

# Mechanistic insight into ligand binding to G-quadruplex DNA

Francesco Saverio Di Leva<sup>1</sup>, Ettore Novellino<sup>2</sup>, Andrea Cavalli<sup>1,3</sup>, Michele Parrinello<sup>4</sup> and Vittorio Limongelli<sup>2,\*</sup>

<sup>1</sup>Department of Drug Discovery and Development, Istituto Italiano di Tecnologia, via Morego, 30, I-16163 Genoa, Italy, <sup>2</sup>Department of Pharmacy, University of Naples “Federico II”, via D. Montesano, 49, I-80131 Naples, Italy, <sup>3</sup>Department of Pharmacy and Biotechnology, Alma Mater Studiorum, University of Bologna, via Belmeloro, 6, I-40126 Bologna, Italy and <sup>4</sup>Department of Chemistry and Applied Biosciences, ETH Zurich, and Facoltà di Informatica, Istituto di Scienze Computazionali, Università della Svizzera Italiana, via G. Buffi, 13, CH-6900 Lugano, Switzerland

Received December 22, 2013; Revised March 10, 2014; Accepted March 13, 2014

## ABSTRACT

Specific guanine-rich regions in human genome can form higher-order DNA structures called G-quadruplexes, which regulate many relevant biological processes. For instance, the formation of G-quadruplex at telomeres can alter cellular functions, inducing apoptosis. Thus, developing small molecules that are able to bind and stabilize the telomeric G-quadruplexes represents an attractive strategy for antitumor therapy. An example is 3-(benzo[d]thiazol-2-yl)-7-hydroxy-8-((4-(2-hydroxyethyl)piperazin-1-yl)methyl)-2H-chromen-2-one (compound 1), recently identified as potent ligand of the G-quadruplex [d(TGGGGT)]<sub>4</sub> with promising *in vitro* antitumor activity. The experimental observations are suggestive of a complex binding mechanism that, despite efforts, has defied full characterization. Here, we provide through metadynamics simulations a comprehensive understanding of the binding mechanism of 1 to the G-quadruplex [d(TGGGGT)]<sub>4</sub>. In our calculations, the ligand explores all the available binding sites on the DNA structure and the free-energy landscape of the whole binding process is computed. We have thus disclosed a peculiar hopping binding mechanism whereas 1 is able to bind both to the groove and to the 3' end of the G-quadruplex. Our results fully explain the available experimental data, rendering our approach of great value for further ligand/DNA studies.

## INTRODUCTION

Nucleic acids contain the genetic information essential for life. In particular, their sequence encodes vital instructions for the cell, such as replication and transcription. The deoxyribonucleic acid (DNA) usually assumes a standard double helix conformation (1). However, in the last decades, DNA has been found to adopt several alternative conformations having specific roles during cell life. Among these are G-quadruplexes (2), that are stacks of G-quartets, also known as G-tetrads, formed by four guanines hydrogen bonded to each other (Figure 1A) (3). These structures are found in important guanine-rich regions of the human genome, such as gene promoters and telomeres (4–9). In particular, telomeres are ensembles of proteins and noncoding DNA that protect chromosomes termini from unwanted events, such as recombination, degradation, and end-to-end fusion. Furthermore, they prevent chromosomes termini to be recognized as DNA double strand breaks (10,11). Thus, they are crucial for cell lifespan. In normal cells, telomeric DNA shortens at each cell cycle and this eventually leads to senescence or apoptotic cell death (12). In tumor cells instead, telomere length is maintained thanks either to a finely tuned mechanism involving the telomerase enzyme (13) or through recombination events between telomeres, a phenomenon known as the alternative lengthening of telomeres (ALT) mechanism (14).

In 1991, Zahler *et al.* found that folding of telomeric DNA into a G-quadruplex structure inhibits *in vitro* telomerase activity (15), and more recently also the ALT cells have been found to be particularly sensitive to the presence of G-quadruplexes (14). Furthermore, in the last few years, the formation of G-quadruplex structures at telomeres has been found to activate the fast DNA-damage response pathway, inducing rapid apoptosis (16,17). Consequently, molecules able to bind and stabilize G-quadruplex structures can se-

\*To whom correspondence should be addressed. Tel: +39 081 678 641; Fax: +39 081 678 107; Email: vittoriolimongelli@gmail.com

lectively alter telomeric functions in cancer cells, thus leading to a marked inhibition of tumor growth (18,19).

To date, many G-quadruplex selective ligands with potential antitumor activity have been developed (20–27). Most of them, called end-stackers, are usually polyaromatic molecules that bind to the ends of the G-tetrads, such as BRACO-19 (28), RHPS4 (29,30), and telomestatin (31,32). Unfortunately, none of these ligands passed to clinical trials because of their poor drug-like properties and selectivity. At variance with end-stackers, another class of ligands can bind to the grooves of the G-quadruplexes. These compounds, known as groove binders, can be more selective than end-stackers by recognizing the specific groove conformations of different DNA structures. However, despite efforts in developing potent and selective groove binders, only few examples have been to date reported in literature (33–38).

In this framework, a number of structural and computational studies have been performed targeting the telomeric G-quadruplex [d(TGGGGT)]<sub>4</sub>. Although this structure is not a representative human telomeric G-quadruplex, it is often used as a working model for its structural simplicity. In fact, it has a 4-fold symmetry, with all parallel strands, four identical grooves, and three monovalent cations at the center of the guanines stacks (Figure 1A and B) (39). In particular, [d(TGGGGT)]<sub>4</sub> has been used to identify ligands able to bind both the groove (33,34,36–38) and the stacking surfaces of G-quadruplex (40). One recent example is a structure-based virtual screening campaign that led to the identification of compound **1** (36), which features a 3-(benzo[d]thiazol-2-yl)-2*H*-chromen-2-one aromatic scaffold with the positively charged N-(2-hydroxyethyl)piperazinyl branch (Figure 1C). This ligand has a marked binding affinity toward [d(TGGGGT)]<sub>4</sub> as revealed by nuclear magnetic resonance (NMR) experiments (36). Furthermore, isothermal titration calorimetry (ITC) experiments showed that **1** binds to [d(TGGGGT)]<sub>4</sub> more tightly than distamycin A (41), that is the reference compound among the G-quadruplex groove binders. These biophysical experiments have been complemented by *in vitro* assays showing that **1** is able to induce DNA damage and cell-cycle arrest in specific cancer cell lines (42).

