

Energetics of Sequence-Specific Protein–DNA Association: Computational Analysis of Integrase Tn916 Binding to Its Target DNA[†]

Alemayehu A. Gorfe and Ilian Jelesarov*

Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

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

ABSTRACT: The N-terminal domain of the bacterial integrase Tn916 specifically recognizes the 11 bp DNA target site by positioning the face of a three-stranded β -sheet into the major groove. Binding is linked to structural adaptation. We have characterized INT-DBD binding to DNA in detail by calorimetry [Milev, S., Gorfe, A., Karshikoff, A., Clubb, R. T., Bosshard, H. R., and Jelesarov, I. (2003) *Biochemistry* 42, 3481–3491]. Our thermodynamic analysis has indicated that the major driving force of association is the hydrophobic effect while polar interactions contribute less. To gain more comprehensive information about the binding process, we performed a computational analysis of the binding free energy and report here the results. A hybrid molecular mechanics/continuum approach was followed. The total binding free energy is predicted with reasonable accuracy. The calculations confirm that nonpolar effects stabilize the protein–DNA complex while electrostatics opposes binding. Structural changes optimizing surface complementarity are costly in terms of energy. The energetic consequences from the replacement of nine DNA-contacting residues by alanine were investigated. The calculations correctly predict the binding affinity decrease of eight mutations and the destabilizing effect of one wild-type residue. Bulky side chains stabilize the wild-type complex through packing interactions and favorable nonpolar dehydration, but the net nonpolar energy changes do not correlate with the relative affinity loss upon mutation. Discrete protein–DNA electrostatic interactions may be net stabilizing or net destabilizing depending on the local environment. In contrast to nonpolar energy changes, the magnitude of the electrostatic free energy ranks the mutations according to the experimentally measured $\Delta\Delta G$. Free energy decomposition analysis from a structural perspective leads to detailed information about the thermodynamic strategy used by INT-DBD for sequence-specific DNA binding.

The Tn916 integrase is essential for excision and re-integration of the bacterial Tn916 conjugative transposon, which spreads antibiotic resistance among pathogenic bacteria (1). The protein belongs to the family of three-stranded β -sheet DNA binders. These recently discovered proteins share little sequence homology, yet they adopt a similar fold and bind to the DNA major groove with the face of a β -sheet (2). The solution structure of the N-terminal, minimal DNA-binding domain (INT-DBD)¹ has been solved in both the DNA-free form and in complex with the 13 bp DNA target duplex (1, 3). Binding involves structural rearrangements. Surprisingly, the protein appears more disordered in the DNA-bound state. We have described in detail the thermodynamics of the binding process (4). The energetic role of

discrete protein–DNA contacts has been elucidated by mutational analysis (5). In the present paper, we investigate the energetics of complex formation and the stabilizing effect of protein–DNA contacts by free energy calculations and computational alanine mutagenesis.

Among alternative theoretical approaches to free energy parsing, two computational schemes are finding growing application in calculating binding affinities of protein–DNA complexes. One is the use of purely continuum models within the rigid-body approximation. The nonlinear Poisson–Boltzmann equation (NLPB) is solved to calculate the electrostatic contributions, while nonelectrostatic interactions are computed from semiempirical relationships that are based on solvent accessibility (SA) (6, 7). This approach has been successfully applied to diverse interacting systems (6–10). While computationally efficient, the method requires some approximation(s) taking into account the conformational relaxation when conformational reorganization upon binding is significant (11). Another approach involves all-atom molecular mechanics (MM) calculations in conjunction with NLPB [or the generalized Born (GB) solvent model] for polar solvation and SA for nonpolar solvation [MM-PBSA (12)]. By using MD-generated ensembles of the free components and their complex, the MM-PBSA protocol accounts not only for conformational adaptation processes but also for the structural heterogeneity of the components and of the

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

* To whom correspondence should be addressed. Phone: +41 1 655 5547. Fax: +41 1 635 6805. E-mail: iljel@bioc.unizh.ch.

¹ Abbreviations: D, free DNA duplex; D*, protein-bound DNA duplex; INT-DBD, the minimal DNA-binding domain of Tn916 integrase; MM, molecular mechanics; MM-PBSA, all atom molecular mechanics/nonlinear Poisson–Boltzmann/solvent accessibility model; NLPB, nonlinear Poisson–Boltzmann equation; SA (ASA), atomic surface accessibility; P, free protein (INT-DBD); P*, DNA-bound protein (INT-DBD); P*·DNA*, INT-DBD–DNA complex; γ , vacuum-to-water surface tension coefficient. For an explanation of the superscripts and subscripts associated with G , see the footnotes to Tables 1 and 2.

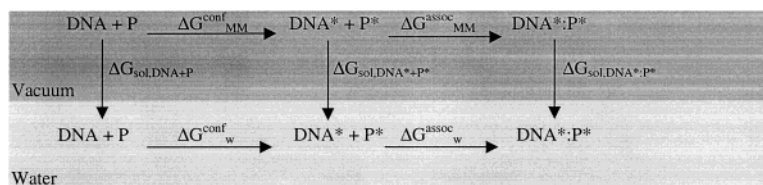


FIGURE 1: Thermodynamic cycle used to calculate the free energy change of INT-DBD binding to DNA. Binding is formally described by the sum of two hypothetical partial processes (see also Figure 2). The left-hand subcycle corresponds to structural transition from the free form of the protein and DNA (P and DNA) to their binding-competent conformations (P* and DNA*). The right-hand subcycle describes rigid-body association of P* and DNA* to the final complex.

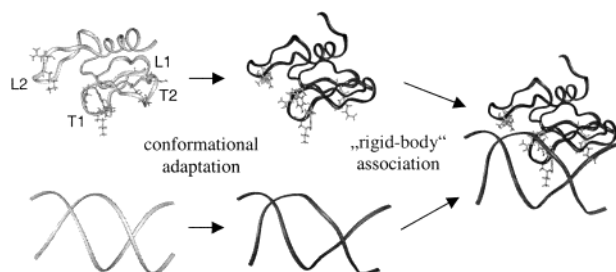


FIGURE 2: Hypothetical binding process for the interaction of INT-DBD with the target DNA. Free protein and the B-DNA duplex (in gray) first undergo conformational transition to their functional form (in black). In a second step, association involves formation of intermolecular contacts across the preformed complementary binding interface. The final structures of 1 ns MD trajectory are shown. Abbreviations: T1, turn 1; L1, loop 1; L2, loop L2. Residues contributing to affinity according to mutagenesis data are shown.

complex as well (12–14). Here we apply the MM-PBSA methodology for free energy analysis of the INT-DBD–DNA complex formation. We explore the balance of energetic contributions to the binding affinity, the sources of complex destabilization by mutations, and the role of electrostatics in discriminating the energetic importance of charged side chains.

COMPUTATIONAL METHODS

Physical Model

The thermodynamic cycle used for the free energy calculation is depicted in Figure 1. It is assumed that the reacting components undergo conformational change (adaptation) to the functional form prior to association. In the MM-PBSA protocol, all structural effects are accounted for explicitly by extensive molecular dynamics (MD) sampling. First, ensembles are generated in explicit water. Second, intermolecular and intramolecular energies are computed using force field parametrization after removing the solvent, i.e., in the gas phase. Third, the gas-to-water transfer free energy (solvation) of each species is calculated by NLPB and SA using appropriate parameters.

