Sliding of Alkylating Anticancer Drugs along the Minor Groove of DNA: New Insights on Sequence Selectivity

Attilio V. Vargiu,* Paolo Ruggerone,† Alessandra Magistrato,* and Paolo Carloni*
*SISSA-ISAS, CNR-INFM Democritos Modeling Center and Italian Institute of Technology (ITT)-SISSA unit, I-34014 Trieste, Italy; and †CNR-INFM SLACS and Dipartimento di Fisica, Università di Cagliari, I-09042 Monserrato, Italy

ABSTRACT Currently, little is known about the molecular recognition pathways between DNA-alkylating anticancer drugs and their targets despite their pharmacological relevance. In the framework of classical molecular dynamics simulations, here we use umbrella sampling to map the potential of mean force (PMF) associated with sliding along the DNA minor groove of two of these compounds. These are an indole derivative of duocarmycin (DSI) and the putative reactive form of anthramycin (anhydroanthramycin, IMI). Twenty-three configurations were considered for each drug/DNA complex, corresponding to a movement along ~3 basepairs. The alkylation site turns out to be the most favorable for DSI, while a barrier of ~6 kcal/mol separates the reactive configuration of IMI·DNA from the absolute minimum. An analysis of various contributions to the PMF reveals that solvent effects play an important role for the largest and more flexible drug DSI. Instead, the PMF of IMI·DNA overall correlates with changes in the binding enthalpy. Implications of these results on the sequence selectivity of the two drugs are discussed.

INTRODUCTION

Most of the successful anticancer drugs targeting DNA are organic molecules which form noncovalent or covalent interactions in the minor groove with different sequence selectivity (1,2). To rationally design such drugs, both the structure and the microscopic mechanisms underlying drug-target interactions should be known. Furthermore, it is very important to design sequence-selective anticancer molecules (hopefully competitive with regulatory proteins) that bind to regions of DNA involved in replication and transcription processes (3–6).

In this respect, a number of noncovalent binders were well characterized, and sequence-specific readout codes were partly deciphered (1,7). In contrast, the important class of DNA-alkylating agents has received less attention, particularly from a theoretical point of view, partly because of the paucity of experimentally determined structural data (8). Binding of these drugs is believed to occur in two steps (9): the formation of the noncovalent adduct $D+T \rightarrow_{K_b} D \cdot T$, followed by the covalent linkage of the drug to the DNA $D \cdot T \rightarrow_{k_r} D-T$. The design of new covalent binders requires identification and characterization of the transition state associated with the rate-determining step, since its subtle tuning affects rate and routes of interactions. Both these factors influence the reactivity and consequently the efficacy of the drug (10).

It is commonly accepted that molecular recognition and formation of the noncovalent complex are driven by nonspecific interactions and sequence-specific structural features along the minor groove (8). Recently, molecular dynamics (MD) studies by us (11,12) have provided information at the molecular level on the noncovalent interaction between the DNA and two covalent binders, anthramycin (13–16) and duocarmycin (17,18). Anthramycin alkylates guanines, showing a modest sequence selectivity for PuG*Pu sequences (16), whereas duocarmycin binds to adenines and is very selective toward AT-rich sequences which have to be at least 4 basepairs (bp) long (17,18). Both anthramycin and duocarmycin are powerful cytotoxic agents interfering with transcription and replication processes, and some of their derivatives have entered clinical test phases (19,20). The natural twist of 35° between phenol and pyrrol rings (respectively A and C in Chart 1) gives to anthramycin the ideal shape to fit into the minor groove (14). In contrast, the largest duocarmycin is formed by two moieties connected via an amide link (Chart 1), and to fit into the minor groove it needs a 40° twist around the amide link relative to its conformation in water (18).

For the purposes of this work, we focus on the putative reactive imine form of anthramycin (hereafter IMI, Chart 1) (13–16) and on duocarmycin-SI (hereafter DSI, Chart 1) (17,18). Noncovalent complexes of both IMI and DSI with DNA are mainly stabilized through hydrophobic interactions (11,12). However, although IMI forms a relatively strong H-bond network with DNA (12,14), only one H-bond is formed between DSI and DNA (11,18). In a previous MD simulation performed in our group (12), IMI turned out to slide along the minor groove of a $d[GCCAACGTTGGC]_2$ duplex, leading to the formation of a nonreactive stable complex. Such a sliding is independent on the initial location of the drug, occurring when IMI sits either at the end or in the middle of a DNA duplex. The same kind of displacement has been observed after docking the molecule to its preferred site, the triplet AGA (16), within a 14-mer duplex (data not shown). Instead, the complex between DSI and the duplex $d[GACTAATTGAC]_2$ is stable during the whole dynamics (11). It is worthwhile to notice that a similar shuffling

Submitted May 21, 2007, and accepted for publication August 29, 2007.
Address reprint requests to Paolo Carloni, E-mail: carloni@sissa.it.
Editor: Tamar Schlick.

0006-3495/08/01/550/12 $2.00
doi: 10.1529/biophysj.107.113308

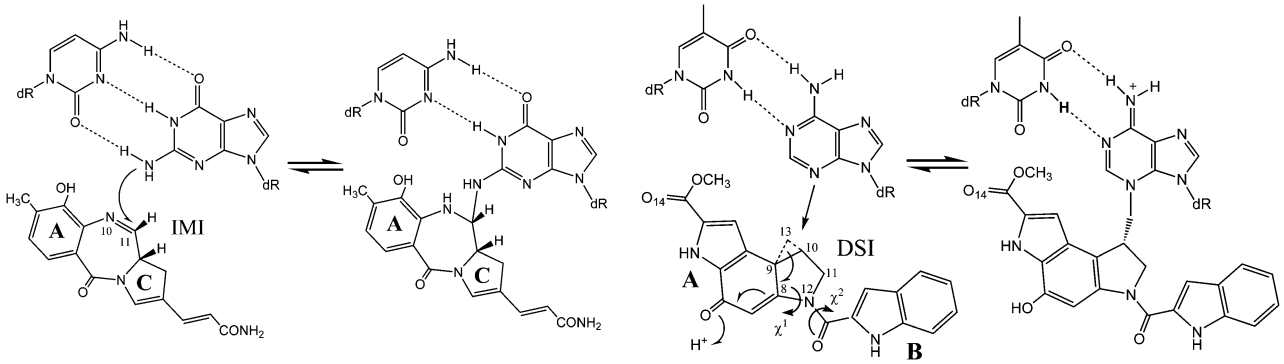

CHART 1 Schematic view of alkylation of guanine by IMI (*left*) and adenine by DSI (*right*).

mechanism was suggested by footprinting experiments for a photoactive derivative of actinomycin, while noncovalent translocation mechanism has been invoked for the covalent binder CC-1065 (21,22).

As the noncovalent recognition of the preferred DNA sequence may involve a sliding of drugs along the minor groove, knowledge of the energetics of such processes may provide useful information on ligand selectivity and molecular recognition. Here we address this issue by investigating quantitatively the mechanism of sliding of IMI and DSI. Because of the increasing reliability of methods for simulating DNA (23–28), nowadays it has become possible to extract free energies of drug/DNA recognition from MD simulations (29,30). The free energy is then dissected at a qualitative level into its enthalpic and entropic contributions, the rank of which has been shown to depend strongly on the chemical structure of the drug and on the DNA sequence (31,32).

We find that for IMI, binding to the reactive site is less favored than at nearby bp, whereas for DSI the noncovalent binding site coincides with the reactive one, in agreement with previous MD simulations (11,12). Although the potential of mean force (PMF) associated to the sliding of IMI can be roughly rationalized through a simple analysis of drug/DNA enthalpic interactions, solvation effects appear to be much more relevant for DSI as a consequence of more stringent requirements for optimal fit of drug into the minor groove. Our findings suggest the need to consider multiple binding pathways in drug design and provide a rationale for the modest selectivity of anthramycin relative to duocarmycin (18,33).