Despite these efforts, a comprehensive elucidation of the way this compound interacts with DNA is still lacking. In fact, NMR data showed the involvement in the ligand binding of residues both at the groove and at the end of the DNA structure. These observations are suggestive of a complex binding mechanism that was only partially described by the docking calculations where compound **1** was found to bind at the center of the groove (41). Thus, a deeper insight is necessary where considering the full DNA flexibility and the solvent effect during the exploration of the sites available for the ligand binding. These investigations are essential for an exhaustive elucidation of the ligand binding mechanism and to provide the molecular information useful for drug design studies.

With this in mind, we decided to investigate at atomic level the binding mechanism of compound **1** to its target DNA [d(TGGGGT)]<sub>4</sub>. Unfortunately, the ligand binding and unbinding processes are typically a long time-scale event, in the order of microseconds to milliseconds, difficult

to sample with standard techniques, such as molecular dynamics (MD). Thus, the use of enhanced-sampling methods is recommended. Among the emerging techniques, metadynamics (43) has been successful in describing long time-scale biological events and complex ligand binding mechanisms, allowing also to reconstruct the free-energy surface (FES) of the investigated processes (44–48). In the nucleic acids research field, metadynamics has been successfully used to study the unbinding process of small duplex DNA intercalators (49,50), and more recently that of a drug from a protein–duplex DNA complex (51). However, so far this technique has never been applied to investigate the binding of a ligand to a G-quadruplex DNA.

Here, we present a fully unbiased binding study of compound **1** to the G-quadruplex [d(TGGGGT)]<sub>4</sub> during which no structural information on ligand binding is used. To this goal, we performed well-tempered metadynamics simulations (52), which enhance the sampling allowing an exhaustive exploration of all the possible ligand binding sites on the DNA structure in an affordable computational time. The whole binding event is thus described taking fully into account the DNA flexibility together with the presence of explicit solvent and ions. At the end of the simulation, we reconstructed the free-energy landscape of the binding event together with an accurate estimate of the absolute ligand/DNA binding free energy. These calculations allowed us identifying the main interaction sites of compound **1** and revealing the lowest energy ligand binding modes. In particular, we show that compound **1** interacts with G-quadruplex through a hopping binding mechanism passing from one DNA site to the other. This process is ruled by specific ligand/DNA interactions, such as hydrophobic and polar contacts between the ligand aromatic scaffold and the nucleobases, and salt–bridge interactions between the ligand charged chain and the phosphate groups of the G-quadruplex. Compound **1** shows two lowest free-energy binding modes, being able at the same time to bind to the groove and to the 3' end of the G-quadruplex. These findings are of great value for the structural optimization of compound **1** and for *de-novo* rational drug design, since different binding modes to G-quadruplex can be exploited to develop new ligands.

Our study also reveals that DNA flexibility plays an important role during the ligand binding and water molecules contribute to stabilize the most relevant ligand binding poses. The results obtained are in agreement with data coming from NMR experiments and allow us to interpret features of the experiments that were previously unclear. For instance, we show that the T6 residue is involved in the ligand binding to the 3' end of DNA, thus explaining the experimental observations. These findings render our protocol of great value for further investigations on ligand/DNA interaction that can assist drug discovery strategies to develop potent and selective DNA binders.

## MATERIALS AND METHODS

### Molecular docking

Molecular docking calculations of **1** in the crystal structure of the G-quadruplex [d(TGGGGT)]<sub>4</sub> (PDB code: 1s45)

(39) were carried out using the AutoDock4.2 (AD4) software package (53) as implemented through the graphical user interface AutoDockTools (ADT 1.5.4) (54). The ligand tridimensional structure was generated with the Maestro Build Panel (55), and its protonation state at physiological pH (7.4) was assigned using Epik (56). The target DNA structure was prepared through the Protein Preparation Wizard of the graphical user interface Maestro 9.3 (55). Water molecules were removed, hydrogen atoms were added and the co-crystallized  $\text{TI}^+$  ions were replaced with the more physiologically relevant  $\text{K}^+$  ions. Ligand and receptor structures were converted to AD4 format files using ADT and the Gasteiger–Marsili partial charges were then assigned to the ligand and nucleic acid atoms. In order to allow the ligand to explore the whole conformational space, the docking area was centered on the Cartesian coordinates of the center of mass of the G-quadruplex structure and defined by a box large enough to include the whole macromolecule. Thus, grid points of  $90 \times 90 \times 90$  with a 0.375 Å spacing were calculated around the docking area for all the ligand atom types using AutoGrid4.2. Thus, 100 separate docking calculations were performed. Each docking run consisted of 10 millions energy evaluations using the Lamarckian genetic algorithm local search method. Otherwise default docking parameters were applied. Docking conformations were clustered on the basis of the root mean square deviation (rmsd) between the Cartesian coordinates of the ligand atoms (cutoff = 2.0 Å) and were ranked based on the AutoDock scoring function (53). Finally, the lowest energy binding pose provided by preliminary docking calculations was selected as the starting conformation for the subsequent metadynamics simulation. However, it is important to underline that the ligand starting conformation does not affect the final results of a well-converged metadynamics simulation, thus we can arbitrarily choose whatever pose predicted by the docking program as starting structure.

Additional docking calculations using the DNA conformation of the free-energy minimum **Aa** obtained from metadynamics simulation were performed applying the same protocol.

Molecular docking calculations on duplex DNA were performed using the structure of the B-DNA dodecamer [d(CGCGAATTCGCG)]<sub>2</sub> (PDB code: 1bna) (57), which has been reported to be recognized by **1** (42).

### Molecular dynamics

All the simulations were carried out using the standard *parm99* Amber force field for nucleic acids modified using the recently developed *parmbsc0* parameters (58–60). The NAMD 2.8 code (61) combined with the external plugin PLUMED 1.2.2 (62) was used to perform the simulations. The 1/[d(TGGGGT)]<sub>4</sub> complex was solvated in a 12.0 Å layer cubic water box using the TIP3P water model parameters (63).  $\text{K}^+$  cations were used to neutralize the system, with three of these ions placed at the center of the G-tetrads. During the simulations, distance constraints were applied between these ions and their coordinating guanine oxygens. Further 11  $\text{K}^+$  and 11  $\text{Cl}^-$  ions were added to reproduce the NMR experimental conditions of 70 mM KCl. A cut-

off of 10 Å was used for short-range interactions. The long-range electrostatic interactions were computed by means of the particle mesh Ewald method using a 1.0 Å grid spacing in periodic boundary conditions. The SHAKE algorithm was applied to constraint bonds involving hydrogen atoms, with a 2 fs integration time step. Amber charges were applied to the DNA and water molecules, whereas ligand charges were computed using the restrained electrostatic potential (RESP) fitting procedure (64). The ESP was first calculated by means of the Gaussian package (65) using a 6–31G\* basis set at Hartree–Fock level of theory, and then the RESP charges were obtained by a two-stages fitting procedure using Antechamber (66). The system was thus equilibrated through 10 ns MD in the isothermal–isobaric ensemble (NPT) at 1 atm and 300 K before running metadynamics simulations in the canonical isothermal–isochoric (NVT) ensemble.

