JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

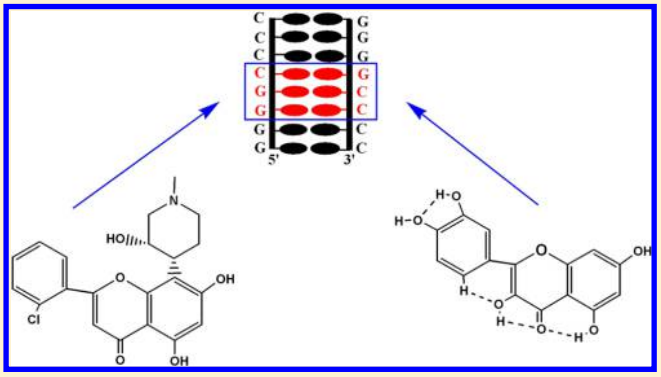
pubs.acs.org/jcim

Sequence-Dependent Binding of Flavonoids to Duplex DNA

Petar M. Mitrasinovic*

Department of Natural Sciences, Belgrade Institute of Science and Technology, 11060 Belgrade, Serbia

ABSTRACT: The whole family of structurally distinct flavonoids has been recognized as a valuable source of prospective anticancer agents. There is experimental evidence demonstrating that some flavonoids, like flavopiridol (FLP) and quercetin (QUE), bind to DNA influencing their key physiological function. FLP is involved in the combined mode of interaction (intercalation and minor groove binding), while QUE is viewed as a minor groove binder. From a physical standpoint, experimental and theoretical studies have not so far provided a sufficiently consistent picture of the nature of interaction with DNA. Herein the sequence-dependent binding of FLP and of QUE (two representative examples of the structurally different flavonoids) with duplex DNA, containing a variety of the sequences of eight nucleotides (**I**: GGGGCCCC, **II**: GGCCGGCC, **III**: AAAATTTT, **IV**: AAGCGCTT, **V**: GCGCGCGC) in the 5′-strand, is investigated using a sophisticated molecular dynamics (MD) approach. For various parts (helix, backbone, bases) of the DNA structure, the change of asymptotic (in terms of an infinite length of MD simulation) configurational entropy, being the thermodynamic consequence of DNA flexibility change due to ligand binding, is explored. As far as the sequence-dependent extent of DNA flexibility change upon QUE (or FLP) binding is concerned, for the entire double helix, increased flexibility is observed for **I** (or **I** ≈ **II**), while increased rigidity is found to be in the order of **V** > **III** > **II** > **IV** (or **III** > **V** > **IV**) for the rest of sequences. For the backbone, increased rigidity in the order of **V** > **III** > **II** > **IV** > **I** (or **III** > **V** > **IV** > **I** > **II**) is generally observed. For the nucleobases, increased flexibility is determined for **I** and **II** (**I** > **II** for both ligands), while increased rigidity in the order of **V** ≈ **III** > **IV** (or **III** > **V** > **IV**) is reported for the other sequences. Of the overall increased rigidity of the DNA structure upon ligand binding that is observed for the sequences **III**, **IV**, and **V**, about 50−70% comes from the sugar−phosphate backbone. Noteworthy is that the increased flexibility of the entire double helix and of the complete system of nucleobases upon ligand binding is only established for sequence **I**. The insights are further subtly substantiated by considering the configurational entropy contributions at the level of individual nucleobase pairs and of individual nucleo-base pair steps and by analyzing the sequence dependent estimates of intra-base pair entropy and inter-base pair entropy. The GGC triplet, which is part of the central tetramer (GGCC) of **I**, is concluded to be critical for binding of flavonoids, while the effect of the presence of ligand to the flexibility of nucleobases is localized through the intra-base pair motion of the intercalation site and its immediate vicinity. G-rich DNA sequences with consecutive Gs going before and/or after the critical GGC code (such as **I**: GGGGCCCC) are proposed to be uniquely specific for flavonoids. The configurational entropy contribution, as an upper bound of the true entropy contribution to the free energy in noncovalent binding, is demonstrated to influence the fundamental discrimination (intercalation vs groove binding) of DNA-flavonoid recognition modes. Some interesting implications for the structure-based design of optimal DNA binders are discussed.



INTRODUCTION

Flavonoids, the major constituents of plant pigments, have important antioxidant, anti-inflammatory, antiallergic, anti-cancer, and antiviral properties and are exploited as drugs and/or dietary supplements. The most common dietary sources of flavonoids are fruits, green or black tea, and soybean. Distinct structure−function relationships at various stages in cancer progression reflect different functional abilities of flavonoids, such as to control the growth of cancer cells, mediate cytotoxic effect, display potent antioxidant and anti-inflammatory activities, stabilize DNA triple helical complexes, affect tubulin polymerization, act as antiprotease compounds, etc. Thus, flavonoids are a versatile source of prospective anticancer drugs.[1]

Flavone containing 15 carbon atoms is the parent molecule out of which the whole family of structurally distinguishable flavonoids is derived by adding a different number of hydroxyl groups at the positions 3, 5, 7, 3′, 4′, and 5′ (Figure 1a). Flavonoids are mostly water soluble due to their polyphenolic nature, and their specific activity primarily depends on the arrangement of functional groups about the core aromatic structure. Therefore, the spatial arrangement of functional groups is more relevant than the scaffold itself for the functional properties of the molecules. The range of possible structural diversities is best conceivable through consideration of QUE[2]

Received: November 21, 2014
Published: January 12, 2015

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433

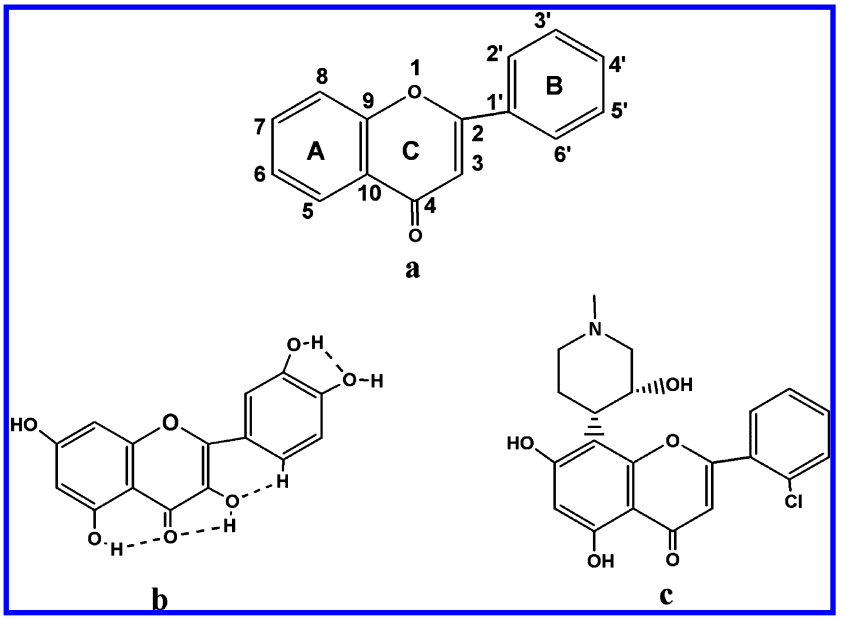


Figure 1. (a) Flavone, the parent molecule from which the whole family of structurally different flavonoids is derived. A, B, and C denote three aromatic rings,. (b) Quercetin (QUE) with its characteristic intramolecular hydrogen bonds that are denoted by dashed line. (c) Flavopiridol (FLP).

and FLP,[3,4] two representative examples of the structurally different flavonoids that have entered clinical domain as prospective anticancer agents. QUE has five hydroxyl groups attached to the carbons 3, 5, 7, 3′, and 4′, while its intramolecular hydrogen bonds are responsible for the conspicuous planarity of the QUE structure (Figure 1b). The flavone ring system of FLP consists of both the chlorophenyl portion and the heterocyclic portion with the C5 and C7 hydroxyl groups, while the C8 piperidinyl substituent (Figure 1c) structurally makes FLP more complex in comparison to QUE having a sole hydrogen at the position C8 (Figure 1b). It is important to note that in contrast to QUE (Figure 1b) FLP (Figure 1c) does not possess the C3−OH hydroxyl group, thus being more inclined than QUE to adopt a nonplanar binding geometry, as an extreme planarity of flavonoid structures was frequently found to be a consequence of the intramolecular bonding in between the oxygen of the C3−OH group and the hydrogen attached to the C6′ atom.[5]

Even though FLP was distinguished from several cytotoxic flavones that were reported to inhibit topoisomerase I, no evidence for the topoisomerase-mediated DNA effects was found.[6] In fact, the rapid inhibition of RNA synthesis and the up-regulation of p53 indicated that FLP may be affecting the integrity of DNA.[6] There is experimental evidence showing that FLP, the first potent cyclin-dependent kinase inhibitor to reach the clinic, directly interacts with duplex DNA as the second target, providing a possible explanation for the capability of the agent to kill noncycling cancer cells.[3] Nuclear magnetic resonance (NMR) studies demonstrated the ability of FLP to be nonspecifically intercalated between each adjacent set of DNA base pairs.[3] FLP was shown using high-performance liquid chromatography (HPLC) to be associated with an intermediate level of binding to all tested oligonucleotides (AAAATTTT, AATTAATT, ATATATAT, AAGCGCTT, GGGGCCCC GGCCGGCC, GCGCGCGC), and no clear specificity for any two-nucleotide sequence was observed.[3] However, the possibility that an optimal FLP-DNA binding mode requires more specific sequences of three or more nucleotides, a structural signature that is more typical for groove binders than intercalating agents, was not ruled out.[3] In the case of QUE interacting with monomeric and dimeric G-quadruplexes (G4-DNAs), absorption, fluorescence, circular dichroism (CD), and ^{1}H NMR spectroscopies[7−9] revealed that QUE is stacked with the terminal tetrads of monomeric G4-DNAs due to intercalation and is bound to dimeric G4-DNAs due to groove binding.[7] These experimental observations raise the following issue: Does the optimal binding of flavonoids to DNA involve the elements of both intercalation and groove binding?