SYSTEMS AND METHODS

Free energy profiles were calculated as a function of the position of the drug along the minor groove for the noncovalent complexes of IMI with d(GCCAACGTTG*GC)-d(GCCAACGTTGGC) (hereafter **IMI·DNA**) and DSI with d(GACTAATTGAC)-d(GTCAATTA*GTC) (hereafter **DSI·DNA**). These are built starting from experimental structures (14,18) by cutting the covalent bond between the carbon of the drug and the nitrogen of the nucleobase and manually pulling out the drugs from the minor groove until the distance d[C-N] was ~3.3 Å. As DSI moves toward the nearest end of the duplex upon sliding, test calculations were performed to investigate the possible influence of end effects (12). To this end, we constructed an additional 14-mer d(GACGACTAATTGAC)-d(GTCAATTA*GTCGTC) (hereafter **DSI·DNAc**), and DSI was placed with its reactive carbon C13 in front of the base T24 (corresponding to T21 in **DSI·DNA**). A 7 ns MD simulation was performed on this system, and analyses were performed on the last 2 ns.

All simulations were carried out using the GROMACS package (34–36). AMBER/gaff force fields (25,26,37,38) were used for the parameterization of oligonucleotides and drugs (see Spiegel et al. (11) and Vargiu et al. (12) and Supplementary Material Tables S1 and S2). Drug structures were optimized by means of DFT calculations at the B3LYP/6-31G(d,p) (39,40) level, using the Gaussian03 package (41). Atomic restrained electrostatic potential (RESP) charges (42) were derived using the *resp* module of AMBER after wave function relaxation. Potassium ions, modeled with the AMBER-adapted Aqvist potential (43), were added to achieve charge neutrality (22, 20, and 28 in **IMI·DNA** and **DSI·DNA** and **DSI·DNAc**, respectively). Systems were solvated with a cubic box of TIP3P water molecules (44), ensuring that the solvent shell would extend for at least 12 Å around the DNA. Periodic boundary conditions were used, and constant temperature-pressure ($T = 300$ K, $P = 1$ atm) dynamics were performed through the Nosé-Hoover (45,46) and Andersen-Parrinello-Rahman (47,48) coupling schemes ($\tau = 1$ ps). Electrostatic interactions were treated using the particle mesh Ewald algorithm (49) with a real space cutoff of 10 Å, the same as for van der Waals interactions. The pair list was updated every 10 steps, and Lincs constraints (50) were applied to all bonds involving hydrogen atoms, allowing us to use a time step of 2 fs. Coordinates were saved each 500 steps, corresponding to 1,000 snapshots per ns. The DNA minor groove width was defined as the distance between sugar C4′ atoms, and it was calculated with the program Curves (51–53).

The PMFs (54) associated to drug sliding along the minor groove of **IMI·DNA** and **DSI·DNA** were calculated with the umbrella sampling method (55). The distance between the reactive atoms of drug and DNA base (specifically the C11 atom of IMI to the N2 atom of guanine 10 in the DNA and atom C13 of DSI to atom N3 in adenine 19 of the DNA) was chosen as the reaction coordinate. This simple choice is well suited to describe movements of the drugs that are a few bps long, since both compounds cause no appreciable bending of DNA (11,12). To sample the various conformations corresponding to different positions of the drug along the minor groove, we imposed a harmonic constraint of 15 kcal/(mol·Å^2) to the distance d[C-N] from 2.9 Å to 12.1 Å with a step of 0.4 Å. Doing so the total movement of 9.2 Å (slightly less than 3 bps steps) is partitioned into 23 windows (see movies in the Supplementary Material). Initial configurations of each window were generated starting from the reactive one and increasing the distance d[C-N] by steps of 0.2 Å, after 70 ps of equilibration for each d[C-N] value. The weighted histogram analysis method (56) was used to recombine PMF obtained from different windows. As can be seen from insets in Figs. 1 and 2, the error on the calculated PMFs is almost constant, meaning that the reaction coordinate was sampled in a fairly uniform manner.

Extended MD simulations of oligonucleotides in water (57) indicate that some of the DNA conformational and helicoidal parameters have relaxation times of ~0.5 ns. Thus, at least multi-ns trajectories must be collected to obtain well-converged free energy profiles (58,59). Systems investigated

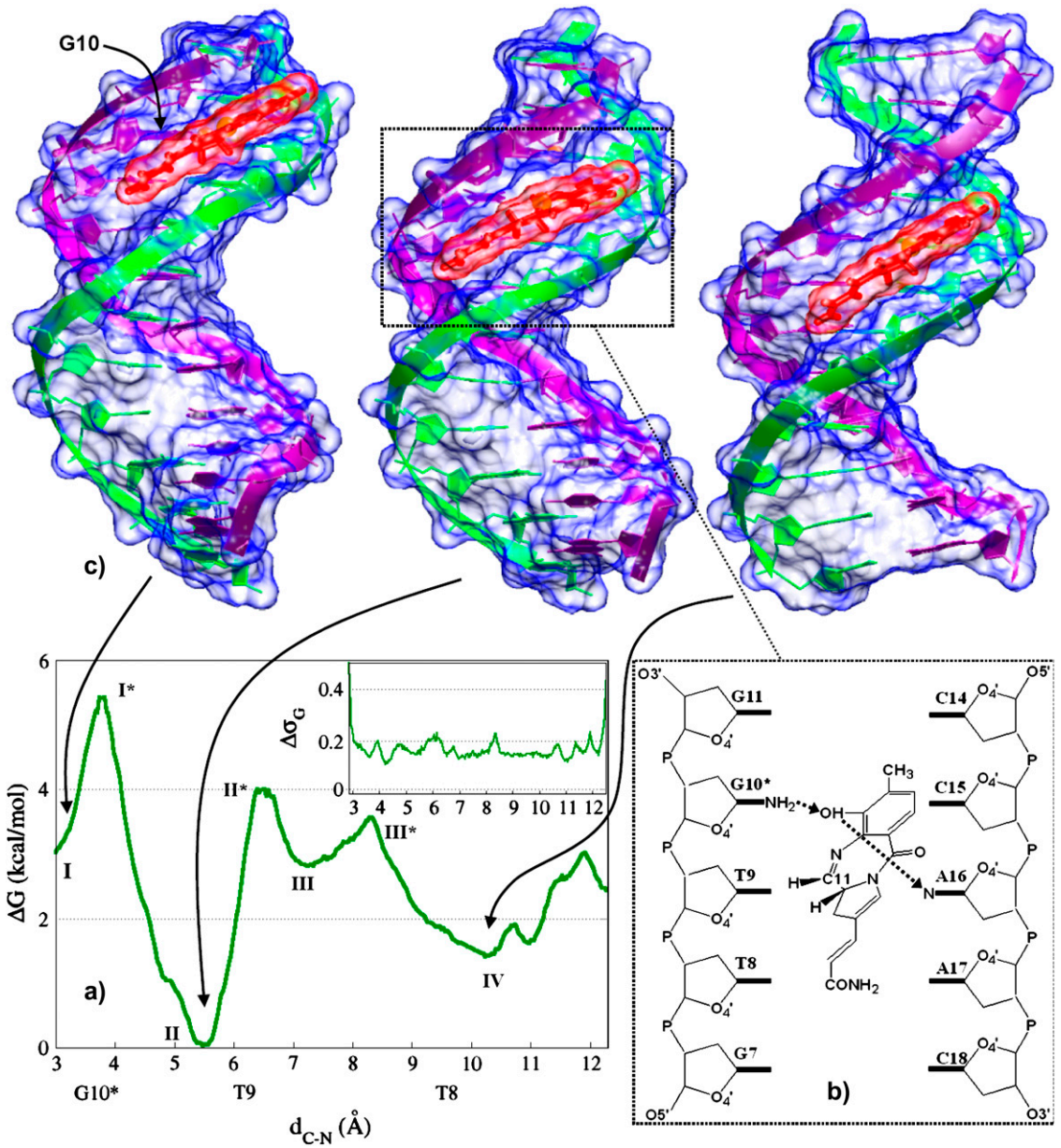

FIGURE 1 (*a*) PMF associated to IMI sliding along the DNA minor groove. Standard deviation is reported in the inset. (*b*) Schematic view of the complex in the configuration corresponding to the free energy minimum. (*c*) Structures of three relevant conformations of the drug/DNA complex. The second and third structures (from the *left*) were rotated respectively by ~30° and ~100° about the helical axis to center the drug for visualization. Molecular surfaces of drug and DNA are depicted in transparent red and blue, respectively. Figures created with the program VMD (76).

here needed almost 4 ns of equilibration phase, although the PMF landscape roughly converged after 2 ns (see Supplementary Material Fig. S1 for the profile of **DSI·DNA**; similar results are found for **IMI·DNA**). A further check on the RMSD of **drug·DNA** complexes indicates that convergence was reached for all of the windows after ~4 ns (Supplementary Material Fig. 2S, *A* and *B*). Thus, we performed 6 ns of MD simulation on each window (for a total of ~140 ns), and we used the last 2 ns for extracting the PMF and for the structural analysis of the complexes.