Free Energy of Conformational Adaptation. According to the left-hand side subcycle of Figure 1, the structural “adaptation” (see also Figure 2) is described as a transition from the native, nonbound conformation (P and DNA) to the functional, “binding-competent” form (P* and DNA*). The total conformational free energy in the aqueous phase (ΔG_w^{conf}) is

$$\Delta G_w^{\text{conf}} = \Delta G_{\text{MM}}^{\text{conf}} + \Delta G_{\text{sol,DNA*+P*}} - \Delta G_{\text{sol,DNA+P}} \quad (1)$$

The molecular mechanics energy ($\Delta G_{\text{MM}}^{\text{conf}}$) describing the change in intramolecular energies of the protein and DNA

is calculated in the gas phase from snapshots of MD trajectories using the CHARMM force field. It is composed of the difference in the ensemble mean values of the electrostatic ($\langle E_{\text{ele}}^{\text{conf}} \rangle$), van der Waals ($\langle E_{\text{vdW}}^{\text{conf}} \rangle$), and internal ($\langle E_{\text{int}}^{\text{conf}} \rangle$) energies for the transitions $P \rightarrow P^*$ and $\text{DNA} \rightarrow \text{DNA}^*$:

$$\Delta G_{\text{MM}}^{\text{conf}} = \langle \Delta E_{\text{ele}}^{\text{conf}} \rangle + \langle \Delta E_{\text{vdW}}^{\text{conf}} \rangle + \langle \Delta E_{\text{int}}^{\text{conf}} \rangle \quad (2)$$

The last two terms of eq 1 describe the changes in solute–solvent interactions of the protein and DNA upon gas-to-water transfer. Thus, $\Delta G_{\text{sol,DNA+P}}$ is the solvation free energy for the free forms, while $\Delta G_{\text{sol,DNA*+P*}}$ is the corresponding value for the binding-competent conformations. For simplicity of notation, we introduce the term $\Delta G_{\text{sol}}^{\text{conf}} = \Delta G_{\text{sol,DNA*+P*}} - \Delta G_{\text{sol,DNA+P}}$. $\Delta G_{\text{sol}}^{\text{conf}}$ can be further partitioned into contributions from polar solvation ($\Delta G_{\text{psol}}^{\text{conf}}$) and nonpolar solvation ($\Delta G_{\text{npsol}}^{\text{conf}}$):

$$\Delta G_{\text{sol}}^{\text{conf}} = \Delta G_{\text{sol,DNA*+P*}} - \Delta G_{\text{sol,DNA+P}} = \Delta G_{\text{psol}}^{\text{conf}} + \Delta G_{\text{npsol}}^{\text{conf}} \quad (3)$$

$\Delta G_{\text{psol}}^{\text{conf}}$ is calculated by NLPB and $\Delta G_{\text{npsol}}^{\text{conf}}$ is estimated from the changes in solvent accessibility according to

$$\Delta G_{\text{npsol}}^{\text{conf}} = \gamma_{\text{vw}}(\Delta \text{ASA}^{\text{conf}}) + 0.92 \text{ kcal mol}^{-1} \quad (4)$$

where $\Delta \text{ASA}^{\text{conf}}$ is the total solvent-accessible surface area difference between conformations X and X* and γ_{vw} is the macroscopic surface tension coefficient relating the solvent-accessible surface area to the transfer free energy of a macromolecule from vacuum to water (15, 16). The total electrostatic contribution ($\Delta G_{\text{ele}}^{\text{conf}}$) can be calculated as the sum of $\Delta G_{\text{psol}}^{\text{conf}}$ and $\langle \Delta E_{\text{ele}}^{\text{conf}} \rangle$ and the nonpolar contribution ($\Delta G_{\text{np}}^{\text{conf}}$) as the sum of $\Delta G_{\text{npsol}}^{\text{conf}}$ and $\langle \Delta E_{\text{vdW}}^{\text{conf}} \rangle$. Therefore, eq 1 can also be written as

$$\Delta G_w^{\text{conf}} = \Delta G_{\text{ele}}^{\text{conf}} + \Delta G_{\text{np}}^{\text{conf}} + \langle \Delta E_{\text{int}}^{\text{conf}} \rangle \quad (5)$$

Free Energy of Association of the Structurally Adapted Protein and DNA. The association process is shown in the right-hand side subcycle of Figure 1. Since all of the energetic consequences of the major conformational changes associated with binding are included in the ΔG_w^{conf} , the association of P* and DNA* is assumed to represent a “rigid-body” binding reaction (7). In strict analogy with the formalism described above, the association free energy is given by

$$\Delta G_w^{\text{assoc}} = \Delta G_{\text{MM}}^{\text{assoc}} + \Delta G_{\text{sol,DNA*+P*}} - \Delta G_{\text{sol,DNA*+P*}} \quad (6)$$

$\Delta G_{\text{MM}}^{\text{assoc}}$ corresponds to the gas-phase intermolecular interaction energy, whereas the difference $\Delta G_{\text{sol,DNA}^* \cdot \text{P}^*} - \Delta G_{\text{sol,DNA}^* + \text{P}^*}$ describes the change in solvation due to complex formation ($\Delta G_{\text{sol}}^{\text{assoc}}$). Analogously to eq 5

$$\Delta G_{\text{w}}^{\text{assoc}} = \Delta G_{\text{ele}}^{\text{assoc}} + \Delta G_{\text{np}}^{\text{assoc}} \quad (7)$$

where $\Delta G_{\text{ele}}^{\text{assoc}}$ is the sum of intermolecular electrostatic interactions in vacuum ($\langle \Delta E_{\text{ele}}^{\text{assoc}} \rangle$) and the change in electrostatic solvation of the complex ($\Delta G_{\text{psol}}^{\text{assoc}}$) and $\Delta G_{\text{np}}^{\text{assoc}}$ is the nonelectrostatic contribution calculated as the sum of the intermolecular vdW interactions ($\langle \Delta E_{\text{vdW}}^{\text{ass}} \rangle$) and nonpolar solvation ($\Delta G_{\text{npol}}^{\text{assoc}}$). The latter was estimated using eq 4, substituting $\Delta \text{ASA}^{\text{conf}}$ by $\Delta \text{ASA}^{\text{assoc}}$, i.e., the difference in ASA between the complex DNA*·P* and the isolated components DNA* + P*. Note that the change in internal energy is zero for noncovalent association of the binding-competent components.

Total Binding Free Energy. The total binding free energy was calculated as the sum of the conformational and association free energies corrected for entropic effects:

$$\Delta G_{\text{w}}^{\text{bind}} = \Delta G_{\text{w}}^{\text{assoc}} + \Delta G_{\text{w}}^{\text{conf}} - T\Delta S_{\text{other}} \quad (8)$$

where ΔS_{other} includes entropic contributions that are not related to solvent effects (see Results).

Structural Models

The average energy-minimized structures of the free protein and of the complex with a 13 base pair DNA duplex were obtained from the Brookhaven Protein Data Bank [PDB accession numbers 2bb8 and 1b69 (1, 3)]. Residues that are not defined in the structures (Ser2, His72, Asp73, and Gly74 in the complex; Ser2 and Gly74 in the free protein) were manually built. X-to-Ala mutants were generated by CHARMM, and the resulting structures were minimized following standard protocols.