Standard MD simulations to verify the stability of the metadynamics poses were performed in the NPT ensemble at 1.0 atm and 300 K.

### Metadynamics

The estimation  $F(s, t)$  at time  $t$  of the FES  $F(s)$  as a function of the collective variable (CV)  $s$  was determined by metadynamics (43) in its well-tempered variant (52), using the following formula:

$$F(s, t) = -\frac{T + \Delta T}{\Delta T} V(s, t), \quad ()$$

where  $V(s, t)$  is the bias potential added to the system and  $T$  is the temperature of the simulation.  $\Delta T$  is the difference between the temperature of the CV and that of the simulation. The bias potential is made up by the sum of the Gaussians deposited along the trajectory of the CV. The exploration of the CV space can be increased by tuning  $\Delta T$ . A Gaussian deposition rate of 0.5 kcal/mol/ps was initially used and gradually decreased on the basis of the adaptive bias with a  $\Delta T$  of 2,700 K. Two CVs were chosen to describe the different ligand conformations during the metadynamics run: (i) the distance ( $d$ ) between the center of mass of [d(TGGGGT)]<sub>4</sub> and that of the 3-benzothiazol-2-yl-chromen-2-one scaffold of compound **1**; (ii) the dihedral angle (torsion  $-\varphi$ ) defined by the major inertia axis of the ligand and that of DNA (Supplementary Figure S1A and Supplementary Table S1). Gaussian widths of 0.23 Å and 0.05 rad were used for the  $d$  and  $\varphi$  CVs, respectively. During the metadynamics simulations, an upper limit constraint at 27.0 Å was applied to the  $d$  CV to reduce the conformational space to explore in the fully unbound state.

### The reweighting procedure

To identify the main ligand binding sites on the G-quadruplex [d(TGGGGT)]<sub>4</sub> together with the lowest energy ligand binding conformations, we applied a recently developed reweighting procedure (67). This algorithm allows reconstructing the Boltzmann distribution relative to CVs different from those biased in the metadynamics run. In fact, once the metadynamics simulation is converged, using the



newly computed probability distribution, the FES can be reconstructed as a function of the newly selected CVs. In the present case, two CVs were chosen to distinguish the main ligand binding sites on [d(TGGGGT)]<sub>4</sub>: (i) the projection of the center of mass of the 3-benzothiazol-2-yl-chrom-en-2-one scaffold of compound **1** on the major inertia axis of the target DNA (projection on axis—POA) and (ii) the distance of the same center of mass from this axis (distance from axis—DFA) (Figure 3A and Supplementary Table S1). Finally, to identify the lowest energy ligand binding conformations, the FES was computed as function of (i) the POA CV described above, and (ii) a torsion CV defined by the dihedral angle ( $\psi$ ) between the major inertia axis of the N-(2-hydroxyethyl)piperazinyl tail of compound **1** and that of the G-quadruplex (Supplementary Figure S1B and Supplementary Table S1).

All figures were rendered with PyMOL ([www.pymol.org](http://www.pymol.org)) and GnuPlot.

## RESULTS

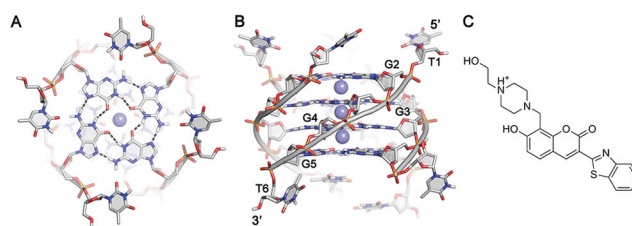
As discussed before, we used well-tempered metadynamics to simulate the binding process of **1** to [d(TGGGGT)]<sub>4</sub>. In any metadynamics simulations, one is required to define a set of properly chosen degrees of freedom, called CVs (52), with whom an adaptive bias is constructed. In such a way, the binding process can be sampled in a relatively short computational time, overcoming large free-energy barriers and reconstructing the FES of the whole binding event (43,52). The choice of the CVs represents a crucial point in this kind of simulations, as they should describe the most relevant slow modes of the investigated process to achieve a reliable estimation of the free energy (68,69).

In the present study, we used a distance ( $d$ ) and a dihedral angle (torsion —  $\varphi$ ) CV (see ‘Materials and Methods’, Supplementary Figure S1A and Supplementary Table S1). In particular, we purposely chose for  $d$  the distance between the center of mass of [d(TGGGGT)]<sub>4</sub> and that of the 3-benzothiazol-2-yl-chrom-en-2-one scaffold of compound **1**. In fact, since [d(TGGGGT)]<sub>4</sub> presents symmetric binding sites relative to the center of mass of DNA, this choice assigns to these sites the same  $d$  values, thus considering them equivalent during the simulation. In addition to the distance  $d$ , the torsion CV describes the different orientations of the ligand relative to DNA during the sampling.

A movie showing the exploration of the DNA binding sites by the ligand during metadynamics simulation can be found as Movie S1.

### The ligand/G-quadruplex binding sites

The whole sampling took  $\sim 1 \mu\text{s}$  of metadynamics simulations. Looking at the FES computed as function of  $d$  and  $\varphi$  (Figure 2), one can note that while these CVs take into account the 4-fold symmetry of the target and they are able to explore all the binding conformations, they have the disadvantage of not being able to discriminate between the different ligand binding modes (see Figure 2 and Supplementary Discussion for details). Therefore, we decided to reconstruct the free-energy landscape along different CVs using the reweighting algorithm of Bonomi *et al.* (67). This algo-

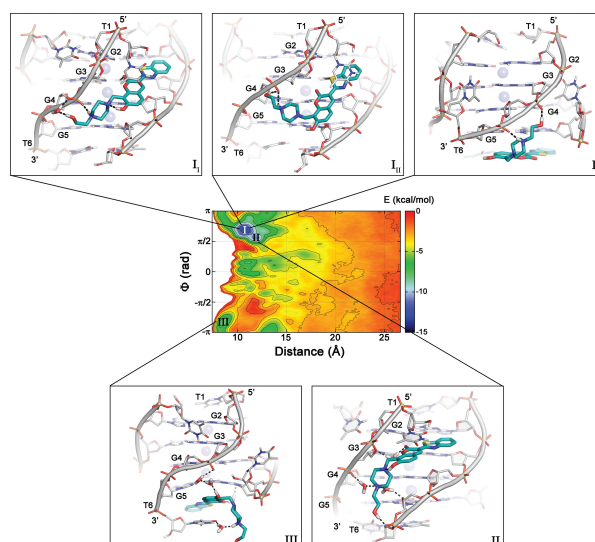


**Figure 1.** (A) Top and (B) side view of the crystal structure of the tetrameric G-quadruplex [d(TGGGGT)]<sub>4</sub> (PDB code 1s45). Nucleotides are shown as sticks colored by atom type and backbone as gray cartoon. K<sup>+</sup> ions are depicted as purple spheres. Hydrogen bonds are represented as dashed black lines. (C) Chemical structure of compound **1** at physiological pH 7.4 (see ‘Materials and Methods’).