Molecular modeling has been dealing with the question,[1,3,10] but it has not provided a sufficiently consistent answer. The nature of the interactions was investigated by molecular docking of FLP into a randomly chosen crystal structure of a duplex DNA (5′-TCGTCAAA-3′).[3] A total of eight interactions characterized the binding orientation of FLP relative to the DNA target, one-half of which was the staking interactions of the chlorophenyl and heterocyclic portions (Figure 1c) of FLP with the DNA bases, while the rest was four electrostatic interactions (three hydrogen bonds and one van der Waals interaction) involving the C8 piperidinyl substituent (Figure 1c).[3] In other words, the energetically feasible intercalation of FLP between the nucleobases is associated with the chlorophenyl and heterocyclic aromatic rings of FLP, while the piperidinyl moiety of flavopiridol stretches into the DNA minor groove through additional hydrogen bonding and van der Waals interactions. The data provided by molecular modeling essentially reconciled NMR and HPLC experimental results advocating the intercalation and groove binding conformations of FLP, respectively.[3] In addition, QUE was docked in a quadruplex DNA structure (PDB ID: 1NP9) and in a duplex DNA structure (PDB ID: 3IXN).[10] The minor groove binding between QUE and the DNA was observed, and a substantial distortion of the aromatic ring B around the C2−C1′ bond (Figure 1a) with respect to the rest of QUE structure was detected.[10] The very nonplanar conformation of QUE was stabilized by means of electrostatic interactions with the DNA residues, providing the specificity pattern GGGG for the quadruplex DNA and CGG for the duplex DNA.[1] The clear

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433

majority of Gs in the specificity patterns is in agreement with the highest preference of QUE for G relative to any other single nucleobase, as established by density functional theory (DFT) calculations.[10] The minor groove binding of QUE to the quadruplex DNA (PDB ID: 1NP9)[10] is not in agreement with the experimental result showing QUE stacked with the terminal tetrads of monomeric G-quadruplexes.[7] The clear difference may be accounted for by the use of different DNA sequences in the computational[10] and experimental[7] studies. Although useful, the computational results are limited to some extent in the sense that the results are based on the molecular docking of QUE and FLP in the particular crystal structures of DNA. By knowing that various conformations in static crystal structures may be due to differences in crystallization conditions or procedures, the dynamics of molecular systems may be additionally explored with the aim to see into the consequences of large DNA domain motions and substantial active site rearrangements.

To achieve the objective, we investigate the sequence-dependent change of flexibility of the duplex DNA upon ligand binding. We in fact explore the thermodynamic consequence of DNA flexibility change upon binding of FLP and of QUE, the change of configuartional entropy referring to the entropy of the solute, usually without translational and rotational contributions. Whereas the separation of translational entropy of the solute is mainly acceptable without considering any kind of approximation, the removal of rotational entropy is in line with assuming a negligible correlation between the internal degrees of freedom and the global rotation of the solute. Thus, the configurational entropy, estimated by way of the quasi-harmonic approximation that considers a system of atoms with many degrees of freedom as a system of uncoupled harmonic oscillators and assumes a multivariate normal distribution of the atomic fluctuations, gives an upper bound of the true entropy contribution to the free energy of noncovalent binding. Special care needs to be taken in the cases of very flexible targets (like proteins) where the quasi-harmonic approximation may be invalid.

METHODS

The 5′-DNA sequences of eight nucleotides (GGGGCCCC, GGCCGGCC, AAAATTTT, AAGCGCTT, GCGCGCGC), which were previously explained in the literature,[3] were the starting point. By using each of the 5′-sequences as an input, atomic coordinates that define the three-dimensional (3-D) models of double-stranded DNA were generated using the NAB molecular manipulation language, which facilitates the prediction of complex nucleic acid structures. The NAB tool is included as part of the Amber11 package.[11,12] Before performing MD simulations, the initial 3-D structures of the duplex DNA targets in complex with every single small ligand molecule were generated as the lowest energy poses of the flexible docking of each ligand in the DNA receptors via the program AutoDock4.2.[13,14] The Lamarckian Genetic Algorithm in combination with a grid-based energy evaluation method was employed to calculate grid maps, while atomic potential grid map was computed by AutoGrid4 with a 0.536 Å spacing in a 44 Å × 44 Å × 60 Å (1 Å = 1 × 10^{-10} m) box centered on the macromolecule. All the other default options were used by the AutoDock4 tools[14] in preparing the molecular docking runs.

For MD simulations, the molecular system under study was set up using the Amber11[11,12] program tLeap in association with the ff99sb force field.[15] Inclusion of the ligands in the simulations was done by parametrizing the atom types, charges, and connectivity of QUE and FLP. The molecular geometry was optimized by Gaussian98[16] at the MP2/6-31G* level of theory, while the molecular electrostatic potential was calculated by Gaussian98[16] at the HF/6-31G* level of theory and was used to derive the atomic charges by means of the RESP fitting technique[17] that is included as part of AmebrTools 1.5.[11,12] The rest of the parameters was assigned from the General Amber Force Field (GAFF)[18] containing the parameters for almost all the organic molecules and being entirely compatible with the ff99SB macromolecular force field.[15]

Before performing production MD, every solute was solvated using a 8.5 Å pad of about 3000 TIP3P water molecules, and the counterions Na^+ were added to neutralize each system subjected to the process of minimization and equilibration via a number of stages. At the outset, the positions of the solute atoms were kept fixed, while the positions of the water atoms were minimized by gradually reducing an initial harmonic restraint of 2 kcal mol^{-1} $Å^{-2}$ on all nonhydrogen nonwater atoms through 100 combined steepest descent and conjugate gradient minimization steps. Afterward, the entire system was minimized without restrains by means of 200 combined steepest descent and conjugate gradient minimization steps. Following minimization, a 50 ps (1 ps = 1 × 10^{-12} s) MD run was performed in order to linearly heat the water up to 300 K in the canonical NVT ensemble (constant number of particles, N; constant volume, V; constant temperature, T) using a Langevin thermostat, with a collision frequency of 2 ps^{-1} and harmonic restraints of 1 kcal mol^{-1} $Å^{-2}$. By releasing the position restraints, the entire system was heated to 300 K via a subsequent 50 ps constant-pressure simulation. To equilibrate the system, a further 500 ps long run at 300 K was conducted in the NPT ensemble, without positional restraints and with a pressure relaxation time of 2 ps. A production run was then made for 10 ns (1 ns = 1 × 10^{-9} s) duration in the NPT ensemble at 300 K. Temperature was controlled by way of a Langevin thermostat with a 2 ps^{-1} collision frequency. The time step used throughout all stages was 2 fs (1 fs = 1 × 10^{-15} s), and all hydrogen atoms were constrained using the Shake algorithm.[19] The long-range interactions were included on every step using the Particle Mesh Ewald algorithm[20] with a fourth order B-spline interpolation, a grid spacing of <1 Å, and a direct space cutoff of 12 Å. The minimizations and MD simulations were done using the Sander module of Amber11.[11,12] The production trajectories were processed using the Ptraj analysis tool included in the Amber11 program suite.[11,12]

The estimation of configurational entropy was based on the approximation of Andricioaei and Karplus (A&K),[21] treating a system of atoms with many degrees of freedom as a system of uncoupled harmonic oscillators, assuming a multivariate normal distribution of the atomic fluctuations and providing an upper bound of the true entropy (eq 1).

$$S_{A\&K} = k \sum_{i}^{3N-6} \frac{\hbar\omega_i/kT}{e^{\hbar\omega_i/kT} - 1} - \ln(1 - e^{-\hbar\omega_i/kT}) \quad (1)$$

where $\hbar$ is the reduced Planck constant, k is the Boltzmann constant, and the sum runs over all 3N-6 vibrational degrees of freedom that are associated with the harmonic oscillator frequencies (ω_i) determined by converting the nonzero eigenvalues (λ_i) of the mass-weighted covariance matrix (ω_i =