The PMF was decomposed as the sum of individual components, each with a physical meaning and evaluated via the force field terms. In particular, we examined variations in the drug-DNA electrostatic and van der Waals interactions, solute adaptation energy, solute configurational entropies, and solvation free energies. Energetic contributions were evaluated as MD averages, using terms in the force field (25,26,38). Van der Waals interactions were roughly estimated by summing the number of hydrophobic contacts for each atom-type pair. For each of these, the equilibrium distance d_0 of the Lennard-Jones potential was estimated (Supplementary Material Table S3, *A* and *B*), and we considered a contact if $d < (d_0 + 0.2$ Å). Solute vibrational entropies were calculated within the harmonic approximation (60), using the *nmode* module of AMBER 9 (61). As customary (62–64), we selected a subset of structures for this analysis (20 for each relevant configuration, extracted every 0.1 ns from the last 2 ns). These were minimized in the absence of solvent, using instead a dielectric constant $\varepsilon = 4r$ (r is the interatomic distance in Å) to mimic solvent effects. Then, up to 20,000 steps of minimization with no cutoff for all the interactions were performed, of which the first 100 are steepest descent followed by conjugate gradient, until the RMS of gradient

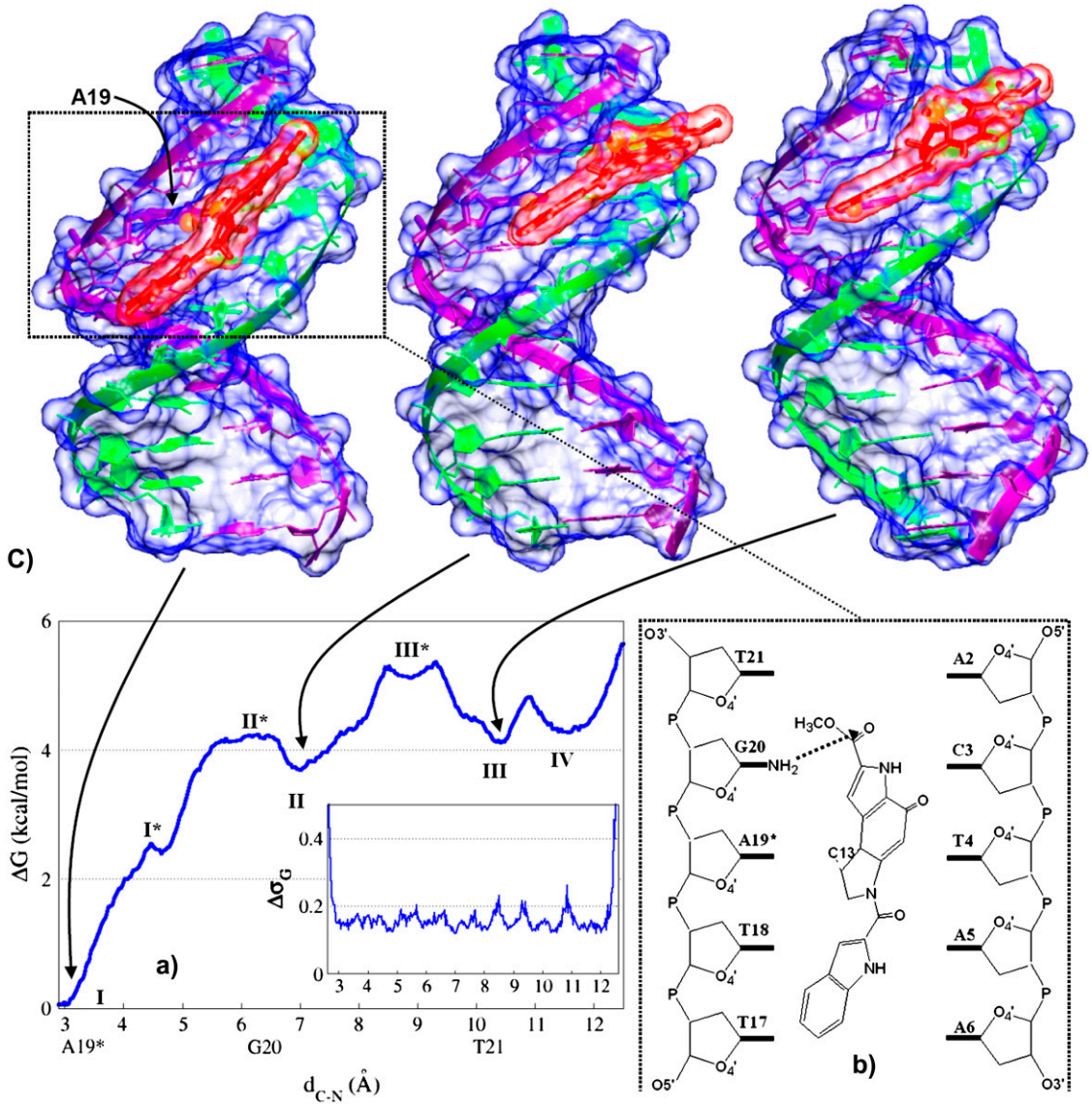

FIGURE 2 (*a*) PMF associated to DSI sliding along the DNA minor groove. Standard deviation is reported in the inset. (*b*) Schematic view of the configuration corresponding to the free energy minimum. (*c*) Structures of three relevant conformations of the complex. The third structure (from the *left*) was rotated ~40° about the helical axis to center the drug for visualization. Molecular surfaces of drug and DNA are depicted in transparent red and blue, respectively.

drops below 10^{-4} kcal/(mol·Å). The estimate of relative solute conformational entropies is not a trivial task (63,65), in particular for very flexible systems such as oligonucleotides. Moreover, it is not possible to guess a priori how the ''spreading'' of the dynamics among the various basins of the free energy surface depends on the conformation (i.e., the position of the drug). Note that also using the quasiharmonic model relative entropies are sensitive to the number of independent snapshots considered (63). Therefore, results on conformational entropies should be taken only as suggestive.

Solvent contributions to free energy were calculated using the molecular mechanics-Poisson Boltzmann surface area (MM-PBSA) methodology (62,63); for each relevant configuration we saved 200 snapshots from the last 2 ns of MD runs. Since $\Delta G_{\text{solv}} = G_{\text{solv}}^{\text{complex}} - G_{\text{solv}}^{\text{drug}} - G_{\text{solv}}^{\text{DNA}}$ and the last two terms in the right-hand side are the same for each configuration of the complex, ΔG_{solv} is directly proportional to $G_{\text{solv}}^{\text{complex}}$. Based on this proportionality, only this latter term was evaluated for each relevant configuration of the two complexes. The electrostatic contribution to solvation was evaluated using the Poisson-Boltzmann continuum method (66), as implemented in the module *pbsa* (67) of AMBER 9. We set the values of internal and external dielectric constants to 1 and 80, respectively, the grid twice as long as the linear dimensions of the solute, and a grid spacing of 0.25 Å. The dielectric boundary is the molecular surface defined by a 1.4 Å probe sphere and by spheres centered on each atom with radii taken from the PARSE (68) parameter set ($H = 1.0$, $C = 1.7$, $N = 1.5$, and $O = 1.4$ Å, with a value of 2.0 Å for the phosphorus). The boundary dielectric constants were set as the harmonic sum of solvent and solute Debye-Huckel values. Salt effects were not included implicitly in the continuum model. The hydrophobic component of solvation free energy is assumed proportional to the change of the solvent accessible surface area (SASA), $\Delta G_{\text{np}} = \gamma \Delta\text{SASA} + b$ where $\gamma = 0.00542$ kcal/Å^2 and $b = 0.92$ kcal/mol (68). For comparison, we performed MM-PBSA calculations also using a new approach available in the *pbsa*

module of the AMBER 9 package (61,67,69). In this method the nonpolar contribution is cast into two terms, a repulsive one (cavity), correlated to the SASA, and an attractive one (dispersion), calculated through a surface-integration approach (70).

