatic interactions, considering both intramolecular and solvation effects, prove unfavorable. Analysis of individual contributions to the standard free energy of complexation at the nucleotide and amino

[1] To whom correspondence should be addressed.

0021-9991/99 $30.00

acid residue level highlights the role of contact interactions as well as context effects. Some patterns in compensation effects among the various terms are identified and discussed.

I. INTRODUCTION

The nature of binding and specificity in intermolecular interactions is a problem of central interest in chemistry and biology. In chemical systems, binding energies and free energies can be calculated accurately by molecular quantum mechanics in the gas phase [1] and by molecular dynamics and free energy simulations in condensed phases [2–5]. In biological systems, the situation is somewhat more complex, since interactions among charged and conformationally labile macromolecules in solution must usually be considered. For this class of problems, molecular quantum mechanics is intractable. Free energy simulations [6], while possible in principle, remain computationally intensive, and results may be subject to convergence problems and statistical uncertainties. Although major advances in structure determination have occurred in recent years and reliable measurements of binding constants have been reported, making the link between structure and thermodynamics in protein–DNA complexes continues to be a challenge. In particular, a number of key quantities required to develop a causal chain of inference are not directly measurable in experiments, and must be obtained from molecular modeling and structure-based theoretical calculations.

Molecular mechanics (MM) calculations and molecular dynamics (MD) simulations based on all-atom models including counterions and water are just becoming feasible for systems of the size and complexity of macromolecular complexes. Most MD studies on this class of systems reported to date, with notable exceptions [7], have addressed dynamical structure but not free energies, for reasons as indicated above. Considerable progress has been achieved using continuum electrostatics on the theory of salt-dependent components of the standard free energies of solvation [8, 9]. The next stage of the problem, incorporating both intramolecular and intermolecular interactions consistently into a thermodynamic treatment of complexation, is presently a developing area of research with a number of alternative strategies being investigated [10]. Recent improvements in the description of intermolecular interactions using empirical force fields [11, 12] and new methodology for obtaining estimates of the free energy of solvation simply but accurately using the "generalized Born–solvent accessibility" (GBSA) model [13–18] present to us a basis for simple, phenomenological free energy analyses of macromolecular binding processes. Using these developments, we describe herein a theoretical "component analysis" of the standard free energy of binding for a protein–DNA complex in solution at 298 K. Protein-DNA interactions are a class of systems fundamental to regulatory and catalytic processes in biology, in which the nature of the exquisite specificity is yet to be fully explained at the molecular level. Observed values of binding constants for specific complexes are well differentiated from those of nonspecific complexes [19], but a proper theoretical account of the thermodynamics of binding is a necessary prerequisite to understanding specificity in general.

The binding of the restriction enzyme EcoRI endonuclease to its cognate DNA sequence in solution (Fig. 1) is the focus of our current investigation. This complex is particularly propitious for study, since the structure of the binary complex and the uncomplexed DNA have been determined from crystallography [20–22], and extensive experiments on binding equilibria have been carried out [23]. We describe herein a detailed analysis of the

binding free energy for the EcoRI DNA complex based on a thermocycle of seven steps, with the contributions to free energy for each step of the cycle delineated with regard to the current "toolchest" of techniques for making reasonable theoretical estimates. The free energies of the seven steps are calculated from a total of 24 primary contributions, using an eclectic mix of all-atom, discretized solvent MM and MD studies, GBSA models of hydration, Debye–Huckel (DH) theory of added salt effects, and physicochemical estimates. Our laboratory vantage point is all-atom MD models of biological molecules in solution [24]. Here we seek to explore the viable articulation of fully discrete MM and MD models with phenomenological yet accurate treatments of solvation, and to obtain a robust, theoretical reduction of macromolecular binding processes in solution into terms that can be independently calculated, assessed, and subjected to successive improvements.

The methodology adopted for this study, "free energy component analysis," is attractive in its simplicity, but problematic for a number of reasons. The analysis of errors, some of which are not readily quantifiable, assumptions about additivity, and nonuniqueness of partitioning of terms mitigate the results. Taking the EcoRI DNA complex considered here as a case study, we conclude Section VI with some perspectives on the uncertainties that arise in the application of free energy component analysis to ligand binding and the implications thereof.

II. BACKGROUND

Structural features of the DNA in complexes have been recently reviewed [19, 25, 26], but the quantitative link between structure and free energy of binding remains, for reasons cited above, largely uncharted at the molecular level. The energetic factors involved in protein–DNA complexes include van der Waals interactions, electrostatic interactions of functional moieties with each other and with mobile ions, and environmental effects. In macromolecular complexes, the van der Waals attractions arising from the contact surfaces of the constituents serve to quantify the contribution of shape complementarity to binding. The sugar–phosphate backbone of DNA, polyanionic at physiological pH, was initially not ascribed a role in specificity; but this view is currently under revision (see below). Potentially specific interactions were identified some years ago by Seeman *et al*. [27], who correlated the pattern of hydrogen bonding of donor and acceptor sites in the major and minor grooves of DNA with that of side chains of the amino acid residues on proteins. The thymine methyl group along with the H-bond pattern projected by the bases in the major groove enables a unique identification of each of the four base pairs (AT, TA, GC, and CG), while the minor groove offers only two distinguishable arrangements (AT/TA versus GC/CG). This aspect of recognition is local in nature; i.e., interactions involving each base pair in isolation dictate specificity. The direct effect of the neighboring base pairs as well as DNA conformation leads to a magnification or diminution of the local interactions, manifesting as a reduced or enhanced steric accessibility of the functional groups in the grooves or improved complementarity of electrostatic potentials.

The predictions of Seeman *et al*. were essentially confirmed in the first crystal structure of a protein–DNA complex, EcoRI endonuclease-DNA, in which some 12 hydrogen bonds formed between 6 amino acid residues and the six base pairs in the recognition site were implicated to contribute to recognition [20]. The "helix swap" experiments of Wharton and Ptashne [28, 29] and mutational studies in a similar vein provide further support of the role

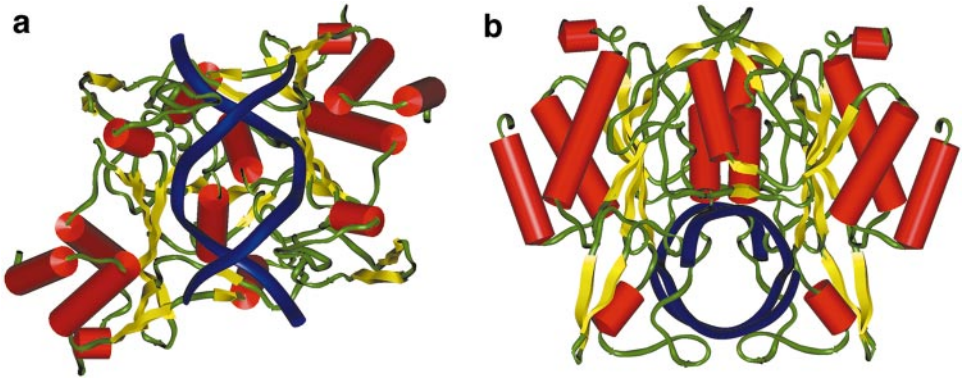

FIG. 1. Two cartoon views of the structure of EcoRI endonuclease–DNA specific complex [20]: (a) structure oriented looking into the active site with the DNA helical axis in the plane of the paper; (b) structure oriented with the DNA helical axis perpendicular to the plane of the paper.

of direct interactions between the protein and DNA as contributory to specific binding. On the other hand, the crystal structure of the trp repressor–operator [30] specific complex and experiments of Koudelka *et al*. [31] on the 434 repressor–operator system, among others, point out the inadequacies of a recognition model based solely on direct hydrogen bonding between base pairs and amino acid residues. Examining the crystal structures available at that time, Matthews [32] concluded that a simple code for recognition at the amino acid base pair level could not exist. Subsequent discussions in the literature on direct interactions include intermolecular complementarity ("snug fit") in addition to hydrogen bonding implicating (dynamical) structure as well as local interactions in specificity.

The role of DNA conformation in protein–DNA complexes was articulated particularly by Dickerson [33], who classified the information content in DNA as extrinsic and intrinsic, the former referring to hydrogen bond pattern presented by the base pairs in the grooves and the latter alluding to sequence-dependent conformational features of the DNA. Travers [19] categorized the recognition as (a) *digital* or *direct code*, i.e., direct hydrogen bonding between protein side chains and exposed edges of the base pairs mainly in the major and to a lesser extent in the minor groove, providing complementarity to correct sequence (van der Waals interactions between the protein and the DNA are included under direct code in subsequent discussions in the literature), and (b) *analog* or *indirect code*, i.e., structural deformation of the DNA to provide sequence selectivity by virtue of the ability of the nucleotide base sequence to assume a particular conformation (intrinsic or induced), required for binding to a protein at lower free energy cost than other sequences. The recent structures of IHF–DNA [34] and CAP–DNA [35, 36] complexes vividly illustrate the extent of deformation achieved by DNA to accomplish specific binding.