Ensembles of thermally averaged structures were generated from 1 ns MD simulations. The solutes were immersed in a preequilibrated TIP3P (17) water-containing box of appropriate dimensions so that in all cases the minimal distance from any solute atom to the edge of the box was 10 Å. Sodium and chloride ions were added to neutralize the total charge of the macromolecular system (+5, −24, and −19 for the protein, DNA, and protein–DNA complex, respectively). The system was first relaxed by 50 steps of steepest descent minimization, followed by 200 steps of the adopted-basis Newton–Raphson (ABNR) minimization with harmonic constraints of force constants of 1.0 and 2.0 kcal mol^{−1} Å^{−2} applied to water oxygen atoms and sodium atoms and to solute heavy atoms, respectively. This was followed by heating to the target temperature in 15 ps. The harmonic constraint force constant was progressively decreased from 2.0 to 0.0 kcal mol^{−1} Å^{−2} in the initial 10 ps, and all atoms were free to move during the last 5 ps of the heating process. The system was then subjected to a 20 ps equilibration at constant temperature and volume using Gaussian distribution for the assignment of atomic velocities. The production simulations were performed for 1 ns at 298 K. A shift function with a cutoff at 12 Å for the van der Waals interactions was used. Long-range electrostatic interactions were treated with the particle mesh Ewald method (18). The

real space contribution was truncated at 14 Å. The simulations were performed in the isothermal, isobaric ensemble using the leapfrog integrator. The pressure and temperature were kept constant using a Langevin piston of mass 600 amu and a Hoover thermostat with a thermal piston of mass 1000 kcal ps.² All bonds involving hydrogen atoms were constrained by the SHAKE algorithm (19). The integration time step was 2 fs. The MD simulations were carried out with the CHARMM program (20) using the all-atom parameter set of CHARMM27 (21).

Computational Procedures and Parameters

For the MM-PBSA calculations, the solvent was removed from snapshots sampled at 10 ps intervals (12–14). No cutoff was applied in the MM calculation of solute–solute interactions terms (i.e., electrostatic and van der Waals). In all calculations, CHARMM charges, neutral pH, and 300 K were used. All NLPB calculations were done using the DelPhi program (22). The solute was modeled as a low dielectric ($\epsilon_s = 1$) immersed into a region of high dielectric ($\epsilon_s = 80$) defined by a probe radius of 1.4 Å. The solvent medium also contains univalent salt ions corresponding to an ionic strength of 0.15 M (with a Stern layer of 2 Å radius). Charges and molecular surface were mapped onto a 229³. Accuracy was improved by using a three-step focusing procedure, where the solute fills 24% of the grid in the initial coarse calculations and 97% of the grid in the final calculation (23).

The nonpolar contributions to the free energies were calculated using a $\gamma_{\text{vw}} = 0.00542$ kcal mol^{−1} (10, 15). Solvent-accessible surfaces were calculated with the program NACCESS using a probe radius of 1.4 Å (24).

RESULTS

Free Energy Calculations

To gain insight into the balance of forces that drive the formation of the INT-DBD–DNA complex, we performed free energy computations based on the physical model sketched in Figure 1 (see also Figure 2). INT-DBD binds its target DNA sequence with a modest affinity of −9.5 kcal mol^{−1} in 100 mM NaCl-containing buffer at room temperature and pH 6 (4). To facilitate direct comparisons with the thermodynamic information available, it is necessary to parametrize the calculations so as to mimic the experimental conditions. For an electrostatically intensive system like the present one, the assignment of charges is a particularly important step. To this end, we calculated the pK_a's of all 36 titratable protein groups, applying the continuum electrostatics model. According to the calculation, all groups but two titrate outside the pH region 4.5–8, both in the unbound and bound state. His72 has a pK_a of 6.3. The pK_a of Glu35 is slightly upshifted to 4.7 in the free protein and rises to 6.0 in the protein–DNA complex. For this reason the charges in the MD simulations and in free energy calculations were assigned so as to correspond to pH 7.0 (all carboxylates and the single His being deprotonated, all Lys and Arg side chains as well as the N-terminal amine being protonated). The net

² All computed free energy values, e.g., $\Delta G_{\text{w}}^{\text{conf}}$, are in fact $\Delta \Delta G$'s, e.g., $\Delta \Delta G_{\text{w}}^{\text{conf}}$, since they represent energy differences between processes on the corresponding horizontal or vertical limbs of the thermodynamic cycles in Figure 1. For simplicity, we use Δ 's throughout.

charges of protein and DNA are +5 and −24, respectively. In the absence of thermodynamic data collected at pH 7.0, equilibrium-binding experiments were performed by fluorescence titration (not shown). The binding affinity decreases by ~ 1 kcal mol $^{-1}$ between pH 4.5 and pH 9.5. The difference between pH 6 and pH 7 is only 0.1–0.2 kcal mol $^{-1}$. Thermal unfolding experiments indicate that the stability of the protein is insensitive to pH from pH 4.5 to pH 9.0 and there is no spectroscopic evidence for local structural changes in this pH interval.

Structure Calculations. The MM-PBSA calculation involves explicit computation of conformational free energy changes and hence requires extensive conformational sampling, which can be achieved by MD simulations. MD trajectories in explicit water were started from the average NMR structures of the complex and its components. The time evolution of the root mean square deviations (RMSD) from the starting structures along the MD trajectories is shown in Figure 3A,B. Figure 3C presents the effective energy ($\Delta G_{\text{MM}} + \Delta G_{\text{psol}} + \Delta G_{\text{npol}}$). According to both structural and energetic criteria, all trajectories reached a plateau after approximately 200 ps. Free energy calculations were therefore performed using the stable part of the trajectories from 200 to 1000 ps. The structure of unbound DNA is not known. However, the DNA ensemble generated by simulations that were started from the structure extracted from the complex converged to the canonical B-form DNA. The structures from this trajectory are hereupon used to represent the conformation of free DNA duplex. The results of the MM-PBSA calculations are shown in Table 1.

Free Energy Changes of Rigid-Body Association.³ The total association free energy change ($\Delta G_{\text{w}}^{\text{assoc}}$) of structurally adapted protein (P*) and DNA (D*) is -74.7 kcal mol $^{-1}$. It is dominated by intermolecular van der Waals interactions ($\Delta E_{\text{vdW}}^{\text{assoc}}$). Dehydration of the nonpolar surface ($\Delta G_{\text{npol}}^{\text{assoc}}$) favors binding as well. The total electrostatic free energy ($\Delta G_{\text{ele}}^{\text{assoc}}$) opposes binding by 8 kcal mol $^{-1}$. The modest magnitude of the electrostatic effects is the consequence of the balance between favorable Coulombic interaction ($\langle \Delta E_{\text{ele}}^{\text{assoc}} \rangle$) and unfavorable polar solvation ($\Delta G_{\text{psol}}^{\text{assoc}}$). The free energy components estimated here are all close to the mean values calculated for 40 protein–DNA complexes based on a similar approach (25). The calculated association free energy is significantly more negative than the experimental value of ~ -10 kcal mol $^{-1}$, suggesting that there is a large free energy contribution from the conformational changes of the protein and/or DNA, as well as from entropic effects.

Energetic Costs of Conformational Adaptation. The free energy associated with conformational changes of reacting species were computed as

$$\Delta G_{\text{X}}^{\text{conf}} = G(\text{complex})_{\text{X}} - G(\text{free})_{\text{X}} \quad (9)$$

where X designates protein or DNA, $G(\text{complex})_{\text{X}}$ is the free energy calculated from the simulation of the complex, and

³ “Rigid-body association” is clearly inappropriate for calculations according to MM-PBSA, where energies were computed on ensembles of conformations. We use this term throughout as a simple literal description of association between the structurally adapted protein and DNA (see energy terms indexed as in Figure 1 and the second partial process in Figure 2).

