

# Small-Molecule Interaction with G-Quadruplex DNA: Context of Anti-Cancer Drug Design

Petar M. Mitrasinovic

Center for Biophysical and Chemical Research, Belgrade Institute of Science and Technology, 11060 Belgrade, Serbia

Author's e-mail address: pmitrasinovic.ist-belgrade.edu.rs@tech-center.com

RECEIVED: November 25, 2018 \* REVISED: January 25, 2019 \* ACCEPTED: January 28, 2019

**Abstract:** Targeting G-quadruplex (G4) DNA structures by small molecules is a potential strategy for directing gene therapy of cancer disease. Herein, novel insights into non-covalent interactions between a structurally diversified spectrum of ligands and a G-quadruplex DNA (formed in the c-Myc oncogene promoter region) are reported. Solvation-induced effects on and entropic contributions to the binding free energy are explored. In addition, the correlation of G4 domain motions and active site rearrangements with the binding of highest affinity ligands, being associated with the fundamentally distinguishable modes of interaction (external stacking: BRACO-19, TMPyP<sub>4</sub>, and CX-3543; groove binding: Sanguinarine, Tetrahydropalmatine, and Hoechst 33258), is quantitatively evaluated and elaborated by observing thermodynamic consequences of the receptor conformational flexibility changes in the asymptotic regime ( $t \rightarrow \infty$ ) of molecular dynamics (MD) simulation. BRACO-19 and Tetrahydropalmatine are identified as unique (thermodynamically favorable and highly selective) G4-DNA binders. Implications of the present study for experimental research are elucidated.

**Keywords:** anti-cancer drug design, BRACO-19, c-Myc oncogene promoter, G-quadruplex DNA, molecular dynamics, Tetrahydropalmatine.

## INTRODUCTION

BESIDES folding into different duplex structures, highly dynamical DNA molecules can fold into hairpin, triplex, G-quadruplex, and i-motif structures that contain noncanonical base pairs. G-quadruplex (or G-tetraplex) structures stem from guanine (G)-rich tracts, and are composed of two or more stacked G-tetrads (or G-quartets) assembled from a single strand of DNA in an intramolecular (backfolded) way or from two-, three-, or four DNA strands that can associate in intermolecular G4-DNAs. G4s adopt a great diversity of conformations and folding energies, and their thermodynamic stabilities are often comparable to those of corresponding duplex structures. The presence and function of G4s *in vivo* are not quite clear. Being located in several significant genomic regions (human telomeres, oncogene promoter regions, immunoglobulin switch regions, ribosomal DNA, and some regions of RNA), G4s are hypothesized to be involved in important biological phenomena, such as telomere maintenance, end-capping and protection, chromosome stability, gene expression, viral

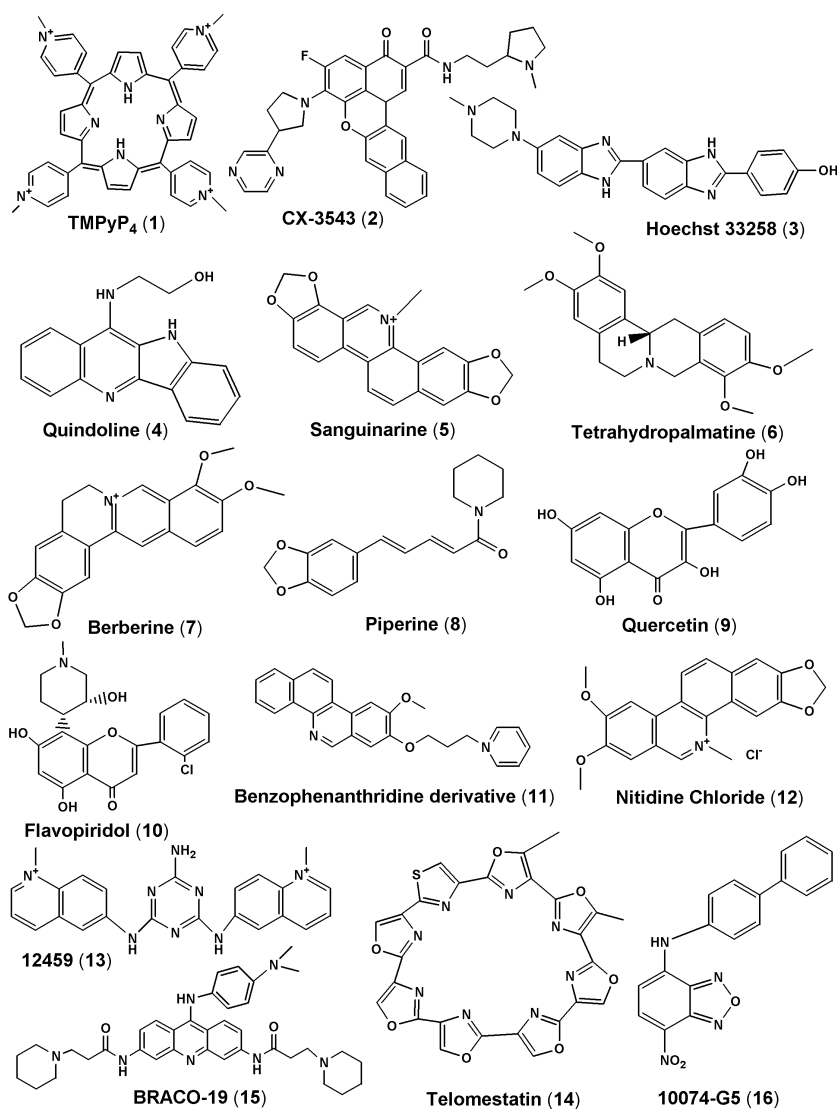
integration, and recombination.<sup>[1,2]</sup> For example, the inhibition of telomere elongation by telomerase in cancer cells is the consequence of the formation of quadruplexes in telomeric DNA.<sup>[3,4]</sup> As a growing number of proteins with a strong propensity to bind to G4s are being identified indicates that their interactions with G4s may be quite relevant for the cellular processes. Thus, targeting G4-DNA by small molecules, with the aim to disrupt the recognition of G4-DNA by the binding proteins, emerges as a potential anti-cancer strategy.<sup>[5]</sup> For this approach, the structural design and development of highly selective small ligand molecules is of vital importance.

G4-DNAs have more compact structures than duplex DNAs and display well-defined binding sites supposed to accommodate ligands that are complementary in shape and charge to the biological target.<sup>[1,2]</sup> Regardless of this fact, the way in which the ligand/G4 interaction is mediated currently seeks a more physically grounded elucidation in order to get closer to the successful identification of optimal G4-DNA binders. The main reason is related to finding optimal ligand/receptor configurations and predicting binding

free energy (BFE) in an efficient (without relying on formal statistical mechanics procedures) and consistent (without sacrificing accuracy) fashion. Therefore, as the literature suggested,<sup>[6]</sup> sophisticated computational methods are herein employed to explore solvation-induced effects on and entropic contributions to the binding free energy, as well as the correlation between substantial G4 active site rearrangements and ligand binding in the asymptotic regime ( $t \rightarrow \infty$ ) of MD simulation.

AG-rich element (5'-AGGGTGGGGAGGGTGGGGAAGG-3'), located between -137 and -115 bp upstream of the P1 promoter in the c-Myc oncogene, can form intramolecular G-quadruplex structure, which consists of repeated sequences with three or four guanine residues to suppress c-Myc transcription in a silenced form.<sup>[7]</sup> The element is a potential target to down-regulate c-Myc overexpression in

tumor cells.<sup>[8,9]</sup> A set of structurally different ligand molecules (Figure 1), having propensity to bind to the particular G4,<sup>[9]</sup> is chosen with intention to be representative (in terms of exhibited structural diversities) as much as possible. Their interactions with the G4 are characterized from a biophysical point of view. As a result, the highest affinity ligands, associated with external stacking interaction and groove binding respectively, are identified. Given that various conformations in static crystal structures may be due to differences in crystallization conditions or procedures, the dynamics of binding of the highest affinity ligands is investigated. For the various parts (complete structure, sugar-phosphate backbone, system of all bases, G-tetrad bases) of the G4-DNA structure, MD simulations are employed to evaluate the asymptotic configurational entropy change - the thermodynamic consequence of DNA



**Figure 1.** Chemical structures of ligand molecules.

flexibility change upon ligand binding. The configurational entropy is conceivable as an upper bound of the true entropy contribution to the free energy in non-covalent binding. It is the entropy of the solute, mainly without translational and rotational contributions. While the separation of translational entropy of the solute is acceptable without any approximation, the removal of rotational entropy is based on the assumption of a negligible correlation between the internal degrees of freedom and the global rotation of the solute. The configurational entropy is determined using the quasi harmonic approximation, considering a system of atoms with many degrees of freedom as a system of uncoupled harmonic oscillators.

## METHODS

A convenient review of distinct groups of small molecules with natural inclination to bind to G-quadruplex DNA was previously given in the literature.<sup>[9]</sup> Particular ligands (Figure 1) were selected to be typical structural representatives of the ligand families and adequate (in terms of atomic composition) to be treated by current computational methods. To obtain the initial coordinates of the target atoms, the experimental structure of the monomeric parallel-stranded G-quadruplex (PDB ID: 2A5P) was retrieved from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB).<sup>[10]</sup> The assembly of G-tetrads is shown in Figure 2.

Flexible docking of each ligand in the receptor was performed by AutoDock 4.2.<sup>[11,12]</sup> The lowest energy and physically meaningful (in terms of the ligand spatial orientation with respect to the compact binding sites of G4) conformations were extracted from docking experiments. 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 65 Å × 65 Å × 65 Å (1 Å = 10<sup>-10</sup> m) box centered on the macromolecule.

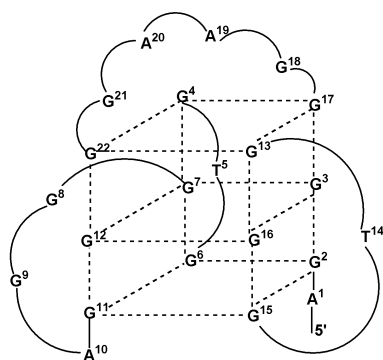


Figure 2. Assembly of G-tetrads in G-quadruplex structure.