Ion and water release is a factor generally considered favorable to complex formation [37]. Binding constants measured as a function of ionic strength are frequently interpreted in theory set forth by Manning [38] and Record [39], from which the thermodynamic equivalents of ions released can be obtained empirically. The magnitude of this quantity is treated as a reflection of the strength of electrostatic effects, the number of phosphate contacts, and also the (favorable) entropic contribution of ion release to free energy of binding. Misra *et al*. [8] present a persuasive argument that the net electrostatics, when both enthalpic and entropic components are considered, disfavors complexation and that the unfavorable

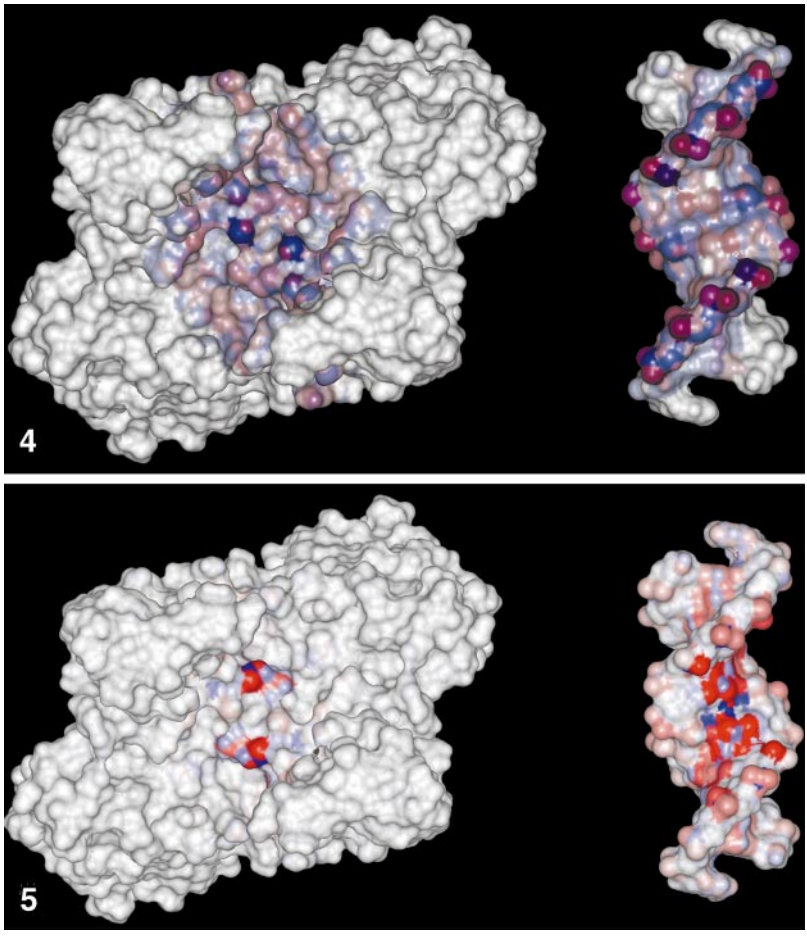

FIG. 4. An "open book" view of the EcoRI–DNA complex, with nonelectrostatic interaction components color coded. Colored regions are atomic van der Waals (pure blue) and hydrophobic (pure red) contributions favorable to binding. Intensity of colors is proportional to the size of the contribution, and blending of colors is proportional to relative contribution of van der Waals and hydrophobic interactions.

FIG. 5. An "open book" view of the EcoRI–DNA complex, with total electrostatic interaction (coulombic and solvation) components color coded. Colors represent favorable (negative) contribution (red) and unfavorable (positive) contributions (blue) to binding. The intensity of the colors is proportional to the size of the contribution to binding.

contribution from ion–molecule electrostatics rather than the entropy of ion reorganization (including but not limited to "release") dominates the salt-dependent solvation effects on complexation. The large, negative change in heat capacity observed on complex formation has been taken as evidence for a significant contribution of the hydrophobic effect to binding in protein–DNA complexes [40–42]. The role of the hydrophobic effect, with its origins in the entropy of water release from nonpolar surfaces, remains today a debated topic with respect to diverse biomolecular interaction processes [43].

Theoretical studies of protein DNA complexes based on molecular dynamics simulation are just beginning to appear [44–49], having been hindered by the enormous computational resources required. Brownian dynamics studies [50], quantifying the free energy of a nonspecific λ Cro repressor protein–DNA complex formation, have been reported recently.

Favorable electrostatic interactions were found to be partially offset by a loss of entropy, as the incoming protein dimer orientations become increasingly restrictive. MD simulations on the DNA EcoRI and DNA–EcoRV complexes [51] provided an estimate of the configurational and vibrational entropy change of DNA upon complexation in EcoRI binding, which was placed at +63.0 kcal/mol.

Free energy simulations specifically aimed at the ion release component [52, 53] suggested that this contribution in the λ system may not contribute favorably to binding process in which the two interacting species are brought from infinity to the final state of complexation. However, the contribution from counterion release to association of the DNA and protein from short range (7 Å) to its final state was favorable, but computed to be only ~ -1 kcal for the binding of the N-terminal fragment of λ repressor to its cognate operator site. Finite difference Poisson–Boltzmann (FDPB) calculations [8, 54] account for the thermodynamic equivalents of charge quantitatively in both λ and EcoRI systems. For the case of EcoRI, Misra *et al*. observed that salt effects destabilized the complex by almost +27 kcal at 0.1 M.

Footprinting titrations and quantitative binding assays have been a valuable source of experimental information on the issue of binding and specificity from an energetic viewpoint. The observed standard free energies of formation for specific protein–DNA complexes are typically in the range of −9 to −17 kcal/mol [23a]. For EcoRI DNA with the specific recognition site GAATTC, the free energy of binding was observed to be −15.2 kcal under the conditions of 0.18 M NaCl, pH 7.3, and temperature 22°C. For a nonspecific complex of the same enzyme, the free energy of formation decreased to −4.8 kcal. Some 12 thermodynamic equivalents of counterions were released upon formation of the specific complex. The EcoRI* activity studies strongly indicated that the observed decrease in binding constants upon mutation of the nucleotide base pairs in the recognition site could not be rationalized simply in terms of loss of hydrogen bonds [23b].

In an illuminating and comprehensive exposition of the thermodynamics of protein– DNA specificity, Jen-Jacobson [23] recently presented a critical analysis of the energetics of complexation of DNA sequences with EcoRI, based on results from the structural perturbation method [55]. The magnitudes for the different components proposed for this case (with a caveat about estimated uncertainties) were ~ -60 kcal for base–phosphate interactions with the enzyme including hydrogen bonding, van der Waals contacts, and ion release. Desolvation was inferred to contribute −60 kcal and another +105 kcal assigned to structural deformation and entropic loss. Binding studies on base analogs, with some assumptions about additivity and polyelectrolyte effects, led to the estimate of the base–phosphate interactions. The desolvation estimates were based on loss in nonpolar surface area calculations, which multiplied, by 25 cal/Å^2 [56] provided an estimate of the hydrophobic component. The value of structural deformation and entropic loss was deduced from observed binding constants and the preceding estimates of components. The nonspecific component of binding was proposed to involve weaker direct interactions, a smaller entropic loss, little desolvation gain, and no deformation loss. Jen-Jacobsons's treatment is the most extensive to date on the nature of specificity from an energetic perspective based on experimental binding data.

We present herein an alternative view of the energetic scenario of protein–DNA complexation, considering the complex formation in EcoRI endonuclease–DNA system as a case study. The approach as described is complementary to that of Jen-Jacobson, who deduced or inferred values for selected contributions from experimental data. We aim at providing

essentially theoretical estimates of all components which either are based on or can be validated by calibrations carried on small molecule prototype systems and do not rely on disposable parameters obtained to fit reported protein–DNA binding energies. Our objective is to formulate and characterize a detailed yet accessible, computationally tractable methodology with potential use as a bioinformatics tool in conjunction with the Nucleic Acids Data Base (NDB) [57] and other sources of structure for biomolecular complexes. A variation of this formalism is being used for the free energy analysis of an ensemble of oligonucleotide structures from various MD simulations to investigate the nature of conformational preferences of DNA oligonucleotides [58].

Previous studies in this vein in chemistry and biochemistry are cited in the review by Gilson *et al*. [10]. This project has specific biophysical precedents in the area of protein–protein binding energetics [59–62] but little has been reported to date on protein–DNA binding. We explore how the primary terms in our model can be compounded systematically to define a hierarchical reductionist scenario that responds more directly to questions typically of interest to the field. We note particularly the recent calculations of Honig and co-workers on the binding free energies of MHC Class I protein–peptide interactions using continuum electrostatics [63]. Their conclusion that the net electrostatics opposes formation and nonpolar interactions favor complexation foreshadows the results independently derived and presented here for the EcoRI–DNA complex, and for the consensus view of binding obtained from a similar study on a number of other protein–DNA complexes [64].

III. METHODOLOGY

Analysis

This study is carried out in the theoretical framework of "free energy component analysis," in which additivity is assumed [65] and the net free energy change is treated as a sum of selected individual contributions for which best estimates are obtained. The relationship of component analysis to a formal statistical mechanical treatment of binding affinities has been described by Gilson *et al*. [10]. We have chosen the individual contributions as a somewhat extensive list here, defined in a way that strategically isolates various contributions to the standard free energy of binding accessible to theoretical calculations via empirical energy functions and simplified models of solvation. Some of the terms in this model may in fact be decomposed further in subsequent studies (see below). With the assumption of additivity and an arbitrary, albeit rational, selection of terms, component analysis is *not* theoretically rigorous, and one can expect at best only a semi-quantitative account [60, 61]; expectations must be framed accordingly [62]. However, for complex processes such as protein–DNA binding, no viable alternative currently exists, and simple enumeration of the important terms, estimates of their relative magnitudes, and determination of whether they are favorable or unfavorable contributions to the free energy of complexation provides potentially useful new knowledge. The results can thus form a useful basis for a conceptual understanding and explanation, provided the capabilities and limitations intrinsic to this approach are clearly appreciated. The values assigned to components are subject to uncertainties, which in some cases are of a magnitude comparable to the net result. This problem has been encountered and discussed in several recent applications of this genre of calculations [60–63]. The level of confidence we place in the individual terms in our analysis is provided under Discussion.

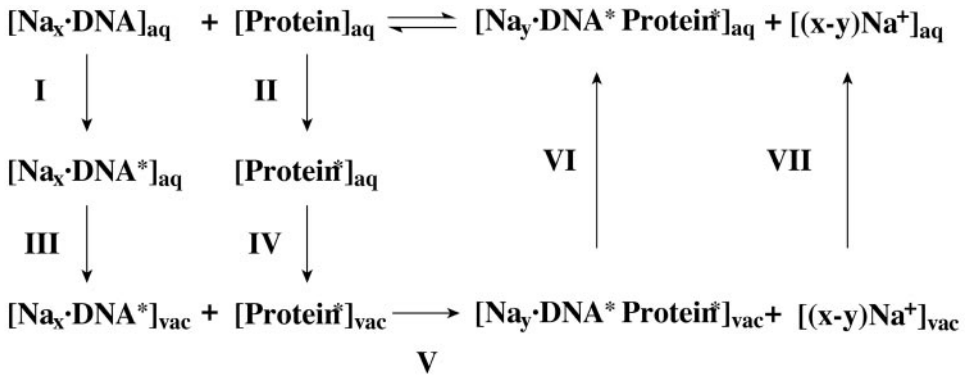

FIG. 2. The thermodynamic cycle used for the analysis of binding free energies of protein–DNA complexes.

