Journal of Molecular Graphics and Modelling 38 (2012) 279–289

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Early stage intercalation of doxorubicin to DNA fragments observed in molecular dynamics binding simulations

Hongxing Lei[a,*,1], Xiaofeng Wang[a,b,1], Chun Wu[c,**]

[a] CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China
[b] CAS Key Lab for Biomedical Effects of Nanomaterial and Nanosafety, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China
[c] Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106, USA

ARTICLE INFO

Article history:
Accepted 15 May 2012
Available online 13 June 2012

Keywords:
Anti-cancer drug
Intercalation process
Molecular dynamics

ABSTRACT

The intercalation mode between doxorubicin (an anticancer drug) and two 6-base-pair DNA model fragments (d(CG**AT**CG)$_2$ and d(CG**TA**CG)$_2$) has been well studied by X-ray crystallography and NMR experimental methods. Yet, the detailed intercalation pathway at molecular level remains elusive. In this study, we conducted molecular dynamics binding simulations of these two systems using AMBER DNA (parmbsc0) and drug (GAFF) force fields starting from the unbound state. We observed outside binding (minor groove binding or end-binding) in all six independent binding simulations (three for each DNA fragment), followed by the complete intercalation of a drug molecule in two simulations (one for each DNA fragment). First, our data directly supported that the minor groove binding is the dominant pre-intercalation step. Second, we observed that the opening and flipping of a local base pair (A3–T10 for d(CGATCG)$_2$ and C1–G12 for d(CGTACG)$_2$) in the two intercalation trajectories. This locally cooperative flipping–intercalation mechanism was different from the previously proposed rise–insertion mechanism by which the distance between two neighboring intact base pairs increases to create a space for the drug insertion. Third, our simulations provided the first set of data to support the applicability of the AMBER DNA and drug force fields in drug–DNA atomistic binding simulations. Implications on the kinetics pathway and drug action are also discussed.

1. Introduction

The anthracyclines doxorubicin (Fig. 1) and daunomycin, consisting of an anthraquinone ring and an amino sugar group, are two anticancer drugs that are effective in more types of cancer than any other classes of cancer chemotherapy agents [1]. The anti-cancer activity of these drugs is likely due to their intercalation into DNA, which may disrupt replication and transcription of genomic DNA and lead to the death of cancer cells [2]. Most early studies have been focused on cytotoxicity, sequence specificity and binding affinity of this intercalation mode [3–5]. For instance, it has been shown that the anthracyclines have stronger binding toward alternating purine–pyrimidine sequences over non-alternating sequences [6,7] and a slight binding preference for G–C base pair over A–T base pair [8]. Only after the determination of the structures of the DNA–anthracycline complexes by X-ray diffraction method [9–12], the detailed structural information of the intercalation mode was finally revealed: the anthraquinone ring is sandwiched between two neighboring base pairs. Based on these structures, the intercalation process has been assumed to follow a rise–insertion mechanism: the distance between two consecutive base pairs increases to create a space for drug insertion while the H-bond pairing within the two base pairs remains intact [13].

Yet, this rise–insertion mechanism cannot explain the complicated dynamic behavior observed in the kinetics studies by ultrafast methods such as stopped-flow or temperature-jump relaxation methods using absorption or fluorescence detection. For example, based on the binding kinetics data between daunomycin and calf thymus DNA, Chaires et al. [14] proposed that the drug–DNA binding process consists of three sequential steps (a three-step model): a rapid "outside" binding, drug intercalation, and slow conformational adjustment of the DNA–drug complex. Rizzo et al. [15] further suggested that two additional branching steps take place at the first and the third step of the three-step model, corresponding to the formation of a weak off-pathway complex and an additional conformational rearrangement of the bound complex, respectively. These experiments all pointed to a more complicated dynamics of the intercalation process. However, due to the low-resolution nature of these experiments, it was unfeasible to elucidate the detailed structural information at each step.

* Corresponding author.
** Corresponding author.
E-mail addresses: leihx@big.ac.cn (H. Lei), cwu@chem.ucsb.edu (C. Wu).
[1] Contributed equally to this work.

1093-3263/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2012.05.006

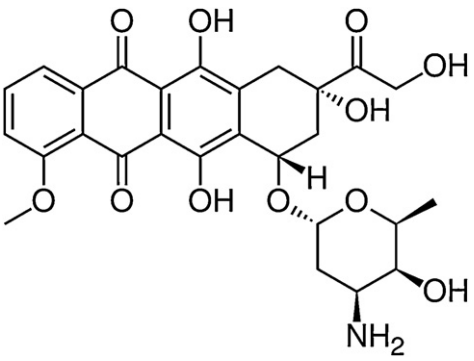

Fig. 1. Chemical structure of doxorubicin.