All the other default options were chosen for the set up of runs by the AutoDock 4 tools.<sup>[12]</sup>

For the calculations done by the Amber 11 suite of programs,<sup>[13,14]</sup> the solute was prepared using the Amber11 utility program tLeap in association with the ff99sb force field.<sup>[15]</sup> Every ligand was initially prepared by parametrizing its atom types, charges, and connectivity in order to be treated as part of the solute. The molecular geometry was optimized by Gaussian 98 at the MP2/6-31G\* level of theory.<sup>[16]</sup> The molecular electrostatic potential was calculated by Gaussian 98 at the HF/6-31G\* level of theory,<sup>[16]</sup> while the atomic charges were derived by means of the RESP fitting technique<sup>[17]</sup> that is part of AmberTools 1.5.<sup>[13,14]</sup> Remaining parameters were assigned from the General Amber Force Field (GAFF),<sup>[18]</sup> being entirely compatible with the ff99sb macromolecular force field<sup>[15]</sup> and containing the parameters for almost all the organic molecules.

The single point free energy change upon complex formation was calculated using the Sander module under Amber11<sup>[13,14]</sup> on the basis of the following equation:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - [G_{\text{ligand}} + G_{\text{quadruplex}}] \quad (1)$$

where  $G_{\text{complex}}$ ,  $G_{\text{ligand}}$ , and  $G_{\text{quadruplex}}$  are the Molecular Mechanics-Generalized Born Solvent-Accessible Surface Area (MM-GBSA) free energies [Eq. (2)] for the complex, ligand, and quadruplex respectively. The free energies ( $G_{\text{MM-GBSA}}$ ) were calculated by taking into account molecular mechanics contribution ( $G_{\text{MM}}$ ) and solvation-induced effects ( $G_{\text{solv}}$ ) through the MM-GBSA approach,<sup>[19–23]</sup> as implemented in Amber 11.<sup>[13,14]</sup> The MM contribution ( $G_{\text{MM}}$ ) consists of the internal (bond, angular, and dihedral) energy ( $E_{\text{internal}}$ ), the van der Waals interaction ( $E_{\text{vdW}}$ ), and the Coulombic (electrostatic) interaction ( $E_{\text{elec}}$ ).  $\Delta E_{\text{internal}}$  was set to zero. Solvation was included implicitly through GBSA model treating solvent as a continuous medium instead of individual (explicit solvent) molecules. The solvation free energy ( $G_{\text{solv}}$ ) comprises the electrostatic contribution ( $E_{\text{GB}}$ )<sup>[24]</sup> and the nonpolar (SA-dependent) contribution ( $E_{\text{surf}}$ ) - the attractive (dispersion) and repulsive (cavity) interactions.<sup>[25]</sup> Thus, it follows:

$$\begin{aligned} \Delta G_{\text{MM-GBSA}} &= \Delta G_{\text{MM}} + \Delta G_{\text{solv}} \\ &= (\Delta E_{\text{internal}} + \Delta E_{\text{vdW}} + \Delta E_{\text{elec}}) + (\Delta E_{\text{GB}} + \Delta E_{\text{surf}}) \end{aligned} \quad (2)$$

The nonpolar solvation term was computed in the following way:

$$\Delta G_{\text{surf}} = \gamma \Delta \text{SASA} + b \quad (3)$$

where  $\gamma$  is the surface tension that was set to  $0.0072 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ , while  $b$  is a constant that was set to 0.<sup>[25]</sup> SASA means the Solvent-Accessible Surface Area ( $\text{\AA}^2$ ) that was predicted using the Molsurf algorithm. The GBSA terms are formally identical to the free energies of hydration that can be directly compared with experimental data.<sup>[19]</sup> To define the dielectric boundary around the molecular surface, the solvent probe radius was set to  $1.4 \text{ \AA}$ . The dielectric constant was set to 1 for the solute and to 80 for the surrounding solvent molecules.

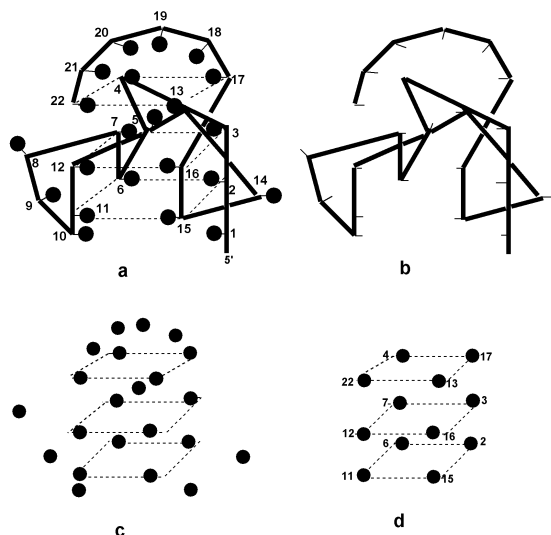
The solute entropy change at 300 K ( $T \Delta S$ ) was estimated using the normal (translational, rotational, and vibrational) modes for the complex, receptor, and ligand, which were calculated by the Nmode module inside Amber 11.<sup>[13,14]</sup>

Before running production MD, every solute was solvated using a  $8 \text{ \AA}$  pad of about 3220 TIP3P water molecules and the counter ions  $K^+$  were added to neutralize each system, what was followed by minimization and equilibration through 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  $2 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  on all non-hydrogen non-water atoms *via* 100 combined steepest descent and conjugate gradient minimization steps. Afterwards, the entire system was minimized without restraints by means of 200 combined steepest descent and conjugate gradient minimization steps. Following minimization, a 50 ps ( $1 \text{ ps} = 10^{-12} \text{ s}$ ) MD run was performed 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  $2 \text{ ps}^{-1}$  and harmonic restraints of  $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . 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 executed in the NPT ensemble, without positional restraints and with a pressure relaxation time of 2 ps. Production run was then made for 10 ns ( $1 \text{ ns} = 10^{-9} \text{ s}$ ) in the NPT ensemble at 300 K. Temperature was controlled by way of a Langevin thermostat with a  $2 \text{ ps}^{-1}$  collision frequency. A time step used throughout all stages was 2 fs ( $1 \text{ fs} = 10^{-15} \text{ s}$ ) and all hydrogen atoms were constrained using the Shake algorithm.<sup>[26]</sup> The long-range interactions were included on every step using the Particle Mesh Ewald algorithm<sup>[27]</sup> with a 4th order B-spline interpolation, a grid spacing of  $< 1 \text{ \AA}$ , and a direct space cutoff of  $12 \text{ \AA}$ . The minimizations and MD simulations were performed using the Sander module of Amber11.<sup>[13,14]</sup> The production trajectories were processed using the Ptraj analysis tool included in the Amber11 program suite.<sup>[13,14]</sup>

For the various structural parts of the receptor (Figure 3), the configurational entropy was estimated using the approximation of Andricioaei and Karplus - A&K,<sup>[28]</sup> 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. (4)].

$$S_{\text{A\&K}} = k \sum_i^{3N-6} \frac{\hbar \omega_i / kT}{e^{\hbar \omega_i / kT} - 1} - \ln(1 - e^{-\hbar \omega_i / kT}) \quad (4)$$



**Figure 3.** Different parts of G4-DNA (5'-A<sup>1</sup>G<sup>2</sup>G<sup>3</sup>G<sup>4</sup>T<sup>5</sup>G<sup>6</sup>G<sup>7</sup>G<sup>8</sup>G<sup>9</sup>A<sup>10</sup>G<sup>11</sup>G<sup>12</sup>G<sup>13</sup>T<sup>14</sup>G<sup>15</sup>G<sup>16</sup>G<sup>17</sup>G<sup>18</sup>A<sup>19</sup>A<sup>20</sup>G<sup>21</sup>G<sup>22</sup>) that were separately subjected to configurational entropy calculation: (a) Complete Structure, (b) Backbone, (c) Bases, and (d) G-tetrads.

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 ( $\omega_i$ ). The frequencies were determined by converting the nonzero eigenvalues ( $\lambda_i$ ) of the mass-weighted covariance matrix ( $\omega_i = \sqrt{kT/\lambda_i}$ ). The eigenvalues were obtained by means of the quasi-harmonic analysis, consisting of: (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*.<sup>[13,14]</sup> The frequencies and entropies were numerically computed by specifying the key word *thermo* on the *analyze matrix* command line. Another computationally less expensive formula for the evaluation of configurational entropy was proposed by Schlitter.<sup>[29]</sup> However, in case of DNA, both procedures<sup>[28,29]</sup> yielded equal upper bounds of the true entropy with numerical precision.<sup>[30]</sup>

**Table 1.** Various contributions to the free energy associated with the initial mode of ligand-G4 interaction.

Ligand (no.) <sup>(a)</sup>	Mode of interaction	$\Delta G_{MM}^{(b)}$ / kcal mol <sup>-1</sup>	$\Delta G_{solv}^{(c)}$ / kcal mol <sup>-1</sup>	$\Delta G_{MM-GBSA}^{(d)}$ / kcal mol <sup>-1</sup>	$T \Delta S^{(e)}$ / kcal mol <sup>-1</sup>
BRACO-19 ( <b>15</b> )	Stacking	-42.47	9.19	-33.28	-21.91
TMPyP <sub>4</sub> ( <b>1</b> )	Stacking	-43.57	11.61	-31.96	-24.75
CX-3543 ( <b>2</b> )	Stacking	-49.31	18.74	-30.57	-27.39
10074-G5 ( <b>16</b> )	Stacking	-30.18	7.42	-22.76	-15.87
Telomestatin ( <b>14</b> )	Stacking	-25.44	16.26	-9.18	-23.07
Sanguinarine ( <b>5</b> )	Groove	-34.54	11.66	-22.88	-16.95
Tetrahydropalmatine ( <b>6</b> )	Groove	-36.94	14.84	-22.10	-18.21
Hoechst 33258 ( <b>3</b> )	Groove	-37.87	17.04	-20.83	-19.53
Benzophenanthridine derivative ( <b>11</b> )	Groove	-21.72	4.61	-17.11	-22.32
Nitidine Chloride ( <b>12</b> )	Groove	-21.51	4.84	-16.67	-13.53
Piperine ( <b>8</b> )	Groove	-21.17	5.08	-16.09	-20.31
12459 ( <b>13</b> )	Groove	-33.24	19.37	-13.87	-21.09
Quercetin ( <b>9</b> )	Groove	-39.55	29.67	-9.88	-18.69
Quindoline ( <b>4</b> )	Groove	-19.32	10.13	-9.19	-15.06
Berberine ( <b>7</b> )	Groove	-24.17	17.51	-6.66	-17.88
Flavopiridol ( <b>10</b> )	Groove & Stacking	-9.27	5.58	-3.69	-12.84

<sup>(a)</sup> Taking into account that the current computational methods do not enable a routine determination of enthalpy, the data for different ligands need to be analyzed by observing the trends of  $\Delta G_{MM-GBSA}$  and  $T \Delta S$  separately.