The thermodynamic cycle for protein–DNA binding in solution used in this study is presented in Fig. 2. Here the net binding process is decomposed into seven steps. Step I describes the process of converting uncomplexed DNA, denoted "D," and counterions (some number of which are considered "condensed" on the DNA and treated explicitly) to the form D*, in which the DNA has adapted its structure to that of the bound form. Perturbations to both the structure of counterions and water as D is converted to D* are included in this step. The free energy of this step is

$$\Delta G_{\mathrm{I}}^{0} = \Delta G_{1}^{\mathrm{adpt.D}}. \tag{1}$$

Individual contributions to the free energy are numbered sequentially for later reference, and labeled with mnemonic superscripts. The labels are defined explicitly in Table 1, and details on their computation are given in the following section. Step II is the corresponding adaptation required of the protein, converting the uncomplexed form P to the complexed form P* in solution. The free energy is thus

$$\Delta G_{\mathrm{II}}^{0} = \Delta G_{2}^{\mathrm{adpt.P}}. \tag{2}$$

The next two steps (III and IV) involve desolvation of D* and P* from aqueous medium to vacuum. The free energy of each of these steps is written as a sum of five components,

$$\Delta G_{\mathrm{III}}^{0} = \Delta G_{3}^{\mathrm{el.D}} + \Delta G_{4}^{\mathrm{el.ci.D}} + \Delta G_{5}^{\mathrm{vdW.D}} + \Delta G_{6}^{\mathrm{cav.D}} + \Delta G_{7}^{\mathrm{DH.D}} \tag{3}$$

$$\Delta G_{\mathrm{IV}}^{0} = \Delta G_{8}^{\mathrm{el.P}} + \Delta G_{9}^{\mathrm{el.ci.P}} + \Delta G_{10}^{\mathrm{vdW.P}} + \Delta G_{11}^{\mathrm{cav.P}} + \Delta G_{12}^{\mathrm{DH.P}}, \tag{4}$$

with contributions from electrostatic effects of desolvating the macromolecule, electrostatic effects of desolvating the counterions, the van der Waals interactions with solvent, elimination of the solvent cavity in which the molecule is accommodated, and the change in the added salt effects. The transfer from aqueous medium to vacuum in steps III and IV involves the loss of favorable electrostatic and van der Waals interactions with solvent and a gain from the cavity term, the latter being, of course, the reverse of the free energy of cavity formation. The free energy of interaction with added salt (that over and above salt ions treated explicitly) is also lost on desolvation.

TABLE 1
Calculated Values of the Various Contributions to the Standard Free Energy of Binding for the DNA EcoRI Endonuclease Complex at 298 K

Step	Term	Component	Value (kcal/mol)	Method[a]
		Step I: Structural adaptation of DNA		
1	$\Delta G_1^{\text{adpt.D}}$	Free energy change for the process $D \rightarrow D^*$	+63.1	FF
		Step II: Structural adaptation of protein		
2	$\Delta G_2^{\text{adpt.P}}$	Free energy change for the process of $P \rightarrow P^*$	∼0.0	est
		Step III: Desolvation of DNA		
3	$\Delta G_3^{\text{el.D}}$	Electrostatic component of D^* desolvation	+6892.6	GB
4	$\Delta G_4^{\text{el.ci.D}}$	Counterion effect on D^* desolvation	−3575.5	GB
5	$\Delta G_5^{\text{vdW.D}}$	vdW component of D^* desolvation	+217.7	SA
6	$\Delta G_6^{\text{cav.D}}$	Cavity component of D^* desolvation	−257.1	SA
7	$\Delta G_7^{\text{DH.D}}$	Loss of added salt interactions with $Na \cdot D^*$	+17.4	DH
		Step IV: Desolvation of protein		
8	$\Delta G_8^{\text{el.P}}$	Electrostatic component of P^* desolvation	+5832.1	GB
9	$\Delta G_9^{\text{el.ci.P}}$	Counterion effect on P^* desolvation	∼0.0	GB
10	$\Delta G_{10}^{\text{vdW.P}}$	vdW component of P^* desolvation	+899.3	SA
11	$\Delta G_{11}^{\text{cav.P}}$	Cavity component of P^* desolvation	−1061.9	SA
12	$\Delta G_{12}^{\text{DH.P}}$	Loss of added salt interactions with P^*	+43.4	DH
		Step V: Complex formation *in vacuo*		
13	$\Delta H_{13}^{\text{el.C}}$	Electrostatic interactions of P^*D^*	−1538.0	FF
14	$\Delta H_{14}^{\text{ci.C}}$	Change in counterion interactions on P^*D^* binding	+1541.6	FF
15	$\Delta H_{15}^{\text{vdW.C}}$	vdW interactions of P^*D^*	−271.0	FF
16	$-T\Delta S_{16}^{\text{ci.C}}$	Entropy of (complex–DNA–protein) counterions	−22.2	
17	$-T\Delta S_{17}^{\text{tr\&rot}}$	Rotational and translational entropy change	+32.1	PF
18	$-T\Delta S_{18}^{\text{vib\&cnf}}$	Vibrational and configurational entropy change	+17.4	
		Step VI: Solvation of complex		
19	$\Delta G_{19}^{\text{el.C}}$	Electrostatic component of complex solvation	−11045.0	GB
20	$\Delta G_{20}^{\text{el.ci.C}}$	Counterion effect on complex solvation	+2664.6	GB
21	$\Delta G_{21}^{\text{vdW.C}}$	vdW component of complex solvation	−945.1	SA
22	$\Delta G_{22}^{\text{cav.C}}$	Cavity component of complex solvation	+1116.1	SA
23	$\Delta G_{23}^{\text{DH.C}}$	Added salt interactions with complex	−65.3	DH
		Step VII: Solvation of freed counterions		
24	$\Delta G_{24}^{\text{r.ci}}$	Solvation free energy of released counterions	−567.8	GB
	ΔG^0	Net binding free energy	−11.5	

[a] The theoretical method employed for the calculation of each term is indicated with the following abbreviations: FF, force field; GB, generalized Born; DH, Debye–Huckel; SA, surface area; est, estimated.

In step V, the structurally adapted D^* and P^* associate as a non-covalently bound complex. The thermodynamics of this step can be described as

$$\Delta G_{\text{V}}^0 = \Delta H_{13}^{\text{el.C}} + \Delta H_{14}^{\text{ci.C}} + \Delta H_{15}^{\text{vdW.C}} - T\Delta S_{16}^{\text{ci.C}} - T\Delta S_{17}^{\text{tr\&rot}} - T\Delta S_{18}^{\text{vib\&conf}}. \quad (5)$$

Complexation involves introducing the electrostatic and van der Waals interactions between the protein and the DNA *in vacuo*. A reorganization of the counterion atmosphere occurs, with some number of ions are "released" from associations on the contact surface. These are considered explicitly in our model. A change in external entropy due to a loss of translational

and rotational degrees of freedom enters this step, which always disfavors complexation. The lost external modes are converted into low-frequency internal vibrational and configurational degrees of freedom in the complex, and are reflected along with motional changes occurring as a consequence of burial of amino acid side chains on complexation in the corresponding change in vibrational and configurational entropy [10].

In Step VI, the complex is transferred from vacuum back to aqueous solution,

$$\Delta G^0_{\mathrm{VI}} = \Delta G^{\mathrm{el.C}}_{19} + \Delta G^{\mathrm{el.ci.C}}_{20} + \Delta G^{\mathrm{vdW.C}}_{21} + \Delta G^{\mathrm{cav.C}}_{22} + \Delta G^{\mathrm{DH.C}}_{23}, \tag{6}$$

and the free energy change is due to solvation of the complex and explicit counterions. Here again an electrostatic component, a van der Waals component, and a cavity formation term are involved, the first two being favorable to complexation in solution and the latter unfavorable. In Step VII, the counterions released on complex formation (r.ci) are transferred back into solvent, with a contribution to free energy favorable to complexation,

$$\Delta G^0_{\mathrm{VII}} = \Delta G^{\mathrm{r.ci}}_{24}. \tag{7}$$

In summary, the binding process in solution consists of seven well-defined thermodynamic steps, each of which can be decomposed into physically meaningful thermodynamic components. The total number of individual contributions to the free energy of binding in this model is 24. Following Holtzer [66] and Gilson *et al*. [10], no additional entropy of mixing terms are included explicitly; all momentum-based terms must cancel in forming a standard free energy change.

Theory

The theoretical estimates of values for the various contributions proceed as follows. We write the standard free energy of a given macromolecular structure (chemical potential) in solution, G^0, as

$$G^0 = G^0_{\mathrm{int}} + g^0_{\mathrm{solv}}, \tag{8}$$

where G^0_{int} is the free energy intrinsic to the molecule or complex and g^0_{solv} is the standard free energy of solvation; the upper and lower case notation for the intrinsic and solvation components, respectively, is introduced to clearly distinguish these terms. Values of G^0 will be used to obtain free energy differences ΔG^0 between initial and final thermodynamic states defined for complexation in Fig 2.

The underlying energetics intrinsic to macromolecules and complexes thereof is written in the conventional form of an empirical energy function,

$$E_{\mathrm{int}} = E_{\mathrm{bonds}} + E_{\mathrm{angles}} + E_{\mathrm{dihedrals}} + E_{\mathrm{nb}}, \tag{9}$$

where E_{bonds}, E_{angles}, and $E_{\mathrm{dihedrals}}$ describe bond stretching, angle bending, and dihedral displacements. The nonbonded interaction term, E_{nb}, is written as a sum of electrostatic (el) and van der Waals (vdW) terms,

$$E_{\mathrm{nb}} = E_{\mathrm{el}} + E_{\mathrm{vdW}}. \tag{10}$$

Each of these terms is pairwise additive over atoms explicitly considered in this model, viz.

$$E_{\mathrm{es}} = \sum_{i<j} \frac{q_i q_j}{r_{ij}}, \tag{11}$$

where q_i and q_j are any two net atomic charges separated by a distance r_{ij}, and

$$E_{\mathrm{vdW}} = \sum_{i<j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{12}$$

where σ and ε are the Lennard–Jones collision diameter and binding energy from dispersion forces, respectively, for each atom pair. Note that by expressing the energy of any molecular entity in the form of Eq. (10), a decomposition of the total energy into contributions from bonded and nonbonded interactions follows in a straightforward manner. The nonbonded interactions partition further into atomic pairwise additive energies. Various combinations of these primary terms provide well-defined contributions to the total energy in terms of functional groups, subunits such as nucleotide bases or amino acid residues, various elements of secondary structure (helices), or morphological features of a macromolecule structure such as grooves in DNA, crevices in proteins, or binding motifs such as helix–turn–helix or zinc fingers.