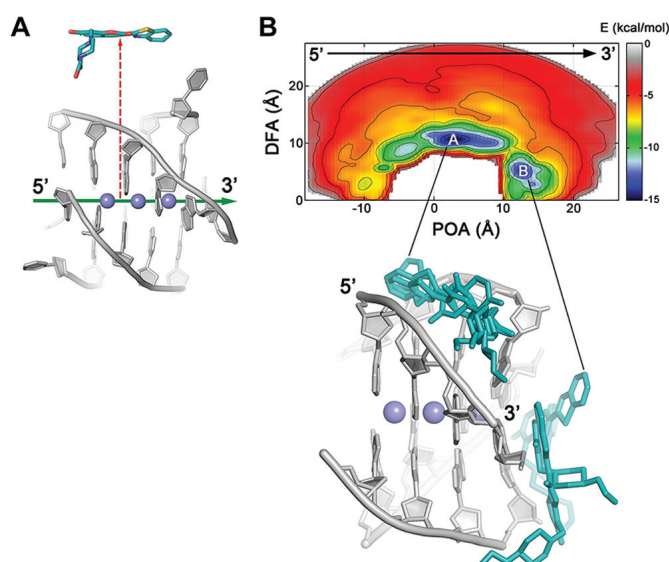
ritm allows computing the FES as function of CVs different from those biased during the metadynamics simulation. As new CVs, we chose a “POA” and “DFA” CVs, where the axis is the major inertia axis of the G-quadruplex (see ‘Materials and Methods’, Figure 3A and Supplementary Table S1). In such a way, one can discriminate among the different ligand binding sites. The reweighted FES shows that the ligand binds to DNA through a hopping mechanism passing from one binding site to the other. In particular, two main free-energy minima, **A** and **B** were found, with basin **A** approximately 1.5 kcal/mol deeper than **B** (Figure 3B). In **A**, the ligand binds to the groove of the G-quadruplex structure, while in **B** compound **1** stacks at the 3' region of DNA, thus showing a peculiar dual binding mode. In fact, to the best of our knowledge, no other G-quadruplex ligand has a similar behavior and this information might be of great value for drug design, since one can exploit both binding modes to develop new ligands. From the FES in Figure 3B, one can see that the ligand binds preferably to the 3' end rather than to 5'. This preference can be explained by the electrostatic properties of [d(TGGGGT)]<sub>4</sub> and in particular by the electric dipole moment that is oriented toward the 5' end (see Supplementary Figure S2). The dipole arises from the different topology at the 3' and 5' end. In particular, the orientation of the negative phosphate groups, which point toward the 3' end of the G-quadruplex, favors the binding of **1** to the 3' terminus.

### The ligand binding modes

Although this FES representation is a marked improvement over that shown in Figure 2, it does not fully separate the different ligand binding modes. In fact, a cluster analysis of basin **A** and **B** poses shows a high heterogeneity of ligand conformations, suggesting the necessity of using CVs different from POA and DFA to distinguish the various ligand binding modes at the groove and at the 3' end. To this end, we recomputed the FES as a function of the POA CV and the torsion CV  $\psi$ . In particular, the latter was purposely chosen to distinguish the different orientations of the N-(2-hydroxyethyl)piperazinyl tail of the ligand relative to the G-quadruplex axis (see ‘Materials and Methods’, Supplementary Figure S1B and Supplementary Table S1). At variance with the previously computed FES, now the free-energy landscape shows three main energy minima, **Aa**, **Ab**, and **B** (Figure 4A), with the lowest energy minimum **Aa**



**Figure 2.** Representation of the binding FES of **1** to  $[d(TGGGGT)]_4$  as function of the distance ( $d$ ) and dihedral ( $\varphi$ ) with isosurfaces displayed every 1.5 kcal/mol. The conformations representing the main free-energy minima are shown as insets. One can note that using this FES representation the groove and the 3'-end binding modes are not unequivocally identified by the free-energy minima and the use of more appropriate CVs is necessary (see the main text for details). DNA is displayed as gray sticks and cartoon, while the ligand is shown as cyan sticks.  $K^+$  ions are depicted as purple spheres. Hydrogen bonds are highlighted as dashed black lines. Nonpolar hydrogens are omitted for clarity.