To probe the minimum binding free energy pathway, Mukherjee et al. [16] recently constructed an intercalative structure (i.e. daunomycin + DNA fragment) from the crystal complex structure as the bound state and a hypothetical minor groove-bound state from docking as the "outside bound" state. Using AMBER ff99 DNA force field and TIP3P water model, they simulated the unbinding process (i.e. transforming the bound state to the unbound state via the outside bound state) using umbrella sampling to probe the binding free energy landscape. This unbinding simulation provided a good estimation of the intercalation free energy barrier, and further supported that the minor groove-bound state is the "outside bound" intermediate state towards the final intercalation. Yet, the dynamic adjustments of the intercalated drug–DNA complex in the latter steps could not be obtained from this equilibrium thermodynamics methodology, and the detailed intercalation process at atomic level remains elusive. In addition, the validity of AMBER DNA force field [17] in drug–DNA binding simulation [18] remains to be established.

In this study, starting from an unbound state (a B-DNA fragment + two free doxorubicin molecules), we performed all-atom molecular dynamics (MD) binding simulations with explicit water. We studied two model DNA sequences (d(CG**AT**CG)$_2$ and d(CG**TA**CG)$_2$) which have been well studied by X-ray diffraction [10,11] and solution NMR method [19,20]. These experimental studies have shown that both sequences share the same binding sites at CpG sites (Fig. 2b) in spite of the order change of the two middle nucleotides (AT vs. TA). The simulations allowed us to validate the force fields and to probe the structural and energetic nature of the dynamic binding process with high spatial and temporal resolution. From the simulation trajectories, we observed multiple binding modes including end-stacking, minor groove binding and intercalation modes. We assessed the structural and energetic properties of these binding modes. The structural deformations of DNA in these binding modes were also compared with those in the simulations with the experimental complex structure and DNA-only system. Next, we analyzed the pathways in the two trajectories with the complete intercalation of a drug molecule. Our in-depth analyses showed that the insertion of the drug was directly coupled with a local base flipping after an outside binding. This observed flipping–intercalation mechanism is completely different from the rise–insertion mechanism which requires a global rise between the two base pairs to create a space for the insertion of the drug in the absence of any base pair flipping. Finally, implications of our simulation results on simulation force fields and the experimental kinetics models will be discussed.

2. Materials and methods

2.1. Simulation systems

We constructed six simulation systems from the DNA–drug complex (d(CG**AT**CG)$_2$ + doxobubincin) solved by X-ray diffraction (PDB ID: 1D12) [11], each solvated in a water box of truncated octahedron with Na^+ as counter ions to neutralize the system (Table 1). The crystal symmetry information in the pdb file was used to generate the double stranded DNA structure. The first two were DNA-only systems (sequences d(CG**AT**CG)$_2$ and d(CG**TA**CG)$_2$) (Fig. 2a for sequence d(CG**AT**CG)$_2$), in which doxobubincin was removed from the crystal structure and the DNA fragment was relaxed to B-form. The structure for d(CG**TA**CG)$_2$ was obtained by switching the AT bases of the X-ray structure of d(CG**AT**CG)$_2$. The DNA fragment had six base pairs with a total charge of −10, thus 10 Na^+ were added as counter ions to neutralize the system. The third and fourth systems were the crystal complex with one drug molecule (Fig. 2b for sequence d(CG**AT**CG)$_2$). Since the net charge of the drug was +1, additional 9 Na^+ were added to neutralize the system. The first four systems were used as reference systems. The fifth system includes the DNA fragment (Fig. 2c for sequence d(CG**AT**CG)$_2$) plus two free drug molecules that were 10 Å away from the DNA, thus requiring only 8 Na^+ as counter ions. Given two bound drugs observed in the X-ray structure, we added two drug molecules to enhance binding chance as compared to systems with only a single drug molecule. This 3:1 base pair–drug

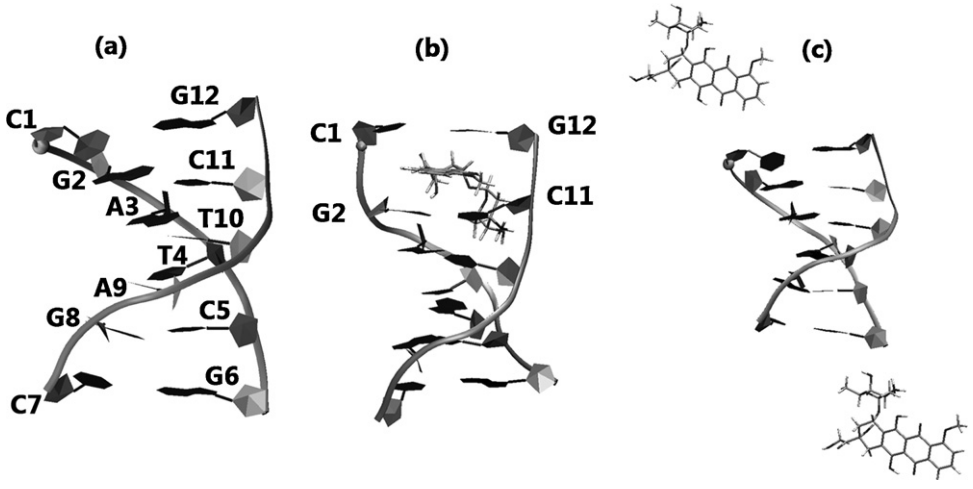

Fig. 2. Initial structures of the simulated systems. (a) The six base pair DNA fragment. (b) The DNA fragment with one drug molecule from the X-ray complex structure (PDB code: 1D12). (c) The DNA fragment with two free drug molecules.