<sup>(b)</sup>  $\Delta G_{MM} = \Delta E_{internal} + \Delta E_{vdw} + \Delta E_{elec}$ ; the molecular mechanical (MM) contributions are:  $E_{internal}$  - the internal (bond, angular, and dihedral) energy,  $E_{vdw}$  - van der Waals interaction from MM and  $E_{elec}$  - electrostatic energy calculated by the MM force field.

<sup>(c)</sup>  $\Delta G_{solv} = \Delta E_{GB} + \Delta E_{surf}$ ;  $E_{GB}$  - the electrostatic contribution to the solvation free energy calculated by Generalized Born (GB) method;  $E_{surf}$  - the nonpolar contribution to the solvation free energy determined by solvent-accessible Surface Area (SA)-dependent terms underlying the attractive (dispersion) and repulsive (cavity) interactions.

<sup>(d)</sup>  $\Delta G_{MM-GBSA} = \Delta G_{MM} + \Delta G_{solv}$ .

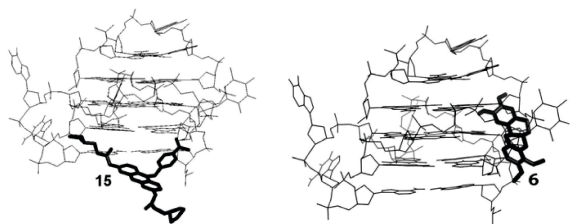
<sup>(e)</sup> The solute entropy change (multiplied by 300 K) is estimated using the normal (translational, rotational, and vibrational) modes for the complex, receptor, and ligand in Eq. (1).

## RESULTS AND DISCUSSION

### Initial Binding Modes of Highest Affinity Ligands

The MM-GBSA parameters that characterize the initial G4 : ligand interfaces are given in Table 1. Even though this approximate method is known to have many limitations (such as overestimation of the binding free energy),<sup>[31]</sup> this method is usable for single point calculation<sup>[5]</sup> in terms of scoring the initial structures that are further subjected to some, more rigorous, theoretical and computational methods. The question how inclusion of solvation in the computational protocol affects the binding free energy, defines the main criterion ( $\Delta G_{MM-GBSA}$  in Table 1) for determining highest affinity ligands. In practice, entropy contributions ( $T \Delta S$ ) obtained by performing normal mode analysis on the three species (complex, receptor, and ligand) can be neglected if a comparison of states of similar entropy is sought such as two ligands binding to the same receptor. The

reason for this is that normal mode analysis calculations are computationally expensive and their large margin of error introduces significant uncertainty in the result.<sup>[13,14,30]</sup> Thus, the values of  $\Delta G_{MM-GBSA}$  illustrate that BRACO-19 (**15**), TMPyP<sub>4</sub> (**1**), and CX-3543 (**2**) nearly equally stabilize the G4-DNA through external stacking interaction, while Sanguinarine (**5**), Tetrahydropalmatine (**6**), and Hoechst 33258 (**3**) do the same through groove binding. Their BFEs, of about  $-32 \pm 1.4$  kcal mol<sup>-1</sup> for external stacking and  $-22 \pm 1.3$  kcal mol<sup>-1</sup> for groove binding, are substantially more negative than those for the other ligands in the specific mode of binding respectively (Table 1), keeping in mind an average thermo chemical accuracy of around 2.0 kcal mol<sup>-1</sup>.<sup>[32-41]</sup> It means that the complex with the ligand bound to the G-tetrad is more stable than that with the ligand in the groove-binding mode by circa 10 kcal mol<sup>-1</sup>. The spatial orientations of BRACO-19 (**15**) and Tetrahydropalmatine (**6**) with respect to the G4 are shown in Figure 4, while those of the other highest affinity ligands are given in Figure S1 (Supplementary Information).



**Figure 4.** Initial modes of G4 interaction with BRACO-19 (**15**) and Tetrahydropalmatine (**6**) respectively.

Complex formation is generally followed by decreasing entropy, what is demonstrated by the negative value of the entropic term  $T \Delta S$  (at 300 K) for the every solute (Table 1). For different ligands, a lowering trend of  $\Delta G_{\text{MM-GBSA}}$  is expected to be associated with an increasing trend of  $T \Delta S$ . In the case of the highest affinity ligands involved in external stacking, the values of  $\Delta G_{\text{MM-GBSA}}$  ( $-30.57$ ,  $-31.96$ , and  $-33.28$  kcal mol $^{-1}$ ) are properly lined up with the values of  $T \Delta S$  ( $-27.39$ ,  $-24.75$ , and  $-21.91$  kcal mol $^{-1}$ ) for CX-3543 (**2**), TMPyP $_4$  (**1**), and BRACO-19 (**15**) respectively (Table 1). In the case of the highest affinity ligands involved in groove binding, the values of  $\Delta G_{\text{MM-GBSA}}$  ( $-20.83$ ,  $-22.10$ , and  $-22.88$  kcal mol $^{-1}$ ) are appropriately lined up with the values of  $T \Delta S$  ( $-19.53$ ,  $-18.21$ , and  $-16.95$  kcal mol $^{-1}$ ) for Hoechst 33258 (**3**), Tetrahydropalmatine (**6**), and Sanguinarine (**5**) respectively (Table 1).

In the context of predicting the solute entropy change ( $T \Delta S$  in Table 1) using the normal (translational, rotational, and vibrational) modes of the complex, receptor, and ligand, it is interesting to acquire some knowledge on which portion (in kcal mol $^{-1}$ ) of  $T \Delta S$  goes to ligand structural alterations upon complex formation specifically. Flexible docking of ligand using the AutoDock 4.2 algorithm is based on active torsions in ligand structure,<sup>[11,12]</sup> which can be viewed as particular sp $^3$  bonds being directly involved in finding the lowest energy ligand:receptor configurations. Entropy of ligand binding is essentially related to the loss of its degrees of freedom upon binding. The torsional potential, as determined by AutoDock 4.2,<sup>[11,12]</sup> indirectly takes care of it by being proportional to the number of active torsions in ligand structure. An active torsion is estimated to cost 0.3 kcal mol $^{-1}$  energetically.<sup>[11,12]</sup> It means that the structures of BRACO-19 (**15**), TMPyP $_4$  (**1**), and CX-3543 (**2**), having 11, 4, and 6 active torsions, experience the negative entropy changes that roughly cost  $-3.3$ ,  $-1.2$ , and  $-1.8$  kcal mol $^{-1}$  upon external stacking interaction with G4, respectively. By not having any active torsion in its structure, the groove binder Sanguinarine (**5**) does not entropically associate with a negative energetic amount. The entropy reductions

of about  $-1.2$  and  $-0.3$  kcal mol $^{-1}$  correspond to the groove binding of Tetrahydropalmatine (**6**) with 4 active torsions and Hoechst 33258 (**3**) with 1 active torsion, respectively.

The trend of  $\Delta G_{\text{MM-GBSA}}$  and the trend of  $T \Delta S$  are separately analyzed above, because the enthalpy calculation is often impractical and hard to converge. There is a proposal of calculating the binding free energy by summing  $\Delta G_{\text{MM-GBSA}}$  and  $(-T \Delta S)$ .<sup>[42]</sup> If this proposal is conditionally taken as a correct one, the values from Table 1 give the following BFEs: i)  $-11.37$ ,  $-7.21$ , and  $-3.18$  kcal mol $^{-1}$  for (**15**), (**1**), and (**2**) upon external stacking interaction and ii)  $-5.93$ ,  $-3.89$ , and  $-1.30$  kcal mol $^{-1}$  for (**5**), (**6**), and (**3**) upon groove binding respectively. All the negative values mean that the particular G4:ligand complexes are favorable thermodynamically.

The results suggest the  $\pi$ - $\pi$  stacking of ligand at the end of the G-quadruplex to be the favorable mode of interaction, what is in agreement with experimental<sup>[9,43]</sup> and theoretical<sup>[5]</sup> reports. Also, the groove or loop is suggested to be a viable site, which is of vital importance for blocking the interaction between the G4 and its binding proteins in aqueous solution and which is not related to crystal packing effects.<sup>[5]</sup> Even though nonspecific ligand-groove or ligand-loop binding is not inherently stable and depends on the particular topology of the loop,<sup>[5,9]</sup> the groove or loop may be seen as a recognition motif of interest for the structure-based drug design.

### Stable Regime of MD Simulation and Asymptotic Configurational Entropy

The reference structure for each MD run was calculated as an average over all frames fitted to the first frame of the specific trajectory. The configurational entropy was evaluated by considering vibrational degrees of freedom.<sup>[28]</sup> The translational entropy is removed without making any approximation, while the removal of rotational entropy is acceptable for a negligible correlation between the internal degrees of freedom and the global rotation of the solute.<sup>[44]</sup> The global translational and rotational entropies were separated using a least square fitting to the reference structure. The computed entropies were found to be insensitive to any iterative improvement of the reference structures. For the different parts of the G4 structure (Figure 3) the configurational entropy was evaluated using a procedure that treats only non-hydrogen atoms as input to the Ptraj command *matrix*,<sup>[13,14]</sup> which is described in the section Methods. The terms used in the caption of Figure 3 mean the following: i) Complete Structure refers to all of the non-hydrogen atoms of the entire G4, ii) Backbone refers to the non-hydrogen atoms (C1', C2', C3', O3', C4', O4', C5', O5', OP1, OP2, P) of the sugar and phosphate moieties of the nucleotides, iii) Bases refers to the non-hydrogen atoms (N1, C2, O2, N2, N3, C4,

O4, C5, C6, O6, C7, N7, C8, N9) of all nucleobases, and iv) G-tetrads refer to the non-hydrogen atoms (N1, C2, N2, N3, C4, C5, C6, O6, N7, C8, N9) of Gs making the G<sup>2</sup>G<sup>6</sup>G<sup>11</sup>G<sup>15</sup>, G<sup>3</sup>G<sup>7</sup>G<sup>12</sup>G<sup>16</sup>, and G<sup>17</sup>G<sup>4</sup>G<sup>22</sup>G<sup>13</sup> tetrads respectively. Not including the fast motion of hydrogen atoms in MD simulation may cause some error in the absolute value of entropy.<sup>[45,46]</sup> As the present focus is on the entropy change upon ligand binding to the apo (ligand-free) G4, the possibility of accumulating this type of inaccuracy was circumvented.