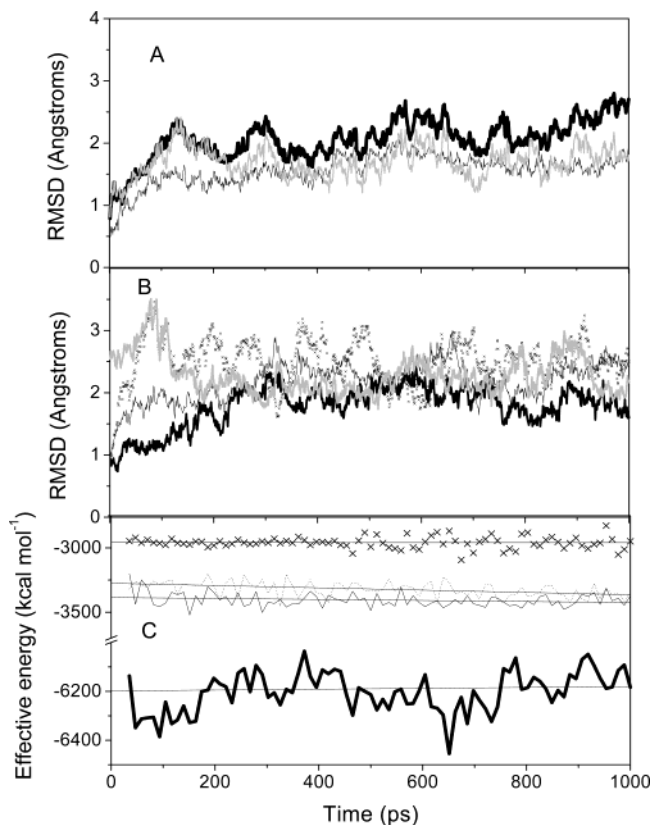


FIGURE 3: Time evolution of the trajectories used to generate the MD ensembles. (A) Simulation of the INT-DBD–DNA complex. Positional RMSD from the starting structure (energy-minimized average structure of the complex; PDB accession code 1B69). Key: black heavy line, protein; gray heavy line, DNA; thin black line, protein–DNA complex. (B) Simulations of nonbound components. RMSD were measured from the starting structures. Key: black heavy line, free protein (started from the average NMR structure; PDB code 2BB8); thin black line, bound protein (started from the average NMR structure of the complex; PDB code 1B69); crosses, bound DNA (started from the average NMR structure of the complex; PDB code 1B69). RMSD along the same DNA trajectory were measured from the canonical B-DNA conformation and are shown with the heavy gray line. Except for the initial ~ 100 ps the DNA duplex is closer to the B-form than to the starting structure. In (A) and (B) RMSD were measured for protein C $_{\alpha}$ atoms and for DNA backbone atoms. (C) Time evolution of the effective energy ($\Delta G_{\text{MM}} + \Delta G_{\text{psol}} + \Delta G_{\text{npol}}$) calculated for free protein (thin line), bound protein (dotted line), bound DNA (crosses), and the protein–DNA complex (heavy line).

$G(\text{free})_{\text{X}}$ is the free energy calculated from the simulation of the isolated components (12). The calculated $\Delta G_{\text{DNA}}^{\text{conf}}$ using this approach is 22 kcal mol $^{-1}$ (Table 1). The intramolecular electrostatic energy ($\langle \Delta E_{\text{ele}}^{\text{conf}} \rangle$) strongly favors the bending of DNA while solvation, van der Waals, and internal energies disfavor the bent form. The van der Waals interactions and nonpolar solvation contribution are relatively small. For the P \rightarrow P* transition ($\Delta G_{\text{protein}}^{\text{conf}}$), the calculated free energy using this approach was found to be unrealistically high (~ 70 kcal mol $^{-1}$). The reason is not clear. Although the simulation of the complex appears equilibrated between 200 and 1000 ps (Figure 3), the protein is perhaps exploring a high-energy region of the conformational space. Indeed, the trajectory exhibits a step at about 1400 ps where the protein appears to attain a second equilibration phase. Equilibration problems in 1 ns trajectories have been commented (26). A more realistic estimate was obtained by

Table 1: Breakdown of the Energetic Contributions to the Free Energy of INT-DBD Binding to DNA^a

	$\langle \Delta E_{\text{ele}} \rangle$	ΔG_{psol}	ΔG_{ele}	$\langle \Delta E_{\text{vdW}} \rangle$	ΔG_{npol}	ΔG_{np}	$\langle \Delta E_{\text{int}} \rangle$	ΔG_{calc}	ΔG_{expt}
ΔG^{assoc}	-2942.7	2950.7	8.0	-70.7	-12.0	-82.7	0.0	-74.7	
$\Delta G^{\text{conf}}_{\text{DNA}}$	-77.7	92.9	15.2	2.8	0.3	3.1	3.5	21.8	
$\Delta G^{\text{conf}}_{\text{protein}}$	77.5	-60.5	17.0	-0.3	0.7	0.4	6.8	24.2	
ΔG^{conf}	-0.2	32.4	32.2	2.5	1.0	3.5	10.3	46.0	
ΔG^{bind}								-12.9	-9.5

^a All values are in kcal mol⁻¹. To improve legibility, superscripts and subscripts otherwise associated with ΔG values are omitted. ΔG 's in the leftmost column refer to the ΔG 's calculated for the particular process in water and thus would be indexed with w, as, for example, ΔG^{assoc} (table) vs $\Delta G^{\text{assoc}}_{\text{w}}$ (text). In the header row, ΔG 's should be understood with superscripts assoc or conf according to whether they refer to association or to conformational free energy changes, respectively. Indices of ΔG are defined as follows (see also Computational Methods): int, internal (bond + angle + torsion) energy calculated by the CHARMM molecular mechanical force field; ele, electrostatic energy ($\Delta G_{\text{ele}} = \langle \Delta E_{\text{ele}} \rangle + \Delta G_{\text{psol}}$); np, nonpolar energy ($\Delta G_{\text{np}} = \Delta E_{\text{vdW}} + \Delta G_{\text{npol}}$); vdW, van der Waals energy; psol, polar solvation; npol, nonpolar solvation; assoc, association of structurally adapted ("binding-competent") components; conf, transition between the free and bound form; calc, total computed energy for a particular process ($\Delta G_{\text{calc}} = \Delta G_{\text{ele}} + \Delta G_{\text{np}} + \Delta E_{\text{int}}$) excluding the entropic effect of side chain freezing and of changes of rotational/translational degrees of freedom ($T\Delta S_{\text{other}}$); bind, binding of free components to the final complex ($\Delta G^{\text{bind}} = \Delta G^{\text{assoc}} + \Delta G^{\text{conf}} - T\Delta S_{\text{other}}$; eq 8); expt, experiment.

running a separate trajectory starting from the bound conformation extracted from the complex:

$$\Delta G^{\text{conf}}_{\text{protein}} = G(\text{bound})_{\text{protein}} - G(\text{free})_{\text{protein}} \quad (9a)$$

The ensembles generated for the bound form and free form of the protein do not converge to each other, indicating sampling of different local regions of the conformational space (A. Gorfe and I. Jelesarov, submitted for publication). The use of eq 9a instead of eq 9 is consistent with the thermodynamic cycle in which the conformational change is considered as a separate process preceding association (Figure 1). Using this approach, a reasonable estimate of ~ 24 kcal mol⁻¹ was obtained for $\Delta G^{\text{conf}}_{\text{protein}}$ (Table 1). The electrostatic interactions and the internal energies oppose the conformational transition to the bound form. Contrary to the DNA, the bound conformation is characterized by a severe deoptimization of Coloumbic interactions accompanied by a compensating gain of polar solvation free energy. The sum of van der Waals interactions and nonpolar solvation also disfavors the bound conformation, yet it is small. The total conformational adaptation free energy ($\Delta G^{\text{conf}}_{\text{w}} = \Delta G^{\text{conf}}_{\text{protein}} + \Delta G^{\text{conf}}_{\text{DNA}}$) is significant, 46 kcal mol⁻¹. However, since the protein structure is more relaxed in the bound conformation compared with the free form (3), an increase in conformational entropy can be expected during adaptation, reducing the energetic penalty of the conformational change.