RESULTS AND DISCUSSION

Here we first provide a description of the PMF profiles corresponding to the sliding of IMI and DSI inside the minor groove of oligonucleotides d(GCCAACGTTG*GC)-d(GCCAACGTTGGC) and d(GACTAATTGAC)-d(GTCAATTA*GTC) (**IMI·DNA** and **DSI·DNA**) respectively, considering d[C-N] as the reaction coordinate (see Systems and Methods). Note that the standard deviation is very small along the entire sampling interval (≈0.2 kcal/mol, see *insets* in Figs. 1 and 2), which makes the values of the free energy extracted from our simulations quantitatively reliable. We then analyze, at the qualitative level, the enthalpic and entropic contributions to the PMF from the solute and the effects of the solvent. We close this section by assessing the relevance of end effects (12) on the energetics of the sliding process.

PMF profiles

The PMF of **IMI·DNA** features four minima (I–IV, Fig. 1). In I, the complex is in its reactive configuration, i.e., carbon C11@IMI faces N2@G10 (d[C-N] ≈ 3 Å). This corresponds to the average structure assumed by the complex during the first 10 ns of a 20 ns long MD simulation (12). From I, a small barrier of 1.5 kcal/mol has to be overcome at transition state I* to reach the absolute minimum II, where d[C-N] ≈ 5.6 Å. In II, C11@IMI sits slightly before the T9 plane (along the direction $3' \rightarrow 5'$), and IMI sits at the same location as in the last 10 ns of a 20 ns long MD simulation (12). Interestingly, the barrier from II to I (5.5 kcal/mol) is much larger than that from I to II, consistently with Vargiu et al. (12), which suggests that IMI is stable in II after its departure from I. At larger d[C-N] distances, we find III (7.2 Å) and IV (10.1 Å), in which C11@IMI is located immediately beyond T9 and in front of O2@T8, respectively.

The PMF landscape of **DSI·DNA** is remarkably different (Fig. 2). Indeed, sliding is hindered by a barrier of ~4 kcal/mol which traps the drug in the reactive configuration I (corresponding to the absolute minimum in the PMF, along the investigated path). This minimum is followed by three other ones (higher in energy by ~4 kcal/mol) virtually isoenergetic. In such minima, the reactive group of the drug (in particular C13@DSI) is in front of G20 (II, a very shallow minimum), T21 (III), and between T21 and C22 (IV).

The PMF of **IMI·DNA** appears to be much rougher than that of **DSI·DNA** and, in particular, features two configurations corresponding to a larger gain in binding free energy than in the reactive conformation. Results obtained for the two adducts are in line with those extracted from previous MD simulations (11,12) and may partly explain the higher selectivity of duocarmycins as compared to anthramycins (17,33). For instance, in **IMI·DNA** the reactive configuration should be markedly less populated with respect to II (and competitive with III and IV). In fact, within Arrhenius's theory, and assuming the same prefactor at configurations I and II, the former will be $e^{(3/0.6)} \approx 150$ times less likely. In contrast, in **DSI·DNA** there is a definite larger gain in the binding free energy at the reactive configuration, with an almost flat PMF landscape elsewhere.

Dissection of the PMFs

The PMF can be interpreted as the change in binding free energy ΔG_b upon drug sliding. To identify the relevant contributions to this process, we can express ΔG_b changes as a sum of individual terms (5,30,71,72). Here we decompose ΔG_b in a term due to solute conformation and interactions between the two moieties (solute terms) and a second one due to the presence of water and counterions (solvent effects):

$$\Delta G_b = \Delta G_{solute} + \Delta G_{solv}, \quad (1)$$

where

$$\Delta G_{solute} = \Delta H_{adapt} + \Delta H_{el} + \Delta H_{vdW} - T\Delta S_{vib} - T\Delta S_{r+t} \quad (2)$$

$$\Delta G_{solv} = \Delta G_{solv,p} + \Delta G_{solv,np}. \quad (3)$$

In Eq. 2, ΔH_{adapt} represents the enthalpic term due to DNA and drug structural deformations upon binding, and ΔH_{el} and ΔH_{vdW} are contributions from electrostatic and van der Waals interactions, respectively. We approximately estimated ΔH_{adapt} and ΔH_{el} using terms of the AMBER force field (25,26,37,38), whereas ΔH_{vdW} was assumed to be roughly proportional to the variation in the number of hydrophobic contacts (see Systems and Methods and Supplementary Material Table S3). $T\Delta S_{vib}$ and $T\Delta S_{r+t}$ are the free energy contributions due to vibrational and translational+rotational entropy changes upon binding. The former was evaluated through normal mode analysis (60), whereas $T\Delta S_{r+t}$ was assumed to be sequence independent (32). In Eq. 3, $\Delta G_{solv,p}$ and $\Delta G_{solv,np}$ are respectively the polar (electrostatic) and nonpolar (hydrophobic) contributions to solvation, evaluated here with the MM-PBSA method (see Systems and Methods). We point out that although the PMFs are quantitatively reliable, their dissection into enthalpic, entropic, and solvation terms was carried out using approximated and/or strongly sampling-dependent methods. Thus, these contributions have to be considered qualitatively, and are here used to gain insights into the sources of drug selectivity. In the following, all the values we discuss refer to the reactive configuration in each complex, which is indicated as I.

Solute terms

First, we report the analysis of enthalpic terms for **IMI·DNA**, summarized in Table 1 and in Fig. 3, *a–c*, for ΔH_{adapt}, ΔH_{el}, and number of contacts, respectively. As shown in Vargiu et al. (12), at I the DNA is distorted in the central C_6G_7 tract

TABLE 1 Selected values of solute ΔH_{adapt} (column II), drug/DNA ΔH_{el} (column III), number of hydrophobic contacts (column IV), average number of H-bonds (column V), polar (column VI), and nonpolar (column VII) contributions to ΔG_{solv}

	$d_{C\text{-}N}$ (Å)	ΔH_{adapt}	ΔH_{el}	No. contacts	No. H-bonds	$\Delta G_{solv,p}$	$\Delta G_{solv,np}$
IMI·DNA	3.2 (I)	0.0(22.7)	0.0(1.6)	15(4)	2.0(0.8)	0.0(34.2)	0.0(2.0)
	3.9 (I*)	+9.6(22.3)	+3.3(1.3)	12(3)	0.9(0.6)	−38.7(50.2)	−2.1(2.4)
	5.6 (II)	−7.0(21.6)	−2.4(1.1)	19(3)	1.2(0.6)	−8.8(47.3)	−4.6(1.8)
	6.5 (II*)	−3.8(21.6)	−1.0(1.8)	13(3)	0.6(0.5)	−17.5(35.4)	+2.0(2.2)
	7.2 (III)	−4.9(22.2)	+1.2(1.1)	11(3)	1.0(0.2)	+11.4(36.7)	+4.2(2.4)
	8.2 (III*)	−1.4(21.8)	+2.2(1.3)	9(2)	0.5(0.3)	−16.4(31.6)	+4.8(1.7)
	10.1 (IV)	−1.5(22.4)	−1.4(1.0)	15(3)	0.5(0.5)	−1.5(53.2)	−2.1(3.0)
DSI·DNA	3.0 (I)	0.0(18.1)	0.0(1.9)	63(10)	0.5(0.3)	0.0(36.0)	0.0(2.0)
	4.3 (I*)	+18.3(18.2)	+6.3(2.0)	47(9)	0.0(0.0)	−60.2(45.6)	+7.4(2.6)
	6.0 (II*)	+18.6(18.1)	+3.6(1.7)	63(11)	0.0(0.0)	−41.3(29.6)	+0.3(2.8)
	7.0 (II)	+17.6(18.2)	+2.2(5.2)	53(12)	0.8(0.4)	−25.6(35.0)	+1.3(2.4)
	8.9 (III*)	+19.9(18.5)	+3.7(2.4)	56(11)	0.0(0.0)	−34.5(40.0)	+1.4(2.5)
	10.3 (III)	+15.7(18.1)	−1.2(2.2)	58(10)	0.9(0.6)	−44.4(41.4)	+3.9(2.0)
	11.9 (IV)	+14.1(18.5)	+5.3(1.5)	54(11)	0.0(0.0)	−49.5(39.1)	+2.9(2.0)