Table 1
Summary of the molecular dynamics simulations performed in this study.

System ID	DNA	No. of ligand	No. of run	Pre-equilibrium	NPT equ. (ns)	NVT equ. run (ns)	Total time (ns)
1	d(CGATCG)2	–	3	No	1	99	300
2	d(CG**TA**CG)2	–	3	No	1	99	300
3	d(CGATCG)2	1× Doxorubicin	1	No	1	99	100
4	d(CG**TA**CG)2	1× Doxorubicin	1	No	1	99	100
5	d(CGATCG)2	2× Doxorubicin	3	Yes	1	99	300
6	d(CG**TA**CG)2	2× Doxorubicin	3	Yes	1	99	300

ratio is also comparable to the experimental value (3.6:1) derived from the binding kinetics data between daunomycin and calf thymus DNA [14]. The sixth system is almost the same as the fifth system except that the DNA sequence d(CG**AT**CG)$_2$ was replaced with d(CG**TA**CG)$_2$ [19,20]. The minimal distance from any atom of the solute to the water box wall was set to 10 Å for all six systems, leading to 2481, 2348, 2668, 2251, 7811 and 7778 water molecules for the six systems, respectively.

A refined version of the AMBER ff99 (parmbsc0) [17] was chosen and correctly applied to represent the DNA fragments, TIP3P model [21] was chosen to represent water, and the default parameters for Na^+ were adapted from an early study by Aqvist [22]. This force field is commonly used in the simulation of nucleic acids [16,23–25]. The parameters for doxobubincin were developed by following standard AMBER protocol: the electrostatic potential of doxobubincin molecule was obtained at the HF/6-31G* level after geometry optimization at the same level; the partial charges were derived by fitting the electrostatic potential using the RESP (Restrained ElectroStatic Potential) method [26] and other parameters were taken from the AMBER GAFF [27] parameter set. The parameter file in AMBER PREP format is included in the supporting information (SI).

2.2. Simulation protocols

The simulations were conducted using the AMBER 9 simulation package [28]. After energy minimization for the initial structure of the six systems (Fig. 2), a total of 14 runs (Table 1) were performed with different initial random velocities. A production run (100.0 ns), performed at 310 K, included a short 1.0 ns molecular dynamics in the NPT ensemble mode (constant pressure and temperature) to equilibrate the system density and 99.0 ns dynamics in the equivalent NVT ensemble mode (constant volume and temperature). The initial positions of the two drugs are shown in Fig. 2C. To randomize the orientations and positions of the two free drug molecules an additional 10.0 ps pre-run at 500 K was conducted for the last two unbound complex systems (i.e. DNA with two free drug molecules) in the NPT mode, in which the DNA was subjected to Cartesian restraints (1.0 kcal/mol/Å). Indeed, their positions and orientations diverged and distributed widely (see Fig. S5A), SHAKE [29] was applied to constrain all bonds connecting hydrogen atoms, which enables a 2.0 fs time step in the simulations. The particle-mesh Ewald method [30] was used to treat long-range electrostatic interactions under periodic boundary conditions (charge grid spacing of ~1.0 Å, the fourth order of the B-spline charge interpolation; and direct sum tolerance of 10^{-5}). The cutoff distance for short-range non-bonded interactions was 10 Å, with the long-range van der Waals interactions based on a uniform density approximation. To reduce the computation, non-bonded forces were calculated using a two-stage RESPA approach [31] where the short range forces were updated every step and the long range forces were updated every two steps. Temperature was controlled by using Berendsen's algorithm [32] with a coupling constant of 2.0 ps. To eliminate the "block of ice" problem [13,33], the translation and rotation of the center of mass were removed every 500 steps. The trajectories were saved at 2.0 ps intervals for analyses.

2.3. Convergence of simulations

For the stability simulations of the DNA-only and the crystals complex, the root mean square deviation (RMSD) of DNA backbone (+ligand) heavy atoms was calculated against the starting structure. The flat and small RMSDs indicated that these systems were stable and the simulations reached a steady state. For example, the RMSD profiles for the two sequences are shown in Figs. S1 and S2.