Because the entropy used for the analysis should converge with increasing trajectory length, the simulation trajectory was divided into nonoverlapping segments of 0.5, 1, 2, 4, and 10 ns. Then, the entropy was averaged over each time interval. The reported entropy values were extrapolated to an infinitely long simulation by the following empirical equation:

$$S(t) = S_{\text{inf}} - (p / t^q) \quad (5)$$

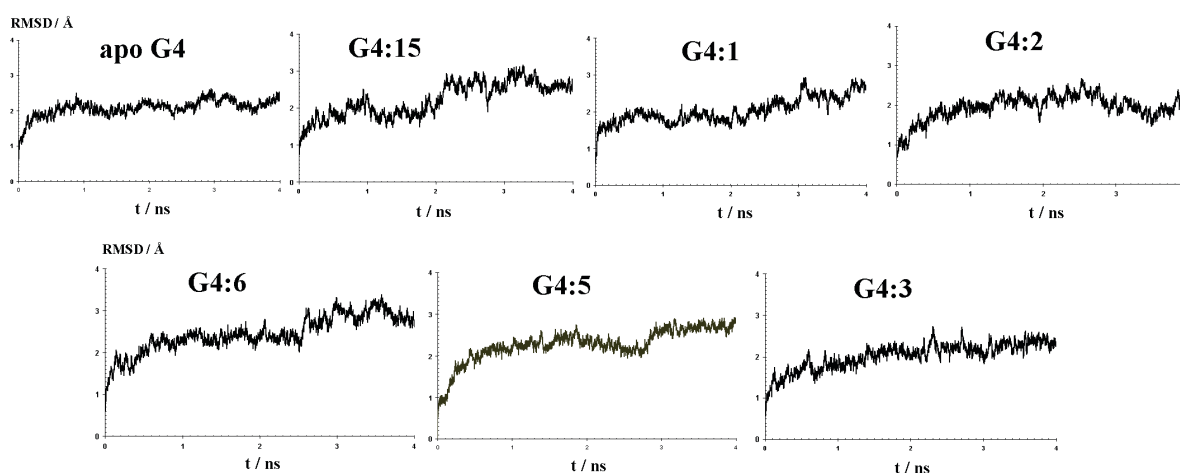
where  $S_{\text{inf}}$  is the estimation of entropy for an infinite time length of MD simulation,<sup>[30,47]</sup> while the physical meanings of the parameters  $p$  and  $q$  have not yet been determined.<sup>[38]</sup> The specific values of  $p$  and  $q$  were obtained by fitting  $S(t)$  to the sets of the entropy values obtained by 0.5, 1, 2, 4, and 10 ns long simulations (Figures 7 and 9). The fitting procedure was separately done for each part of G4 (Figure 3) in order to eliminate unwanted correlation of the particular entropy with the entropies of the other parts.  $S_{\text{inf}}$  multiplied by 300 K is reported in Tables S1 and S3 (Supplementary Information) that contain the absolute (asymptotic) configurational entropies  $\pm$  standard deviations

(SDs). To calculate the relative entropies (Tables 2 and 4) as the difference of the absolute entropies for a complex and the ligand-free G4, it was necessary to use the same type of  $S_{\text{inf}}$  that provides a best fit [Eq. (5)] for each system respectively. This condition was satisfied by  $S_{\text{inf}}$  that takes into account the upper bound (+SD). A satisfactory convergence of the entropies was seen through the correlation coefficients being in between 0.9 and 1. In the case of apo G4, the parameter  $q$  correctly converged to 0.7 - an expected value that was previously reported.<sup>[47]</sup> In the case of the G4:ligand complexes, the parameter  $q$  was in between 0.5 and 0.6, what is in agreement with the literature recommendation.<sup>[30]</sup>

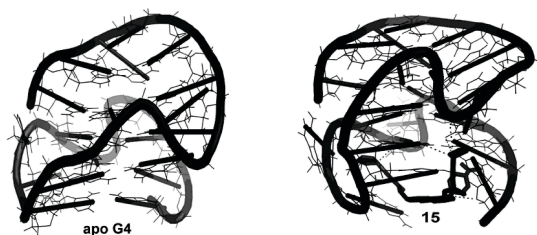
The simulations of both ligand-free (apo) G4 and its complexes with the highest affinity ligands reveal a stable trend around 4 ns, as it is demonstrated by the Root Mean Square Deviation (RMSD) relative to the first frame of the trajectory (Figure 5). The RMSD of apo G4 is stabilized around 2 Å, while the RMSD of its complexes with the ligands is stabilized in between 2.5 and 3 Å (Figure 5). Representative snapshots of the systems in the stable

**Table 2.** Change of configurational entropy ( $T \Delta S$  in kcal mol<sup>-1</sup>,  $T = 300$  K) for various parts of G4 structure upon external stacking interaction. Ligand no.: BRACO-19 (15), TMPyP<sub>4</sub> (1), and CX-3543 (2).

G4:Ligand no. - G4	Complete structure	Backbone	Bases	G-tetrads
G4:15 - G4	17.22	7.95	5.39	0.00
G4:1 - G4	-6.36	2.73	-10.58	-8.16
G4:2 - G4	-35.38	-16.23	-15.13	-13.69



**Figure 5.** Root Mean Square Deviation (RMSD) of non-hydrogen atoms (relative to the first frame of MD trajectory) vs. time for apo (ligand-free) G4, G4:15 (BRACO-19), G4:1 (TMPyP<sub>4</sub>), G4:2 (CX-3543), G4:6 (Tetrahydropalmatine), G4:5 (Sanguinarine), and G4:3 (Hoechst 33258) respectively.



**Figure 6.** Cartoon display of the complex formed upon external stacking of BRACO-19 (**15**) to apo (ligand-free) G4 in the stable regime of MD simulation around 4 ns.

regime of simulation are given in Figures 6 and 8 (see Figures S2 and S3 in Supplementary Information too). Numerical comparisons between the entropies generated by 4 ns-long simulations (Figures 7 and 9) and the entropies extrapolated to infinitely long simulations (Tables S1 and S3) confirm that  $92 \pm 3\%$  of the asymptotic ( $t \rightarrow \infty$ ) values was reached within 4 ns for each part of the G4 structure.

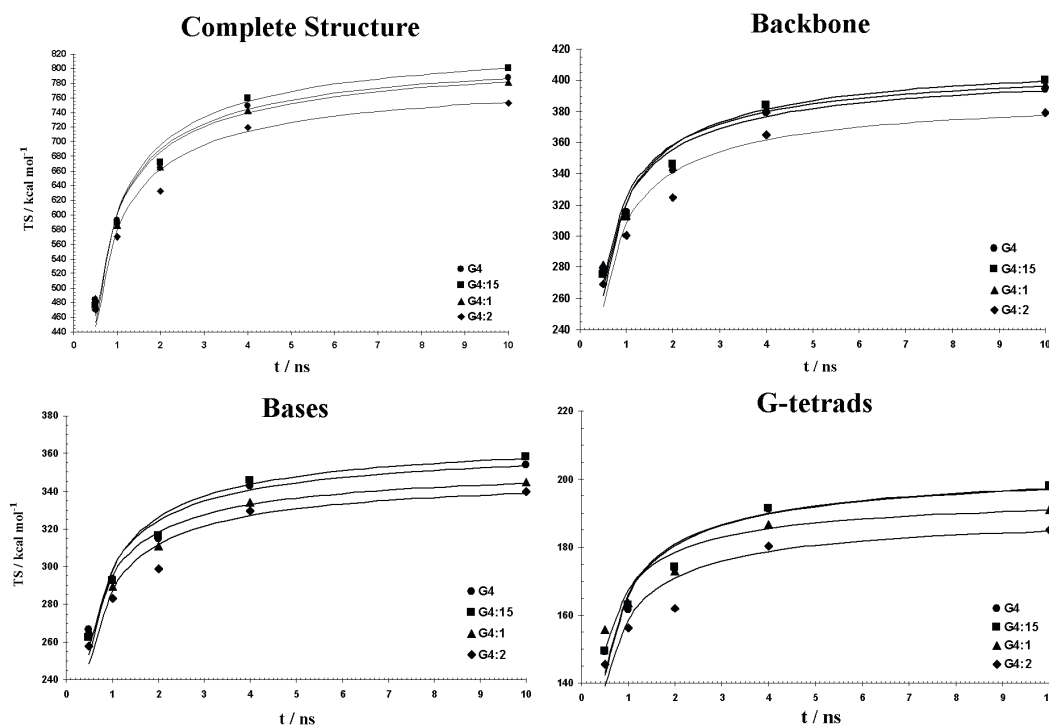
### G4 Flexibility Change Upon External Stacking Interaction

The configurational entropies extrapolated to infinitely long MD simulations are reported in Table S1 (Supplementary Information), while their changes are summarized in

Table 2. The different structural parts of G4 (complete structure, sugar-phosphate backbone, complete system of bases, and G-tetrad bases), for which the configurational entropy change was computed, are displayed in Figure 3.

Binding of BRACO-19 (**15**) is associated with a configurational entropy change of  $17.22 \text{ kcal mol}^{-1}$ , indicating a remarkable amount of increased flexibility of the entire G4-DNA structure. The increased conformational flexibility of G4 contributes favorably to the free energy of non-covalent binding. The backbone itself gets more flexible as an entropic increase of  $7.95 \text{ kcal mol}^{-1}$  is observed. The complete system of nucleobases itself retains the same trend as a relevant entropy change of  $5.39 \text{ kcal mol}^{-1}$  is found, of which almost nothing goes to the system of the G-tetrad nucleobases. The increased flexibility of both the sugar-phosphate backbone and the non-tetrad nucleobases may be associated with creating binding site, while the ability of the central channel of apo G4 to accommodate a solvent cation between each pair of tetrads does not get affected by binding of BRACO-19 (**15**).