Energies are in kcal/mol and referred to the values at configuration I. Standard deviations are reported in parentheses.

with respect to its structure in bulk solvent. The adaptation energy further increases going from I to the maximum I*, whereas drug/DNA interactions weaken. Both this and the previous works indicate that the whole structure relaxes at II (indeed, the conformation assumed by the oligonucleotide is very close to that of canonical B-DNA in aqueous solution), and both electrostatic and van der Waals interactions strengthen. Actually, such interactions are the strongest along the investigated path, whereas the adaptation expense is the lowest. In contrast, the (average) number of direct H-bonds decreases. The subsequent increase in free energy at III* is partially due to the steric clash between the drug acrylamide tail and the amino group of G7. The latter hinders the sliding of the drug, causing a rotation of its principal axis with respect to the oligonucleotide and slightly exposing the molecule to the solvent (see Supplementary Material SM_II). Correspondingly, the number of contacts reaches a minimum in III*, and the minor groove widens in the tract C6...T8 (average width 8 Å with respect to 4 Å at II). Once the guanine amino group is overcome by the drug, a wide free energy minimum is reached in the proximity of T8 (IV). Here, the drug recovers some of the contacts with the DNA backbone.

The (solute) enthalpic contributions fairly correlate with the PMF (Fig. 3, *a–c*), although deviations are present, pointing to the role of solute entropic and/or solvation terms. This is particularly evident at PMF transition states; for example, the sum of the enthalpic terms indicates II* to be lower than III*, whereas in the PMF III* is slightly lower than II* (Table 1). Nevertheless, we can conclude that for **IMI·DNA** the (solute) ΔH landscape has the same overall trend shown by the PMF.

Concerning solute vibrational entropies, the calculated values at minima of the PMF turn out to scarcely differ (Supplementary Material Table S4), consistent with previous results (12). Given the limitations of the methodology used here (63), we take these values as an indication only that solute entropy does not vary significantly at different configurations.

In **DSI·DNA**, the free energy absolute minimum I, corresponding to the reactive configuration, shows the lowest ΔH_{adapt} (Fig. 3 *e* and Table 1) and the largest number of hydrophobic contacts (Fig. 3 *g* and Table 1), whereas ΔH_{el} is very similar to the values in II and III (Fig. 3 *f* and Table 1). The major source of stabilization in I is the extended pattern of hydrophobic contacts formed between the drug and the DNA, in agreement with experimental suggestions (17,18). The drug is significantly distorted when leaving the reactive configuration: the twist between the two drug moieties increases from ~20° at I to ~110° at I*, causing a significant widening of the minor groove (see SM_III). Consequently, the block A (Chart 1), along with the minor groove floor (see also next section and Fig. 3 *h*), is more solvent exposed, the electrostatic interactions weaken, and the number of drug/DNA hydrophobic contacts decreases drastically. Additionally, the H-bond which forms discontinuously at the reactive configuration is here definitively lost. At II*, both the DNA and the drug are less distorted than in I*, and the number of hydrophobic contacts is the same as in I. Surprisingly, the adaptation cost is essentially the same as in I*. II is characterized by a small gain in electrostatic and adaptation energies as compared to II* and III*, along with the recovery of one H-bond. However, it features fewer hydrophobic contacts than II* and III*. Apart from the reactive configuration, III is the lowest enthalpic minimum; however, its free energy value is almost identical to those at II and IV (Fig. 2 *a*). In summary, the (solute) enthalpic analysis shows that several points lack correspondence with the PMF, indicating a key role of entropy and/or solvent effects for the sliding process of DSI. Nevertheless, enthalpy contributions unambiguously pinpoint the configuration I as the most stable one, with ΔH_{adapt} playing a more important role than in **IMI·DNA** (max $\Delta\Delta H_{adapt} = 18$ and 7 kcal/mol for DSI and IMI, respectively).

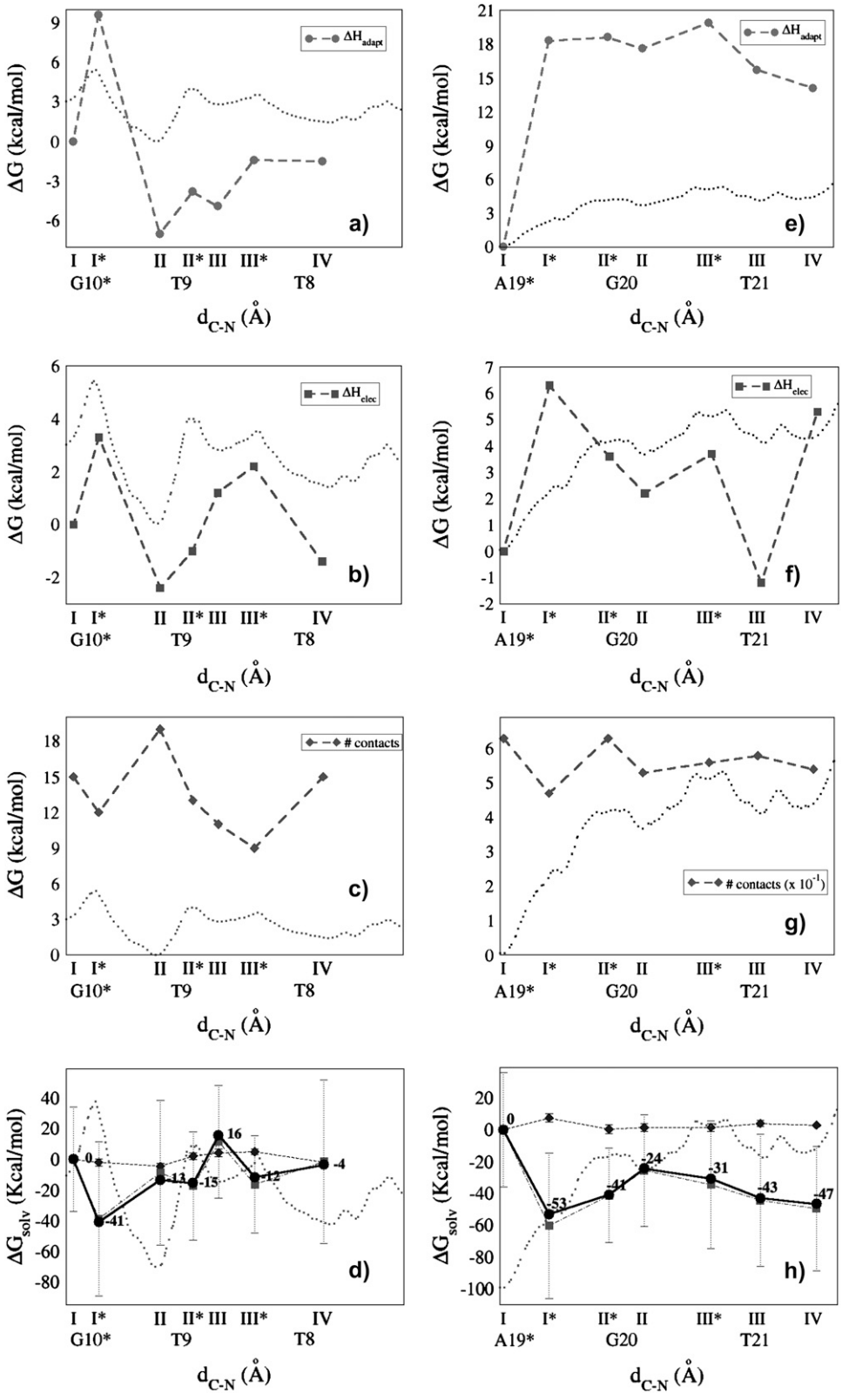

FIGURE 3 Selected values of (*a* and *e*) ΔH_{adapt}; (*b* and *f*) ΔH_{elec}; (*c* and *g*) number of hydrophobic contacts; (*d* and *h*) electrostatic (*squares*, *dotted-dashed line*), hydrophobic (*rhombus*, *dashed line*), and total (*circles*, *solid line*) ΔG_{solv} along the reaction coordinate d[C-N], compared to the PMF (*dotted line*, and rescaled in (*c*, *d*, *g*, and *h*) to allow for an easier comparison with ΔG_{solv} profiles). Data on the left (*a–d*) refer to **IMI·DNA**, those on the right (*e–h*) to **DSI·DNA**. Solvation free energies were obtained using the method of Luo and co-workers. For a comparison with data obtained using PARSE radii, see Supplementary Fig. S3.