For the binding simulations, atom contacts between the DNA fragment and each of the drug molecules were calculated using an atom-to-atom distance cutoff of 3.0 Å. The simulation systems reached a steady state, as indicated by the stable contact number. For example, the contact number vs. time for sequence d(CG**AT**CG)$_2$ is shown in Fig. S3. We defined stable complex as a complex with the number of atom contacts greater than 12. Two free drugs bound to different sites as shown in the last snapshots for six simulations (Fig. 3), indicating a good sampling of binding.

2.4. Binding mode identification

Because the DNA backbone was relatively stable in the binding process, we aligned the DNA backbone of the complexes from the trajectories by a least square fitting. The aligned complexes were clustered into different structural families based on the 2 Å pairwise RMSD cutoff of the drug molecule using Daura algorithm [34], in which the number of neighboring structures was calculated for every structure based on the RMSD cutoff, the structure with the largest number of neighbors plus its neighboring structures were removed to form a structure family and the process continued for the remaining structures until all structures had been assigned into structural families. The centroid structure (i.e. the structure having the largest number of neighbors in the structural family) was used to represent the family. As an example, the centroid structures of populated structural families (>1% of total structure population) for sequences d(CG**AT**CG)$_2$ and d(CG**TA**CG)$_2$ are shown in Table S1a and S1b, respectively. Based on visual inspection, the centroid structures were further merged into super-families corresponding to major binding modes (end-binding, minor groove binding and intercalation).

2.5. Geometrical parameters for characterizing DNA conformation

We calculated six base-base parameters, four base pair-axis parameters and six base pair-step parameters for the DNA fragment in a Cartesian coordinate system (X: short axis of paired base plane, Y: the long axis of paired base plane and Z: the DNA helix direction.) using the CURVES+ program [35]. The six base-base parameters were shear, stretch, stagger, buckle, propeller, and opening, describing the relative orientation of one base with respect to the other in a base pair. The four base pair-axis parameters were *X*-displacement, *Y*-displacement, inclination and tip, describing the relative orientation of a rigid base pair with respect to the helix axis. The six local base pair step parameters