Binding of TMPyP<sub>4</sub> (**1**) is associated with a configurational entropy change of  $-6.36 \text{ kcal mol}^{-1}$ , indicating a reasonable amount of increased rigidity of the entire G4-DNA structure. It means that the reduced conformational flexibility of G4-DNA contributes unfavorably to the free energy of non-covalent binding. This is mainly reflected through



**Figure 7.** Dependence of configurational entropy ( $TS$  in  $\text{kcal mol}^{-1}$ ,  $T = 300 \text{ K}$ ) on the length of MD trajectory for the various parts of the G4 structure in complex with the external stacking ligands BRACO-19 (**15**), TMPyP<sub>4</sub> (**1**), and CX-3543 (**2**) respectively. Solid lines are fitted functions based on Eq. (5).



the complete system of nucleobases as an entropy decrease of  $-10.58 \text{ kcal mol}^{-1}$  is established. Of all the nucleobases, those making the G-tetrads have a key contribution of  $-8.16 \text{ kcal mol}^{-1}$ . At the same time, the backbone does not experience any notable flexibility change as a slight entropic increase ( $2.73 \text{ kcal mol}^{-1}$ ) is detected. The reduced flexibility of G4 may be seen through the strengthening of Hoogsteen hydrogen bonds between the G-tetrad base pairs, thereby narrowing down the central channel that is supposed to accommodate a solvent cation between each pair of tetrads.

Binding of CX-3543 (**2**) is associated with a configurational entropy change of  $-35.38 \text{ kcal mol}^{-1}$ , indicating a substantial amount of increased rigidity of the entire G4-DNA structure. It means that the reduced conformational flexibility of G4-DNA contributes unfavorably to the free energy of non-covalent binding. This is reflected through both the sugar-phosphate backbone and the complete system of nucleobases. The backbone itself is associated with an entropic decrease of  $-16.23 \text{ kcal mol}^{-1}$ , while the complete system of nucleobases itself is associated with an entropic decrease of  $-15.13 \text{ kcal mol}^{-1}$ . Among the nucleobases, the G-tetrad bases have a key entropy contribution of  $-13.69 \text{ kcal mol}^{-1}$ . In comparison to binding of TMPyP<sub>4</sub> (**1**), a considerably larger increase of the rigidity of G4 due to binding of CX-3543 (**2**) may be conceivable by means of both a severe curvature of the backbone (Figure S2) and a conspicuous strengthening of Hoogsteen hydrogen bonds between the G-tetrad base pairs, thereby making the space of the central channel vertically distorted, more narrowed down, and less capable of accommodating a solvent cation between each pair of tetrads.

The configurational entropy calculated for each base pair (bp) of the G-tetrads is given in Table S2 (Supplementary Information). The values obtained by 4 ns simulations are presented because the calculated entropy converges very quickly with the length of trajectory for these systems, reaching 98 % of  $S_{\text{inf}}$  in a 4 ns run. This measure quantifies the entropy within a specific base pair and is uncorrelated with the motion of the other base pairs. All the individual intra-bp entropies remain almost unchanged ( $\pm 2 \text{ kcal mol}^{-1}$ , Table S2) by external stacking interaction of every single ligand with the G4. The sum of all the intra-bp entropies estimates the total intra-bp entropy of G-tetrads, which changes by 6.74,  $-14.20$ , and  $-22.79 \text{ kcal mol}^{-1}$  (Table 3) after having the G4 complexed with BRACO-19 (**15**), TMPyP<sub>4</sub> (**1**), and CX-3543 (**2**) respectively. Only in case of BRACO-19 (**15**), the total intra-base pair conformational flexibility of G-tetrads contributes favorably to the free energy of binding, and the entropic change of  $6.74 \text{ kcal mol}^{-1}$  (Table 3) is comparable with the entropic change of  $5.39 \text{ kcal mol}^{-1}$  (Table 2) that refers to the system of all nucleobases.

**Table 3.** Configurational entropy ( $T S$  in  $\text{kcal mol}^{-1}$ ,  $T = 300 \text{ K}$ ) evaluated for all individual base pairs making G-tetrads (intra-base pair entropy) and difference from entropy of system of G-tetrad bases (inter-base pair entropy).<sup>(a)</sup> Ligand no.: BRACO-19 (**15**), TMPyP<sub>4</sub> (**1**), and CX-3543 (**2**).

G-tetrads	$G^2G^6G^{11}G^{15}$ , $G^3G^7G^{12}G^{16}$ and $G^{17}G^4G^{22}G^{13}$
Bases (G4) <sup>(b)</sup>	205.49
Bases (G4:15)	205.49
Bases (G4:1)	197.33
Bases (G4:2)	191.80
<b>Bases (G4:15) – Bases (G4)</b>	<b>0.00</b>
<b>Bases (G4:1) – Bases (G4)</b>	<b>-8.16</b>
<b>Bases (G4:2) – Bases (G4)</b>	<b>-13.69</b>
Base pairs (G4) <sup>(c)</sup>	481.10
Base pairs (G4:15)	487.84
Base pairs (G4:1)	466.90
Base pairs (G4:2)	458.31
<b>Base pairs (G4:15) – Base pairs (G4)</b>	<b>6.74</b>
<b>Base pairs (G4:1) – Base pairs (G4)</b>	<b>-14.20</b>
<b>Base pairs (G4:2) – Base pairs (G4)</b>	<b>-22.79</b>
Difference (G4) <sup>(d)</sup>	-275.61
Difference (G4:15)	-282.35
Difference (G4:1)	-269.57
Difference (G4:2)	-266.51
<b>Difference (G4:15) - Difference (G4)</b>	<b>-6.74</b>
<b>Difference (G4:1) - Difference (G4)</b>	<b>6.04</b>
<b>Difference (G4:2) - Difference (G4)</b>	<b>9.10</b>

<sup>(a)</sup> The change of entropies upon external stacking interaction, given by bold typeface, need to be compared with the data in Table 2.

<sup>(b)</sup> The entropy values for the entire system of bases of the G-tetrads correspond to the asymptotic regime ( $t \rightarrow \infty$ ) of MD simulations.

<sup>(c)</sup> The sum of all individual base pair entropies gives the estimate of the total intra-base pair entropy.

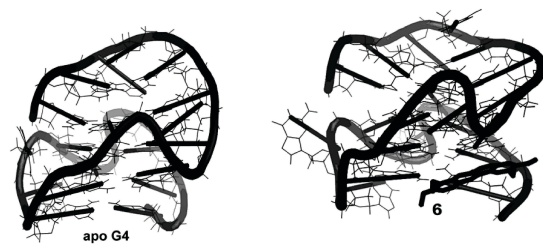
<sup>(d)</sup> The difference of the values Bases and Base pairs gives the estimate of the total inter-base pair entropy.

The sum of entropies of all individual bps (denoted by Base Pairs in Table 3) can be conveniently observed with respect to the entropy of the system of all G-tetrad nucleobases (denoted by Bases in Table 3). By subtracting the Base Pairs values from the Bases values, it is possible to quantitatively evaluate the thermodynamic purpose of total inter-base pair motion (denoted by Difference in Table 3). Only binding of BRACO-19 (**15**) is associated with a negative value of  $-6.74 \text{ kcal mol}^{-1}$ , indicating that the reduced inter-base pair conformational flexibility of G-tetrads contributes unfavorably to the binding free energy. In contrast, the total inter-bp entropy changes of 6.04 and  $9.10 \text{ kcal mol}^{-1}$  contribute favorably to the binding free

energy upon binding of TMPyP<sub>4</sub> (**1**) and CX-3543 (**2**) respectively.

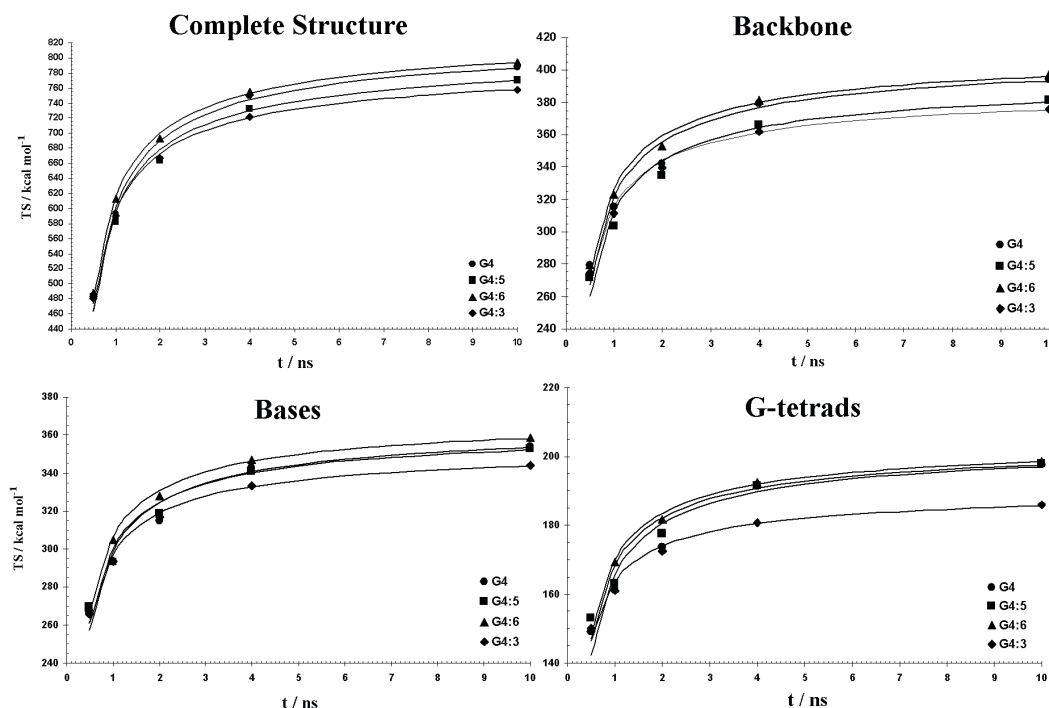
Noteworthy is another uniqueness of BRACO-19 (**15**) among these three ligands. The reduced inter-base pair conformational flexibility of G-tetrads ( $-6.74 \text{ kcal mol}^{-1}$ , Table 3), which is unfavorable thermodynamically, is lined up with a comparable increase of the intra-base pair conformational flexibility of G-tetrads ( $6.74 \text{ kcal mol}^{-1}$ , Table 3) that is favorable thermodynamically.