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433

$(kT/\lambda_i)^{1/2}$). The eigenvaules were obtained by means of the quasi-harmonic analysis, which included (i) a calculation of the mass-weighted covariance matrix using the Ptraj command *matrix* in combination with the keyword *mwcovar* and (ii) a diagonalization of the mass-weighted covariance matrix using the Ptraj command *analyze matrix*.[11,12] The frequencies and entropies were numerically computed by specifying the key word *thermo* on the *analyze matrix* command line. Even though another computationally less expensive formula for the estimation of configurational entropy was proposed by Schlitter,[22] in the case of DNA, both procedures[21,22] were shown to yield equal upper bounds of the true entropy with numerical precision.[23]

RESULTS AND DISCUSSION

Infinitely Long Molecular Dynamics Simulation. The configurational entropy was evaluated using a procedure[21] that takes only vibrational degrees of freedom into account, as described in the previous section. The separation of translational entropy of the solute is generally acceptable without considering any kind of approximation, while the removal of rotational entropy is justified if a negligible correlation between the internal degrees of freedom and the global rotation of the solute is assumed.[24] The entropies associated with the global translational and rotational motions were removed by means of a least-squares fitting to a reference structure. The reference for every MD run was a structure calculated as an average over all frames fitted to the first frame of the specific trajectory. Even though an average structure might generally be improved using an iterative procedure until convergence is reached, the obtained entropies are insensitive to the iterative improvement of the reference structures. The configurational entropy was computed for the variety of structural parts of the DNA double helix (Figure 2) and was based on the calculations treating only

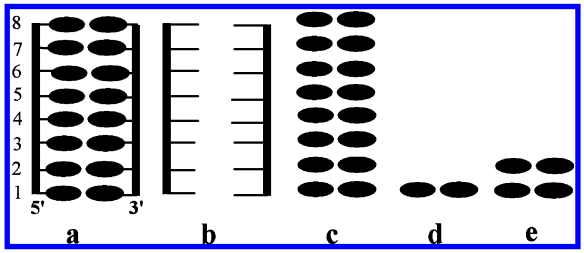


Figure 2. Various parts of the DNA structure for which the configurational entropies were evaluated: (a) helix, (b) backbone, (c) bases, (d) base pair, and (e) base pair step (Adapted from ref 25. Copyright 2010 American Chemical Society).

nonhydrogen atoms. In Figure 2, the following apply: (i) helix refers to all of the nonhydrogen atoms of the entire DNA double helix. (ii) backbone refers to the nonhydrogen atoms of the sugar and phosphate moieties of the nucleotides. (iii) bases refers to the nonhydrogen atoms of the nucleobases. (iv) base pair refers to the nonhydrogen atoms of any two hydrogen-bonded nucleobases. (v) base pair step refers to the nonhydrogen atoms of the two neighboring base pairs. Because of not considering the fast motion of hydrogen atoms by the computational protocol, some error (not substantial) may be incorporated into the absolute values of obtained entropies for the individual systems, as previously mentioned in the literature.[26,27] However, because a particular emphasis was placed on the entropy changes caused by binding of lignads to the free forms of DNA targets rather than on the absolute values, the possibility of accumulating this type of inaccuracy was circumvented. Every single MD run was performed by means of both dividing the simulation trajectory into nonoverlapping segments of 0.5, 1, 2, 4, and 10 ns and averaging the computed entropy over each respective time interval. As the computed entropy is expected to asymptotically converge with increasing the length of MD trajectory, the reported entropy values were extrapolated using the empirical equation of Harris et al.[23,28] to an infinitely long simulation (eq 2).

$$S(t) = S_{\mathrm{inf}} - (p/t^q) \tag{2}$$

where S_{inf} is the estimation of entropy for an infinite time length of MD simulation, while the physical meanings of the parameters p and q have not yet been determined.[16]

Throughout this work, specific values of the parameters were established by fitting $S(t)$ to the sets of the entropy values obtained by way of 0.5, 1, 2, 4, and 10 ns long simulations (Figure 5). It is important to note that the fitting procedure was carried out for each part of the DNA structure (Figure 2) separately, so that unwanted correlation of the particular entropies with other parts of the system was eliminated. The values of S_{inf} multiplied by 300 K for various parts of the DNA structures are reported in Tables 1 and 3. The accomplished

Table 1. *TS* (kcal mol^{-1}) for Various Parts of DNA Structure (Figure 2) at 300 K[a]

5′-DNA sequence	system	helix	backbone	bases
GGGGCCCC	DNA	608.3	330.1	249.9
	DNA:QUE	614.8	325.4	258.8
	DNA:FLP	620.1	322.4	274.1
GGCCGGCC	DNA	593.7	322.4	243.2
	DNA:QUE	573.3	305.7	244.5
	DNA:FLP	606.4	320.2	261.7
AAAATTTT	DNA	663.2	338.6	225.7
	DNA:QUE	622.2	315.2	213.2
	DNA:FLP	578.2	296.2	197.0
AAGCGCTT	DNA	644.9	337.2	241.6
	DNA:QUE	626.3	324.1	239.2
	DNA:FLP	602.2	307.6	242.0
GCGCGCGC	DNA	640.7	348.9	261.1
	DNA:QUE	576.4	302.0	248.6
	DNA:FLP	585.5	309.7	251.3

[a]Reported values are based on configurational entropy extrapolated to an infinitely long molecular dynamics simulation of molecular systems (eq 2).

convergence of entropies was quite satisfactory and was reflected by the correlation coefficients between 0.9 and 1. In the case of bare DNA, the parameter q was found to converge toward a value around 0.67, as previously reported,[28] and no significant sequence dependence of the specific value was established. In the case of the complexes DNA:QUE and DNA:FLP, the converged parameter q had slightly lower values being between 0.56 and 0.65, as generally expected.[23]

For all investigated sequences, the simulations of both ligand-free DNA and its complex with either QUE or FLP displayed quite a stable behavior within 4 ns. This trend is shown by the Root Mean Square Deviation (RMSD) calculated relative to the first frame of the trajectory (Figure 3). It is notable that the RMSD of unbound DNA is stabilized around 3 Å for

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433

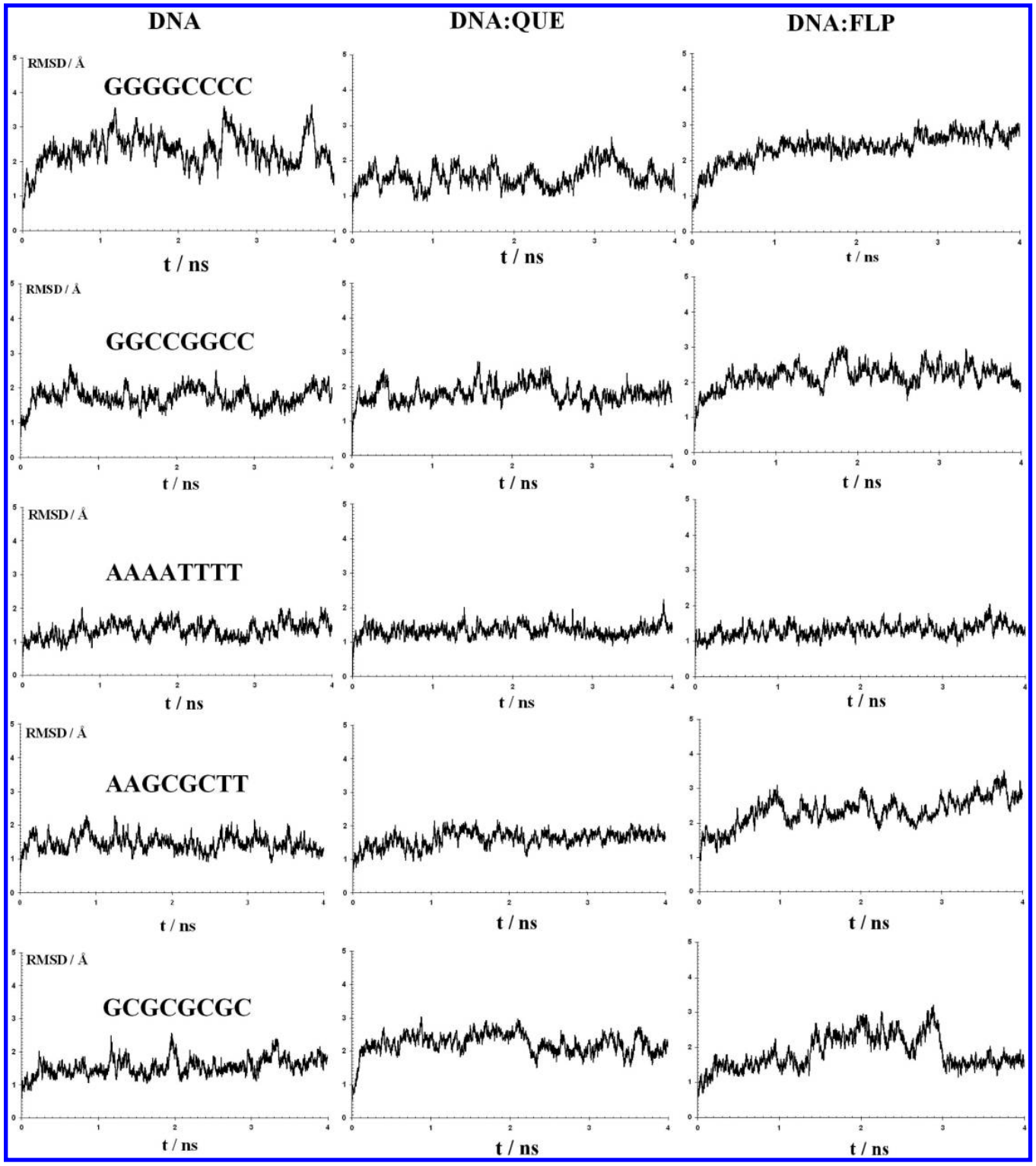


Figure 3. Root Mean Square Deviation (RMSD) of nonhydrogen atoms vs time. RMSD is calculated with reference to the first frame of the trajectory.