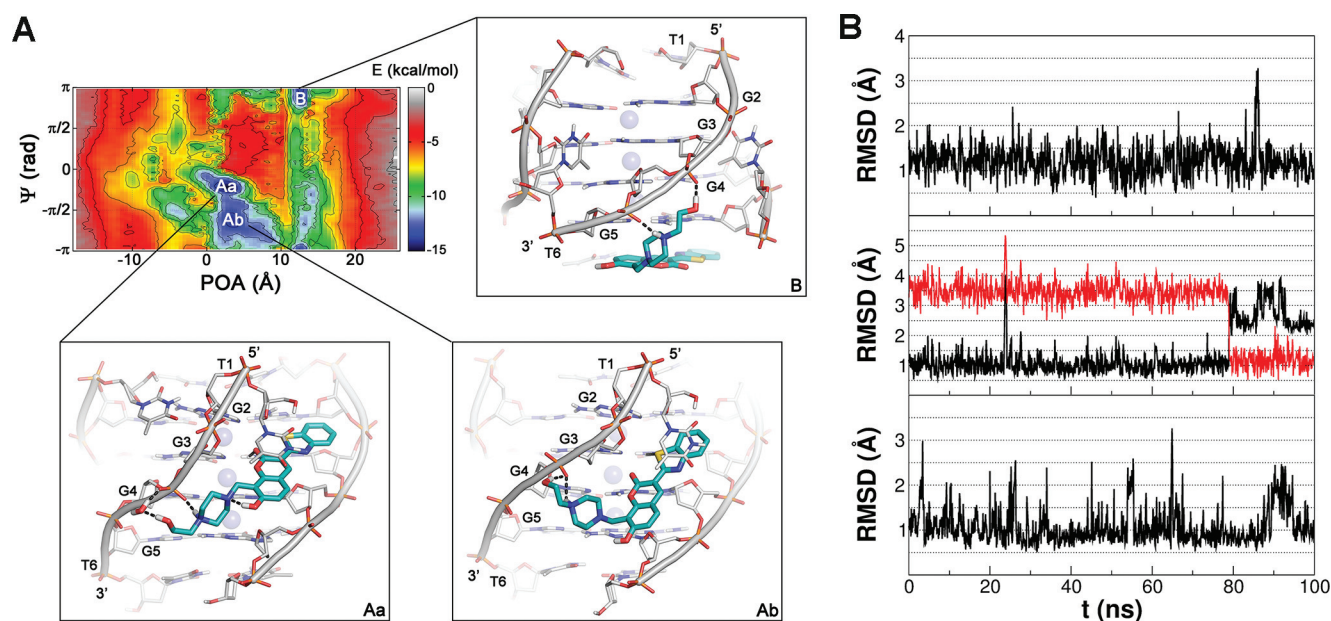


**Figure 3.** (A) Illustration of the “POA” (green arrow) and “DFA” (red dashed arrow) CVs. (B) Representation of the binding FES of **1** to  $[d(TGGGGT)]_4$  as function of the POA and DFA CVs with the most representative ligand binding conformations found at the deepest energy minima **A** and **B**. Isosurfaces of 1.5 kcal/mol are used. DNA is displayed as gray sticks and cartoon, while the ligand is shown as cyan sticks.  $K^+$  ions are depicted as purple spheres. Thymidines and hydrogens are omitted for clarity.

about 0.9 and 1.4 kcal/mol deeper than **Ab** and **B**, respectively. We stress that to obtain the FESs in Figures 3 and 4A we did not perform a new metadynamics calculation, but simply represented in a different and more physically transparent way the sampling data obtained using the variables  $d$  and  $\varphi$ .

### Groove binding

In basin **Aa**, **1** adopts a groove binding conformation in which it interacts mainly with residues G2, G3, G4, and G5 (Figure 4A) on two neighboring strands. In particular, an H-bond is established between the carbonyl oxygen of the ligand coumarin ring and the amino group of G3, while the N-(2-hydroxyethyl)piperazinyl branch of **1** forms a salt-bridge and a water-mediated H-bond with the phosphate group of G4 on the adjacent strand. The presence of this water-mediated interaction was further validated by 100 ns of unbiased MD simulation (Figure 4B), during which the pose resulted stable. Furthermore, it is interesting to note that in the **Aa** pose the conformation of the N-(2-hydroxyethyl)piperazinyl branch is stabilized by an intramolecular H-bond between the 7-hydroxyl group of the coumarin ring and the proximal nitrogen atom of the piperazinyl ring. Furthermore, residue T1 at the 5' end points down and stacks on the ligand aromatic scaffold establishing favorable  $\pi$ -stacking interactions. This pose is fully in agreement with the NMR experiments showing a major involvement of residues G2, G3, G4, and G5 in the binding of **1**. In particular, the NMR spectra show a marked variation of the protons chemical shifts of these guanines in the presence of the ligand (36,41). Further evidence on the involvement of these guanines in the binding of **1** comes from additional NMR experiments that show the decrease of the chemical shift variations of all the guanines when **1** is made to interact with mutated G-quadruplexes, namely,  $[d(TGG^{Br}GGT)]_4$  and  $[d(TGGGG^{Br}T)]_4$  (41). In this case, the presence of the bulky bromine atom inside the grooves hampers the binding of **1**. Finally, the “sandwich-like” conformation adopted by T1 in basin **Aa** (Figure 4A) is in line with the involvement of this residue in the ligand binding shown by the NMR studies (36,41). It is worth mentioning that the **Aa** pose is similar to one of the two binding modes



**Figure 4.** (A) Representation of the binding FES of **1** to [d(TGGGGT)]<sub>4</sub> with the lowest free-energy conformations shown as insets, **Aa**, **Ab**, and **B**. The FES is represented as function of the POA and  $\psi$  CVs using isosurfaces of 1.5 kcal/mol. DNA is shown as gray sticks and cartoon, while the ligand is shown as cyan sticks.  $K^+$  ions are depicted as purple spheres. Hydrogen bonds are represented as dashed black lines. Nonpolar hydrogens are omitted for clarity. (B) Plots of the average rmsd of the heavy atoms of **1** during over 100 ns of MD simulation with the ligand in the **Aa** (upper), **Ab** (middle), and **B** (lower) binding conformation as starting pose. Upper: the ligand binding mode **Aa** is very stable with a low average rmsd value (1.215 Å) and all the ligand/DNA interactions conserved along the whole simulation, including the water-bridge interaction between the ligand N-(2-hydroxyethyl)piperazinyl branch and the phosphate group of G4. Middle: plot of the ligand rmsd in the binding mode **Ab** relative to its initial pose (black line) and to the **Aa** average conformation (red line). The red plot shows that the **Ab** conformation changes in **Aa** after approximately 80 ns. Lower: the ligand binding mode at pose **B** is very stable with a low average rmsd value (1.076 Å) conserving all the ligand/DNA interactions during the whole simulation.

proposed in previous docking studies (Supplementary Figure S3) (41). However, few differences between the docking pose and **Aa** can be noticed. In particular, the docking binding mode is slightly shifted toward the 5' end if compared with **Aa**. In fact, at variance with **Aa**, in the pose predicted by docking the N-(2-hydroxyethyl)piperazinyl group establishes favorable contacts with the phosphate group of G2 instead of that of G4. These differences mainly arise from the fact that the docking algorithm used in this study does not use an explicit solvent model and treats DNA as a rigid body. Thus, water mediated interactions and DNA conformational rearrangement to better accommodate the ligand are not described. As a proof of concept, we redid the docking calculations with the same program using the DNA conformation found in minimum **Aa** (see 'Materials and Methods'). In this test, the docking program finds the metadynamics pose **Aa** to be one of the most favorable binding modes (Supplementary Figure S4). These outcomes confirm the soundness of the previously reported binding mode of **1** but suggest at the same time to take exhaustively into account the target flexibility and solvent effects when more rigorous calculations on ligand binding are needed.

The second energy basin **Ab** (Figure 4A) corresponds to another groove binding conformation in which the ligand aromatic scaffold is slightly shifted toward the 3' end. At variance with **Aa**, in this pose, the sulfur atom of the benzothiazole moiety H-bonds with the amino group of G3. On the other hand, the N-(2-hydroxyethyl)piperazinyl es-

tablishes two polar contacts with the phosphate group of G4 in a way comparable to **Aa**.