By considering the structures of the complexes in the stable regime of MD simulation (Figure S2), the values of  $\Delta G_{\text{MM-GBSA}}$  are established to be  $-75.19$ ,  $-53.31$ , and  $-44.82 \text{ kcal mol}^{-1}$  for G4:BRACO-19 (**15**), G4:CX-3543 (**2**), and G4:TMPyP<sub>4</sub> (**1**) respectively. By matching the specific values with those associated with the initial structures of the complexes (Table 1), it turns out that the binding of BRACO-19 (**15**), CX-3543 (**2**), and TMPyP<sub>4</sub> (**1**) with the G4 lowers  $\Delta G_{\text{MM-GBSA}}$  by  $41.91$ ,  $22.74$ , and  $12.86 \text{ kcal mol}^{-1}$  within 4 ns respectively. The stability of the complexes that is based solely on  $\Delta G_{\text{MM-GBSA}}$  may sometimes be overestimated, so that a physically more realistic estimate may be obtained by subtracting  $T \Delta S_{\text{configurational}}$  from  $\Delta G_{\text{MM-GBSA}}$ .<sup>[48]</sup> Subtracting the entropic terms ( $17.22$ ,  $-6.36$ , and  $-35.38 \text{ kcal mol}^{-1}$ , Table 2) from  $\Delta G_{\text{MM-GBSA}}$  ( $-75.19$ ,  $-53.31$ , and  $-44.82 \text{ kcal mol}^{-1}$ ) gives  $-92.41$ ,  $-46.95$ , and  $-9.44 \text{ kcal mol}^{-1}$  for G4:BRACO-19 (**15**), G4:CX-3543 (**2**), and G4:TMPyP<sub>4</sub> (**1**) respectively.



**Figure 8.** Cartoon display of the complex formed upon groove binding of Tetrahydropalmatine (**6**) to apo (ligand-free) G4 in the stable regime of MD simulation around 4 ns.

BRACO-19 (**15**), therefore, shows a far greatest affinity for the G4 in the stable regime of MD simulation. This trend is in line with the observation that the complexes of G4 with BRACO-19 (**15**) and CX-3543 (**2**) are, besides  $\pi$ - $\pi$  stacking, simultaneously stabilized by a certain number of electrostatic interactions - polar contacts (shown by dashed line in Figures 10 and S2). According to the default PyMol definition,<sup>[49]</sup> a polar contact is a hydrogen-bonding interaction having a donor-acceptor atomic distance that is less than or equal to a cut-off value of  $3.6 \text{ \AA}$ . BRACO-19 (**15**) is involved in such five polar contacts: 1 with the sugar moiety of A<sup>1</sup>, 1 with the nucleobase of G<sup>6</sup>, 1 with the nucleobase of G<sup>11</sup>, and 2 with the nucleobase of G<sup>15</sup> respectively (Figure 10).



**Figure 9.** Dependence of configurational entropy ( $TS$  in  $\text{kcal mol}^{-1}$ ,  $T = 300 \text{ K}$ ) on the length of MD trajectory for the various parts of the G4 structure in complex with the groove binding ligands Tetrahydropalmatine (**6**), Sanguinarine (**5**), and Hoechst 33258 (**3**) respectively. Solid lines are fitted functions based on Eq. (5).

CX-3543 (**2**) is involved in such three polar contacts with the nucleobases of G<sup>11</sup>, G<sup>15</sup>, and G<sup>18</sup> respectively (Figure S2). In case of TMPyP<sub>4</sub> (**1**), no such electrostatic interactions are detected.

Based on the discussion in this subsection, BRACO-19 (**15**) can be viewed as a thermodynamically favorable and highly selective G4-DNA binder that is of special interest for the structure-based design of anti-cancer agents.

There is an experimental evidence of the activity of BRACO-19 at the viral G4 level, supporting the use of G-quadruplexes as new anti-HIV-1 targets.<sup>[50]</sup> On the basis of both the present study and our recent report,<sup>[2]</sup> BRACO-19 comes out to fit the context of anti-HIV-1 drug design as a G4 binder that is functionally compatible with a benzophenanthridine derivative (**11**).

### G4 Flexibility Change Upon Groove Binding

The configurational entropies extrapolated to infinitely long MD simulations are given in Table S3 (Supplementary Information), while their changes are reported in Table 4.

Binding of Tetrahydropalmatine (**6**) is associated with a configurational entropy change of 5.54 kcal mol<sup>-1</sup>, indicating some increase of the flexibility of the entire G4-DNA structure. The increased conformational flexibility of G4 contributes favorably to the free energy of non-covalent binding. At the same time, the backbone and the complete system of nucleobases themselves become more flexible as positive entropic changes of 1.98 and 4.27 kcal mol<sup>-1</sup> are observed respectively. Of the increased flexibility of all nucleobases, a slight portion (0.72 kcal mol<sup>-1</sup>) is reflected through the G-tetrad nucleobases.

Groove binding of Sanguinarine (**5**) and Hoechst 33258 (**3**) increases the rigidity of the entire G4-DNA structure as negative configurational entropy changes of -19.85 and -34.96 kcal mol<sup>-1</sup> are observed respectively. It means that Sanguinarine (**5**) contributes less unfavorably than Hoechst 33258 (**3**) to the free energy of non-covalent interaction by about 15.11 kcal mol<sup>-1</sup>. The same trend holds for both the sugar-phosphate backbone and the complete system of nucleobases. The backbone itself shows an entropic decrease of -14.25 kcal mol<sup>-1</sup> for (**5**) and -21.27 kcal mol<sup>-1</sup>

**Table 4.** Change of configurational entropy ( $T\Delta S$  in kcal mol<sup>-1</sup>,  $T = 300$  K) for various parts of G4 structure upon groove binding. Ligand no.: Tetrahydropalmatine (**6**), Sanguinarine (**5**), and Hoechst 33258 (**3**).

G4:Ligand no. - G4	Complete structure	Backbone	Bases	G-tetrads
G4:6 - G4	5.54	1.98	4.27	0.72
G4:5 - G4	-19.85	-14.25	-1.94	-0.09
G4:3 - G4	-34.96	-21.27	-11.73	-13.89

for (**3**). The complete system of nucleobases itself displays an entropic decrease of -1.94 kcal mol<sup>-1</sup> for (**5**) and -11.73 kcal mol<sup>-1</sup> for (**3**).

The configurational entropy for each base pair (bp) of the G-tetrads is given in Table S4 (Supplementary Information). All the individual intra-bp entropies are essentially unchanged ( $\pm 2$  kcal mol<sup>-1</sup>, Table S4) by groove binding of each ligand to the G4. The total intra-bp entropy of G-tetrads (the sum of all the intra-bp entropies) changes by 0.60, -3.85, and -25.31 kcal mol<sup>-1</sup> (Table 5) after having the G4 complexed with Tetrahydropalmatine (**6**), Sanguinarine (**5**), and Hoechst 33258 (**3**) respectively. While the overall intra-bp conformational behavior is essentially unchanged by

**Table 5.** Configurational entropy ( $TS$  in kcal mol<sup>-1</sup>,  $T = 300$  K) evaluated for all individual base pairs making G-tetrads (intra-base pair entropy) and difference from entropy of system of G-tetrad bases (inter-base pair entropy).<sup>(a)</sup> Ligand no.: Tetrahydropalmatine (**6**), Sanguinarine (**5**), and Hoechst 33258 (**3**).

G-tetrads	G <sup>2</sup> G <sup>6</sup> G <sup>11</sup> G <sup>15</sup> , G <sup>3</sup> G <sup>7</sup> G <sup>12</sup> G <sup>16</sup> and G <sup>17</sup> G <sup>4</sup> G <sup>22</sup> G <sup>13</sup>
Bases (G4) <sup>(b)</sup>	205.49
Bases (G4:6)	206.21
Bases (G4:5)	205.40
Bases (G4:3)	191.60
<b>Bases (G4:6) - Bases (G4)</b>	<b>0.72</b>
<b>Bases (G4:5) - Bases (G4)</b>	<b>-0.09</b>
<b>Bases (G4:3) - Bases (G4)</b>	<b>-13.89</b>
Base pairs (G4) <sup>(c)</sup>	481.10
Base pairs (G4:6)	481.70
Base pairs (G4:5)	477.25
Base pairs (G4:3)	455.79
<b>Base pairs (G4:6) - Base pairs (G4)</b>	<b>0.60</b>
<b>Base pairs (G4:5) - Base pairs (G4)</b>	<b>-3.85</b>
<b>Base pairs (G4:3) - Base pairs (G4)</b>	<b>-25.31</b>
Difference (G4) <sup>(d)</sup>	-275.61
Difference (G4:6)	-275.49
Difference (G4:5)	-271.85
Difference (G4:3)	-264.19
<b>Difference (G4:6) - Difference (G4)</b>	<b>0.12</b>
<b>Difference (G4:5) - Difference (G4)</b>	<b>3.76</b>
<b>Difference (G4:3) - Difference (G4)</b>	<b>11.42</b>

<sup>(a)</sup> The change of entropies upon groove binding, given by bold typeface, need to be compared with the data in Table 4

<sup>(b)</sup> The entropy values for the entire system of bases of the G-tetrads correspond to the asymptotic regime ( $t \rightarrow \infty$ ) of MD simulations.

<sup>(c)</sup> The sum of all individual base pair entropies gives the estimate of the total intra-base pair entropy.

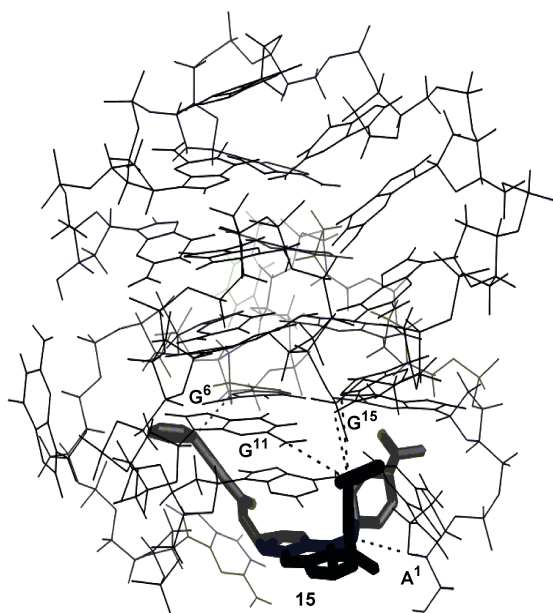
<sup>(d)</sup> The difference of the values Bases and Base pairs gives the estimate of the total inter-base pair entropy.

binding of **(6)**, the reduced total intra-bp conformational flexibility of G-tetrads upon binding of **(5)** and **(3)** is unfavorable to the binding free energy.