GGGGCCCC and about 2 Å for the other sequences (Figure 3, left column), indicating that the GGGGCCCC oligonucleotide is a more flexible target than the other considered receptors in apo (or ligand-free) conformation. Besides, a careful inspection of both the entropy values obtained for 4 ns after the outset of simulations (Figure 5) and the entropy values extrapolated to infinitely long simulations (Table 1) illustrates that 93 ± 3% of the asymptotic ($t \rightarrow \infty$) entropies was reached in a 4 ns period of time for each part of the DNA structure. The representative snapshots of the studied systems in the stable regime of simulation are given in Figure 4. While QUE is a clear minor groove binder (Figure 4b), FLP is intercalated in between the nucleobase pairs 3 and 5, thereby pulling the nucleobase pair 4 away from its usual position in the double helix (Figure 4c).

Entropy of DNA Double Helix. The configurational entropies multiplied by 300 K for five different DNA sequences

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433

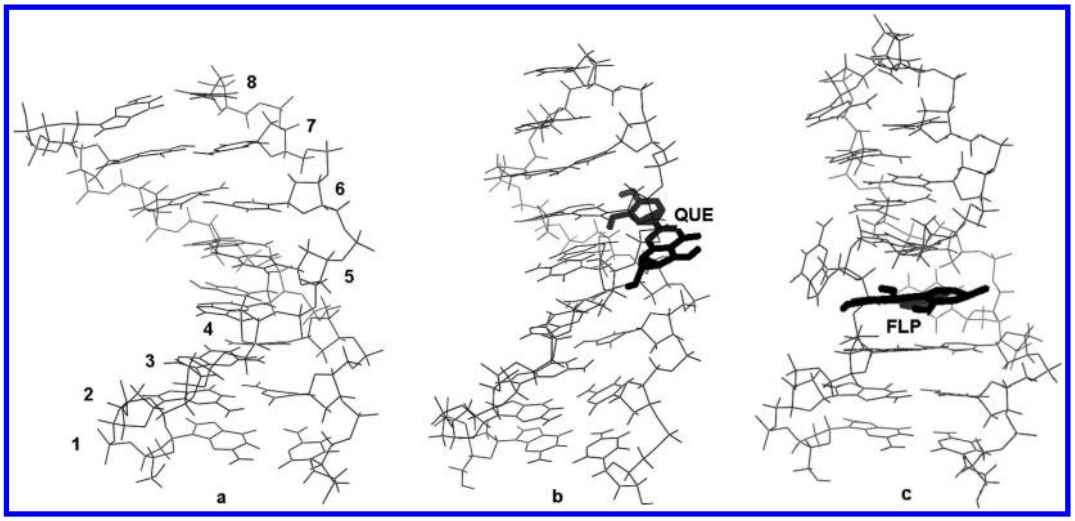


Figure 4. Representative snapshots of the investigated systems with 5′-GGGGCCCC (1-8) that are taken from the simulations in quite a stable regime around 4 ns. (a) DNA double helix. (b) DNA:QUE. (c) DNA:FLP.

are shown in Figure 5 (left column). The absolute asymptotic ($t \rightarrow \infty$) configurational entropies multiplied by 300 K are summarized in Table 1 (third column), while the entropy changes for an infinitely long simulation are given in Table 2.

The groove binding of QUE is accompanied by the configurational entropy changes in the range from −64.3 to 6.5 kcal mol^{-1} (Table 2). The increase (6.5) in configurational entropy of the DNA is only observed for GGGGCCCC, while the decrease in configurational entropy of the DNA is established for all other studied sequences (−64.3 for GCGCGCGC < −41.0 for AAAATTTT < −20.4 for GGCCGGCC < −18.6 for AAGCGCTT). In the case of FLP that is associated with some degree of intercalation, the configurational entropy contribution to the free energy varies in the range from −85.0 to 12.7 kcal mol^{-1} (Table 2). The positive values (11.8 and 12.7) are for GGGGCCCC and GGCCGGCC, while the negative values are lined up in the order −85.0 < −55.2 < −42.7 and correspond to AAAATTTT, GCGCGCGC, and AAGCGCTT respectively (from left to right).

It is notable that the configurational entropy changes of the whole DNA caused by FLP binding are mainly larger from those caused by QUE binding. Also, for GGCCGGCC, the entropy change of DNA due to FLP binding (12.7 kcal mol^{-1}) has an opposite sign with respect to that due to QUE binding (−20.4 kcal mol^{-1}). This indicates that the magnitude of flexibility changes of the DNA helix depends not only on the targeted sequence but also on the spatial orientation of the ligand toward the receptor. Moreover, the increased flexibility of the DNA helix with 5′-GGGGCCCC upon binding of both ligands is exceptional relative to the mainly increased rigidity of all other explored DNA helixes and is in contrast with the understanding that an increased rigidity of DNA helix is an exclusive structural signature of minor groove binding.[28,29]

Entropy of Backbone and Entropy of Nucleobases. The configurational entropies of the backbone and of the bases are plotted in Figure 5 (middle and right column, respectively). The absolute values of asymptotic configurational entropies are summarized in Table 1 (fourth and fifth column, respectively). The configurational entropy changes upon complex formation are reported in Table 2.

The binding of QUE caused the configurational entropy changes of the backbone in the range from −46.9 to −4.7 kcal mol^{-1} (−46.9 < −23.4 < −16.7 < −13.1 < −4.7), meaning that the increased rigidity of the backbone upon binding of the particular ligand follows the order of GCGCGCGC > AAAATTTT > GGCCGGCC > AAGCGCTT > GGGGCCCC (Table 2). The binding of FLP caused the configurational entropy changes in the range from −42.4 to −2.2 kcal mol^{-1} (−42.4 < −39.2 < −29.6 < −7.7 < −2.2), indicating (from left to right) the backbone increased rigidity order of AAAATTTT > GCGCGCGC > AAGCGCTT > GGGGCCCC > GGCCGGCC (Table 2).

In the case of QUE, the configurational entropy changes of the bases are in the range from −12.5 to 8.9 kcal mol^{-1}. The positive values (8.9 and 1.3) denote an increase of the flexibility of the bases for GGGGCCCC and GGCCGGCC, while the negative values (Table 2) for the rest of sequences are lined up in the order of −12.5 (GCGCGCGC) ≈ −12.5 (AAAATTTT) < −2.4 (AAGCGCTT). For FLP, the values are between −28.7 and 24.2 kcal mol^{-1}. The positive values (24.2 and 18.5) indicate a clear increase of the flexibility of the bases for GGGGCCCC and GGCCGGCC. The negative values (−28.7 and −9.8) indicate a clear increase of the rigidity of the bases for AAAATTTT and GCGCGCGC, while a vanishing flexibility change of the bases for AAGCGCTT is observed (Table 2).

Overall, of the total increase in rigidity of the DNA structure that is observed for three studied sequences (AAAATTTT, AAGCGCTT, GCGCGCGC) through the change in configurational entropy of the whole DNA ("helix") upon binding of QUE (or FLP) around 50−70% stems from the structural rearrangements of the sugar−phosphate backbone (Table 2). In addition, the increased flexibility of DNA upon respective binding of both ligands makes GGGGCCCC exceptional among all sequences explored.