Empirical Estimates of Non-Solvent-Related Entropic Contributions. The free energy changes arising from entropic effects accompanying complex formation ($-T\Delta S_{\text{tot}}$) can be partitioned into

$$T\Delta S_{\text{tot}} = T(\Delta S_{\text{hyd}} + \Delta S_{\text{ion-re}} + \Delta S_{\text{other}}) = T(\Delta S_{\text{hyd}} + \Delta S_{\text{ion-re}} + \Delta S_{\text{sc}} + \Delta S_{\text{vib}} + \Delta S_{\text{tr}}) \quad (10)$$

The entropy of changes in hydration (ΔS_{hyd}) and the entropy of ion reorganization ($\Delta S_{\text{ion-re}}$) have been treated at different steps in the calculation of ΔG_{np} and ΔG_{ele} . The term ΔS_{other} lumps together non-solvent-related contributions: the loss of entropy due to side chain immobilization in the complex (ΔS_{sc}), the entropy of redistribution of vibrational modi (ΔS_{vib}), and the entropy associated with the loss of translational and rotational degrees of freedom when one particle (complex) is formed from two independent kinetic units (ΔS_{tr}). We do not consider ΔS_{vib} explicitly. Hydration entropy, when calculated by semiempirical parametrization

schemes based on organic compound transfer thermodynamic data, necessarily includes, at least to some extent, vibrational contributions as well. It has been argued that for macromolecular interactions, where hydration effects govern the entropy, changes in vibrational content may be relatively small. On average, 12 amino acid side chains make contacts with DNA bases or backbone and are buried more than 60% at the protein–DNA interface relative to the unbound protein. Following Piket and Sternberg, the entropic costs of side chain immobilization is ~ 19 kcal mol⁻¹ (27). The same procedure was applied to the $P \rightarrow P^*$ transition. The side chain entropy contribution to the conformational rearrangement of the protein is favorable, -4.7 kcal mol⁻¹, consistent with the structural evidence that the protein appears more disordered in the complex. Thus, the net total side chain entropy ($-T\Delta S_{\text{sc}}$) is 14.3 kcal mol⁻¹. The exact magnitude of the loss of translational/rotational degrees of freedom (ΔS_{tr}) upon binding has been vigorously debated (e.g., ref 28). Here we prefer to use the experimentally derived estimate of ~ -5 cal K⁻¹ mol⁻¹, contributing 1.5 kcal K⁻¹ mol⁻¹ to the binding free energy at 25 °C. The energetic penalty caused by non-solvent-related entropic effects amounts therefore to 15.8 kcal mol⁻¹, which was used to "correct" the calculated free energies according to the thermodynamic cycles in Figure 1.

Total Binding Free Energy. The sum of the conformational free energies and the association free energies listed in Table 1 amounts to -28.7 kcal mol⁻¹. According to eq 8, the calculated total free energy of binding is thus -12.9 kcal mol⁻¹. In view of the large uncertainty associated with many of the energetic terms, the calculation resulted in a fair agreement with the experimental value of -9.5 kcal mol⁻¹. The presented analysis emphasizes the role of nonpolar interactions as the driving force of association. Electrostatic and conformational free energy contributions oppose binding.

Computational Alanine Scanning

On the basis of the MM-PBSA protocol, a computational alanine scanning methodology has been developed by Kollmann and colleagues (12, 26). Using the same approach, we probed the differential effect to affinity of 9 of the 13 alanine mutations studied experimentally by Clubb and colleagues (5). We assume that no major structural changes occur upon alanine substitutions (no change of the conformational free energy) and that the relative entropic effects are negligible. The effect of the mutation $X \rightarrow A$ was

Table 2: Energetic Decomposition of the Calculated Effect of INT-DBD Alanine Mutants on DNA-Binding Affinity^a

	$\Delta\Delta E_{\text{vdW}}$	$\Delta\Delta G_{\text{np sol}}$	$\Delta\Delta G_{\text{np}}$	$\Delta\Delta E_{\text{int}}$	$\Delta\Delta G_{\text{ele}}$	$\Delta\Delta G^{\text{bind}}$	$\Delta\Delta G^{\text{expt}}$
Y40A	11.3	0.2	11.4	1.8	−9.6	3.6	1.6
K28A	5.5	0.4	5.9	−2.1	−0.9	2.9	1.4
K54A	2.4	−0.1	2.3	2.2	−2.1	2.4	1.4
R24A	5.2	0.1	5.3	−2.4	3.8	6.7	1.3
R55A	7.5	0.8	8.3	0.2	−6.6	1.9	1.2
R5A	−3.5	−0.5	−4.0	−1.3	6.9	2.5	0.8
K21A	5.2	0.3	5.5	2.9	−8.3	0.1	0.7
R20A	8.7	−0.3	8.4	1.9	−9.7	0.6	0.5
F38A	3.5	0.3	3.8	0.1	−6.5	−2.6	−0.3

^a Values are in kcal mol^{−1}. Mutants are listed according to the magnitude of the experimentally measured $\Delta\Delta G$ in decreasing order. Indices of $\Delta\Delta G$ are defined as in Table 1. $\Delta\Delta G > 0$ indicates that the wild-type residue is stabilizing the complex; $\Delta\Delta G < 0$ favors the alanine mutant.

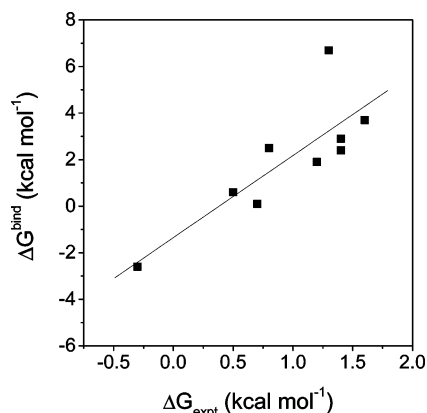


FIGURE 4: Comparison of predicted ($\Delta\Delta G^{\text{bind}}$) and experimentally measured ($\Delta\Delta G^{\text{expt}}$) effects of alanine mutations on binding affinity. Absolute $\Delta\Delta G$ values calculated by MM-PBSA (filled squares) are plotted against the experimental values. The lines represent the linear correlation to experiment of the MM-PBSA prediction [continuous line; $\Delta\Delta G^{\text{bind}} = -1.34 + 3.52(\Delta\Delta G^{\text{expt}})$; $R = 0.82$].

quantified in terms of the free energy difference relative to the wild type:

$$\Delta\Delta G^{\text{bind}} = \Delta G^{\text{X} \rightarrow \text{A}} - \Delta G^{\text{X}} \quad (11)$$

where X designates the corresponding residue in the integrase wild-type protein. The results are summarized in Table 2.

For four of the nine alanine substitutions (R20A, K21A, K54A, and R55A), $\Delta\Delta G^{\text{bind}}$ is predicted within less than 1 kcal mol^{−1} deviation from the experiment (Table 2). Predicted and experimental $\Delta\Delta G$'s differ by 1.5–2.3 kcal mol^{−1} for four other mutations (R5A, K28A, Y40A, and F38A). Only R24A is predicted with large error, the change in free energy being overestimated by 5 kcal mol^{−1}. Despite the fact that almost all of the residues are charged/polar and (partially) buried in the interface, where the MM-PBSA was reported to perform relatively poorly (26), the average unsigned error is 1.6 kcal mol^{−1}. Excluding the outlier R24A, the error reduces to 1.1 kcal mol^{−1}. Moreover, six of the nine substitutions are correctly ranked in terms of their importance to affinity, and the calculation correctly captures the increase of affinity of the F38A mutant. For all nine mutants the final $\Delta\Delta G^{\text{bind}}$ reasonably correlates to the experimental findings ($R = 0.82$; Figure 4).