The sum of entropies of all individual bps (denoted by Base Pairs in Table 5) can be conveniently observed relative to the entropy of the system of all G-tetrad nucleobases (denoted by Bases in Table 5). Subtracting the Base Pairs values from the Bases values evaluates the thermodynamic purpose of total inter-base pair motion (denoted by Difference in Table 5) quantitatively. The total inter-bp entropy of G-tetrads changes by 0.12, 3.76, and 11.42 kcal mol<sup>-1</sup> upon binding of **(6)**, **(5)**, and **(3)** respectively, thereby contributing favorably to the free energy.

Only, in the case of the G4:Sanguinarine complex, the reduced intra-base pair conformational flexibility (unfavorable thermodynamically) of G-tetrads (−3.85 kcal mol<sup>-1</sup>, Table 5) is in balance with the increased inter-base pair conformational flexibility (favorable thermodynamically) of G-tetrads (3.76 kcal mol<sup>-1</sup>, Table 5).

The values of  $\Delta G_{MM-GBSA}$ , underlying the structures of the complexes in the stable regime of MD simulation (Figure S3), are −28.63, −37.96, and −52.88 kcal mol<sup>-1</sup> for G4:6, G4:5, and G4:3 respectively. Subtracting the entropic term  $T \Delta S_{\text{configurational}}$  (5.54, −19.85, and −34.96 kcal mol<sup>-1</sup>, Table 4) from  $\Delta G_{MM-GBSA}$  (−28.63, −37.96, and −52.88 kcal mol<sup>-1</sup>) gives −34.17, −18.11, and −17.92 kcal mol<sup>-1</sup> for G4:6, G4:5, and G4:3 respectively.

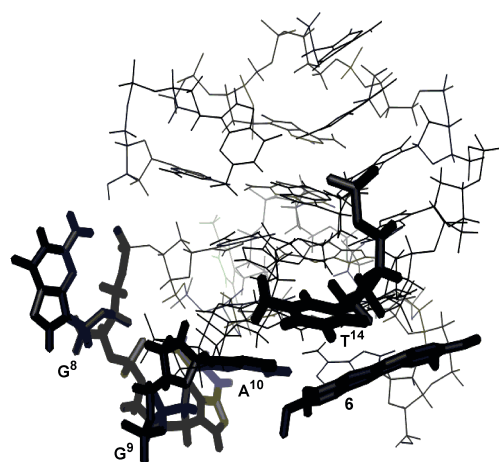


**Figure 10.** BRACO-19 (**15**) is involved not only in  $\pi$ - $\pi$  stacking with the G4, but also in five electrostatic interactions (given by dashed line) with the residues A<sup>1</sup>, G<sup>6</sup>, G<sup>11</sup>, and G<sup>15</sup> in the stable regime of MD simulation around 4 ns.

Overall, Tetrahydropalmatine (**6**) can be seen as a thermodynamically favorable and most specific groove binder.

### Implications for Experimental Research

The structural design of optimal groove (or loop) binders is a challenge, as this mode of interaction is nonspecific in terms of its dependence on the particular arrangement of groove (or loop) residues. Ligand binding in a pure quartet-binding mode is more stable than that in a multiple-binding mode (the simultaneous external stacking and loop binding of ligands). This is likely because loop-binding ligand induces loop rearrangement, thereby influencing the interaction between the side chains of G-quartet-binding ligand and the loops/grooves of G-quadruplex. There are indications that combined ligand loop and G-quartet binding enhances G-quadruplex rigidity and decreases flexibility of both G-quartets and backbone.<sup>[5]</sup> In this light, it is important to highlight the following. BRACO-19, a pure G-quartet-binding ligand, concomitantly takes part in  $\pi$ - $\pi$  stacking with the G<sup>2</sup>G<sup>6</sup>G<sup>11</sup>G<sup>15</sup> tetrad by way of its core aromatic scaffold and in five electrostatic interactions with the residues A<sup>1</sup>, G<sup>6</sup>, G<sup>11</sup>, and G<sup>15</sup> (of which none is loop nucleotide) by way of its side chains (Figure 10). Tetrahydropalmatine, a preferred groove binder, shows inclination to induce some rearrangement of G4 loop (Figure 11). Therefore, we propose the determination of an experimental structure, which would simultaneously embody the external stacking of BRACO-19 and the loop/groove binding of Tetrahydropalmatine to the G4. The necessary structural basis, in combination with remarkable theoretical<sup>[41]</sup> and computational<sup>[51]</sup> advances in the accurate determination of the small



**Figure 11.** Tetrahydropalmatine (**6**) is inclined to induce the rearrangement of loop residues G<sup>8</sup>, G<sup>9</sup>, A<sup>10</sup>, and T<sup>14</sup> by binding to the apo G4 in the stable regime of MD simulation around 4 ns.

ligand/G4 binding free energy, would facilitate the development of effective ligand molecules capable of blocking the interaction of G4 with its binding proteins.

### Final Standpoint

The quasi harmonic approximation only considers the equilibrium structure of molecule, allowing each atomic coordinate to fluctuate around a single equilibrium value [Eq. (4)]. In view of this, it is useful to analyze the calculated entropy change of G4 upon ligand binding by considering an expected enthalpic cost for creating a binding site. The enthalpic cost, or the deformation energy of G4, was estimated to be 38–78 kcal mol<sup>-1</sup>.<sup>[52]</sup> Subtracting the configurational entropy changes (Tables 2 and 4) for the whole G4 structure from this specific value gives the following free energies: a) 20.78–60.78, 44.36–84.36, and 73.38–113.38 kcal mol<sup>-1</sup> for (15), (1), and (2) upon external stacking interaction and b) 32.46–72.46, 57.85–97.85, and 72.96–112.96 kcal mol<sup>-1</sup> for (6), (5), and (3) upon groove binding respectively. The positive values mean that the quasi harmonic approximation holds for the systems under present study. However, the applications of the current methodology are not always straightforward. Some contradictory examples show that the intercalation of aromatic ligand molecules may yield negative free energies because of an overestimation of the true entropy contribution by the quasi harmonic approximation. This inconsistency means that the entire DNA would spontaneously unwind to create an intercalation site.<sup>[53]</sup> When the present methodology is not applicable to very flexible targets (especially proteins), a more relevant theoretical framework, which goes beyond the quasi harmonic approximation,<sup>[48]</sup> is needed.

In general, conformational flexibility of molecules (a large protein or small ligand) is a very important feature. Conformational sampling that generates an ensemble of biologically meaningful conformations is needed for computer aided drug design (CADD). Computational methods, such as the Site Identification by Ligand Competitive Saturation (SILCS) protocol for structure based drug design and the Conformationally Sampled Pharmacophore (CSP) protocol for ligand based drug design, have advantages over other CADD methods that are based on single crystal structure or limited ligand conformations.<sup>[54]</sup> For protein-ligand interactions, more advanced MD techniques, such as replica exchange methods,<sup>[55,56]</sup> can be employed to enhance the sampling efficiency at micro-second scales and reveal new modes of interaction.

### CONCLUSIONS

Among structurally different ligand molecules (Figure 1), those having highest affinity for the G-quadruplex from the c-Myc oncogene promoter region and being involved in different modes of interaction are identified (external

stacking: BRACO-19, TMPyP<sub>4</sub>, and CX-3543; groove binding: Tetrahydropalmatine, Sanguinarine, and Hoechst 33258). The ligands are established to be more prone to bind at the end of G-quadruplex via  $\pi$ - $\pi$  stacking interaction. The groove or loop of G4 is proposed to be a viable site - a recognition motif of interest for the structure-based anti-cancer drug design.

Conformational flexibility changes of the receptor upon ligand binding are observed by evaluating the changes of configurational entropy in the asymptotic regime ( $t \rightarrow \infty$ ) of MD simulation.

The preferred mode of interaction is the external stacking of BRACO-19 as a remarkably increased flexibility of the entire G4-DNA structure is observed, while the backbone and the complete system of nucleobases retain the same trend respectively. The origin of the additional conformational flexibility is dissected by analyzing the configurational entropy contributions at the level of individual base pairs of the G-tetrads and by evaluating the thermodynamic purpose of the total intra-base pair and inter-base pair motions within G-tetrads.

Groove binding of Tetrahydropalmatine is uniquely found to increase the flexibility of the entire G4-DNA structure, thus contributing favorably to the free energy of non-covalent binding.

Implications of the present work for experimental research are discussed.

The way of applying the present methodology to very flexible targets, such as proteins, is clarified.

**Acknowledgment.** Prof. David A. Case of Rutgers University is acknowledged for granting the author an academic license for using the molecular dynamics software package Amber 11 in combination with AmberTools 1.5.

**Supplementary Information.** Supporting information to the paper is attached to the electronic version of the article at: <https://doi.org/10.5562/cca3456>.

PDF files with attached documents are best viewed with Adobe Acrobat Reader which is free and can be downloaded from Adobe's web site.