The binding of QUE and of FLP with the duplex DNA (5′-GGGGCCCC) is associated with the increased flexibility of the whole DNA and of the bases and the increased rigidity of the backbone (Table 2). While QUE was recognized as a minor groove binder,[10] FLP was established to be concomitantly involved in intercalation (mediated by stacking interactions of its chlorophenyl and heterocyclic portions with the nucleo-

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433

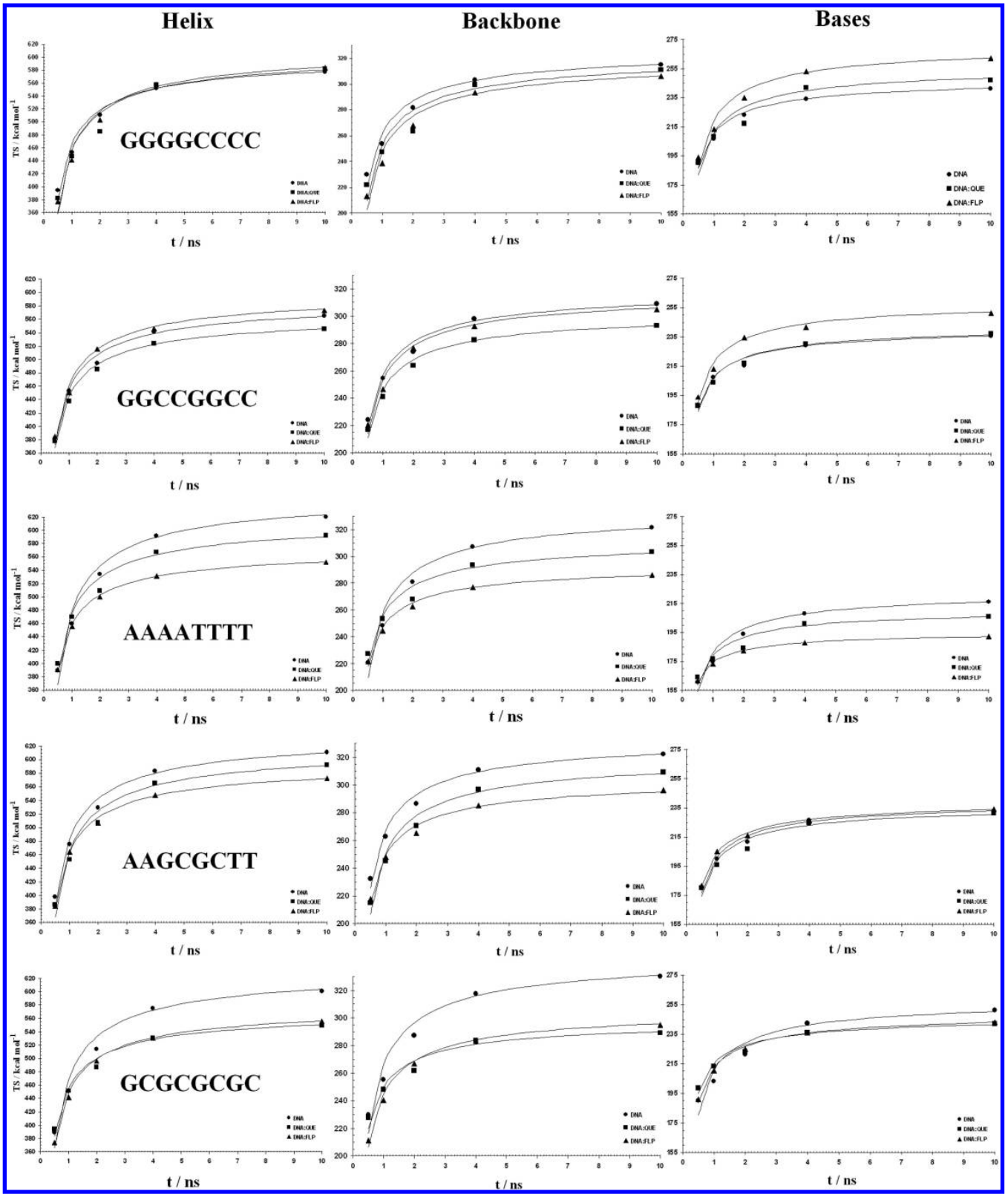


Figure 5. Dependence of calculated configurational entropy on the length of the MD trajectory for double helix, backbone, and bases of the DNA structure at 300 K, as *TS* (kcal mol^{-1}) vs time. Solid lines are fitted functions based on eq 2. Note the identical *y*-axis range throughout each column.

bases) and in groove binding (mediated by additional electrostatic interactions of its piperidinyl moiety that stretches to the minor groove of DNA).[3] Figure 4c illustrates that upon the creation of the binding site by separating the base pair 4 and reorganizing the sugar−phosphate backbone FLP is dominantly intercalated between the base pairs 3 and 5, whereas the simultaneous involvement of FLP in groove binding is obviously negligible. In contrast to QUE, the presence of FLP is much more strongly "felt" by the complete system of nucleobases experiencing an increase in dynamic flexibility

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433

Table 2. Change of Configurational Entropy upon Ligand Binding at 300 K[a]

5′-DNA sequence	system	$T\Delta S$ (kcal mol^{-1})
GGGGCCCC	helix (DNA:QUE)−helix (DNA)	6.5
	helix (DNA:FLP)−helix (DNA)	11.8
	backbone (DNA:QUE)−backbone (DNA)	−4.7
	backbone (DNA:FLP)−backbone (DNA)	−7.7
	bases (DNA:QUE)−bases (DNA)	8.9
	bases (DNA:FLP)−bases (DNA)	24.2
GGCCGGCC	helix (DNA:QUE)−helix (DNA)	−20.4
	helix (DNA:FLP)−helix (DNA)	12.7
	backbone (DNA:QUE)−backbone (DNA)	−16.7
	backbone (DNA:FLP)−backbone (DNA)	−2.2
	bases (DNA:QUE)−bases (DNA)	1.3
	bases (DNA:FLP)−bases (DNA)	18.5
AAAATTTT	helix (DNA:QUE)−helix (DNA)	−41.0
	helix (DNA:FLP)−helix (DNA)	−85.0
	backbone (DNA:QUE)−backbone (DNA)	−23.4
	backbone (DNA:FLP)−backbone (DNA)	−42.4
	bases (DNA:QUE)−bases (DNA)	−12.5
	bases (DNA:FLP)−bases (DNA)	−28.7
AAGCGCTT	helix (DNA:QUE)−helix (DNA)	−18.6
	helix (DNA:FLP)−helix (DNA)	−42.7
	backbone (DNA:QUE)−backbone (DNA)	−13.1
	backbone (DNA:FLP)−backbone (DNA)	−29.6
	bases (DNA:QUE)−bases (DNA)	−2.4
	bases (DNA:FLP)−bases (DNA)	0.4
GCGCGCGC	helix (DNA:QUE)−helix (DNA)	−64.3
	helix (DNA:FLP)−helix (DNA)	−55.2
	backbone (DNA:QUE)−backbone (DNA)	−46.9
	backbone (DNA:FLP)−backbone (DNA)	−39.2
	bases (DNA:QUE)−bases (DNA)	−12.5
	bases (DNA:FLP)−bases (DNA)	−9.8

[a] $T\Delta S$ (kcal mol^{-1}) calculated for various parts of DNA structure (Figure 2).

Table 3. TS (kcal mol^{-1}) Configurational Entropy Evaluated for Base Pairs and Base Pair Steps of DNA Structure (Figure 2) at 300 K

		base pair								base pair step						
5′-DNA sequence	system	1	2	3	4	5	6	7	8	1/2	2/3	3/4	4/5	5/6	6/7	7/8
GGGGCCCC	DNA	40.0	37.8	37.8	37.5	37.7	37.3	37.8	39.3	69.2	66.7	66.8	66.1	66.1	66.1	67.9
	DNA:QUE	41.7	40.2	39.2	39.3	38.6	38.5	39.3	44.6	74.6	71.0	69.7	68.8	68.6	69.8	75.6
	DNA:FLP	43.2	41.2	40.6	47.1	39.7	39.5	42.7	44.3	75.8	72.6	79.6	78.8	70.2	72.6	78.9
GGCCGGCC	DNA	38.3	36.6	36.2	36.5	36.1	36.6	37.6	39.4	68.4	65.1	65.8	65.9	65.3	65.6	69.6
	DNA:QUE	41.1	37.4	37.0	37.0	37.6	36.4	36.3	42.2	70.7	65.4	65.7	66.8	66.2	64.0	70.2
	DNA:FLP	41.6	38.0	38.0	38.2	38.3	39.0	42.7	43.6	71.8	68.0	68.4	69.5	69.7	73.7	79.5
AAAATTTT	DNA	38.4	36.4	35.3	34.6	34.6	35.3	35.8	37.6	66.4	63.2	60.9	60.2	61.2	62.7	65.9
	DNA:QUE	37.4	34.9	33.6	32.9	33.6	33.3	33.9	35.5	64.3	60.1	58.4	58.1	58.6	58.4	61.1
	DNA:FLP	33.3	31.0	30.5	30.0	30.3	30.3	31.8	33.4	57.6	54.6	52.7	52.2	52.8	54.8	58.3
AAGCGCTT	DNA	37.5	36.1	40.0	39.8	39.8	39.9	35.9	38.0	64.7	67.1	70.5	71.1	70.1	67.0	65.2
	DNA:QUE	38.0	35.7	39.2	39.8	38.6	38.6	35.9	39.0	64.9	65.5	69.1	69.3	68.3	65.0	66.7
	DNA:FLP	34.0	32.8	36.2	35.6	47.5	36.4	34.1	41.3	59.3	61.2	63.7	76.4	77.7	62.7	69.3
GCGCGCGC	DNA	40.8	39.4	38.6	38.7	38.6	38.6	39.4	41.4	72.4	71.3	70.5	71.3	69.4	71.2	73.1
	DNA:QUE	38.0	36.7	36.4	35.9	35.4	35.9	42.7	42.0	67.4	65.1	63.6	63.6	62.9	71.5	78.5
	DNA:FLP	39.0	37.6	36.8	36.8	36.7	37.2	40.0	45.4	68.6	66.7	66.3	66.6	65.4	69.0	78.2