For a thermodynamic step with well-defined initial and final states, the intramolecular energy change is

$$\Delta E_{\mathrm{int}} = \Delta E_{\mathrm{bonds}} + \Delta E_{\mathrm{angles}} + \Delta E_{\mathrm{dihedrals}} + \Delta E_{\mathrm{el}} + \Delta E_{\mathrm{vdW}} \tag{13}$$

In the case where a single time-averaged crystal structure is used to represent a Boltzmann ensemble of states, we shall assume

$$\Delta H^0_{\mathrm{int}} \approx \Delta E_{\mathrm{int}}, \tag{14}$$

where ΔH^0 is the standard enthalpy change for the step. In a parallel study involving free energy analysis of A- and B-form DNA structures obtained from molecular dynamics we have calculated intramolecular enthalpies by ensemble averaging [58].

The contribution of explicit counterions to intramolecular enthalpies and entropies is a subject with a considerable history, and no small amount of controversy, but treating this phenomenon is essential to a comprehensive treatment of binding. Several protocols were explored for dealing with this aspect of the problem, and the approach we adopt is to start as simply as possible, and proceed to document the results from this and a series of successive improvements (see below). For the purposes of this study, we proceed according to the following rationale. From the results of Manning theory [38], we expect that condensed monovalent counterions per se neutralize only ~76% of the DNA charge, a result independently supported by recent large-scale MD simulations [67, 68]. Thus, a model with enough fully charged sodium counterions condensed on the DNA to provide local electroneutrality would be unrealistic. We assume a model of the ion atmosphere of the DNA in which counterions neutralize the Manning fraction of the DNA charge and represent this with discrete counterions with Na^+-sized solvatons, each bearing an effective charge of

$$q_{\mathrm{Na}} = 0.76 - (0.25/N_{\mathrm{bp}}) \tag{15}$$

and forming ion pairs with each of phosphates in the uncomplexed DNA. The first term is the Manning fraction, and the second is a correction for oligonucleotides of finite length

[39, 54, 69] with $N_{\rm bp}$ as the number of base pairs on DNA. The solvatons are placed on the OPO bisector of anionic phosphates, 7 Å from the phosphorous atom. The interactions of these fractionally charged particles with all the DNA atoms are computed explicitly, and each condensed ion is considered to have lost an entropic contribution to free energy of 2 kcal/mol relative to the bulk [70, 71]. This procedure is repeated in the complex in the presence of the protein. Ions clashing with atoms of the protein are repositioned within 3 Å if possible, using a short Monte Carlo process. Those ions which could not be accommodated due to clashes are treated "released" into free space in this step, with a concomitant gain in translational entropy (and change in vibrational and configurational entropy of the sodium ions plus DNA). The value for the corresponding solvation free energy is estimated from experimental data [71]; each unit of charge gains a solvation free energy of −98.3 kcal/mol upon transfer to bulk.

The external rotational and translational entropies for D*, P*, and complex are required for the free energy of complexation *in vacuo* (step V of Fig. 2). These quantities are calculated from ideal gas partition functions Q using classical statistical mechanics [72], viz.

$$S_0 = k \ln Q - (\partial \ln Q/\partial\beta)_{\rm v}. \tag{16}$$

The translational partition function is computed as

$$Q_{\rm trans} = V/h(\beta/2\pi m)^{3/2}, \tag{17}$$

where V is the volume, $\beta = (kT)^{-1}$ with k the Boltzmann constant and T the temperature, and m is the mass. For the rotational partition function,

$$Q_{\rm rot} = (\pi^{1/2}/\sigma)(1/hc\beta)^{3/2}(1/ABC)^{1/2}, \tag{18}$$

where the A, B, and C are rotational constants calculated from molecular geometry by standard methods; c is the velocity of light; and σ is the symmetry number. The translational and rotational entropies are introduced into the thermocycle at the step of complexation of the D* and P* in vacuum.

Vibrational and configurational contributions to entropy are indistinguishable in this problem and are considered together. Included in this contribution is the increase in vibrational/configurational entropy as a consequence of the new low-frequency motions that are interconverted from external degrees of freedom on complex formation, and the loss of conformational entropy when an amino acid side chain of the protein is restricted by contacts with DNA on complexation [73].

We write the solvation energy of a structure as

$$g^0_{\rm solv} = g^0_{\rm el} + g^0_{\rm nel}. \tag{19}$$

Here, the electrostatic contribution to the solvation energy, $g^0_{\rm GB}$, is estimated via the generalized Born (GB) model [13–17]. The defining equation of GB is

$$g^0_{\rm el} = -166\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_i q_j}{f_{\rm m2GB}}, \tag{20}$$

where $f_{\rm m2GB}$ is an effective atomic size/distance parameter derived from the Born radii α_i and pairwise distances r_{ij}. With suitable values for the α_i, the solvation energy of a given

molecule in a specified conformation can be computed. The GB solvation energy can be partitioned into contributions from polarization and solvent screening if necessary.

Added salt effects were incorporated into GB theory via Debye–Huckel theory, resulting in the expression

$$g_{\mathrm{el}}^{0} = -\frac{166}{\varepsilon}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_i q_j}{f_{\mathrm{m2GBDH}}}$$
$$f_{\mathrm{m2GBDH}} = (\kappa^{-1} + r_{ij})(f_{\mathrm{m2GB}}/r_{ij}) \qquad \text{for } i \neq j \tag{21}$$
$$f_{\mathrm{m2GBDH}} = (\kappa^{-1} + r_{\mathrm{vdW}})(\alpha_i/r_{ij}) \qquad \text{for } i = j,$$

where f_{m2GBDH} is the effective Born radius parameter, including the Debye–Huckel modification. With this addition, the solvation model becomes a combination of GB and Debye–Huckel theory.

The nonelectrostatic (nel) contributions to the standard free energy are due to van der Waals interactions between the solute and solvent and the work required to alter the cavitation in water in going from initial to final conditions. The total nonelectrostatic free energy is written as a linear function of the solvent accessible (SA) surface area

$$g_{\mathrm{nel}}^{0} = \gamma_{\mathrm{nel}}\Delta A, \tag{22}$$

with an empirical coefficient, γ_{nel}, defining the proportionality. Still and co-workers found that a value of $\gamma_{\mathrm{nel}} = 7.2$ cal/Å^2 gave reasonable results for a large number of cases [13]. The quantity γ_{nel} can be considered as the sum of van der Waals and cavitation terms,

$$\gamma_{\mathrm{nel}} = \gamma_{\mathrm{vdW}} + \gamma_{\mathrm{cav}}, \tag{23}$$

with value of 7.2 cal/Å^2 considered a resultant of $+47$ cal/Å^2 from the cavity term [74, 75] and -39.8 cal/Å^2 from van der Waals interactions of the solute with solvent. An independent check on this partitioning comes from noting that the van der Waals contribution is close to the value of 38.75 cal/mol/Å^2 derived from experimental enthalpies of vaporization of hydrocarbons [76]. The surface area A referred to in Eq. (23) is, however, that of all atoms. Thus the contribution to free energy from nonelectrostatic sources overall or the van der Waals and cavity terms individually can be further decomposed into contributions from charged, polar and nonpolar atoms or groups. The contribution to the cavity term from nonpolar groups is associated with the hydrophobic effect as conventionally defined, leading to the definition

$$g_{\mathrm{hy}\varphi}^{0} = \gamma_{\mathrm{nel}}A_{\mathrm{np}}. \tag{24}$$

Relating the contribution of the hydrophobic effect to nonpolar surface area A_{np} in this manner has been discussed recently as "hydrophobicity regained" [77]; note that our model allows for water release from polar or charged groups as separate contributions originating in the intrinsic size of each structural component. Although there is debate in the literature on the exact value to be employed for hydrophobic coefficient [43], the value of 7.2 cal/Å^2 weighting net nonelectrostatic contributions to solvation has been demonstrated to perform well on small molecules [13] and is the operational quantity in our calculation of standard free energy of binding.

The agreement between GBSA results and the experimental solvation energies for a wide range of molecules is well documented [13–17], and comparable to that obtained with both free energy simulations and finite difference Poisson–Boltzmann calculations while requiring much less computational effort. In a recent study, we have derived α_i parameters consistent with the AMBER parm94 force field [11] and experimental solvation energies of small molecules [17].

IV. CALCULATION

The atomic coordinates of the EcoRI endonuclease–DNA complex crystallography [20, 21] obtained from the NDB (code: PDE001) serve as a point of departure for this study. Our calculations are based on an all-atom model, which necessitated the addition of explicit hydrogen atoms to the crystal structure. The protonation state of ionizable groups was set at that corresponding to pH = 7 and assumed to be constant. Next, energy minimization of the protein–DNA complex was performed using the Sander module of the Amber 4.1 molecular modeling package [12] employing the most recent parameterization of the AMBER empirical energy function, the "parm94" force field [11]. In the energy minimization, we seek only to relieve any unfavorable clashes in the crystal structure and prepare the system for further study. Here 500 steps of minimization restraining heavy atoms (50 steps of steepest descent, SD, followed by 450 steps of conjugate gradient, CG), followed by a further 250 steps (50 SD+200 CG) of free minimization, are carried out to a tolerance of 0.5 kcal/mol Å. The structure obtained at this stage is still very close to the crystal structure (0.25 Å rms) and forms the basis for further analysis of the binding process.

The availability of the crystal structure of the uncomplexed form of the cognate DNA sequence [78], NDB Code (BDL002), enabled us to obtain an estimate of the deformation enthalpy of DNA. As an alternative strategy for cases where the uncomplexed DNA structure was not available, canonical B-DNA (B80) structure of the same sequence as the DNA in the complex was taken through a heating protocol identical to the preparatory steps of an MD simulation to bring the uncomplexed DNA and the complexed form of DNA to 298 K. To this energy difference we added the change in the GBSA free energy of solvation. Lacking corresponding information, we neglect the structural adaptation of the protein, although some rearrangement of the arms encircling the DNA on complex formation is likely, and the free energy change for ΔG_2 is taken to be zero.

The electrostatic contributions to solvation computed using the GB equation employed the effective radii parameters derived by Jayaram *et al.* [17] based on AMBER charges. This permits the calculation of both the intramolecular and solvation electrostatics based on a single set of charges, eliminating a possible inconsistency in the model. Note that usage of $\varepsilon = 1$ in the computation of direct electrostatic interactions between protein and DNA is consistent with the GB methodology for solvation. The molecular surface area calculations required for the nonelectrostatic contribution to the solvation energy were performed using the ACCESS program for solvent accessibility based on the algorithm of Lee and Richards [79] and AMBER parm94 vdW radii. The sequence d(TCGCGAATTCGCG) in the crystal structure of the EcoRI consists of 24 phosphates. Thus 24 explicit counterions of charge q_{net} (Eq. (15)) are provided and 9 of these were found to remain condensed upon complexation. The energies were computed from the Cornell *et al.* AMBER force field. The added salt concentration employed in the Debye–Huckel factor was 0.18 M.