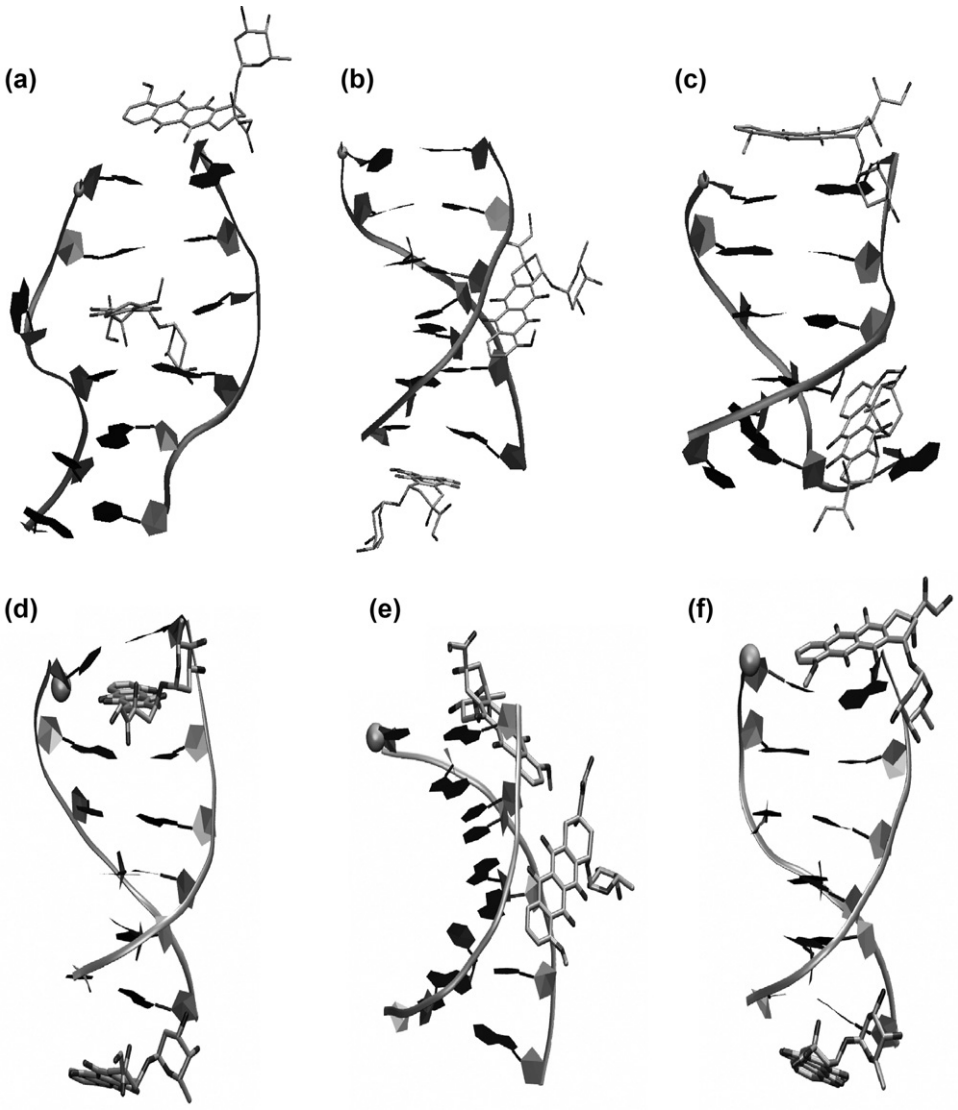

Fig. 3. The last snapshots of the six binding simulations: (a)–(c) d(CG**AT**CG)$_2$ and (d)–(f) d(CG**TA**CG)2.

included shift, slide, rise, tilt, roll and twist, describing translational and rotational motions between two consecutive base pairs. The sugar puckering that characterizes the sugar conformation [36,37] was also obtained using CURVES+ program. As an example, these local structure values for sequence d(CG**AT**CG)$_2$ are shown in Table S2.

To characterize the global structural features, we calculated three additional global parameters: the ratio of the end-to-end distance to the DNA curved length (Bending ratio), the burial of solvent accessible surface (Burial of SAS) upon binding of the drug to the DNA and the number of hydrogen bonds (H-bonds) within DNA and between the DNA and the drug. These parameters characterized the bending of DNA chain, the binding between the DNA and the drug and the H-bonding between base pairs and between the DNA and the drug, respectively. The bending ratio was calculated using the 3DNA program [38], and the SAS was obtained using VMD [39]. The H-bond number was obtained by following the common criteria: the donor–acceptor distance <3.3 Å and the donor–acceptor–hydrogen angle <25°. There are three H-bonds for each C–G base pair and two for each A–T base pair, so the total number of H-bond for this 6-base-pair DNA segment is 16 in the standard B-conformation. As an example, these global structure values for sequence d(CG**AT**CG)$_2$ are shown in Table S3.

2.6. Geometrical parameters for characterizing DNA–drug complex

We calculated two structural parameters for the DNA–drug complexes: the center-to-center distance D and the drug-base dihedral angle θ. The distance D was defined as the distance from the DNA center to the drug molecule center. The dihedral angle θ was defined as the dihedral angle between the plane of the second or fourth base pair of the DNA and the drug's ring plane. For the complex with an intercalated drug, we calculated an additional structural parameter: a drug insertion angle φ which is the angle between the long axis (the H-bond direction) of the unflipped base pair (A3–T10 for d(CG**AT**CG)$_2$ and G2–C11 for d(CG**TA**CG)$_2$) and the long axis of the drug ring plane.

2.7. Corrected binding energy

MM-GBSA [40] (Molecular Mechanics-Generalized Born/Surface Area) module in the AMBER package (iGB1 model with mBondi radii set, salt concentration of 0.2 M, and surface tension of 0.0072 kcal/Å^2) was used to analyze the energetics of the bound complexes to avoid the large energy fluctuation of explicit solvent. A recent benchmark study [41] has shown that

GB models give quite reliable results even for charged molecules when the relative solvation energy is considered. For this analysis, ions were removed from the systems by assuming that GBSA gives a good estimate on the solvation energy of charged DNA systems. Before binding with ligands, DNA fragments usually adopt stable B-form conformation. After the binding with ligands, the B-form conformation can deform a lot. Therefore, the DNA deformation penalty is an important part of the true binding energy started from free B-DNA and ligand. However, the typical MM-GBSA protocol for calculating the binding energy of a complex does not take the DNA deformation energy into the consideration, as the binding energy is calculated from the energy difference between the total energy of the complex and the sum of the individual energy of DNA receptor and ligand, obtained from the bound complex by a simple separation (i.e. the separated DNA may not be in the B-form). We refer this binding energy as the uncorrected binding energy, which does not include the DNA deformation penalty. To estimate the deformation penalty, we simply calculated the energy difference between the deformed DNA from the complex (the separated DNA) and the B-form DNA at the beginning of simulations. Next, the true binding energy of a bound complex structure was obtained by adding this deformation penalty to the uncorrected binding energy. In this paper, we refer this true binding energy as the corrected binding energy. Finally, the three energy terms of each binding mode were estimated from all structures belonging to the corresponding binding mode. Please note that since the solute conformational entropy is not included in our analysis, the binding energies by MM-GBSA may over-estimate the true binding free energy (i.e. the binding affinity). But when the solute conformational entropies in different binding modes are comparable, the relative binding free energy can be estimated from the relative MM-GBSA binding energies [41]. As an example, these energetic values for both sequences are shown in Table S4.