The bar graph in Figure 5 displays calculated $\Delta\Delta G_{\text{ele}}$, $\Delta\Delta G_{\text{np}}$, and $\Delta\Delta G^{\text{bind}}$. Analysis of the free energy components

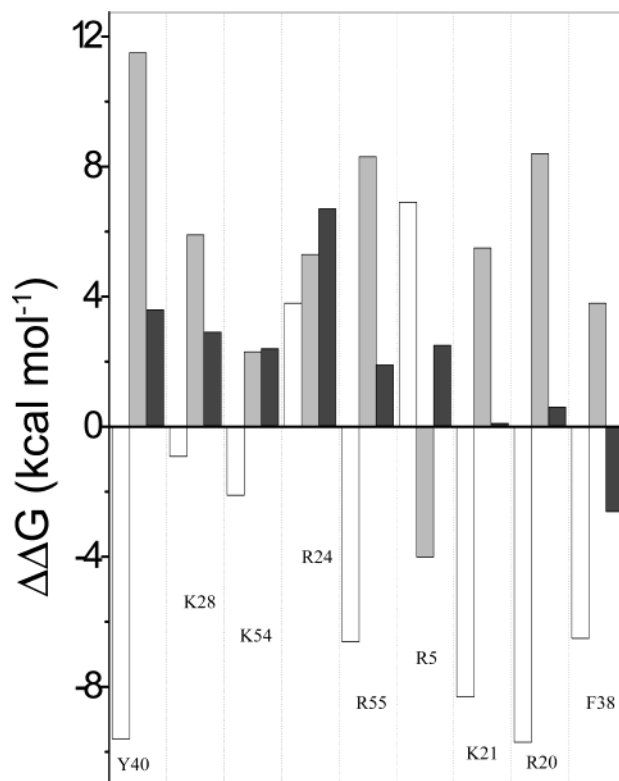


FIGURE 5: Decomposition of the total predicted free energy effects of alanine substitutions (black bars) in terms of electrostatic ($\Delta\Delta G_{\text{ele}}$, white bars) and nonpolar ($\Delta\Delta G_{\text{np}}$, gray bars) free energy contributions. The mutations are indicated by the name of the wild-type residue and are ordered by the absolute experimentally measured $\Delta\Delta G$ in decreasing order (left to right). The rightmost F38A mutation stabilizes the mutant protein–DNA complex.

shows that van der Waals interactions with DNA of all studied residues except for Arg5 stabilize the wild-type complex ($\Delta\Delta E_{\text{vdW}} > 0$; Table 2). Indeed, in the course of MD simulations, all side chains of interest but Arg5 are involved in van der Waals contacts with the DNA (A. Gorfe and I. Jelesarov, submitted for publication). However, $\Delta\Delta E_{\text{vdW}}$ is not correlated to the experimentally observed effect of a given mutation ($R = 0.3$). $\Delta\Delta G_{\text{np sol}}$ is small for all substitutions, and thus $\Delta\Delta G_{\text{np}}$ (sum of $\Delta\Delta E_{\text{vdW}}$ and $\Delta\Delta G_{\text{np sol}}$) does also have no correlation to the experimental trend. Interestingly, electrostatics is found to favor the wild type ($\Delta\Delta G_{\text{ele}} < 0$) only for the R24A and R5A substitutions among three lysines and four arginines in the set. Arg5 makes a hydrogen bond with the well-exposed phosphate oxygen of Thy4 in the minor groove. Thus its electrostatic interaction is less expensive in terms of desolvation. Arg24, on the other hand, is completely buried, but strong attractive Coulombic interactions with DNA dominate over the desolvation cost. Although the binding energetics of these two mutants can be rationalized in structural terms, the numeric values should be considered with caution. Both $\Delta\Delta G_{\text{ele}}$ and $\Delta\Delta G_{\text{np}}$ for R5A fall outside the general trend (Figure 5). Arg5 is located in the flexible N-terminal segment of the protein, and sampling might not have been sufficient. Hence, the very reasonable prediction of $\Delta\Delta G^{\text{bind}}$ may result from cancellation of errors. As to R24A, the likely reason for the large error of the prediction is that Arg24 is involved in an extensive network of hydrogen bonds with DNA. In such cases, the assumption about the structural invariance around the site of mutation

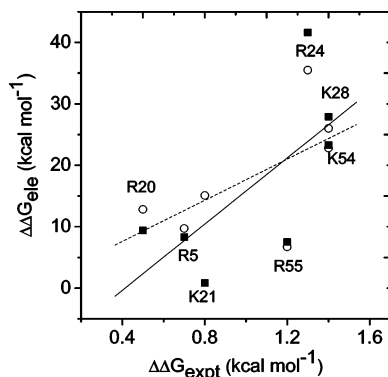


FIGURE 6: Predicted contribution of charge–charge interactions on the loss of DNA-binding affinity upon replacement of charged residues by alanine. Squares represent $\Delta\Delta G_{\text{ele}}$ calculated for charge neutralization of the side chain ($X \rightarrow X^\delta$ mutation). Open circles correspond to $\Delta\Delta G_{\text{ele}}$ calculated with artificial side chains which have the same atomic composition as the wild-type residue but have the charge distribution as in an alanine and carry no charges beyond the C_β atom ($X \rightarrow X^A$ mutation). The names of the seven charged residues are indicated. The correlation with the experimental $\Delta\Delta G$ is displayed for $X \rightarrow X^\delta$ (continuous line) and $X \rightarrow X^A$ (dotted line).

may not hold (26). Treating R24A and R5A separately, $\Delta\Delta G_{\text{ele}}$ for the remaining charged residues appears to correlate with the experimental data ($R = 0.92$ compared with $R = 0.4$ for all charged residues).

Electrostatics Determines the Side Chain's Importance for Affinity. The role of electrostatics in discriminating between residues at the interface was further investigated in the following manner. In a first step, lysine and arginine residues were neutralized so that they carry zero net charge ($X \rightarrow X^\delta$). This facilitates the estimation of the free energy contribution of electrostatic interaction between the +1 charge of a particular residue and all of the other charges (designated as $\Delta\Delta G_{\text{ion}}$). In a second step, all partial charges beyond the C_β atom of the considered side chain were “turned off” ($X \rightarrow X^A$). The resulting mutated side chain has the charge distribution of an alanine, yet there is no change in the local conformation and water accessibility. In this way we may estimate the pure charge effect of mutation to Ala (termed $\Delta\Delta G_{A^*}$) without the interfering effect of structural and solvent readjustment. The difference in electrostatic free energy between steps 1 and 2 quantifies the effect of the partial charges residing on the atoms beyond C_β ($\Delta\Delta G_{\text{partial}} = \Delta\Delta G_{A^*} - \Delta\Delta G_{\text{ion}}$).

The computations with $X \rightarrow X^\delta$ and $X \rightarrow X^A$ demonstrate that charge–charge interactions contribute favorably to the binding free energy (not shown). Compared to the experiment, the relative order of the charge–charge contributions is predicted well, with $R = 0.7$ and $R = 0.6$ for $\Delta\Delta G_{\text{ion}}$ and $\Delta\Delta G_{A^*}$, respectively (Figure 6). Interestingly, Arg20 which is positioned so as to form hydrogen bonds with DNA in the NMR structure, yet experimentally has been found to be less important for affinity (5), provides little electrostatic stabilization. In most cases, partial charges are not very important since $\Delta\Delta G_{\text{partial}} = \Delta\Delta G_{A^*} - \Delta\Delta G_{\text{ion}}$ is small. The most important difference between the artificial charge-deficient mutants ($X \rightarrow X^\delta$, $X \rightarrow X^A$) and the “genuine” $X \rightarrow A$ mutants is the energetic effect of desolvation. Removal of the bulky side chain reduces the energetic penalty paid for desolvation of the polar surface. This effect, which offsets

the favorable Coloumbic interactions, explains why electrostatics favors the $X \rightarrow A$ mutants. It appears that electrostatics is responsible for the relative, not for the absolute, energetic contribution of the studied charged residues to affinity. This agrees with the idea that hydrogen bonds importantly govern the specificity of binding while van der Waals contacts contribute to overall stabilization.