(reflected through a configurational entropy change of 24.2 kcal mol^{-1} in Table 2) than by the sugar−phosphate backbone experiencing an increase in dynamic rigidity (reflected through a configurational entropy change of −7.7 kcal mol^{-1} in Table 2). In addition, the intercalation of small organic molecules, which are composed of several mutually attached aromatic rings, into A/T-rich or G/C-poor oligonucleotides was mainly related to the structural response of the backbone[30] and is in agreement with the data (Table 2) for all explored sequences, except GGGGCCCC and GGCCGGCC.

In the structure-based design of a potential anticancer agent acting on duplex DNA, the initial idea is often to achieve an optimal binding conformation that will include the elements of both groove binding and intercalation, with the aim of

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433

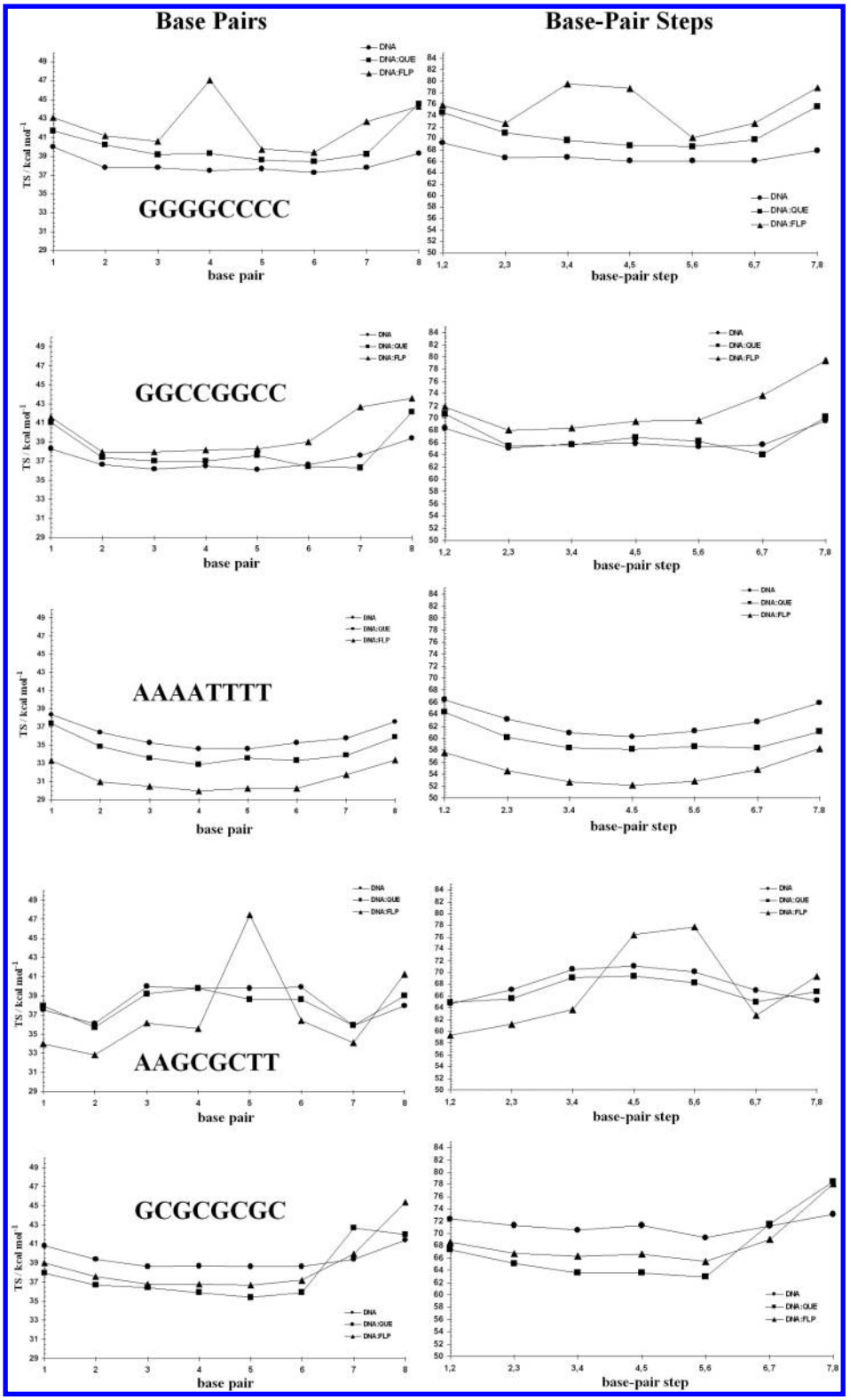


Figure 6. Entropy contributions of base pairs and base pair steps of the DNA structure as *TS* (kcal mol^{-1}) at 300 K. Note the identical *y*-axis range throughout each column.

intercalation to additionally stabilize a DNA:ligand complex by way of the stacking interactions of some parts of the ligand structure with the nucleobases. The specificity issue related to a particular DNA code was shown to be a key factor in obtaining the optimal binding conformation of QUE, and the pattern GGC was established to be the most specific for QUE in terms of maximizing the number of electrostatic interactions between QUE and the duplex DNA (PDB ID: 3IXN).[10] In light of this, Figure 4c demonstrates that the intercalation of FLP is associated with GGC that belong to the base pairs 3, 4, and 5, respectively. In other words, the particular code (GGC), which is part of the central tetramer (GGCC) of GGGGCCCC

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433

Table 4. *TS* (kcal mol^{-1}) Configurational Entropy Evaluated for Individual Base Pairs (Intra-Base Pair Entropy) and Difference from Entropy of System of Bases (Inter-Base Pair Contribution) at 300 K[a]

5′-DNA sequence	GGGGCCCC	GGCCGGCC	AAAATTTT	AAGCGCTT	GCGCGCGC
bases (DNA)[b]	249.9	243.2	225.7	241.6	261.1
bases (DNA:QUE)	258.8	244.5	213.2	239.2	248.6
bases (DNA:FLP)	274.1	261.7	197.0	242.0	251.3
bases (DNA:QUE)−bases (DNA)	**8.9**	**1.3**	**−12.5**	**−2.4**	**−12.5**
bases (DNA:FLP)−bases (DNA)	**24.2**	**18.5**	**−28.7**	**0.4**	**−9.8**
base pairs (DNA)[c]	305.2	297.3	288.0	307.0	315.5
base pairs (DNA:QUE)	321.4	305.0	275.1	304.8	303.0
base pairs (DNA:FLP)	338.3	319.4	250.6	297.9	309.5
base pairs (DNA:QUE)−base pairs (DNA)	**16.2**	**7.7**	**−12.9**	**−-2.2**	**−12.5**
base pairs (DNA:FLP)−base pairs (DNA)	**33.1**	**22.1**	**−37.4**	**−9.1**	**−6.0**
difference (DNA)[d]	−55.3	−54.1	−62.3	−65.4	−54.4
difference (DNA:QUE)	−62.6	−60.5	−61.9	−65.6	−54.4
difference (DNA:FLP)	−64.2	−57.7	−53.6	−55.9	−58.2
difference (DNA:QUE)−difference (DNA)	**−7.3**	**−6.4**	**0.4**	**−0.2**	**0**
difference (DNA:FLP)−difference (DNA)	**−8.9**	**−3.6**	**8.7**	**9.5**	**−3.8**

[a]Change of entropies upon ligand binding, given by bold typeface, need to be compared with the data in Table 2. [b]Entropy values of all base pairs (bases in Figure 2) correspond to infinitely long MD simulations (eq 2). [c]Sum of individual base pair entropies, intra-base pair entropy. [d]Difference of the values for bases and base pairs, estimate of inter-base pair entropy.

in the 5′-strand of DNA, is critical for generating the binding site and for reorganizing the sugar−phosphate backbone (Figure 4c). The observation is in agreement with the previous experimental indication that an optimal FLP-DNA binding mode may require the more specific sequences of three or more nucleotides,[3] but it is in disagreement with the understanding that a three-nucleotide specificity pattern is more typical for groove binders than for intercalators.[3] Moreover, the highest preference of QUE for G relative to any other single nucleobase was established by density functional theory (DFT) methods.[10] The same trends of the changes of configurational entropy upon binding of QUE and of FLP to the GGGGCCCC target (Table 2) allow us to identify G as a single nucleobase for which flavonoids generally show a highest preference. Therefore, the first two Gs, foregoing the critical GGC triplet in GGGGCCCC, certainly make the particular oligonucleotide additionally more specific for flavonoids than the other four sequences. Even though the possibility that an optimal flavonoid-duplex DNA binding mode requires more specific sequences of four or more nucleotides is not ruled out, the major part of highly specific sequences is believed to consist of a GGC triplet that is surrounded by consecutive Gs in its nearest neighborhood.