V. RESULTS

A complete list of our calculated contributions to the standard free energy of binding for the d(CGCGAATTCGCG) EcoRI endonuclease complex is provided in Table 1. Our conventions are uniformly defined such that negative values are favorable and positive values are unfavorable to binding. Of the 24 components listed, 10 are found favorable to binding. As is typical of a component analysis based on terms representative of fundamental aspects of the structural chemistry of a process, the net free energy is seen to be a result of large competing terms. Especially notable in our results is the appearance of various compensation effects between (a) the internal and solvation electrostatics, (b) the loss in counterion–DNA interactions of the released ions and gain in solvation free energy of these ions as they enter the bulk, and (c) the direct van der Waals interactions between protein and DNA and loss in van der Waals interactions with solvent.

The results of Table 1 provide a relatively fine-grained view of the contributions from various chemical forces to complexation. Questions of interest to the field are more typically couched in terms of the contribution of electrostatics, shape complementarity, hydrophobic effects, structural adaptation, counterion release, etc., to the binding. The answers to these types of questions can be obtained from a combination of the values associated with the primary terms in Table 1. Specifically, the contribution of structural adaptation to free energy can be written as

$$\Delta G^{\mathrm{adpt}} = \Delta G_1^{\mathrm{adpt.D}} + \Delta G_2^{\mathrm{adpt.P}}. \tag{25}$$

The contribution of electrostatics (excluding the small ion effects) to the free energy result can be expressed as

$$\Delta G^{\mathrm{el}} = \Delta G_3^{\mathrm{el.D}} + \Delta G_8^{\mathrm{el.P}} + \Delta H_{13}^{\mathrm{el.C}} + \Delta G_{19}^{\mathrm{el.C}}. \tag{26}$$

The van der Waals interactions, effectively the net energetics of shape complementarity, are reflected in the sum

$$\Delta G^{\mathrm{vdW}} = \Delta G_5^{\mathrm{vdW.D}} + \Delta G_{10}^{\mathrm{vdW.P}} + \Delta H_{15}^{\mathrm{vdW.C}} + \Delta G_{21}^{\mathrm{vdW.C}}. \tag{27}$$

The total contribution of cavitation effects to the binding is

$$\Delta G^{\mathrm{cav}} = \Delta G_6^{\mathrm{cav.D}} + \Delta G_{11}^{\mathrm{cav.P}} + \Delta G_{22}^{\mathrm{cav.C}}. \tag{28}$$

The entropy change on complexation is described by the combination

$$\Delta G^{\mathrm{trvc}} = -T\Delta S_{17}^{\mathrm{tr\&rot}} - T\Delta S_{18}^{\mathrm{vib\&conf}}. \tag{29}$$

Small ion effects on free energy, due to both explicit ions and added salt in the model, can be summed as

$$\begin{aligned}\Delta G^{\mathrm{ions}} = {} & \Delta G_4^{\mathrm{el.ci.D}} + \Delta G_7^{\mathrm{DH.D}} + \Delta G_9^{\mathrm{el.ci.P}} + \Delta G_{12}^{\mathrm{DH.P}} + \Delta H_{14}^{\mathrm{ci.C}} - T\Delta S_{16}^{\mathrm{ci.C}} \\ & + \Delta G_{20}^{\mathrm{el.ci.C}} + \Delta G_{23}^{\mathrm{DH.C}} + \Delta G_{24}^{\mathrm{r.ci}}.\end{aligned} \tag{30}$$

The sum of all these terms equals the net standard free energy of binding, viz.

$$\Delta G^0 = \Delta G^{\mathrm{adpt}} + \Delta G^{\mathrm{el}} + \Delta G^{\mathrm{vdW}} + \Delta G^{\mathrm{nel}} + \Delta G^{\mathrm{ions}} + \Delta G^{\mathrm{trvc}}. \tag{31}$$

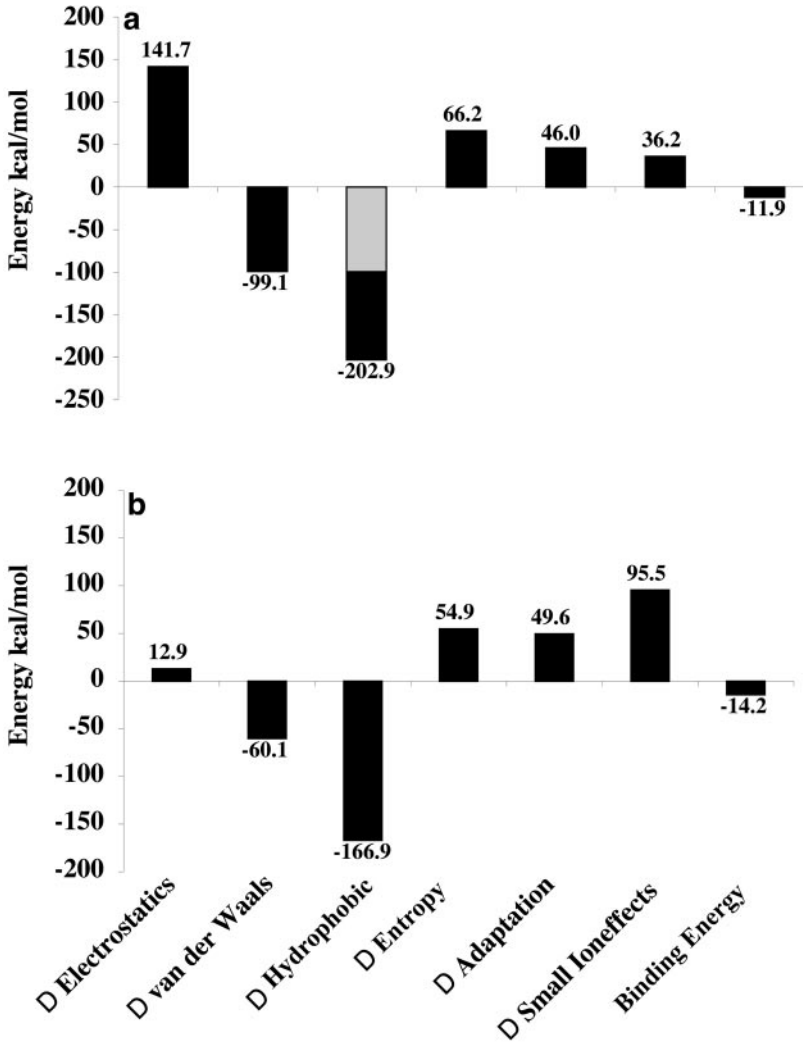

FIG. 3. A histogram view of the calculated contributions to the binding free energy of protein–DNA complexes. The reference state is separated protein and DNA, with negative values favorable and positive values unfavorable to binding: (a) results for the EcoRI complex; (b) results from 40 protein–DNA complexes.

An analysis of the results of Table 1 based on the contributions as defined in Eqs. (25)–(30) is presented schematically in Fig. 3a. Here the differential effects of direct van der Waals interactions between the protein and of cavity formation upon complexation are seen to be favorable to binding. All other terms, including electrostatics, turn out to be unfavorable to binding. In the protein–DNA complex literature, we note an emphasis on hydrophobic contacts, hydrogen bonds, phosphate contacts, and ion release; the role of van der Waals forces is somewhat underplayed except for references to "snug fit." Thus, our results introduce a potentially significant new perspective on the binding phenomena in this class of systems. The change in the size and shape of the solvent cavity on complexation gives rise to water reorganization, a component of which, originating from nonpolar sources, is the hydrophobic effect. In the EcoRI complex, our calculations predict the nonelectrostatic contribution to be 51% from nonpolar atoms and 49% from polar atoms.

This analysis demonstrates that some care is required in the phraseology adopted to describe a (compound) free energy component as "favorable" to complexation. Considering the results of Table 1, the free energy of water release on binding has an electrostatic part (+1679.7 kcal), a van der Waals part (+171.9 kcal), and a cavity part (−202.9 kcal). The sum of these three is positive, making water release by this definition unfavorable to binding; it is only the last term which is favorable, part of which is from nonpolar surface and identified with the hydrophobic effect. The situation with regard to polyelectrolyte effects is similar in nature. Ion release is favorable to binding based on component 24 in Table 1, but the sum effect of all the small ions including those of added salt (Fig. 3a) is unfavorable. The necessity of considering all significant enthalpic and entropic components for both initial and final states in the analysis of free energy results is thus underscored.

VI. DISCUSSION

Our analysis of the binding free energy for the EcoRI DNA complex indicates that the nonelectrostatic contributions, i.e., van der Waals interactions and differential cavitation effects, are favorable to complexation. Electrostatics, structural adaptation, and small ion effects are unfavorable. The interactions resulting from nonelectrostatic interactions are illustrated in Fig. 4, in which it should be noted that the DNA nestles within the arms of the protein in the complex, as may be seen in Fig. 1b. In Fig. 4, the atoms contributing to van der Waals and hydrophobic interactions are color coded, blue for van der Waals and red for hydrophobic, with an intensity proportional to their respective contributions to the net binding free energy. This illustrates the detail that can be applied to a binding analysis based on this model.

A structural view of the electrostatic surface complementarity in the EcoRI DNA complex is shown in Fig. 5. The result that the net electrostatics in this complex is unfavorable to complexation is consistent with the results of Fig. 5, which shows that the electrostatic complementarity of the components in the complex is not dramatic. The role of electrostatics, which includes here a combination of contributions originating in charged group interactions, hydrogen bonding and solvent effects, is consistent with that obtained in studies of drug–DNA complexes by Misra and Honig [8] using FDPB calculations. We hasten to emphasize that a result that electrostatics is net destabilizing to complexation does not imply that electrostatics is unimportant. The net free energy of binding is a fine balance of competing terms and would show a corresponding sensitivity to the magnitude of the electrostatic contribution even if it were destabilizing. Furthermore, in considering relative binding process of a series of molecular or macromolecular ligands, differential effects of electrostatics may still be critical in the result. The fact that the net electrostatics in this model contributes a destabilization to the free energy is nonetheless interesting; see also the study of Proloff *et al*. [63].