DISCUSSION

Protein–DNA binding presents an intriguing case of molecular recognition. Although the DNA molecule is structurally monotonic, proteins use very different thermodynamic and structural strategies for binding. Integrase Tn916 recognizes the major groove with the face of a β -sheet, a novel mode of binding. We have presented a calorimetric study of the thermodynamics of complex formation (4). Calorimetry provides invaluable information on the balance of enthalpic and entropic contributions to binding and directly measures heat capacity changes. However, calorimetric data represent a global picture and contain little information about the magnitude and sign of free energy terms characterizing particular physical processes. Free energy decomposition is currently possible only by theoretical methods. In this work, we explore the free energy profile of the INT-DBD–DNA complex to achieve a more comprehensive picture of the association process. We concentrate more specifically on the analysis of the energetic mechanisms by which particular protein–DNA interactions contribute to binding affinity.

Energetic Determinants of Integrase–DNA Binding. INT-DBD contacts DNA over an interfacial area of $\sim 2300 \text{ \AA}^2$. It is therefore not surprising that intermolecular polar and van der Waals packing interactions contribute a large amount of stabilization energy when evaluated in a vacuum. However, the presented analysis emphasizes the role of nonpolar energy changes as the major driving force of binding. Both van der Waals contacts and dehydration of the nonpolar surface favor association. The total electrostatic free energy in fact counteracts binding, since the favorable charge–charge energy is more than offset by the highly unfavorable contribution from dehydration of polar groups. The stabilizing role of nonpolar interactions and the magnitude of ΔG_{np} are in agreement with similar estimates for other protein–DNA and protein–drug complexes (6, 7, 25, 29, 30). It has also been argued that a net destabilizing effect of electrostatics represents a universal feature of protein–DNA interactions. Our results strengthen this contention. From a computational point of view, we find no principal difference in the energetic signature of major groove binding via β -sheets and α -helices, in spite of the fact that these structural elements intrinsically differ in rigidity and dynamics. The dominant role of the hydrophobic effect as a major affinity determinant is further supported by our calorimetric results (4). First, the total association enthalpy is small and changes sign close to 20°C , a typical enthalpic behavior of an association reaction driven by nonpolar interactions with small net contribution from polar effects. Second, the heat capacity change is large and negative, indicating dehydration of a significant amount of the nonpolar surface.

Binding-induced structural changes are typical for protein–DNA-binding reactions, and their energetics profoundly

influence the thermodynamic profile of protein–DNA complexes (31). Usually, DNA-binding proteins are flexible or partly unfolded and undergo disorder-to-order transition upon binding. The INT-DBD domain represents an interesting exception in that it is more structurally disordered in the bound form (3). Nevertheless, our calorimetric data are consistent with a reduction of the thermal motions of protein and DNA when bound with a concomitant heat release (4). Here we estimate the energetic costs of structural adaptation to disfavor binding by 46 kcal mol⁻¹. Although the magnitude of this term varies much, depending on particular, system-specific details, it is invariably opposing association of proteins with DNA (25). It appears, therefore, that high specificity, which is achieved through optimization of favorable contacts over a large interaction area, is linked to a significant energetic penalty.

Specific Effects of Amino Acid Substitutions. We studied the energetic effect of nine X-to-Ala replacements of INT-DBD residues, for which biochemical data exist (5). Eight side chains are hydrogen-bonded to DNA bases or backbone phosphates and contribute 0.4–1.6 kcal mol⁻¹ to affinity in comparison to their alanine variants. Phe38 is involved in van der Waals contacts with DNA, yet the F38A mutant displays slightly stronger binding than the wild type ($\Delta\Delta G_{\text{expt}} = -0.3$ kcal mol⁻¹). As a general trend, the calculations rank the mutations following the experimentally observed order of affinities (Figures 4 and 5). We find that the electrostatic part of the free energy is the major factor governing the relative energetic importance. Elimination of charges decreases both the screened Coulombic attraction and the polar solvation energy terms, but the latter is affected more. Therefore, positioning of the charged side chains at their defined location is expensive in terms of solvation, which often is not compensated enough by interactions with the DNA charges. Thus, total electrostatics may or may not be stabilizing in absolute terms and is in fact often destabilizing the wild-type complex. However, the more favorable (or less unfavorable) the electrostatics, the larger the loss of affinity upon Ala substitution tends to be. On the other hand, the nonpolar free energy disfavors the mutants in most cases but is “unspecific”.

Energetic decomposition linked to structural details on the atomic level helps to substantiate the available biochemical and structural data on the eight residues that together make up almost all of the net free energy of association. In the following we discuss, from a structural perspective, the energetic role of particular residues, which differ in their structural environment and thus might possibly exemplify typical energetic patterns.

The face of strands $\beta 2$ and $\beta 3$ represents the largest nonpolar patch of INT-DBD–DNA contact area. Lys28 ($\beta 2$), Phe38, and Tyr40 (both $\beta 3$) contact DNA through hydrogen bonds and/or hydrophobic interactions. All three side chains are deeply buried in the interface (relative burial > 0.7). According to biochemical data, Tyr40 and Lys28 are the two major contributors to affinity. Because of the extensive van der Waals contacts of Tyr40 with DNA, $\Delta\Delta G_{\text{np}}$ for the Y40A substitution highly favors the wild type, compensating for the large electrostatic free energy gain calculated for the mutant (11.4 versus -9.6 kcal mol⁻¹). Together with a relatively small change in internal energy (1.8 kcal mol⁻¹), a final $\Delta\Delta G^{\text{bind}}$ of 3.6 kcal mol⁻¹ is obtained. This is the

largest in the set (excluding the outlier Arg24) and compares well with experiment. Contrary to the Y40A, the electrostatic energy only modestly favors the K28A mutant (~ -1 kcal mol⁻¹), but similarly to Y40A, loss of van der Waals contacts renders the mutation unfavorable by 5.5 kcal mol⁻¹, so that, together with a change in the internal energy of -2.1 kcal mol⁻¹, the net calculated $\Delta\Delta G^{\text{bind}}$ is 2.9 kcal mol⁻¹. Phe38 is involved in extensive van der Waals contacts with DNA, and not surprisingly, $\Delta\Delta G_{\text{np}}$ is large and stabilizes the wild-type complex by ~ 4 kcal mol⁻¹. Yet the burial of partial charges is costly (-6.5 kcal mol⁻¹), and the net computed free energy favors the F38A mutant, in agreement with the experiment. It would appear that nonpolar residues in contact with DNA might be energetically preferred in this largely hydrophobic region. However, the polar side chains of Lys28 and Tyr40 make specific hydrogen bonds. Hence, the costly burial of groups with hydrogen-bonding potential confers substantial specificity to binding. It is also interesting to note that the completely hydrophobic Phe38 has lower calculated $\Delta\Delta G_{\text{vdw}}$ in comparison with Lys28 and Tyr40, and its bulky nonpolar side chain is even slightly opposing binding. The difference in net favorable nonpolar energy therefore originates from different packing of the hydrophobic parts of side chains. Oriented hydrogen bonds may assist positioning, may restrict flexibility, and hence may improve packing.