Entropy of Base Pairs. The configurational entropy computed for each base pair is given in Table 3 (left column) and schematically presented in Figure 6 (left panel). The measure quantifies the entropy within a specific base pair and is uncorrelated with the motion of the other base pairs. When analyzing the data, a particular emphasis needs to be placed on the central base pairs 3, 4, 5, and 6 that are responsible for generating the binding site (Figure 4), while the boundary base pairs 1, 2, 7, and 8 are of much less interest because the rest of the DNA is not taken into consideration.

The sequence GCGCGCGC is composed of regularly alternating guanine and cytosine, making the amount of entropy equally distributed over the base pairs 3, 4, 5, and 6 and resulting in a straight line (Figure 6). Even though the other sequences are slightly more heterogeneous, the course of base pair entropies along the strand retains quite a similar behavior for the ligand-free form of DNA.

In the three oligonucleotides having AAAATTTT, AAGCGCTT, and GCGCGCGC in the 5′-strand, the entropy of the base pairs decreases by about 1−3 kcal mol^{-1} per base pair upon QUE binding and by about 1−4 kcal mol^{-1} per base pair (the only exception is base pair 5 in AAGCGCTT) upon FLP binding (Table 3). In the case of both GGGGCCCC and GGCCGGCC, Figure 6 generally illustrates an increase of 1−2 kcal mol^{-1} in the amount of entropy per base pair upon QUE binding and an entropy increase in the range from 2 to 3 kcal mol^{-1} per base pair (the only exception is base pair 4 in GGGGCCCC) upon FLP binding. A dramatic entropy increase of ≈10 kcal mol^{-1} that is experienced by the base pair 4, the intercalation site of FLP in GGGGCCCC, is conceivable through Figure 4c. This means that a small change in the amount of entropy per base pair is relevant for groove binding mediated by electrostatic interactions with the ligand, while a more substantial entropy change per base pair is relevant for intercalation of the ligand involved in stacking interactions with the nucleobases nearest to the intercalation site.

The sum of entropies of the individual base pairs ("base pairs", Table 4) is fairly large and in the range from 288.0 to 315.5 kcal mol^{-1} for the simulations of bare DNA, in the range from 275.1 to 321.4 kcal mol^{-1} for the simulations of the DNA:QUE complex, and in the range from 250.6 to 338.3 kcal mol^{-1} for the simulations of the DNA:FLP complex. The very different values of the "base pairs" entropies were obtained by the calculations performed on eight different molecular systems, and the rest of the DNA was not considered. The magnitude of the different values suggests that the variation of entropy of all nucleobases is not only due to the motion of hydrogen-bonded base pairs with respect to the others (inter-base pair motion) but also due to the motion within some particular isolated base pairs (intra-base pair motion). The change in intra-base pair entropy upon ligand binding is positive (Table 4) for the DNA with 5′-GGGGCCCC (16.2 kcal mol^{-1} for QUE, 33.1 kcal mol^{-1} for FLP) and for the DNA with 5′-GGCCGGCC (7.7 kcal mol^{-1} for QUE, 22.1 kcal mol^{-1} for FLP), meaning that the intercalation of FLP (Figure 4c) increases the intra-base pair flexibility roughly 2−3 times relative to the minor groove binding of QUE (Figure 4b). In

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433

other words, the increased intra-base pair conformational flexibility contributes more favorably to the free energy of intercalation than to the free energy of minor groove binding. Of the total intra-base pair entropy increase of 33.1 kcal mol^{-1} upon intercalation of FLP in the 5′-GGGGCCCC DNA, 16.6 kcal mol^{-1} is distributed to the base pairs 3, 4, 5, and 6, keeping in mind that the intercalation site (base pair 4), by itself, carries about 10 kcal mol^{-1} (Table 3).

The sum of entropies of the individual base pairs ("base pairs", Table 4) can be conveniently compared with the entropy of the system of all nucleobases ("Bases", Table 4). By subtracting the "base pairs" values from the "bases" values, it is possible to quantitatively estimate the thermodynamic purpose of the inter-base pair motion ("difference", Table 4). With five exceptions such as binding of FLP to either the 5′-AAAATTTT DNA or the 5′-AAGCGCTT DNA and binding of QUE to either the 5′-AAAATTTT DNA or the 5′-AAGCGCTT DNA or the 5′-GCGCGCGC DNA, the negative differences show that the decreased inter-base pair conformational flexibility of the DNA does not contribute favorably to the free energy of binding. Even though the differences are mainly much less negative than those of the backbone (Table 2), the 5′-GGGGCCCC is an exception again. The inter-base pair conformational flexibility reflected by the negative values −7.3 and −8.9 kcal mol^{-1} in Table 4 is quite comparable with the conformational flexibility of the backbone reflected by the negative values −4.7 and −7.7 kcal mol^{-1} in Table 2, which correspond to the binding of QUE and of FLP with the 5′-GGGGCCCC DNA, respectively.

Entropy of Base Pair Steps. The configurational entropies of the base pair steps are summarized in Table 3 (right column) and plotted in Figure 6 (right panel). In interpreting the data, the base pair steps of particular interest are those that include the base pairs 3, 4, 5, and 6 making the central tetramer, as explained in the previous section.

The simulations of bare DNA produced the entropies that are symmetric relative to the central 4−5 base pair steps, thereby reflecting the symmetry of the sequences. In all of the sequences, the entropy is nearly equally distributed over the base pair steps (≈ 70, 70, 60, 65, and 66 kcal mol^{-1} per step in GCGCGCGC, AAGCGCTT, AAAATTTT, GGCCGGCC, and GGGGCCCC, respectively, Table 3). Depending on the sequence, upon ligand binding, the entropy contributes to the free energy an amount that is between −8 and 13 kcal mol^{-1} per step (Table 3).

The sequence GGGGCCCC is exceptional among the others because the entropy of steps increases upon binding of both ligands (Table 3, right column). The entropy remains approximately constant upon minor groove binding of QUE. In contrast, an asymmetry of entropy is nicely visible for the intercalated 5′-GGGGCCCC DNA by FLP and is most apparent in the region of the intercalation site (around base pair step 4/5), as given in Figure 6 (right panel, top). Thus, it may be concluded that the change of flexibility of the nucleobases upon intercalation of FLP is spatially localized to the region surrounding the intercalation site.

Final Remark. It is necessary to make a final remark regarding the methodology used in the present study. The quasi-harmonic approximation assumes only one equilibrium structure of the molecule, enabling each atomic coordinate to fluctuate around a single equilibrium value (eq 1). In this context, it is convenient to analyze the calculated entropy changes of the DNA upon intercalation of FLP by considering the enthalpic cost that is expected for the creation of an intercalation site. The enthalpic cost, the deformation energy of DNA, was estimated to be 20−24 kcal mol^{-1}.[30] Subtracting the changes of configurational entropy of the entire double helix (Table 2) from this value would give the following free energies: 105.0−109.0, 75.2−79.2, 62.7−66.7, 8.2−12.2, and 7.3−11.3 kcal mol^{-1} for AAAATTTT, GCGCGCGC, AAGCGCTT, GGGGCCCC, and GGCCGGCC, respectively. The positive values mean that the quasi-harmonic approximation holds for every sequence, and the decreasing trend of the values (from left to right) is in agreement with the increasing trend of dynamical flexibility change of the whole DNA upon FLP binding. In other words, by going from AAAATTTT to GGCCGGCC via GCGCGCGC, AAGCGCTT, and GGGGCCCC, the intercalation of part of the FLP structure becomes the more prevailing mode of interaction than the minor groove binding of the rest of the same ligand structure. These indications should be carefully taken. There are absurd examples showing that the intercalation of aromatic ligand molecules could yield a negative free energy due to the quasi-harmonic approximation-based overestimation of true entropy contribution, meaning that the whole double helix of DNA would spontaneously unwind in order to create an intercalation site.[30] Therefore, in the cases of very flexible targets like proteins where the quasi-harmonic approximation may be invalid, a correct treatment of the magnitude of conformational flexibility changes is not always straightforward.