### REFERENCES

- [1] N. W. Luedtke, *Chimia* **2009**, *63*, 134. <https://doi.org/10.2533/chimia.2009.134>
- [2] P. M. Mitrasinovic, *J. Biomol. Struct. Dyn.* **2018**, *36*, 2292. <https://doi.org/10.1080/07391102.2017.1358670>
- [3] R. J. Harrison, J. Cuesta, G. Chessari, M. A. Read, S. K. Basra, A. P. Reszka, J. Morrell, S. M. Gowan, C. M. Incles, F. A. Tanious, W. D. Wilson, L. R. Kelland, S. Neidle, *J. Med. Chem.* **2003**, *46*, 4463. <https://doi.org/10.1021/jm0308693>

- [4] J. H. Tan, T. M. Ou, J. Q. Hou, Y. J. Lu, S. L. Huang, J. Y. Wu, Z. S. Huang, K. Y. Wong, L. Q. Gu, *J. Med. Chem.* **2009**, *52*, 2825–2835. <https://doi.org/10.1021/jm801600m>
- [5] J. Q. Hou, S. B. Chen, J. H. Tan, H. B. Luo, D. Li, L. Q. Gu, Z. H. Huang, *J. Comput. Aided Mol. Des.* **2012**, *26*, 1355–1368. <https://doi.org/10.1007/s10822-012-9619-1>
- [6] T. E. III. Cheatham, *Curr. Opin. Struct. Biol.* **2004**, *14*, 360–367. <https://doi.org/10.1016/j.sbi.2004.05.001>
- [7] D. J. Cashman, R. Buscaglia, M. W. Freyer, J. Dettler, L. H. Hurley, E. A. Lewis, *J. Mol. Model.* **2008**, *14*, 93–101. <https://doi.org/10.1007/s00894-007-0254-z>
- [8] J. Dai, M. Carver, L. H. Hurley, D. Yang, *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680. <https://doi.org/10.1021/ja205646q>
- [9] B. J. Chen, Y. L. Wu, Y. Tanaka, W. Zhang, *Int. J. Biol. Sci.* **2014**, *10*, 1084–1096. <https://doi.org/10.7150/ijbs.10190>
- [10] A. T. Phan, V. Kuryavii, H. Y. Gaw, D. J. Patel, *Nat. Chem. Biol.* **2005**, *1*, 167–173. <https://doi.org/10.1038/nchembio723>
- [11] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639–1662. [https://doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14%3c1639::AID-JCC10%3e3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14%3c1639::AID-JCC10%3e3.0.CO;2-B)
- [12] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *J. Comput. Chem.* **2009**, *30*, 2785–2791. <https://doi.org/10.1002/jcc.21256>
- [13] D. A. Case, T. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Jr. Merz, A. Onufriev, C. Simmerling, B. Wang, R. Woods, *J. Comput. Chem.* **2005**, *26*, 1668–1688. <https://doi.org/10.1002/jcc.20290>
- [14] D. A. Case, T. A. Darden, T. E. III. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. P. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, *Amber 11*, University of California, San Francisco, CA, **2010**.
- [15] V. Hornak, R. Abei, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, *65*, 712–725. <https://doi.org/10.1002/prot.21123>
- [16] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. A. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, J. A. Pople, *Gaussian 98, revision A.9*, Gaussian, Inc., Pittsburgh, PA, **1998**.
- [17] C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269–10280. <https://doi.org/10.1021/j100142a004>
- [18] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174. <https://doi.org/10.1002/jcc.20035>
- [19] G. D. Hawkins, C. J. Cramer, D. G. Truhlar, *Chem. Phys. Lett.* **1995**, *246*, 122–129. [https://doi.org/10.1016/0009-2614\(95\)01082-K](https://doi.org/10.1016/0009-2614(95)01082-K)
- [20] G. D. Hawkins, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem.* **1996**, *100*, 19824–19839. <https://doi.org/10.1021/jp961710n>
- [21] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, D. A. Case, *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409. <https://doi.org/10.1021/ja981844+>
- [22] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T. E. Cheatham, *Acc. Chem. Res.* **2000**, *33*, 889–897. <https://doi.org/10.1021/ar000033j>
- [23] V. Tsui, D. A. Case, *Biopolymers* **2001**, *56*, 275–291. [https://doi.org/10.1002/1097-0282\(2000\)56:4%3c275::AID-BIP10024%3e3.0.CO;2-E](https://doi.org/10.1002/1097-0282(2000)56:4%3c275::AID-BIP10024%3e3.0.CO;2-E)
- [24] M. K. Gilson, K. A. Sharp, B. Honig, *J. Comput. Chem.* **1987**, *9*, 327–335. <https://doi.org/10.1002/jcc.540090407>
- [25] W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129. <https://doi.org/10.1021/ja00172a038>
- [26] H. C. Andersen, *J. Comput. Phys.* **1983**, *52*, 24–34. [https://doi.org/10.1016/0021-9991\(83\)90014-1](https://doi.org/10.1016/0021-9991(83)90014-1)
- [27] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092. <https://doi.org/10.1063/1.464397>
- [28] I. Andricioaei, M. Karplus, *J. Chem. Phys.* **2001**, *115*, 6289–6292. <https://doi.org/10.1063/1.1401821>
- [29] J. Schlitter, *Chem. Phys. Lett.* **1993**, *215*, 617–621. [https://doi.org/10.1016/0009-2614\(93\)89366-P](https://doi.org/10.1016/0009-2614(93)89366-P)
- [30] S. A. Harris, C. A. Laughton, *J. Phys.: Condens. Matter* **2007**, *19*, 076103. <https://doi.org/10.1088/0953-8984/19/7/076103>

- [31] S. Genheden, U. Ryde, *Expert Opin. Drug Discov.* **2015**, *10*, 449–461.  
<https://doi.org/10.1517/17460441.2015.1032936>
- [32] A. D. Becke, *J. Chem. Phys.* **2002**, *117*, 6935–6938.  
<https://doi.org/10.1063/1.1503772>
- [33] S. Cosconati, S. Forli, A. L. Perryman, R. Harris, D. S. Goodsell, A. J. Olson, *Expert Opin. Drug. Discov.* **2010**, *5*, 597–607.  
<https://doi.org/10.1517/17460441.2010.484460>
- [34] P. M. Mitrasinovic, *Chem. Phys.* **2003**, *286*, 1–13.  
[https://doi.org/10.1016/S0301-0104\(02\)00902-3](https://doi.org/10.1016/S0301-0104(02)00902-3)
- [35] P. M. Mitrasinovic, *Curr. Drug Targ.* **2013**, *14*, 817–829.  
<https://doi.org/10.2174/1389450111314070009>
- [36] P. M. Mitrasinovic, *Med. Chem.* **2014**, *10*, 46–58.  
<https://doi.org/10.2174/157340641001131226122124>
- [37] P. M. Mitrasinovic, *Med. Chem.* **2014**, *10*, 252–270.  
<https://doi.org/10.2174/157340641003140304143442>
- [38] P. M. Mitrasinovic, *J. Chem. Inf. Model.* **2015**, *55*, 421–433. <https://doi.org/10.1021/ci5006965>
- [39] A. Pavlov, P. M. Mitrasinovic, *Curr. Org. Chem.* **2010**, *14*, 129–137. <https://doi.org/10.2174/138527210790069866>
- [40] B. V. Babu, N. K. Konduru, W. Nakanishi, S. Hayashi, P. M. Mitrasinovic, N. Ahmed, *Anti-Cancer Agents Med. Chem.* **2013**, *13*, 307–332.  
<https://doi.org/10.2174/1871520611313020017>
- [41] N. Deng, L. Wickstrom, P. Cieplak, C. Lin, D. Yang, *J. Phys. Chem. B* **2017**, *121*, 10484–10497.  
<https://doi.org/10.1021/acs.jpcc.7b09406>
- [42] J. Zeng, W. Li, Y. Zhao, G. Liu, Y. Tang, H. Jiang, *J. Phys. Chem. B* **2008**, *112*, 2719–2726.  
<https://doi.org/10.1021/jp710029r>
- [43] S. Neidle, *Curr. Opin. Struct. Biol.* **2009**, *19*, 239–250.  
<https://doi.org/10.1016/j.sbi.2009.04.001>
- [44] H. Schäfer, X. Daura, A. E. Mark, W. F. van Gunsteren, *Proteins: Struct., Funct., Genet.* **2001**, *43*, 45–56.  
[https://doi.org/10.1002/1097-0134\(20010401\)43:1%3c45::AID-PROT1016%3e3.0.CO;2-N](https://doi.org/10.1002/1097-0134(20010401)43:1%3c45::AID-PROT1016%3e3.0.CO;2-N)
- [45] H. Schäfer, A. E. Mark, W. F. van Gunsteren, *J. Chem. Phys.* **2000**, *113*, 7809–7818.  
<https://doi.org/10.1063/1.1309534>
- [46] H. X. Zhou, M. K. Gilson, *Chem. Rev.* **2009**, *109*, 4092–4107.  
<https://doi.org/10.1021/cr800551w>
- [47] S. A. Harris, E. Gavathiotis, M. S. Searle, M. Orozco, C. A. Loughton, *J. Am. Chem. Soc.* **2001**, *123*, 12658–12663.  
<https://doi.org/10.1021/ja016233n>
- [48] J. Numata, M. Wan, E. W. Knapp, *Genome Inform.* **2007**, *18*, 192–205.  
[https://doi.org/10.1142/9781860949920\\_0019](https://doi.org/10.1142/9781860949920_0019)
- [49] W. L. DeLano, *PyMol, version 0.99*, De Lano Scientific LLC, South San Francisco, CA, **2006**.
- [50] R. Perrone, E. Butovskaya, D. Daelemans, G. Palù, C. Pannecouque, S. N. Richter, *J. Antimicrob. Chemother.* **2014**, *69*, 3248–3258.  
<https://doi.org/10.1093/jac/dku280>
- [51] F. Moraca, J. Amato, F. Ortuso, A. Artese, B. Pagano, *Proc. Natl. Acad. Sci.* **2017**, *114*(11), E2136–E2145.  
<https://doi.org/10.1073/pnas.1612627114>
- [52] A. N. Lane, J. B. Chaires, R. D. Gray, J. O. Trent, *Nucl. Acids Res.* **2008**, *36*, 5482–5515.  
<https://doi.org/10.1093/nar/gkn517>
- [53] T. Kubar, M. Hanus, F. Ryjacek, P. Hobza, *Chem. – Eur. J.* **2005**, *12*, 280–290.  
<https://doi.org/10.1002/chem.200500725>
- [54] W. Yu, A. D. Mackerell, *Methods Mol. Biol.* **2017**, *1520*, 85–106.  
[https://doi.org/10.1007/978-1-4939-6634-9\\_5](https://doi.org/10.1007/978-1-4939-6634-9_5)
- [55] M. Yang, A. D. Mackerell, *J. Chem. Theory. Comput.* **2015**, *11*, 788–799.  
<https://doi.org/10.1021/ct500993h>
- [56] M. Yang, J. Huang, A. D. Mackerell, *J. Chem. Theory. Comput.* **2015**, *11*, 2855–2867.  
<https://doi.org/10.1021/acs.jctc.5b00243>