Finally, we evaluate $T\Delta S_{vib}$ at the minima of the PMF. The calculated values show a range of variation comparable to that found for **IMI·DNA** (Supplementary Material Table S4), and we assume also in the case of DSI to neglect solute entropic contributions to the sliding process.

Solvent effects

The analysis of ΔG_{solute} carried out above clearly points to the role of solvent for the sliding of the two investigated drugs, in particular for DSI. It is indeed well known that solvent is crucial for nucleic acid structure and stability, and consequently it is important for the binding of drugs to DNA. The development of reliable implicit solvent models and fast finite difference numerical procedures (67–70,73) has made feasible the calculations of solvation free energy via MD simulations. In particular, binding free energies of ligands to DNA calculated with the MM-PBSA method agree fairly well with experimental values (29,30,64).

We evaluate here both electrostatic ($\Delta G_{solv,p}$) and hydrophobic ($\Delta G_{solv,np}$) contributions to ΔG_{solv}, which can be interpreted as the free energy cost for drug desolvation upon binding, using either the ''standard'' setup (63,68) or the alternative approach available in AMBER (61,67,69,70) (see Systems and Methods). We stress that binding is here always associated to an unfavorable—positive—change in solvation free energy. ΔG_{solv} profiles evaluated with the two methodologies are very similar, although the second approach (61,67,69,70) highlights the sequence dependence of nonpolar contributions (Figs. 3, *d* and *h*, and Supplementary Material Fig. S3). The changes in ΔG_{solv} turned out to be very large compared to the PMF and characterized by large standard deviations (Fig. 3, *d* and *h*), as reported by other authors (30,64). For these reasons, calculated ΔG_{solv} are only used for qualitative purposes. Nevertheless, some interesting differences between IMI and DSI are observed: i), The cost in desolvation upon formation of the complex ΔG_{solv} is the highest at the absolute PMF minimum in **DSI·DNA** (Fig. 3 *h*), in agreement with the better fit of the drug into the minor groove at I. Already at I*, where drug distortion results in its enhanced solvation, ΔG_{solv} suddenly decreases and reaches a rough plateau for the other configurations. This behavior mirrors the PMF landscape, where the minimum corresponding to the reactive configuration is the lowest. ii), By contrast, ΔG_{solv} does not show any clear correlation with the PMF in **IMI·DNA**. In particular, the deepest minimum in the PMF is not associated with the highest value of ΔG_{solv} (highest solvation cost). We also notice that density plots for water (SM_II and SM_III) exhibit qualitatively the same trend of the profile of ΔG_{solv}. In particular, in both **IMI·DNA** and **DSI·DNA** I* features, there is more water inside the groove with respect to I, and in **IMI·DNA** III has the largest desolvation cost, consistent with its very poor hydration (see SM_II). Interestingly, the number of hydrophobic contacts in III is lower than that of II, indicating that this number does not necessarily increase with a lower degree of hydration. This might be due, at least in part, to the ''bridging'' of the two DNA strands by the drug, which prevents water to access the groove around the binding region (see SM_II).

Within the limitations of our analysis, we conclude that solvent effects may play a key role in **DSI·DNA**, consistent with the finding that ΔG_{solute} does not correlate with the PMF. Instead, no clear correlation exists between the PMF and ΔG_{solv} in **IMI·DNA**, which is also consistent with the rough correlation found between ΔG_{solute} and the PMF.

Evaluation of end effects

In our PMF calculations, we selected as starting conformation the alkylation sites of IMI and DSI, as those are the only ones definitely visited by the drugs. These sites correspond to the G10 and A19 in **IMI·DNA** and **DSI·DNA**, respectively. (Indicating such nucleobases with the asterisk the sequence are d(GCCAACGTTG*GC)-d(GCCAACGTTGGC) and d(GACTAATTGAC)-d(GTCAATTA*GTC) for **IMI·DNA** and **DSI·DNA**, respectively.) When building up the sliding windows as described in Systems and Methods, DSI turned out to slide toward the 3′ end of the strand containing A19, whereas in **IMI·DNA** the drug moves toward the oligonucleotide center. As a result, the regions explored by the two drugs are slightly different (although it should be noticed that they are comparable, as can be seen from Figs. 1 and 2), and diverse influence of ''end effects'' (12) might hamper a thorough comparison between the calculated PMFs.

In this respect, we have already shown that end effects play only a minor role on the dynamics of **IMI·DNA** (12). Here we perform MD calculations to assess the possible influence of such effects on the interaction between DSI and DNA. Specifically, we compare the last 2 ns from MD simulations of two identical tracts **DSI·d[GACT]$_2$** either embedded in the 11-mer **DSI·DNA (DSI·d[G$_1$A$_2$C$_3$T$_4$]$_2$)** or in the 14-mer **DSI·DNAc (DSI·d[G$_4$A$_5$C$_6$T$_7$]$_2$**; see Systems and Methods).

The structure, the conformational flexibility, the hydration of **DSI·d[G$_1$A$_2$C$_3$T$_4$]$_2$**, and **DSI·d[G$_4$A$_5$C$_6$T$_7$]$_2$** turned out to be rather similar, along with the interactions between DSI and the tracts d[GACT]$_2$. In fact

i. The structures of **DSI·d[G$_1$A$_2$C$_3$T$_4$]$_2$** and **DSI·d[G$_4$A$_5$C$_6$T$_7$]$_2$** are almost superimposable (Supplementary Material Fig. S4), and the calculated RMSD is consistently 0.6 Å. Furthermore, the widths of minor grooves at the binding region are comparable, with or without the presence of three additional nucleobases (Supplementary Material Table S5 and Fig. S4), and, independent from the length of the duplex, DSI does not fit perfectly into the minor groove. In consequence of these structural similarities in both **DSI·DNA** and **DSI·DNAc**, the methoxyl-carbonyl ester (Chart 1) of the drug is fully solvated.

ii. The number of hydrophobic contacts and H-bonds between DSI and d[GACT]$_2$ is almost the same (Supplementary Material Table S6).
iii. The DSI-O$_{wat}$ and DSI-H$_{wat}$ radial distribution functions, which provide information on the solvation of the drug, are rather similar too (Supplementary Material Fig. S5). Consistently, the first and second solvation shells around the drug contain virtually the same number of waters (Supplementary Material Table S5). This is not unexpected due to the excellent superimposition between the two tracts (Supplementary Material Fig. S4) and the comparable hydration of the DSI methoxyl-carbonyl ester. In particular, O14$_{DSI}$ (Chart 1), which in **DSI·DNA** is the drug acceptor (pointing to the minor groove) closest to the duplex end, forms in both cases one H-bond with similar lifetimes (Supplementary Material Table S8). Interestingly, the tracts d[G$_1$A$_2$]$_2$ and d[G$_4$A$_5$]$_2$ also do not show relevant differences in their H-bond patterns (Supplementary Material Table S8). Finally, the density plots of waters extracted from MD simulations are rather similar too (Supplementary Material Fig. S6).
iv. The values of the root mean-square fluctuations do not differ significantly, apart from the terminal sugar moieties, which are obviously more flexible in **DSI·DNA** (Supplementary Material Fig. S7).
v. The electrostatic energy between DSI and tracts d[GACT]$_2$ is exactly the same in both systems (Supplementary Material Table S6).

Summarizing, end effects seem to play a minor role on the interactions between the two drugs and the DNA duplexes, even though they do slightly influence the structure of water within the minor groove. In conclusion, these findings confirm that steric hindrance and optimal interactions with the groove have a prominent role in determining the preferred binding sequence of duocarmycin.