To verify the stability of the binding conformations of basin **Aa** and **Ab**, over 100 ns unbiased MD simulations were performed. Standard MD has been indeed successfully used to describe the interactions engaged by ligands locally at the DNA binding site (70). In our simulations, the results show a different behavior in the two cases. In particular, the binding conformation **Aa** is stable during the whole simulation with low ligand rmsd values (Figure 4B), and conserves all the ligand/DNA interactions. In contrast, **Ab** first maintains its original position in the binding site; then, after approximately 80 ns of simulation, it slightly moves toward the 5' end transforming in **Aa**. This finding is not totally surprising since, looking at the FES in Figure 4A, the energy barrier that separates **Ab** from **Aa** is rather small, approximately 1.5 kcal/mol, and thus it can be overcome in the standard MD time scale. This motion can be appreciated looking at Figure 4B where the rmsd values of the ligand with respect to its position in the **Aa** conformation are shown.

Furthermore, to investigate the G-quadruplex binding selectivity of **1** toward duplex DNA, we performed docking calculations on a B-DNA structure (see 'Materials and Methods'). The docking results show that **1** binds to the duplex minor groove similarly to what observed for G-quadruplex (see Supplementary Discussion and Supplementary Figure S5). These results confirm the experimental data showing that **1** is able to bind the duplex DNA,



interfering with the binding of the known B-DNA ligand distamycin A (41). The structural information coming from these computations can be exploited to develop more potent and selective G-quadruplex ligands.

### End-stacking

At variance with **Aa** and **Ab** where **1** shows groove binding modes, in basin **B**, compound **1** lies at the 3' end where it engages favorable  $\pi$ -stacking interactions with G5 and close contacts with two neighboring thymine residues (T6; Figure 4A). However, the N-(2-hydroxyethyl)piperazinyl moiety partially inserts into the groove interacting with the sugar-phosphate backbone of DNA. In particular, the N4 atom of the piperazinyl ring salt bridges with the phosphate group of G5, while the hydroxyethyl group H-bonds with the phosphate of G4. It is worth mentioning that NMR data show a marked chemical shift variation of T6–H6 and T6–methyl protons in  $^1\text{H}$  NMR titrations on the  $1/[\text{d}(\text{TGGGGT})_4]$  complex (36,41), suggesting a role for T6 in the ligand binding. However, this residue is not involved in the previously reported ligand binding modes, while these data are now for the first time rationalized by our results.

The stability of this binding conformation was checked running an over 100 ns unbiased MD simulation during which the ligand binding mode was stable conserving all the ligand/DNA interactions (Figure 4B).

### DNA-ligand binding free energy

The absolute ligand binding free energy ( $\Delta G_{\text{bind}}$ ) can be computed from the FES as the free-energy difference of the system with the ligand in the bound and unbound state. However, to obtain an accurate estimate of  $\Delta G_{\text{bind}}$ , a number of recrossing events between these two states should be observed during the simulation (48). In our simulations, the ligand visits several times both the bound ( $7.0 \leq d \leq 14.0$  Å) and the unbound states ( $24 < d < 26$  Å) as shown in Supplementary Figure S6, leading to a quantitatively well-characterized FES. At the end of the simulation,  $\Delta G_{\text{bind}}$  is equal to  $-9.4 \pm 1.4$  kcal/mol. Unfortunately, experimental data on the binding free energy of this ligand are not yet available. However, the computed value falls in the range from  $-9.0$  to  $-11.0$  kcal/mol, which are typically the experimental  $\Delta G_{\text{bind}}$  values measured for lead compounds (71).

To provide a picture of the convergence of the binding free-energy estimation, the free-energy difference between the bound and unbound state was computed as a function of the simulation time (Supplementary Figure S7). This figure shows that after 870 ns the free energy is converged. In fact, for the rest of the simulation the free-energy difference between bound and unbound states does not considerably change, while the ligand continues visiting both the bound and unbound states.

### DISCUSSION

In the present study, we used metadynamics simulations to study the binding process of **1** to the G-quadruplex  $[\text{d}(\text{TGGGGT})_4]$ . From the metadynamics calculation, we reconstructed the free-energy landscape of the ligand/DNA

binding process, estimating also the absolute DNA/ligand binding free energy. We have thus identified the preferred ligand binding sites on DNA associated with the lowest energy ligand binding modes. In particular, we have shown that **1** interacts with G-quadruplex through a hopping binding mechanism passing from one DNA site to the other. During this process the ligand assumes two lowest free-energy binding modes, one to the groove and the other to the 3' end of the G-quadruplex. This is a peculiar binding behavior with respect to other G-quadruplex ligands that stimulates further experimental and theoretical investigations. For instance, in drug design one might exploit both the binding modes, through a lead optimization procedure of compound **1**, to develop more potent and selective compounds. Our study also shows that solvent and target flexibility play an important role during the ligand binding and neglecting such aspects might lead to inaccurate results. The relevance of our results is enhanced by the fact that some of the previously reported NMR data (41) are here for the first time rationalized. For instance, we have shown that the T6 residue is involved in the 3'-end binding mode, thus explaining the experimental observations.

Our work is an example of fully unbiased binding study of a ligand to DNA where no structural information on ligand binding is used. Here, metadynamics simulations are applied for the first time to describe the ligand binding to G-quadruplex. In our study, all the possible binding sites on the DNA structure are visited and energetically evaluated. Our findings confirm and well integrate the previously reported experimental data, highlighting the importance of using atomistic simulations to assist, complement, and rationalize experiments on biologically relevant phenomena. The results, achieved at a relatively low computational cost, render our protocol valuable for further investigations on various forms of ligand/DNA interaction and useful even for more automated drug discovery strategies to develop potent and selective DNA ligands.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### ACKNOWLEDGMENTS

The authors thank Antonio Randazzo and Bruno Pagano for useful discussions. Computational resources were provided by the Swiss National Supercomputing Center [project ID s362].

### FUNDING

European Union [ERC-2009-AdG-247075]; the Gabriele Charitable Foundation; the Italian Ministry of Education [MIUR-PRIN 2010/2011 E61J12000210001]. Funding for open access charge: Italian Ministry of Education [MIUR-PRIN 2010/2011 E61J12000210001].

*Conflict of interest statement.* None declared.

### REFERENCES

1. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.