3. Results

3.1. Stability test of the force fields

To validate the force fields for both DNA and the drug, we constructed four control systems from the crystal structure for stability simulations: DNA-only and DNA–drug complex for sequences d(CG**AT**CG)$_2$ and d(CG**TA**CG)$_2$. All these systems were stable at 310 K over the course of the 100 ns simulations. Here we discuss the data from sequence d(CG**AT**CG)$_2$ as an example. For the DNA-only system, the DNA fragment was stable in all three simulations, as indicated by the small root mean square deviation of the backbone from the starting structure (RMSD = 2.4 ± 0.3 Å, Fig. S1 of SI). It retained a good double helix structure, indicated by the normal DNA geometrical parameters for all base pairs except for the two terminal base pairs (Table S2). The fluctuation of the terminal base pairs was expected, as one face of the base pairs was exposed to solvent. For the DNA–drug crystal complex, the DNA structure and the intercalation of the drug between the first two base pair (i.e. C1–G12 and G2–C11) were well maintained over the course of the simulation, judged by the small RMSD of the DNA backbone (averaged RMSD = 1.9 ± 0.3 Å, shown in Fig. S2) and the good DNA geometrical parameters for all base pairs except for the first two between which the drug intercalates (Table S2). Indeed, the force fields for both the DNA and the drug were adequate to describe the stable structure of the DNA and the DNA–drug complex [17].

3.2. Binding simulations

Starting from an unbound state (Fig. 2C), we carried out three independent simulations (100 ns each) for each DNA sequence (i.e. d(CG**AT**CG)$_2$ and d(CG**TA**CG)$_2$). At the end of 100 ns, the two ligands were bound to the DNA fragment in all six simulations (Fig. 3). Among the six ligands in the three simulations of d(CG**AT**CG)$_2$, three were bound to the ends of the DNA fragment, two were bound to the minor groove and the other one was intercalated between base pair G2–C11 and base pair T4–A9. Similar results were observed for the three simulations with d(CGTACG)$_2$) except that the intercalation was between C1–G12 and G2–C11. These data indicated at least three major binding modes (end binding, minor groove binding and intercalation mode). Of particular interest were the two intercalation trajectories. In the following sections, the structural and energetic features of these binding modes and the two intercalation pathways will be discussed. Since there was almost no interaction between the two drugs (see Fig. S4), we separated the 6 simulations into 12 trajectories which contain one DNA and one drug for further analysis.

3.3. Drug binding modes

To visualize the drug binding sites on d(CG**AT**CG)$_2$, we superimposed the DNA chains of the stable complexes from all six binding trajectories (Fig. S5B). Four populated drug clusters were identified: two clusters at the two DNA ends, one cluster spreading in the middle of the DNA, and a long cluster spreading along the minor groove of the DNA. To gain more insight into these binding modes, the stable complexes were classified into different structural families based on the clustering analysis as described in Section 2.

From the clustering analysis, 19 structural families of complexes with population over 1% were identified from the trajectories. The centroid structure and properties of each complex family are presented in Table S1 of SI. By merging these families, four major binding modes were obtained (Fig. 4) which can be further merged into three categories: (a) in mode A, the drug rings stacked on the π surface of the DNA ends. It should be noted that the end-binding is less relevant for long DNA duplex *in vivo*, because the ends are rare. However, this type of end mode may play an important role in stabilizing the G-rich quadruplets located at the ends of the chromosomes [33]. (b) In mode B, the drug ring inserted into the minor groove of the DNA, and the long axis of the drug ring was almost parallel to one of the DNA backbone chain. (c) In modes C and D, the drug intercalated between the base pairs G2–C11 and T4–A9, whereas the base pair A3–T10 opened and the two bases flipped out in mode C, but the T10 base paired with the ring part of the drug and only the base A3 flipped out in mode D.

To characterize the DNA conformation in the four binding modes as well as in the two reference systems, local geometry parameters in these six modes were calculated as described in Section 2 and shown in Table S2. Notable features of the six DNA modes are summarized here: (1) the terminal base pairs had a large fluctuation in all six modes; (2) all base pairs except for the terminal ones maintained a good geometry in modes A and B; (3) in modes C and D, the third base pair (A3–T10) had abnormal values, resulting from the opening of this base pair. In addition, three global geometry parameters (see Section 2) are shown in Fig. S6 and are also listed in Table S3. As expected, the DNA conformation did not change much in binding mode A. The bend ratio, the SAS (solvent accessible surface) and the H-bond number of DNA were comparable to those of the DNA-only system. In mode B, the DNA bended slightly, indicated by a lower bend ratio value (0.81 ± 0.08). But none of the base pairs was opened and the total H-bond number increased by ∼1 due to the formation of ∼1 DNA–drug H-bond. In modes C and D, the opening of base pair A3–T10 led to some notable global conformational changes in the DNA: big DNA bending and loss of ∼2 H-bonds compared to the DNA-only system. But when the drug was taken into consideration, the total burial of SAS and the H-bond number were actually enhanced. For example, the SAS burial in mode D was