■ SUMMARY

The sequence-dependent recognition between duplex DNA and flavonoids has been explored using molecular dynamics simulations aimed to evaluate the magnitude of flexibility changes experienced by various structural parts (entire double helix, sugar−phosphate backbone, and nucleobases) of the DNA upon ligand binding. A compromise among all possible structural differences displayed by the members of the flavonoid family has been considered through the investigation of the sequence-dependent binding between duplex DNA and two representative examples of the structurally different ligand molecules: quercetin (a minor groove binder) and flavopiridol (a combined agent−minor groove binder and intercalator). The configurational entropy contribution, as an upper bound of the true entropy contribution to the free energy, has been shown to quantitatively describe the dependence of the magnitude of DNA flexibility change on the particular receptor sequence and to make the fundamental distinction between the intercalation and the groove binding of the ligands. In this way, the magnitude of the flexibility change of the DNA, depending not only on the targeted sequence but also on the spatial conformation of the ligand, has been characterized.

Of all considered 5′-DNA sequences, GGGGCCCC has been identified as unique because the increased flexibility of the entire double helix and the complete system of nucleobases upon ligand binding has only been observed for the particular 5′-DNA sequence. The GGC triplet, which belongs to the central tetramer (GGCC) of 5′-GGGGCCCC, has been found to be critical for generating the binding site and reorganizing the sugar−phosphate backbone. Because the possibility that an optimal flavonoid−duplex DNA binding mode requires more specific sequences of four or more nucleotides has not been ruled out, the major part of highly specific sequences has been suggested to consist of consecutive Gs foregoing and/or going after the critical GGC triplet as much as possible because G has

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433

been established to be a nucleobase for which flavonoids display the highest preference in comparison to any other single nucleobase. Even though such sequences have been hypothesized as a structural signature that is more typical for groove binders (such as quercetin) than for intercalators, the molecular dynamics of the flavopiridol:DNA complex with 5′-GGGGCCCC has revealed that the prevailing mode of interaction is intercalation. It means that an optimal recognition mode, which simultaneously involves the elements of both intercalation and groove binding, may be very specific and mediated by simultaneous stacking (underlying the intercalation) and electrostatic (underlying the groove binding) interactions. As the intrabase conformational flexibility becomes more substantial due to the presence of a ligand thereby contributing more favorably to the free energy of intercalation than to the free energy of minor groove binding, an optimal DNA binder is more likely to be an intercalating agent localizing the flexibility of nucleobases to the intercalation site and its nearest neighborhood.

In the cases of very flexible targets when the quasi-harmonic approximation underlying the determination of configurational entropy may be invalid, potential applications of the present methodology are to be taken with special care.

AUTHOR INFORMATION

Corresponding Author

*E-mail: director.ist-belgrade.edu.rs@tech-center.com.

Notes

The author declares no competing financial interest.

ACKNOWLEDGMENTS

The present work was supported by Perth Science and Technology Consortium through an Applied Academic Research Carrier Award to the author. Prof. David A. Case of Rutgers University is gratefully acknowledged for granting the author an academic license that is relevant for using the molecular dynamics software package Amber11 in combination with AmberTools 1.5.

REFERENCES

(1) Babu, B. V.; Konduru, N. K.; Nakanishi, W.; Hayashi, S.; Ahmed, N.; Mitrasinovic, P. M. Experimental and Theoretical Advances in Functional Understanding of Flavonoids as Anti-tumor Agents. *Anti-Cancer Agents Med. Chem.* **2013**, *13*, 307−332.

(2) Bible, K. C.; Kaufmann, S. H. Flavopiridol (NSC 649890, L86−8275): A Cytotoxic Flavone That Induces Cell Death in Human Lung Carcinoma Cells. *Cancer Res.* **1996**, *56*, 4856−4861.

(3) Bible, K. C.; Bible, R. H., Jr.; Kottke, T. J.; Svingen, P. A.; Xu, K.; Pang, Y. P.; Hajdu, E.; Kaufmann, S. H. Flavopiridol Binds to Duplex DNA. *Cancer Res.* **2000**, *60*, 2419−2428.

(4) Hirpara, K. V.; Aggarwal, P.; Mukherjee, A. J.; Joshi, N.; Burman, A. C. Quercetin And Its Derivatives: Synthesis, Pharamacological Uses with Special Emphasis on Anti-Tumor Properties and Prodrug with Enhanced Bio-availability. *Anti-Cancer Agents Med. Chem.* **2009**, *9*, 138−161.

(5) Aparicio, S. A Systematic Computational Study on Flavonoids. *Int. J. Mol. Sci.* **2010**, *11*, 2017−2038.

(6) Boege, F.; Straub, T.; Kehr, A.; Boesenberg, C.; Christiansen, K.; Andersen, A.; Jakob, F.; Kohrle, J. Selected Novel Flavones Inhibit the DNA Binding or the DNA Religation Step of Eukaryotic Topoisomerase I. *J. Biol. Chem.* **1996**, *271*, 2262−2270.

(7) Sun, H.; Tang, Y.; Xiang, J.; Xu, G.; Zhang, Y.; Zhang, H.; Xu, L. Spectroscopic Studies of the Interaction between Quercetin and G-quadruplex DNA. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3586−3589.

(8) Sun, H.; Xiang, J.; Tang, Y.; Xu, G. Regulation and Recognization of the Extended G-quadruplex by Rutin. *Biochem. Biophys. Res. Commun.* **2007**, *352*, 942−946.

(9) Hussain, S. T.; Siddiqa, A. Voltammetric and Viscometric Studies of Flavonoids Interactions with DNA at Physiological Conditions. *Euro. J. Chem.* **2011**, *2*, 109−112.

(10) Mitrasinovic, P. M.; Palakshan, P. T.; Tripathi, S.; Tripathi, A. N. On the Affinity and Specificity of Quercetin for DNA. *Med. Chem.* **2013**, *9*, 193−202.

(11) Case, D. A.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(12) Case, D. A.; Darden, T. A.; Cheatham, III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. −J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *Amber 11*; University of California: San Francisco, CA, 2010.

(13) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(14) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785−2791.

(15) Hornak, V.; Abei, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712−725.

(16) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T. A.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.9; Gaussian, Inc.: Pittsburgh, PA, 1998.

(17) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-behaved Electrostatic Potential Based Method Using Charge-restraints for Deriving Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(18) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(19) Andersen, H. C. Rattle: A 'Velocity' Version of the Shake Algorithm for Molecular Dynamics Calculations. *J. Comput. Phys.* **1983**, *52*, 24−34.

(20) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \bullet \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(21) Andricioaei, I.; Karplus, M. On the Calculation of Entropy from Covariance Matrices of the Atomic Fluctuations. *J. Chem. Phys.* **2001**, *115*, 6289−6292.

(22) Schlitter, J. Estimation of Absolute and Relative Entropies of Macromolecules Using the Covariance Matrix. *Chem. Phys. Lett.* **1993**, *215*, 617−621.

(23) Harris, S. A.; Laughton, C. A. A Simple Physical Description of DNA Dynamics: Quasi-harmonic Analysis as a Route to the Configurational Entropy. *J. Phys.: Condens. Matter* **2007**, *19*, 076103.

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433

(24) Schafer, H.; Daura, X.; Mark, A. E.; van Gunsteren, W. F. Entropy Calculations on a Reversibly Folding Peptide: Changes in Solute Free Energy Cannot Explain Folding Behavior. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 45−56.
(25) Kolář, M.; Kubař, T.; Hobza, P. Sequence-Dependent Configurational Entropy Change of DNA Upon Intercalation. *J. Phys. Chem. B* **2010**, *114*, 13446−13454.
(26) Zhou, H. X.; Gilson, M. K. Theory of Free Energy and Entropy in Noncovalent Binding. *Chem. Rev.* **2009**, *109*, 4092−4107.
(27) Schafer, H.; Mark, A. E.; van Gunsteren, W. F. Absolute Entropies from Molecular Dynamics Simulation Trajectories. *J. Chem. Phys.* **2000**, *113*, 7809−7817.
(28) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. Cooperativity in Drug-DNA Recognition: a Molecular Dynamics Study. *J. Am. Chem. Soc.* **2001**, *123*, 12658−12663.
(29) Wang, H.; Laughton, C. A. Molecular Modelling Methods for Prediction of Sequence-Selectivity in DNA Recognition. *Methods* **2007**, *42*, 196−203.
(30) Kubar, T.; Hanus, M.; Ryjacek, F.; Hobza, P. Binding of Cationic and Neutral Phenanthridine Intercalators to a DNA Oligomer Is Controlled by Dispersion Energy: Quantum Chemical Calculations and Molecular Mechanics Simulations. *Chem.—Eur. J.* **2005**, *12*, 280−290.

■ NOTE ADDED AFTER ASAP PUBLICATION

This article was published ASAP on January 12, 2015, with ref 25 and acknowledgment of adaptation of Figure 2 omitted. The corrected version was published ASAP on January 16, 2015.

DOI: 10.1021/ci5006965
J. Chem. Inf. Model. 2015, 55, 421−433