2. Bochman, M.L., Paeschke, K. and Zakian, V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.*, **13**, 770–780.
3. Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U.S.A.*, **48**, 2013–2018.
4. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
5. Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
6. Brooks, T.A., Kendrick, S. and Hurley, L. (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.*, **277**, 3459–3469.
7. Sundquist, W.I. and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.
8. Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
9. Lam, E.Y., Beraldi, D., Tannahill, D. and Balasubramanian, S. (2013) G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.*, **4**, 1796.
10. Blackburn, E.H. (1991) Structure and function of telomeres. *Nature*, **350**, 569–573.
11. de Lange, T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.*, **19**, 2100–2110.
12. Blackburn, E.H. (2001) Switching and signaling at the telomere. *Cell*, **106**, 661–673.
13. Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L. and Shay, J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science*, **266**, 2011–2015.
14. Cesare, A.J. and Reddel, R.R. (2010) Alternative lengthening of telomeres: models, mechanisms and implications. *Nat. Rev. Genet.*, **11**, 319–330.
15. Zahler, A.M., Williamson, J.R., Cech, T.R. and Prescott, D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718–720.
16. Gomez, D., O'Donohue, M.F., Wenner, T., Douarre, C., Macadré, J., Koebel, P., Giraud-Panis, M.J., Kaplan, H., Kolkes, A., Shin-ya, K. *et al.* (2006) The G-quadruplex ligand telomestatin inhibits POT1 binding to telomeric sequences *in vitro* and induces GFP-POT1 dissociation from telomeres in human cells. *Cancer Res.*, **66**, 6908–6912.
17. Denchi, E.L. and de Lange, T. (2007) Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature*, **448**, 1068–1071.
18. De Cian, A., Lacroix, L., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J.F. and Mergny, J.L. (2008) Targeting telomeres and telomerase. *Biochimie*, **90**, 131–155.
19. Balasubramanian, S. and Neidle, S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.*, **13**, 345–353.
20. Monchaud, D. and Teulade-Fichou, M.P. (2008) A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.*, **6**, 627–636.
21. Rodriguez, R., Müller, S., Yeoman, J.A., Trentesaux, C., Riou, J.F. and Balasubramanian, S. (2008) A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.*, **130**, 15758–15759.
22. Hamon, F., Largy, E., Guédin-Beaurepaire, A., Rouchon-Dagois, M., Sidibe, A., Monchaud, D., Mergny, J.L., Riou, J.F., Nguyen, C.H. and Teulade-Fichou, M.P. (2011) An acyclic oligoheteroaryle that discriminates strongly between diverse G-quadruplex topologies. *Angew. Chem. Int. Ed.*, **50**, 8745–8749.
23. Casagrande, V., Salvati, E., Alvino, A., Bianco, A., Ciammaichella, A., D'Angelo, C., Ginnari-Satriani, L., Serrilli, A.M., Iachettini, S., Leonetti, C. *et al.* (2011) N-cyclic bay-substituted perylene G-quadruplex ligands have selective antiproliferative effects on cancer cells and induce telomere damage. *J. Med. Chem.*, **54**, 1140–1156.
24. Collie, G.W., Promontorio, R., Hampel, S.M., Micco, M., Neidle, S. and Parkinson, G.N. (2012) Structural basis for telomeric G-quadruplex targeting by naphthalene diimide ligands. *J. Am. Chem. Soc.*, **134**, 2723–2731.
25. Nicoludis, J.M., Miller, S.T., Jeffrey, P.D., Barrett, S.P., Rablen, P.R., Lawton, T.J. and Yatsunyk, L.A. (2012) Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. *J. Am. Chem. Soc.*, **134**, 20446–20456.
26. Haudecoeur, R., Stefan, L., Denat, F. and Monchaud, D. (2013) A model of smart G-quadruplex ligand. *J. Am. Chem. Soc.*, **135**, 550–553.
27. Di Leva, F.S., Zizza, P., Cingolani, C., D'Angelo, C., Pagano, B., Amato, J., Salvati, E., Sissi, C., Pinato, O., Marinelli, L. *et al.* (2013) Exploring the chemical space of G-quadruplex binders: discovery of a novel chemotype targeting the human telomeric sequence. *J. Med. Chem.*, **56**, 9646–9654.
28. Burger, A.M., Dai, F., Schultes, C.M., Reszka, A.P., Moore, M.J., Double, J.A. and Neidle, S. (2005) The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res.*, **65**, 1489–1496.
29. Phatak, P., Cookson, J.C., Dai, F., Smith, V., Gartenhaus, R.B., Stevens, M.F. and Burger, A.M. (2007) Telomere uncapping by the G-quadruplex ligand RHPS4 inhibits clonogenic tumour cell growth *in vitro* and *in vivo* consistent with a cancer stem cell targeting mechanism. *Br. J. Cancer*, **96**, 1223–1233.
30. Leonetti, C., Scarsella, M., Riggio, G., Rizzo, A., Salvati, E., D'Incalci, M., Staszewsky, L., Frapolli, R., Stevens, M.F., Stoppacciaro, A. *et al.* (2008) G-quadruplex ligand RHPS4 potentiates the antitumor activity of camptothecins in preclinical models of solid tumors. *Clin. Cancer Res.*, **14**, 7284–7291.
31. Kim, M.Y., Vankayalapati, H., Shin-ya, K., Wierzbicka, K. and Hurley, L.H. (2002) Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular G-quadruplex. *J. Am. Chem. Soc.*, **124**, 2098–2099.
32. Tauchi, T., Shin-ya, K., Sashida, G., Sumi, M., Okabe, S., Ohyashiki, J.H. and Ohyashiki, K. (2006) Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: *in vitro* and *in vivo* studies in acute leukemia. *Oncogene*, **25**, 5719–5725.
33. Martino, L., Virno, A., Pagano, B., Virgilio, A., Di Micco, S., Galeone, A., Giancola, C., Bifulco, G., Mayol, L. and Randazzo, A. (2007) Structural and thermodynamic studies of the interaction of distamycin A with the parallel quadruplex structure [d(TGGGGT)]<sub>4</sub>. *J. Am. Chem. Soc.*, **129**, 16048–16056.
34. Pagano, B., Virno, A., Mattia, C.A., Mayol, L., Randazzo, A. and Giancola, C. (2008) Targeting DNA quadruplexes with distamycin A and its derivatives: an ITC and NMR study. *Biochimie*, **90**, 1224–1232.
35. Li, Q., Xiang, J., Li, X., Chen, L., Xu, X., Tang, Y., Zhou, Q., Li, L., Zhang, H., Sun, H. *et al.* (2009) Stabilizing parallel G-quadruplex DNA by a new class of ligands: Two non-planar alkaloids through interaction in lateral grooves. *Biochimie*, **91**, 811–829.
36. Cosconati, S., Marinelli, L., Trotta, R., Virno, A., Mayol, L., Novellino, E., Olson, A.J. and Randazzo, A. (2009) Tandem application of virtual screening and NMR experiments in the discovery of brand new DNA quadruplex groove binders. *J. Am. Chem. Soc.*, **131**, 16336–16337.
37. Cosconati, S., Marinelli, L., Trotta, R., Virno, A., De Tito, S., Romagnoli, R., Pagano, B., Limongelli, V., Giancola, C., Baraldi, P. G. *et al.* (2010) Structural and conformational requisites in DNA quadruplex groove binding: another piece to the puzzle. *J. Am. Chem. Soc.*, **132**, 6425–6433.
38. Pagano, B., Fotticchia, I., De Tito, S., Mattia, C.A., Mayol, L., Novellino, E., Randazzo, A. and Giancola, C. (2010) Selective binding of distamycin A derivative to G-quadruplex structure [d(TGGGGT)]<sub>4</sub>. *J. Nucleic Acids*, **2010**, 247137.
39. Cáceres, C., Wright, G., Gouyette, C., Parkinson, G. and Subirana, J.A. (2004) A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tl<sup>+</sup>/Na<sup>+</sup> ions. *Nucleic Acids Res.*, **32**, 1097–1102.
40. Gai, W., Yang, Q., Xiang, J., Jiang, W., Li, Q., Sun, H., Guan, A., Shang, Q., Zhang, H. and Tang, Y. (2013) A dual-site simultaneous binding mode in the interaction between parallel-stranded G-quadruplex [d(TGGGGT)]<sub>4</sub> and cyanine dye 2, 2'-diethyl-9-methyl-selenacarbocyanine bromide. *Nucleic Acids Res.*, **41**, 2709–2722.
41. Trotta, R., De Tito, S., Lauri, I., La Pietra, V., Marinelli, L., Cosconati, S., Martino, L., Conte, M.R., Mayol, L., Novellino, E. *et al.* (2011) A more detailed picture of the interactions between virtual