The standard free energy of binding can be partitioned into contributions identified with the various atoms of the protein and DNA, and summed into composite contributions from amino acid and nucleotides. The binding free energy partitioned into contributions from amino acids is shown for one monomer of the Eco RI dimer in Fig. 6, with the elements of secondary structure, degree of solvent accessibility, and protein–DNA contacts indicated along the abscissa. Again, a component analysis at this level of resolution shows a large number of positive and negative competing terms, but some trends are evident. The contributions most favorable to complexation are clearly associated with protein–DNA

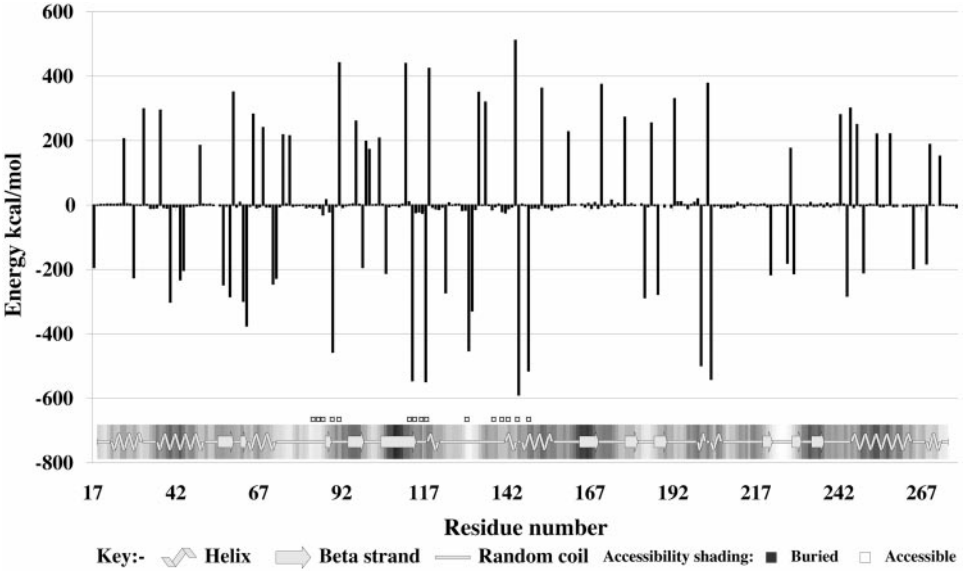

FIG. 6. Histogram of interaction energies of P* with D* partitioned with respect to amino acids. The presentation is for one monomer of the EcoRI protein dimer. The elements of secondary structure in the protein sequence are indicated along the abscissa, along with a shading, which is proportional to the solvent accessibility of each residue as provided by the program PROCHEK [83]. The data points for amino acids making P*D* contacts in the complex as identified using NUCPLOT [82] are denoted just above the secondary structure diagram with "□."

contacts, as identified in the crystal structure of the complex. However, the results of Fig. 6 show that a number of amino acid residues not indicated to be direct contacts also contribute significantly in the decomposition. In addition, inspection of Fig. 6 shows that in a remarkable number of cases strongly unfavorable interactions are immediately juxtaposed with favorable interactions, as if the protein structure evolved so as to assure binding is strong in specific associations but not too strong to allow for the requisite dissociation.

A corresponding analysis referenced to the nucleotides of the DNA is shown in Fig. 7. Here, as expected, strongly favorable contributions to complexation are associated with the nucleotides in recognition site GAATTC. However, here as well the contribution to binding from residues surrounding the recognition site is clearly not negligible, and supports a role for "context effects" as well as contacts in binding, supporting the essence of arguments made for some time now by Jen-Jacobson [55]. The C4 position, which produces an extraordinarily favorable contribution to complexation, is the "clamp" position referred to in the original crystal structure [20].

The level of confidence we place on each of the calculated values presented in Table 1, and by inference in Fig. 3, is as follows. The final estimate of the net free energy, found to be within 5 kcal of experiment, is necessary but not sufficient to provide confidence in the method. The starting point of our study is the crystal structure of the protein–DNA. Duan *et al*. [51] have shown that the rms deviations from the starting X-ray coordinates in the MD simulations were reported to be under 2 Å for the specific complexes, giving hope that single point (in the configuration space of the complex) energy calculations on the protein–DNA complexes based on crystal structures may be a good approximation. Any variations in the theoretical protocols involving addition of hydrogen atoms and subsequent minimization may alter the exact magnitudes somewhat, but in our experience the effects

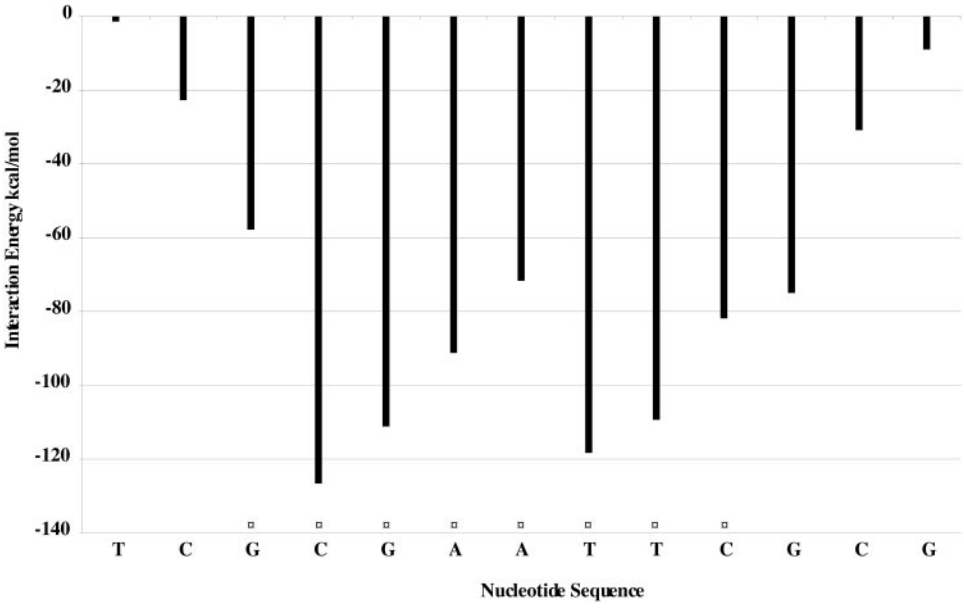

FIG. 7. Histogram of interaction energies of D^* with P^* partitioned with respect to nucleotides of the DNA. The presentation is for one strand of the (palindromic) DNA duplex. The recognition site is the central six nucleotides of the sequence, GAATTC. Nucleotides making P^*D^* contacts in the complex as identified using NUCPLOT [82] are denoted just above the sequence specification as "□."

tend to be compensatory in nature and are unlikely to affect the conclusions. Our assumption regarding the ionization state of amino acid residues is oversimplified [80].

The free energy change for the adaptation of the DNA structure, computed relative to canonical B-DNA at 300 K, results in +63.1 kcal including solvation. Most of this effect arises due to the torsional and van der Waals terms; the difference in GBSA solvation free energy is only 0.8 kcal/mol. An improved estimate of these quantities (components 1 and 2 of Table 1) could be obtained from full scale MD on free protein, free DNA followed by an analysis of the intramolecular and solvation energetics. However, based on a similar study carried out on the λ repressor–operator complex, we would not expect this result to differ too much if MD were applied. An estimate of the DNA structural adaptation energy relative to Drew–Dickerson crystal structure at 298 K after adding hydrogen atoms, without any minimization, led to a similar value.

The direct van der Waals and electrostatic interactions (components 13 and 15) are obtained from the latest version of the AMBER force field, parm94 [11], which is calibrated to a well-defined level of accuracy with respect to experimental data and quantum mechanical calculations on small molecule prototypes. We assume these parameters are transferable, as do macromolecular MD calculations. The solvation/desolvation is based on a force field compatible parameterization of GB theory, the level of accuracy of which is documented against the solvation free energies of small molecules [17] and is generally found to be within 5%. Component 24 is based on experimental solvation free energy of sodium ions. Its validity depends on whether counterions, considered explicitly, are released at all. In a molecular view of the process, this appears justified. The entropy terms (components 17 and 18) for the structures in the gas phase are based on ideal gas statistical mechanics and other considerations [73], and have a well-defined theoretical basis [10]. The calculation

of vibrational and configurational entropy change on complexation follows Janin's procedure [60, 61], which includes contributions due to the motional restriction of amino acid side chains at the contact surface and from the additional low-frequency vibrational modes which arise on complex formation. This latter quantity, amounting to ~50% that of $\Delta S^{\text{tr.rot}}$ (with opposite sign), assumes that the shape of the potential energy hypersurface of the complex is similar to that of the uncomplexed forms. MD results indicate that the surface is such that motions in the complex are of lesser amplitude [44, 51]; this point will likely require further detailed study and refinement.

Added salt contributions (components 7, 12, and 23) are estimated using Debye–Huckel free energy expressions, implemented in a form consistent with the GB solvent model. An increase in added salt in this approach has an unfavorable effect on binding, in agreement with the FDPB trends [8]. The explicit counterion effects (components 4, 9, 14, 16, and 24) are phenomenological in nature and by necessity approximate in this treatment of the problem.

The counterion effects are based on a model of fractional charge for ions condensed on DNA, and upon complexation are either redistributed near their original location or lost to bulk due to clashes. We tried out various protocols, including several variations on the counterion Monte Carlo method [53] around the DNA, the protein, and the complex. We did not find an exact balance between the explicit ion effects and solvation. Because the GB model when applied to MD trajectories with explicit ions and waters shows a fine balance in the energetics of ion solvation versus ion–DNA interactions, some calibration may be necessary for an improved treatment of explicit ion effects in protein–DNA when used together with the GB model for solvation. The estimated net counterion and added salt effect on binding free energy is +36.2 kcal (0.18 M), which on the scale of numbers involved is close to the finite difference Poisson–Boltzmann estimate of +27 kcal (0.10 M). The amount of released charge on complexation with this protocol is 11.1, close to the experimental value of 12.0 [23].