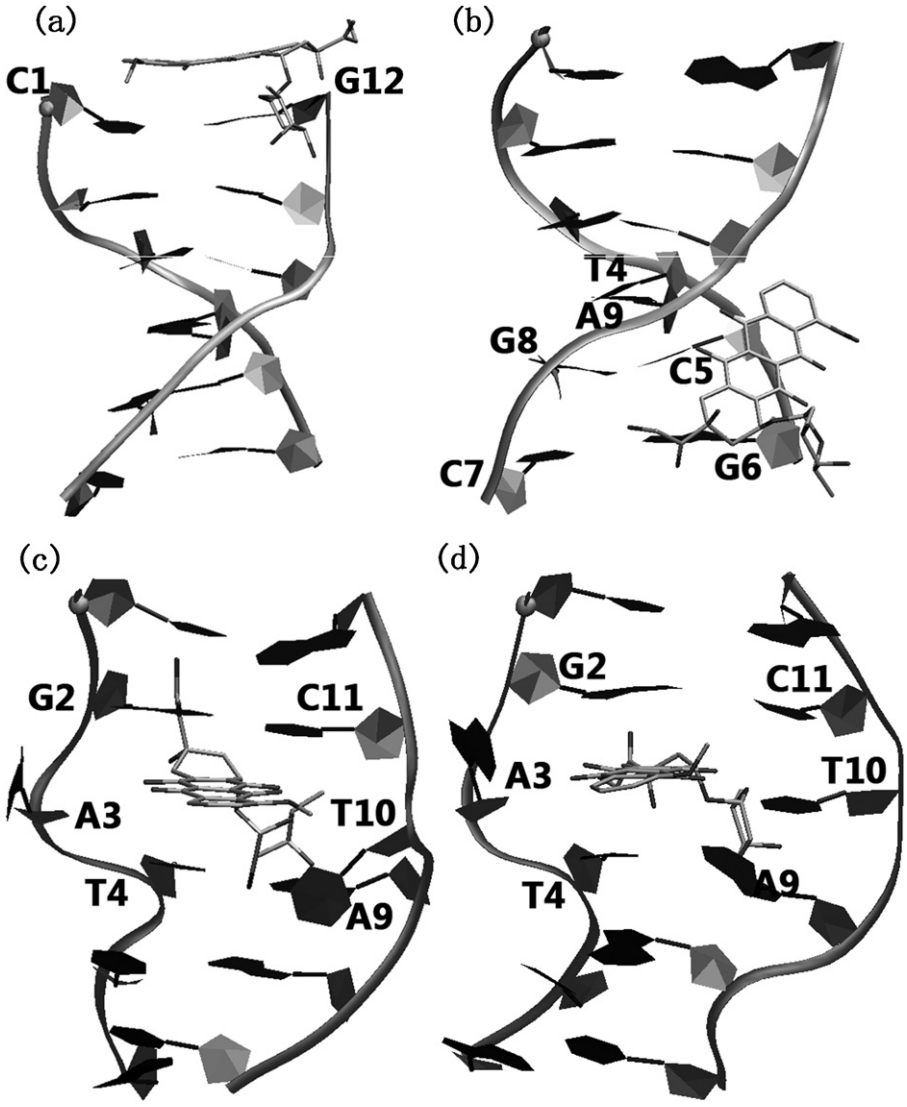

Fig. 4. Representative complex structures of the four major binding modes from the binding simulations to d(CG**AT**CG)$_2$.

~550 Å^2 and the H-bond number was increased by ~2. In the crystal complex, the DNA helix was stretched and untwisted, indicated by the increase of bend radio, decrease of twist and increase of rise in base pair step 1C/2G. In the mean time, the base pairs were intact, indicated by the similar H-bond number as in the DNA-only system.

To examine the relative stability of the bound complexes with reference to the free B-form DNA and ligand, the corrected binding energy including the DNA deformation penalty as described in Section 2 was evaluated for each binding mode as well as the crystal complex. The order of the stability was as follows (Fig. S7 and Table S4): the crystal structure with one drug (-21.5 ± 5.7 kcal/mol) > intercalation with one base flipping out (mode D, -20.6 ± 3.2 kcal/mol) > intercalation with two bases flipping out (mode C, -18.2 ± 2.1 kcal/mol) > minor groove binding (mode B, -15.8 ± 4.7 kcal/mol) > end-stacking (mode A, -12.1 ± 3.1 kcal/mol).