CONCLUSION

The PMFs associated with the sliding of IMI and DSI along the DNA minor groove differ remarkably (Figs. 1 and 2). The reactive configuration is the most favorable for **DSI·DNA**, whereas in **IMI·DNA** significant activation energy is required to reach the reactive site moving from the absolute PMF minimum (corresponding to a nonreactive configuration). Results are consistent with previous MD simulations of **IMI·DNA** (12), showing that the reactive configuration becomes unstable after a few ns and with those of **DSI·DNA** (11), in which the drug is stable in the reactive configuration for the whole dynamics. Moreover, our findings indicate that the higher specificity of DSI, compared to IMI, correlates with a higher cost for moving the drug from the preferred site, at least for the sequences considered here. End effects turn out not to play a relevant role for the binding.

For both complexes, configurations associated with the absolute minima of the PMF are those with the smallest adaptation cost and the better packing (Table 1). This is consistent with the usual assumption of negligible ΔH_{adapt} upon drug binding to the preferred sequence in the minor groove (5,30,32). Apart from this common feature, the various contributions to the PMF have different relative weights for the two adducts. In **IMI·DNA** the changes in enthalpic terms correlate fairly well with the PMF, particularly at the four minima, whereas the ΔG_{solv} landscape is apparently unrelated to the free energy profile. In **DSI·DNA** we found instead a rough anticorrelation between ΔG_{solv} and the PMF profile, whereas enthalpy contributions do not have an overall correspondence with the free energy profile, except for the coincidence of absolute minima.

In summary, our calculations suggest that for alkylating agents, shape complementarity and packing are also significant factors in determining the preferred site of noncovalent binding (30,71,74,75). In addition, they give insights into the way differences in chemical structure, size, and flexibility may influence solvation and molecular recognition along the minor groove. In this respect DSI, due to its larger size relative to IMI, covers more DNA bps and needs to rotate around the linkage between groups *A* and *B* (Chart 1) to fit in the preferred sequence. The cost of such a rotation (which was supposed to be a key factor for reactivity) depends on the flexibility and structure of the target DNA sequence. Furthermore, the lack of optimal docking results in a larger drug exposure to the solvent, a feature that we do not find for **IMI·DNA**, in which drug desolvation cost is uncorrelated to sequence selectivity. This is consistent with the fact that IMI already has the right conformation to fit snugly into the minor groove at different sequences (indeed the reactive configuration does not correspond to a minimum in ΔH_{adapt}) and features a lower sequence selectivity compared to DSI.

Finally, although speculative, different molecular recognition mechanisms can be proposed for the two drugs investigated here. DSI might first bind to DNA at those sequences characterized by the lowest desolvation cost (the nonpreferred ones) and then easily slide toward the preferred site, corresponding to a funnel in the PMF. In this scenario, solvent effects could be critical for the molecular recognition of DNA by duocarmycins. In contrast, recognition of DNA by anthramycin appears to be more complicated, since, according to our results, no preferred route exists. Moreover, the reactive site can be reached only upon crossing a significant free energy barrier.

Clearly, calculations of free energy barriers to form noncovalent complexes at different sequences, also with drugs other than those considered here, are needed to achieve a general picture of the molecular recognition. However, our results help to rationalize the higher selectivity of duocarmycins compared to anthramycins and point out clearly the possibility of multiple patterns for molecular recognition.

SUPPLEMENTARY MATERIAL

To view all of the supplemental files associated with this article, visit www.biophysj.org.

The authors thank Katrin Spiegel, Stefano Piana, and Cristian Micheletti for useful discussions and Thomas E. Cheatham 3rd for his critical reading of the manuscript.

Computational resources were granted by CINECA (INFM grant) and CASPUR (SLACS collaboration). This project represents a scientific collaboration between Trieste and Cagliari units of the CNR-INFM Democritos Modeling Center. This work makes use of results produced by the Cybersar Project managed by the Consorzio COSMOLAB, a project cofunded by the Italian Ministry of University and Research (MIUR) within the Programma Operativo Nazionale 2000–2006 ''Ricerca Scientifica, Sviluppo Tecnologico, Alta Formazione per le Regioni Italiane dell'Obiettivo 1 (Campania, Calabria, Puglia, Basilicata, Sicilia, Sardegna) Asse II, Misura II.2 Società dell'Informazione, Azione a Sistemi di calcolo e simulazione ad alte prestazioni''. More information is available at http://www.cybersar.it.

REFERENCES

1. Dervan, P. B. 2001. Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.* 9:2215–2235.
2. Reddy, B. S. P., S. K. Sharma, and J. W. Lown. 2001. Recent developments in sequence selective minor groove DNA effectors. *Curr. Med. Chem.* 8:475–508.
3. Chaires, J. B. 1997. Energetics of drug-DNA interactions. *Biopolymers.* 44:201–215.
4. Chaires, J. B. 1998. Drug-DNA interactions. *Curr. Opin. Struct. Biol.* 8:314–320.
5. Haq, I. 2002. Thermodynamics of drug-DNA interactions. *Arch. Biochem. Biophys.* 403:1–15.
6. Hurley, L. H. 2002. DNA and its associated processes as targets for cancer therapy. *Nat. Rev. Cancer.* 2:188–200.
7. Buchmueller, K. L., A. M. Staples, C. M. Howard, S. M. Horick, P. B. Uthe, N. M. Le, K. K. Cox, B. Nguyen, K. A. O. Pacheco, W. D. Wilson, and M. Lee. 2005. Extending the language of DNA molecular recognition by polyamides: unexpected influence of imidazole and pyrrole arrangement on binding affinity and specificity. *J. Am. Chem. Soc.* 127:742–750.
8. Neidle, S. 2002. Nucleic Acid Structure and Recognition. Oxford University Press, Oxford.
9. Warpehoski, M. A., and L. H. Hurley. 1988. Sequence selectivity of DNA covalent modification. *Chem. Res. Toxicol.* 1:315–333.
10. Gregersen, B. A., X. Lopez, and D. M. York. 2004. Hybrid QM/MM study of thio effects in transphosphorylation reactions: the role of solvation. *J. Am. Chem. Soc.* 126:7504–7513.
11. Spiegel, K., U. Rothlisberger, and P. Carloni. 2006. Duocarmycins binding to DNA investigated by molecular simulation. *J. Phys. Chem. B.* 110:3647–3660.
12. Vargiu, A. V., P. Ruggerone, A. Magistrato, and P. Carloni. 2006. Anthramycin-DNA binding explored by molecular simulations. *J. Phys. Chem. B.* 110:24687–24695.
13. Teijeiro, C., E. de la Red, and D. Marín. 2000. Electrochemical analysis of anthramycin: hydrolysis, DNA-interactions and quantitative determination. *Electroanalysis* 12:963–968.
14. Kopka, M. L., D. S. Goodsell, I. Baikalov, K. Grzeskowiak, D. Cascio, and R. E. Dickerson. 1994. Crystal structure of a covalent DNA-drug adduct: anthramycin bound to C–C-A-A-C-G-T-T-G-G and a molecular explanation of specificity. *Biochemistry.* 33:13593–13610.
15. Barkley, M. D., S. Cheatham, D. E. Thurston, and L. H. Hurley. 1986. Pyrrolo[1,4]benzodiazepine antitumor antibiotics: evidence for two forms of tomaymycin bound to DNA. *Biochemistry.* 25:3021–3031.
16. Kizu, R., P. H. Draves, and L. H. Hurley. 1993. Correlation of DNA sequence specificity of anthramycin and tomaymycin with reaction kinetics and bending of DNA. *Biochemistry.* 32:8712–8722.
17. Eis, P. S., J. A. Smith, J. M. Rydzewski, D. A. Case, D. L. Boger, and W. J. Chazin. 1997. High resolution solution structure of a DNA duplex alkylated by the antitumor agent duocarmycin SA. *J. Mol. Biol.* 272:237–252.
18. Schnell, J. R., R. R. Ketchem, D. L. Boger, and W. J. Chazin. 1999. Binding-induced activation of DNA alkylation by duocarmycin SA: insights from the structure of an indole derivative-DNA adduct. *J. Am. Chem. Soc.* 121:5645–5652.
19. Alley, M. C., M. G. Hollingshead, C. M. Pacula-Cox, W. R. Waud, J. A. Hartley, P. W. Howard, S. J. Gregson, D. E. Thurston, and E. A. Sausville. 2004. SJG-136 (NSC 694501), a novel rationally designed DNA minor groove interstrand cross-linking agent with potent and broad spectrum antitumor activity: part 2: efficacy evaluations. *Cancer Res.* 64:6700–6706.
20. Small, E. J., R. Figlin, D. Petrylak, D. J. Vaughn, O. Sartor, I. Horak, R. Pincus, A. Kremer, and C. Bowden. 2000. A phase II pilot study of KW-2189 in patients with advanced renal cell carcinoma. *Invest. New Drugs.* V18:193–198.
21. Gunz, D., and H. Naegeli. 1996. A noncovalent binding-translocation mechanism for site-specific CC-1065-DNA recognition. *Biochem. Pharmacol.* 52:447–453.
22. Bailly, C., D. E. Graves, G. Ridge, and M. J. Waring. 1994. Use of a photoactive derivative of actinomycin to investigate shuffling between binding sites on DNA. *Biochemistry.* 33:8736–8745.
23. Beveridge, D. L., and K. J. McConnell. 2000. Nucleic acids: theory and computer simulation, Y2K. *Curr. Opin. Struct. Biol.* 10:182–196.
24. Cheatham 3rd, T. E. 2004. Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.* 14:360–367.
25. Cheatham 3rd, T. E., P. Cieplak, and P. A. Kollman. 1999. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* 16:845–862.
26. Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.
27. Reddy, S. Y., F. Leclerc, and M. Karplus. 2003. DNA polymorphism: a comparison of force fields for nucleic acids. *Biophys. J.* 84:1421–1449.
28. Soares, T. A., P. H. Hünenberger, M. A. Kastenholz, V. Kräutler, T. Lenz, R. D. Lins, C. Oostenbrink, and W. F. van Gunsteren. 2005. An improved nucleic acid parameter set for the GROMOS force field. *J. Comput. Chem.* 26:725–737.
29. Spackova, N., T. E. Cheatham 3rd, F. Ryjacek, F. Lankas, L. van Meervelt, P. Hobza, and J. Sponer. 2003. Molecular dynamics simulations and thermodynamics analysis of DNA-drug complexes. Minor groove binding between 4′,6-diamidino-2-phenylindole and DNA duplexes in solution. *J. Am. Chem. Soc.* 125:1759–1769.
30. Shaikh, S. A., S. R. Ahmed, and B. Jayaram. 2004. A molecular thermodynamic view of DNA-drug interactions: a case study of 25 minor-groove binders. *Arch. Biochem. Biophys.* 429:81–99.
31. Breslauer, K. J., D. P. Remeta, W.-Y. Chou, R. Ferrante, J. Curry, D. Zaunczkowski, J. G. Snyder, and L. A. Marky. 1987. Enthalpy–entropy compensations in drug-DNA binding studies. *Proc. Natl. Acad. Sci. USA.* 84:8922–8926.
32. Lah, J., and G. Vesnaver. 2004. Energetic diversity of DNA minor-groove recognition by small molecules displayed through some model ligand-DNA systems. *J. Mol. Biol.* 342:73–89.
33. Hertzberg, R. P., S. M. Hecht, V. L. Reynolds, I. J. Molineux, and L. H. Hurley. 1986. DNA sequence specificity of the pyrrolo[1,4]benzodiazepine antitumor antibiotics. Methidiumpropyl-EDTA-iron(II) footprinting analysis of DNA binding sites for anthramycin and related drugs. *Biochemistry.* 25:1249–1258.

34. Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. 1995. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43–56.

35. Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* V7:306–317.

36. van der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.

37. Wang, J., P. Cieplak, and P. A. Kollman. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 21:1049–1074.

38. Wang, J., R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. 2004. Development and testing of a general amber force field. *J. Comput. Chem.* 25:1157–1174.

39. Becke, A. D. 1988. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A.* 38:3098–3100.

40. Lee, C., W. Yang, and R. G. Parr. 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B.* 37:785–789.

41. Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. 2004. Gaussian 03. Revision C.02 ed. Gaussian, Wallingford, CT.

42. Bayly, C. I., P. Cieplak, W. Cornell, and P. A. Kollman. 1993. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 97:10269–10280.

43. Aaqvist, J. 1990. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* 94:8021–8024.

44. Jorgensen, W. L. 1981. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.* 103: 335–340.

45. Hoover, W. G. 1985. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A.* 31:1695–1697.

46. Nosé, S. 1984. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 52:255–268.

47. Andersen, H. C. 1980. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* 72:2384–2393.

48. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.

49. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

50. Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.

51. Lavery, R., and H. Sklenar. 1988. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* 6:63–91.

52. Lavery, R., and H. Sklenar. 1989. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* 6: 655–677.

53. Ravishankar, G., S. Swaminathan, D. L. Beveridge, R. Lavery, and H. Sklenar. 1989. Conformational and helicoidal analysis of 30ps of molecular dynamics on the d(CGCGAATTCGCG) double helix: ''curves'', dials and windows. *J. Biomol. Struct. Dyn.* 6:669–699.

54. Kirkwood, J. G. 1935. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* 3:300–313.

55. Torrie, G. M., and J. P. Valleau. 1974. Monte Carlo free energy estimates using non-Boltzmann sampling: application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* 28:578–581.

56. Kumar, S., J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13:1011–1021.

57. Ponomarev, S. Y., K. M. Thayer, and D. L. Beveridge. 2004. Ion motions in molecular dynamics simulations on DNA. *Proc. Natl. Acad. Sci. USA.* 101:14771–14775.

58. Beveridge, D. L., G. Barreiro, K. S. Byun, D. A. Case, T. E. Cheatham 3rd, S. B. Dixit, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai, and M. A. Young. 2004. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys. J.* 87:3799–3813.

59. Haile, J. M. 1992. Molecular Dynamics Simulations: Elementary Methods. John Wiley and Sons, New York.

60. McQuarrie, D. A. 1976. Statistical Mechanics. Harper and Row, New York.

61. Case, D. A., T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross, and P. A. Kollman. 2006. AMBER 9. University of California, San Francisco.

62. Jayaram, B., D. Sprous, M. A. Young, and D. L. Beveridge. 1998. Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Am. Chem. Soc.* 120:10629–10633.

63. Srinivasan, J., T. E. Cheatham 3rd, P. Cieplak, P. A. Kollman, and D. A. Case. 1998. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* 120:9401–9409.

64. Yan, S., M. Wu, D. J. Patel, N. E. Geacintov, and S. Broyde. 2003. Simulating structural and thermodynamic properties of carcinogen-damaged DNA. *Biophys. J.* 84:2137–2148.

65. Frenkel, D., and S. Berend. 2005. Understanding Molecular Simulations: From Algorithms to Applications. Academic Press, Amsterdam.

66. Sharp, K. A., and B. Honig. 1990. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Bioeng.* 19:301–332.

67. Luo, R., L. David, and M. K. Gilson. 2002. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comput. Chem.* 23:1244–1253.

68. Sitkoff, D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98:1978–1988.

69. Tan, C., L. Yang, and R. Luo. 2006. How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? A quantitative analysis. *J. Phys. Chem. B.* 110:18680–18687.

70. Gallicchio, E., L. Y. Zhang, and R. M. Levy. 2002. The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. *J. Comput. Chem.* 23:517–529.

71. Chaires, J. B. 2006. A thermodynamic signature for drug-DNA binding mode. *Arch. Biochem. Biophys.* 453:26–31.

72. Dill, K. A. 1997. Additivity principles in biochemistry. *J. Biol. Chem.* 272:701–704.

73. Lu, Q., and R. Luo. 2003. A Poisson-Boltzmann dynamics method with nonperiodic boundary condition. *J. Chem. Phys.* 119:11035–11047.

74. Han, F., N. Taulier, and T. V. Chalikian. 2005. Association of the minor groove binding drug Hoechst 33258 with d(CGCGAATTCGCG)$_2$: volumetric, calorimetric, and spectroscopic characterizations. *Biochemistry.* 44:9785–9794.

75. Haq, I., J. E. Ladbury, B. Z. Chowdhry, T. C. Jenkins, and J. B. Chaires. 1997. Specific binding of Hoechst 33258 to the d(CGCAAATTTGCG)$_2$ duplex: calorimetric and spectroscopic studies. *J. Mol. Biol.* 271:244–257.

76. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* 14:33–38.