In this study, the objective was to carry an analysis of protein–DNA complexation as far as possible based on crystal structure data. This necessitates an ad hoc model for counterion release that, no matter how plausible, remains a simplified assumption. The development of dynamical models for counterion behavior around DNA from MD simulations including explicit consideration of all solvent has been reported recently [67, 68]. Scaled up to protein DNA complexes, MD holds the promise of providing an *ab initio* model for counterion release. This refined model could subsequently be subjected to free energy analysis in the form proposed here. MD modeling [58] can also contribute vibrational entropies via the quasi-harmonic approximation [81], along with improved estimates of the energy and entropy of structural adaptation. Further explorations of these issues as they relate to theoretical studies of protein–DNA complexes are in progress.

Finally, the results presented so far are only for one system, and raise the questions of whether an analysis of this type can be successfully extended to other systems and whether the results presented for the EcoRI complex are indicative of a general trend. We have obtained, with no essential changes in our methodology, preliminary results on 40 protein–DNA complexes with diverse binding motifs [64]. The results are summarized schematically in Fig. 3b, to be compared with the corresponding results of EcoRI complex in Fig. 3a. The consensus view from the 40 systems is that van der Waals and cavity terms favor complexation. There are, of course, fluctuations at the individual level and cases for which electrostatics is favorable. The trends in the consensus view are nevertheless similar to those calculated for EcoRI.

In concluding this section, we return to the issue of errors and uncertainties in the process of free energy component analysis applied to a complex system. The eclectic choice of theoretical methods used to estimate the various terms comes at the price of possible incommensurabilities among the various terms. The quantifiable errors in certain estimates plus the qualitative approximations in others lead to a net result that has an uncertainty of the order of the calculated net binding energy, which indicates that neither the magnitude- nor the sign of this quantity is secure. One can argue that this is a fatal flaw in free energy component analysis, applied not only to this case but generally to studies of binding in complex systems. The consequences of this are significant: hopes for a reductionist approach to understanding biological processes in terms of chemical forces may founder on a practical limitation, not a theoretical one! On the other hand, the issue of the uncertainty in the net result merely underscores what we concede at the outset, that agreement with experiment does not unequivocally prove the analysis is correct, a well-known limit in theoretical modeling. Thus one must view the results described in this article in the context of the expected uncertainties, but in addition consider what can be learned despite this problem, such as ideas about relative magnitudes of various contributions, and considerations of both initial and final states in estimating thermodynamic components. Nonuniqueness of the partitioning into components is mitigated partially but not fully by the plausibility of the analysis. These results may suggest that additional experiments and theoretical studies should be carried out that ultimately improve and enhance a scientific understanding of the problem. The net benefit of free energy component anaylsis applied to such complex binding problems as we consider here is to be judged in qualitative, not quantitative, terms, and we conclude with the caveat that while the results of this study are presented in quantitative form, the take-home lesson is qualitative in nature.

VII. SUMMARY AND CONCLUSIONS

A detailed theoretical analysis of the thermodynamics and functional energetics of protein–DNA binding in the EcoRI endonuclease–DNA complex has been described. The standard free energy of complexation is considered in terms of a thermodynamic cycle of seven distinct steps decomposed into a total of 24 well-defined components. The model we employ involves explicit all-atom accounts of the energetics of structural adaptation of the protein and DNA on complex formation; the van der Waals and electrostatic interactions between the protein and the DNA; and the electrostatic polarization and screening effects, van der Waals components, and cavitation effects of solvation. The ion atmosphere of the DNA is described in terms of a counterion condensation model, which permits estimates of the ion release upon complexation and a Debye–Huckel treatment of added salt effects. Estimates of entropy loss due to decreased translational and rotational degrees of freedom in the complex relative to the unbound species based on classical statistical mechanics are included, as well as corresponding changes in the vibrational and configurational entropy. The magnitudes and signs of the various components are estimated from the AMBER parm94 force field, generalized Born theory, solvent accessibility measures, and empirical estimates of quantities related to ion release. The calculated standard free energy of formation, -11.5 kcal/mol, agrees with experiment to within 5 kcal/mol. Analysis of the results shows that the calculated binding free energy of the EcoRI endonuclease-DNA complex is the resultant of a balance of competing contributions associated with chemical forces as conventionally defined, with 10 of 24 terms favoring complexation. Contributions to binding

compounded from subsets of the 24 terms provide a basis for analysis of contributions due to structural adaptation, electrostatics, van der Waals interactions, hydrophobic effects, and small ion reorganization and release on complexation. The van der Waals interactions and water release favor complexation, while electrostatic interactions, considering both intramolecular and solvation effects, prove unfavorable. Analysis of individual contributions to the standard free energy of complexation from nucleotides of the DNA and amino acid residues of the protein shows that some contact interactions disfavor complexation and that context, as well as contact interactions, is important.

ACKNOWLEDGMENTS

Funding for this project was provided by the National Institutes of Health under Grant GM-37909. B.J. acknowledges the financial support and encouragement received from the Indian Department of Science & Technology, the Council of Scientific & Industrial Research, and the Indo–French Centre for the Promotion of Advanced Research. K.J.M. was supported by a traineeship in molecular biophysics from the NIH Training Grant GM-08271. Discussions with Dr. M. A. Young, Dr. D. Sprous, and Professors B. Honig and I. Russu are gratefully acknowledged. The facilities of the Pittsburgh Supercomputer Center and the National Center for Supercomputer Applications at University of Illinois were utilized in this study.

Note added in proof. We acknowledge additional helpful comments and criticisms of this work by Professors L. Jen-Jacobson and J. Rosenberg. Professor Rosenberg points out that one of the assumptions in our model is the neglect of the effect of explicit waters residing at the protein–DNA interface. We concur that this issue needs further consideration in any refinement of this study.

REFERENCES

1. W. J. Hehre, L. Radom, P. V. R. Schleyer, *et al.*, *Ab Initio Molecular Orbital Theory* (Wiley, New York, 1986).
2. D. L. Beveridge and F. M. DiCapua, Free energy via molecular simulation applications to chemical and biomolecular systems, *Ann. Rev. Biophys. Biophys. Chem.* **18**, 431 (1989).
3. W. L. Jorgensen, Free energy calculations: A breakthrough for modeling organic chemistry in solution, *Acc. Chem. Res.* **22**, 184 (1989).
4. J. A. McCammon, Free energy from simulations, *Curr. Opinion Struct. Biol.* **1**, 196 (1991).
5. P. A. Kollman, Free energy calculations: Applications to chemical and biochemical phenomena, *Chem. Rev.* **93**, 2395 (1993).
6. W. F. van Gunsteren and H. J. C. Berendsen, Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry, *Angew. Chem. Int. Ed. Engl.* **29**, 992 (1990).
7. E. J. Simon, Dissertation, Harvard University, 1996.
8. V. K. Misra, J. L. Hecht, K. A. Sharp, *et al.*, Salt effects on protein–DNA interactions. The lambda cI repressor and EcoRI endonuclease, *J. Mol. Biol.* **238**, 264 (1994).
9. V. K. Misra, K. A. Sharp, R. A. Friedman, *et al.*, Salt effects on ligand–DNA binding: Minor groove binding antibiotics, *J. Mol. Biol.* **238**, 245 (1994).
10. M. K. Gilson, J. A. Given, B. L. Bush, *et al.*, The statistical thermodynamic basis for computation of binding affinities: A critical review, *Biophys. J.* **72**, 1047 (1997).
11. W. D. Cornell, P. Cieplak, C. I. Bayly, *et al.*, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc.* **117**, 5179 (1995).
12. D. A. Pearlman, D. A. Case, J. W. Caldwell, *et al.*, AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, *Comput. Phys. Commun.* **91**, 1 (1995).
13. W. C. Still, A. Tempczyk, R. C. Hawley, *et al.*, Semianalytical treatment of solvation for molecular mechanics and dynamics, *J. Am. Chem. Soc.* **112**, 6127 (1990).

14. G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium, *J. Phys. Chem.* **100**, 19824 (1996).
15. G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, Pairwise solute descreening of solute charges from a dielectric medium, *Chem. Phys. Lett.* **246**, 122 (1995).
16. B. Jayaram, Y. Liu, and D. L. Beveridge, A modification of the generalized Born theory for improved estimates of solvation energies and pKa shifts, *J. Phys. Chem.* **109**, 1465 (1998).
17. B. Jayaram, D. Sprous, and D. L. Beveridge, Solvation free energy of biomacromolecules: Parameters for a modified generalized Born model consistent with the AMBER force field, *J. Phys. Chem. B* **102**, 9571 (1998).
18. M. R. Reddy, M. D. Erion, A. Agarwal, *et al.*, Solvation free energies calculated using the GB/SA model, *J. Comput. Chem.* **19**, 769 (1998).
19. A. Travers, *DNA–Protein Interactions* (Chapman and Hall, London, 1993).
20. J. A. McClarin, C. A. Frederick, B.-C. Wang, *et al.*, Structure of the DNA–Eco RI endonuclease recognition complex at 3 Å resolution, *Science* **234**, 1526 (1986).
21. Y. Kim, J. C. Grable, R. Love, *et al.*, Refinement of the Eco RI endonuclease crystal structure: A revised protein chain tracing, *Science* **249**, 1307 (1990).
22. H. R. Drew, R. M. Wing, T. Takano, *et al.*, Structure of a B DNA dodecamer. I. Conformation and dynamics, *Proc. Nat./Acad. Sci. USA* **78**, 2179 (1981).
23a. L. Jen-Jacobson, Protein–DNA recognition complexes: Conservation of structure and binding energy in the transition state, *Biopolymers* **44**, 153 (1997).
23b. D. R. Lesser, M. R. Kurpiewski, and L. Jen-Jacobson, The energetic basis of specificity in the EcoRI endonuclease-DNA interaction, *Science* **250**, 776 (1990).
24. D. L. Beveridge, M. A. Young, and D. Sprous, Modeling of DNA via molecular dynamics simulation: Structure, bending, and conformational transitions, in *Molecular Modeling of Nucleic Acids*, edited by N. B. Leontis and J. Santa Lucia (American Chemical Society, Washington, DC, 1998), Vol. 1, p. 260.
25. R. E. Dickerson, in *Oxford Handbook of Nucleic Acid Structure*, edited by S. Neidle (Oxford Univ. Press, Oxford UK, 1998).
26. T. A. Steitz, Structural studies of protein–Nucleic acid interactions: The sources of sequence-specific binding, *Quart. Rev. Biophys.* **23**, 205 (1990).
27. N. C. Seeman, J. M. Rosenberg, and A. Rich, Sequence-specific recognition of double helical nucleic acids by proteins, *Proc. Natl. Acad. Sci. USA* **73**, 804 (1976).
28. R. P. Wharton, E. L. Brown, and M. Ptashne, Substituting an alpha-helix switches the sequence-specific DNA interactions of a repressor, *Cell* **38**, 361 (1984).
29. R. P. Wharton and M. Ptashne, Changing the binding specificity of a repressor by redesigning an alpha-helix, *Nature* **316**, 601 (1985).
30. Z. Otwinowski, R. W. Schevitz, R. G. Zhang, *et al.*, Crystal structure of trp repressor/operator complex at atomic resolution, *Nature* **335**, 321 (1988).
31. G. B. Koudelka, S. C. Harrison, and M. Ptashne, Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and Cro, *Nature* **326**, 886 (1987).
32. B. W. Matthews, Protein–DNA interaction, No code for recognition [news], *Nature* **335**, 294 (1988).
33. R. E. Dickerson, The DNA helix and how it is read, *Sci. Am.* **249**, 94 (1983).
34. P. A. Rice, S. Yang, K. Mizuuchi, *et al.*, Crystal structure of an IHF–DNA complex: A protein-induced DNA U-turn, *Cell* **87**, 1295 (1996).
35. G. Parkinson, C. Wilson, A. Gunasekera, *et al.*, Structure of the CAP–DNA complex at 2.5 angstroms resolution: A complete picture of the protein–DNA interface, *J. Mol. Biol.* **260**, 395 (1996).
36. S. C. Schultz, G. C. Shields, and T. A. Steitz, Crystal structure of a CAP–DNA complex: The DNA is bent by 90 degrees, *Science* **253**, 1001 (1991).
37. M. T. Record, Jr., C. F. Anderson, P. Mills, *et al.*, Ions as regulators of protein–nucleic acid interactions in vitro and in vivo, *Adv. Biophys.* **20**, 109 (1985).