- screening-derived hits and the DNA G-quadruplex: NMR, molecular modeling and ITC studies. *Biochimie*, **93**, 1280–1287.
42. Cosconati, S., Rizzo, A., Trotta, R., Pagano, B., Iachettini, S., De Tito, S., Lauri, I., Fotticchia, I., Giustiniano, M., Marinelli, L. *et al.* (2012) Shooting for selective drug-like G-quadruplex binders: evidence for telomeric DNA damage and tumor cell death. *J. Med. Chem.*, **55**, 9785–9792.
  43. Laio, A. and Parrinello, M. (2002) Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12562–12566.
  44. Limongelli, V., De Tito, S., Cerofolini, L., Fragai, M., Pagano, B., Trotta, R., Cosconati, S., Marinelli, L., Novellino, E., Bertini, I. *et al.* (2013) The G-triplex. *Angew. Chem. Int. Ed.*, **125**, 2325–2329.
  45. Grazioso, G., Limongelli, V., Branduardi, D., Novellino, E., De Micheli, C., Cavalli, A. and Parrinello, M. (2012) Investigating the mechanism of substrate uptake and release in the glutamate transporter homologue Glt(Ph) through metadynamics simulations. *J. Am. Chem. Soc.*, **134**, 453–463.
  46. Limongelli, V., Bonomi, M., Marinelli, L., Gervasio, F.L., Cavalli, A., Novellino, E. and Parrinello, M. (2010) Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 5411–5416.
  47. Limongelli, V., Marinelli, L., Cosconati, S., La Motta, C., Sartini, S., Mugnaini, L., Da Settimo, F., Novellino, E. and Parrinello, M. (2012) Sampling protein motion and solvent effect during ligand binding. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 1467–1472.
  48. Limongelli, V., Bonomi, M. and Parrinello, M. (2013) Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 6358–6363.
  49. Vargiu, A.V., Ruggerone, P., Magistrato, A. and Carloni, P. (2008) Dissociation of minor groove binders from DNA: insights from metadynamics simulations. *Nucleic Acids Res.*, **36**, 5910–5921.
  50. Wilhelm, M., Mukherjee, A., Bouvier, B., Zakrzewska, K., Hynes, J.T. and Lavery, R. (2012) Multistep drug intercalation: molecular dynamics and free energy studies of the binding of daunomycin to DNA. *J. Am. Chem. Soc.*, **134**, 8588–8596.
  51. Coletta, A. and Desideri, A. (2013) Role of the protein in the DNA sequence specificity of the cleavage site stabilized by the camptothecin topoisomerase IB inhibitor: a metadynamics study. *Nucleic Acids Res.*, **41**, 9977–9986.
  52. Barducci, A., Bussi, G. and Parrinello, M. (2008) Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, **100**, 020603.
  53. Huey, R., Morris, G.M., Olson, A.J. and Goodsell, D.S. (2007) A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.*, **28**, 1145–1152.
  54. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. and Olson, A.J. (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.*, **30**, 2785–2791.
  55. Maestro, version 9.3, Schrödinger, LLC, New York, NY, 2012.
  56. Epik, version 2.3, Schrödinger, LLC, New York, NY, 2012.
  57. Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.E. (1981) Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, **78**, 2179–2183.
  58. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M. Jr, Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
  59. Cheatham, T.E. III, Cieplak, P. and Kollman, P.A. (1999) A modified version of the Cornell *et al.* force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, **16**, 845–862.
  60. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E. III, Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of  $\alpha/\gamma$  conformers. *Biophys. J.*, **92**, 3817–3829.
  61. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L. and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
  62. Bonomi, M., Branduardi, D., Bussi, G., Camilloni, C., Provasi, D., Raiteri, P., Donadio, D., Marinelli, F., Pietrucci, F., Broglia, R.A. *et al.* (2009) PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comp. Phys. Comm.*, **180**, 1961–1972.
  63. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
  64. Bayly, C.I., Cieplak, P., Cornell, W.D. and Kollman, P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: The RESP model. *J. Phys. Chem.*, **97**, 10269–10280.
  65. Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford, CT.
  66. Wang, J., Wang, W., Kollman, P.A. and Case, D.A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.*, **25**, 247–260.
  67. Bonomi, M., Barducci, A. and Parrinello, M. (2009) Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.*, **30**, 1615–1621.
  68. Gear, C.W., Kevrekidis, I.G. and Theodoropoulos, C. (2012) ‘Coarse’ integration/bifurcation analysis via microscopic simulators: micro-Galerkin methods. *Comput. Chem. Eng.*, **26**, 941–963.
  69. Hummer, G. and Kevrekidis, I.G. (2003) Coarse molecular dynamics of a peptide fragment: free energy, kinetics, and long-time dynamics computations. *J. Chem. Phys.*, **118**, 10762–10773.
  70. Khalid, S., Hannon, M.J., Rodger, A. and Rodger, P.M. (2006) Simulations of DNA coiling around a synthetic supramolecular cylinder that binds in the DNA major groove. *Chemistry*, **12**, 3493–3506.
  71. Freire, E. (2009) Thermodynamic approach to the affinity optimization of drug candidates. *Chem. Biol. Drug. Des.*, **74**, 468–472.