The protein–DNA complex is also stabilized by interactions from loop regions. Ala substitution of Arg5 from loop 1 and of Lys54 and Arg55 from loop 2 lowers binding affinity by 0.8, 1.4, and 1.2 kcal mol⁻¹, respectively. In contrast to the residues discussed above, all three side chains are well exposed, with a relative surface burial of 0.25–0.4. Arg5 and Arg55 are disordered according to NMR data. As mentioned before, the solvent-exposed Arg5 is one of two residues studied here (together with Arg24) which stabilize the wild type electrostatically. Since the nonpolar part of the side chain is exposed, removal of alkyl–aqueous contacts upon mutation to Ala is favorable in terms of van der Waals energy as well, yet the electrostatic effect dominates. In contrast, Arg55 is electrostatically unfavorable ($\Delta\Delta G_{\text{ele}} = -6.6$ kcal mol⁻¹). Biochemical data have suggested that Arg55 possibly makes direct or water-mediated hydrogen bond(s) in the minor groove. Indeed, we observe fast exchanging water molecules that bridge the guanidinium group of Arg55 with Thy11 O3'. Additionally, Arg55 participates in dynamic hydrophobic interactions with the sugar carbons of several nucleotides. This is manifested in the calculated $\Delta\Delta G_{\text{np}}$ of ~ 8 kcal mol⁻¹. Since the calculations do not explicitly take waters into account, the energetic significance of the bridging water molecule(s) is not clear. However, the good correspondence between calculations ($\Delta\Delta G^{\text{bind}} = 1.9$ kcal mol⁻¹) and experiment indicates that the continuum representation has captured the energetics of this site correctly. The same is true for Lys54 (loop 2), whose highly populated hydrogen bond with the DNA backbone contributes little to affinity while the nonpolar and internal energies contribute significantly. Overall, it appears that residues contacting the DNA duplex from the rather flexible loops and turn T1 employ a “simple” balance-of-forces principle: destabilizing electrostatics is opposed by favorable nonpolar contribution and vice versa. One residue stands out. Arg24 stabilizes the complex by both electrostatic and nonpolar interactions. The likely reason is that it participates

in an extended hydrogen bond network (3) and is involved in extensive hydrophobic interactions.

Our analysis emphasizes the complex thermodynamic role of charged residues at the protein–DNA interface. The alkyl portion of polar side chains contacts DNA and thereby strengthens binding through hydrophobic effects. Desolvation of charged groups opposes binding. The total electrostatic contribution is often destabilizing but depends sensitively on the local environment. Electrostatics favors association when a network of hydrogen bonds is formed or when protein–DNA hydrogen bonds are more solvent-exposed at the periphery of the binding site. Moderate or small loss of binding energy is the consequence of fine energetic tuning between opposing energetic factors.

CONCLUSION

We used an all-atom computational model to calculate the free energy of integrase–DNA interaction. The calculated net free energy as well as the energetic effects of alanine mutations is in reasonable agreement with experimental data. The combination of computational methods with experimental information and structural data provides a more general and comprehensive picture of INT-DBD association to its specific target DNA site. At room temperature and moderate salt concentration, the hydrophobic effect is the main source of binding affinity. The net gain of binding energy through nonpolar contacts is linked to deoptimization of the net electrostatic energy and is achieved at high energetic costs paid for maximizing surface complementarity by conformational adaptation. Protein–DNA hydrogen bonds are expensive in terms of electrostatics, yet confer specificity and facilitate local packing (optimal van der Waals interactions). Our results illustrate that computational free energy parsing analysis is an indispensable tool toward a deeper understanding of the thermodynamic strategies used by proteins for sequence-specific DNA binding.

ACKNOWLEDGMENT

We thank Amedeo Caflisch for advice, Terence Hale for computer support, and Hans Rudolf Bosshard for encouragement and discussions.

REFERENCES

- Connolly, K. M., Wojciak, J. M., and Clubb, R. T. (1998) *Nat. Struct. Biol.* 5, 546–550.
- Wojciak, J. M., Sarkar, D., Landy, A., and Clubb, R. T. (2002) *Proc. Natl. Acad. Sci. U.S.A.* 99, 3434–3439.
- Wojciak, J. M., Connolly, K. M., and Clubb, R. T. (1999) *Nat. Struct. Biol.* 6, 366–373.
- Milev, S., Gorfe, A. A., Karshikoff, A., Clubb, R. T., Bosshard, H. R., and Jelesarov, I. (2003) *Biochemistry* 42, 3481–3491.
- Connolly, K. M., Ilangovan, U., Wojciak, J. M., Iwahara, M., and Clubb, R. T. (2000) *J. Mol. Biol.* 300, 841–856.
- Baginski, M., Fogolari, F., and Briggs, J. M. (1997) *J. Mol. Biol.* 274, 253–267.
- Misra, V. K., Hecht, J. L., Yang, A. S., and Honig, B. (1998) *Biophys. J.* 75, 2262–2273.
- Froloff, N., Windemuth, A., and Honig, B. (1997) *Protein Sci.* 6, 1293–1301.
- Fogolari, F., Elcock, A. H., Esposito, G., Viglino, P., Briggs, J. M., and McCammon, J. A. (1997) *J. Mol. Biol.* 267, 368–381.
- Friedman, R. A., and Honig, B. (1995) *Biophys. J.* 69, 1528–1535.
- Olson, M. A. (2001) *Biophys. J.* 81, 1841–1853.
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E., III (2000) *Acc. Chem. Res.* 33, 889–897.
- Wang, W., Donini, O., Reyes, C. M., and Kollman, P. A. (2001) *Annu. Rev. Biophys. Biomol. Struct.*, 211–243.
- Kombo, D. C., Jayaram, B., McConell, K. J., and Beveridge, D. L. (2002) *Mol. Simul.* 28, 187–211.
- Sitkoff, D., Sharp, K. A., and Honig, B. (1994) *Biophys. Chem.* 51, 397–403.
- Sharp, K. A., Nicholls, A., Fine, R. F., and Honig, B. (1991) *Science* 252, 106–109.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) *J. Chem. Phys.* 79, 926–935.
- Darden, T. A., York, D. M., and Pedersen, L. J. (1993) *J. Chem. Phys.* 98, 10089–10092.
- Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) *J. Comput. Phys.* 23, 327–341.
- Brooks, B. R., Brucorleri, R. E., Olafson, B. D., David, J. S., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem.* 4, 187–217.
- Foloppe, N., and MacKerell, J. (2000) *J. Comput. Chem.* 21, 186–194.
- Nicholls, A., and Honig, B. (1991) *J. Comput. Chem.* 12, 435–445.
- Klapper, I., Hagstrom, R., Fine, R., Sharp, K., and Honig, B. (1986) *Proteins* 1, 47–59.
- Hubbard, S. J., and Thornton, J. M. (1993) NACCESS, Computer Program, Department of Biochemistry and Molecular Biology, University College, London.
- Jayaram, B., McConnell, B., Dixit, S., Das, A., and Beveridge, D. (2002) *J. Comput. Chem.* 23, 1–14.
- Huo, S., Massova, I., and Kollman, P. (2002) *J. Comput. Chem.* 23, 15–27.
- Pickett, S. D., and Sternberg, M. J. (1993) *J. Mol. Biol.* 231, 825–839.
- Yu, Y. B., Privalov, P. L., and Hodges, R. S. (2001) *Biophys. J.* 81, 1632–1642.
- Jayaram, B., McConnell, K. J., Surgit, B. D., and Beveridge, D. L. (1999) *J. Comput. Phys.* 151, 333–357.
- Misra, V. K., and Honig, B. (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92, 4691–4695.
- Jen-Jacobson, L., Engler, L. E., and Jacobson, L. A. (2000) *Structure* 8, 1915–1923.

BI026937P