38. G. S. Manning, The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides, *Quart. Rev. Biophys.* **11**, 179 (1978).
39. M. T. Record, Jr., C. F. Anderson, and T. M. Lohman, Thermodynamics analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: The role of ion association and release, screening, and ion effects on water activity, *Quart. Rev. Biophys.* **11**, 103 (1978).
40. R. S. Spolar and M. T. Record, Jr., Coupling of local folding to site-specific binding of proteins to DNA [see comments], *Science* **263**, 777 (1994).
41. J. W. Schwabe, The role of water in protein–DNA interactions, *Curr. Opinion Struct. Biol.* **7**, 126 (1997).
42. T. Lundback, H. Hansson, S. Knapp, *et al.*, Thermodynamic characterization of non-sequence-specific DNA-binding by the Sso7d protein from Sulfolobus solfataricus, *J. Mol. Biol.* **276**, 775 (1998).
43. H. S. Chan and K. A. Dill, Solvation: How to obtain microscopic energies from partitioning and solvation experiments, *Annu. Rev. Biophys. Biomol. Struct.* **26**, 425 (1997).
44. D. L. Beveridge, K. J. McConnell, R. Nirmala, *et al.*, Molecular dynamics simulations of DNA and protein–DNA complexes including solvent: Recent progress, *ACS Symp. Seri.* **568**, 381 (1994).
45. M. Eriksson, T. Haerd, and L. Nilsson, Molecular dynamics simulation of a DNA binding protein free and in complex with DNA, *NATO ASI Ser. C* **426**, 441 (1994).
46. L. F. Harris, M. R. Sullivan, P. D. Popken-Harris, *et al.*, Molecular dynamics simulations in solvent of the glucocorticoid receptor protein in complex with a glucocorticoid response element DNA sequence, *J. Biomol. Struct. Dyn.* **12**, 249 (1994).
47. L. F. Harris, M. R. Sullivan, and P. D. Popken-Harris, Molecular dynamics simulation in solvent of the estrogen receptor protein DNA binding domain in complex with a non-consensus estrogen response element DNA sequence, *J. Biomol. Struct. Dyn.* **15**, 407 (1997).
48. M. G. Cooney, K. Miaskiewicz, M. Shibata, *et al.*, in *Biomacromol.: 3-D Appl., Hanford Symp. Health Environ., 34th*, edited by R. L. Ornstein (Battelle Press, Columbus, OH, 1997), p. 145.
49. G. Roxstroem, I. Velazquez, M. Paulino, *et al.*, Molecular dynamics simulation of a Zif268–DNA complex in water: Spatial patterns and fluctuations sensed from a nanosecond trajectory, *J. Phys. Chem. B* **102**, 1828 (1998).
50. K. A. Thomasson, I. V. Ouporov, T. Baumgartner, *et al.*, Free energy of nonspecific binding of Cro repressor protein to DNA, *J. Phys. Chem. B* **101**, 9127 (1998).
51. Y. Duan, P. Wilkosz, and J. M. Rosenberg, Dynamic contributions to the DNA binding entropy of the EcoRI and EcoRV restriction endonucleases, *J. Mol. Biol.* **264**, 546 (1996).
52. B. Jayaram, F. M. DiCapua, and D. L. Beveridge, A theoretical study of polyelecrolyte effects in protein–DNA interactions: Monte Carlo free energy simulations on the ion atmosphere contribution to the thermodynamics of lambda repressor–operator complex formation, *J. Am. Chem. Soc.* **113**, 5211 (1991).
53. B. Jayaram, S. Swaminathan, and D. L. Beveridge, Monte-Carlo simulation studies of the structure of the counteranion atmosphere of B-DNA: Variations on the primitive dielectric model, *Macromolecules* **23**, 3156 (1990).
54. M. Zacharias, B. A. Luty, M. E. Davis, *et al.*, Poisson–Boltzmann analysis of the lambda repressor–operator interaction, *Biophys. J.* **63**, 1280 (1992).
55. L. Jen-Jacobson, Structural-perturbation approaches to thermodynamics of site–specific protein–DNA interactions, *Methods Enzymol.* **259**, 305 (1995).
56. C. Chothia, Structural invariants in protein folding, *Nature* **254**, 304 (1975).
57. H. M. Berman, W. K. Olson, D. L. Beveridge, *et al.*, The nucleic acid database: A comprehensive relational database of three-dimensional structures of nucleic acids, *Biophys. J.* **63**, 751 (1992).
58. B. Jayaram, D. Sprous, M. A. Young, *et al.*, Free energy analysis of the conformational preferences of A and B forms of DNA in solution, *J. Am. Chem. Soc.* **120**, 10629 (1998).
59. T. Lazaridis, G. Archontis, and M. Karplus, Enthalpic contribution to protein stability: Insights from atom-based calculations and statistical mechanics, *Adv. Protein Chem.* **47**, 231 (1995).
60. J. Janin, Elusive affinities, *Proteins* **21**, 30 (1995).
61. J. Janin, Protein–protein recognition, *Prog. Biophys. Mol. Biol.* **64**, 145 (1995).

62. J. Novotny, R. E. Bruccoleri, M. Davis, *et al.*, Empirical free energy calculations: A blind test and further improvements to the method, *J. Mol. Biol.* **268**, 401 (1997).
63. N. Proloff, A. Windemuth, and B. Honig, On the calculation of binding free energies using continuum methods: Application to MHC class I protein–Peptide interactions, *Protein Sci.* **6**, 1293 (1997).
64. B. Jayaram, K. McConnell, S. B. Dixit, *et al.*, Analysis of the binding free energy of protein–DNA complexes based on 40 X-ray crystal structures, submitted for publication.
65. K. A. Dill, Additivity principles in biochemistry, *J. Biol. Chem.* **272**, 701 (1997).
66. A. Holtzer, The "Cratic Correction" and related fallacies, *Biopolymers* **35**, 595 (1994).
67. M. A. Young, B. Jayaram, and D. L. Beveridge, Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: Fractional occupancy of electronegative pockets, *J. Am. Chem. Soc.* **119**, 59 (1997).
68. M. A. Young, G. Ravishanker, and D. L. Beveridge, A 5-nanosecond molecular dynamics trajectory for B-DNA: Analysis of structure, motions and solvation, *Biophys. J.* **73**, 2313 (1997).
69. M. C. Olmsted, C. F. Anderson, and M. T. Record, Jr., Importance of oligoelectrolyte end effects for the thermodynamics of conformational transitions of nucleic acid oligomers: A grand canonical Monte Carlo analysis, *Biopolymers* **31**, 1593 (1991).
70. G. A. Krestov, *Thermodynamics of Solvation* (Ellis–Horwood, New York, 1991).
71. H. L. Friedman and C. V. Krishnan, in *Water: A Comprehensive Treatise*, edited by F. Franks (Plenum, New York, 1973), Vol. 3.
72. P. W. Atkins, *Physical Chemistry* (Feeman, New York, 1990).
73. A. V. Finkelstein and J. Janin, The price of lost freedom: Entropy of biomolecular complex formation, *Protein Eng.* **3**, 1 (1989).
74. K. Sharp, R. Fine, and B. Honig, Computer simulations of the diffusion of a substrate to an active site of an enzyme, *Science* **236**, 1460 (1987).
75. K. A. Sharp, A. Nicholls, R. F. Fine, *et al.*, Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects, *Science* **252**, 106 (1991).
76. H. Ohtaki and N. Fukushima, A structural study of saturated aqueous solutions of some alkali halides by X-ray diffraction, *J. Sol. Chem.* **21**, 23 (1992).
77. P. A. Karplus, Hydrophobicity regained, *Protein Sci.* **6**, 1302 (1997).
78. H. R. Drew, S. Samson, and R. E. Dickerson, Structure of a B-DNA dodecamer at 16 K, *Proc. Natl. Acad. Sci. USA* **79**, 4040 (1982).
79. B. K. Lee and F. M. Richards, The interpretation of protein structures: Estimation of static accessibility, *J. Mol. Biol.* **55**, 379 (1971).
80. W. R. Forsyth, M. K. Gilson, J. Antosiewicz, *et al.*, Theoretical and experimental analysis of ionization equilibria in ovomucoid third domain, *Biochemistry* **37**, 8643 (1998).
81. M. Karplus and J. N. Kushick, Method for estimating the configurational entropy of macromolecules, *Macromolecules* **14**, 325 (1981).
82. N. M. Luscombe, R. A. Laskowski, and J.M. Thornton, NUCPLOT: A program to generate schematic diagrams of protein–nucleic acid interactions, *Nucleic Acids Res.* **25**, 4940 (1997).
83. R. A. Laskowski, M. W. MacArthur, D. S. Moss, *et al.*, PROCHECK: A program to check the stereochemical quality of protein structures., *J. Appl. Crystallogr.* **26**, 283 (1993).