To gain further insight, the two components of the corrected binding energy for each mode are compared as below:

In the intercalation modes C and D, two faces of the drug rings interacted with the bases above and below, and the positively charged sugar part had a favorable electrostatic interaction with the negatively charged phosphate group in the minor groove. Thus, the two modes should have very similar uncorrected binding energy between the drug and the deformed DNA. Indeed modes C and D, respectively, had -44.5 ± 5.2 kcal/mol and -41.2 ± 2.3 kcal/mol of the uncorrected binding energy. However, the corrected bind energy should be different due to the differences in the DNA deformation penalty. In mode C, base pair A3–T10 opened and flipped out. In contrast, only base A3 flipped out and base T10 partially came back and paired with the drug ring part in mode D. Thus, the DNA in mode D was less deformed than the DNA in mode C, leading to a lower energetic penalty for mode D (20.6 ± 3.4 kcal/mol) than that for mode C (25.9 ± 4.8 kcal/mol). Put together, mode D is more stable than mode C by ~-2.4 kcal/mol. In the crystal structure, although no base pair opens, the distance between C1–G12 and G2–C11 needs to increase to accommodate the drug (up to 7.7 Å). Such deformation leads to a global cascade of deformation as the base pairs below and above all have to adjust accordingly to accommodate the drug, causing higher energetic penalty (23.1 ± 6.6 kcal/mol) than that in mode D (20.6 ± 4.8 kcal/mol). But the uncorrected binding energy for the crystal structure (-44.5 ± 4.5 kcal/mol) was more favorable than mode D (-41.2 ± 2.3 kcal/mol). In total, the crystal complex was more favorable than mode D by 0.9 kcal/mol.

In mode B, the corrected binding energy of this minor-groove binding was between the end-stacking and the intercalation modes

(modes C, D and the crystal mode). The minor groove of the DNA allowed a good fit by the ring part of the drug (i.e. the aromatic rings laid parallel to the DNA minor groove walls, stretching across three to four base pairs), and the sugar part was outside of the double helix. Thus, the DNA fragment had a slight bending and maintained a good double helix conformation, leading to a smaller deformation penalty (5.9 ± 2.2 kcal/mol) than the intercalation modes (20–25 kcal/mol). On the other hand, the uncorrected binding interactions (-15.8 ± 4.7 kcal/mol) was much weaker than the intercalation modes (~ -20 kcal/mol). Altogether, the minor-groove binding was less stable than the intercalation modes by 3–5 kcal/mol.

In mode A, end-stacking had the least favorable corrected binding energy, likely due to the least atomic contacts with the DNA.

The above binding mode analysis were also performed on the other sequence d(CG**TA**CG)$_2$ (Fig. S7 and Table S4). Although the site of the intercalation mode was different, the trend of stability was the same: intercalation mode > minor groove binding > end-stacking.

The binding energy was decomposed to van der Waals energy (VDW), the surface energy term (SUR) and the electrostatic term (GBELE: GB + ELE) for both sequences (Fig. S7 and Table S4). Interestingly, whereas the VDW and SUR interactions contributed to the binding, GBELE was against the binding. The VDW interactions contributed most in all binding modes for these charged systems.

3.4. The intercalation pathway

Out of the 12 binding trajectories to the two DNA sequences, there were two leading to the intercalation (one for each sequence). These trajectories enabled us to monitor the detailed dynamics of the intercalation process (movies of these trajectories are included in SI). The two intercalation routes of sequences d(CG**AT**CG)$_2$ and d(CG**TA**CG)$_2$ are shown by some representative structures with their appearance time in Figs. 5 and 6, respectively.

The intercalation pathway of d(CG**AT**CG)$_2$ was as follows (Fig. 5): (a) unbound state → (b) minor groove binding (mode B) → (c) initial insertion → (d) deep insertion with two bases flipping out (mode C) → (e) deep insertion with one base flipping out (mode D). Basic structural features for modes B, C and D have been described in the previous section; some extra structural features are summarized as below. In the first state of the binding route (Fig. 5a), the distance D of the drug to the DNA center was more than 17 Å, corresponding to the unbound state. In the second state (Fig. 5b), the ring part of the drug inserted into the minor groove, stretching across three base-pairs (G2–C11, A3–T10 and A9–T4) of the DNA, whereas the sugar part of the drug was out of the DNA double helix. The third state (Fig. 5c) may serve as a transition state from minor groove-bound state to the intercalation state. The drug went back and forth to find a good orientation for insertion, whereas the base pair A3–T10 started to open and close under the influence of the drug. The fourth state (Fig. 5d) was in mode C, the distance D was 3.59 Å and the drug-base dihedral angle θ was 14.21°, indicating the drug ring part inserted deeper and the ring plane was almost parallel to the planes of the base pair above or below. It is worth to note that bases T10 and A3 completely flipped out of the double helix (Fig. 5d, right). The last state (Fig. 5e) was in binding mode D. Base T10 partially flipped back and formed one H-bond with the drug ring and another H-bond with the drug sugar (Fig. 5e, right). The drug insertion angle to the DNA was 86.73°, while it was 47.88° in the previous cluster (Fig. 5d and e, right). The continuous changes of energetic and structural properties are shown and described in Fig. S8 of SI.

The intercalation pathway of d(CG**TA**CG)$_2$ is as follows (Fig. 6): (a) unbound state → (b) breaking base pair C1–G12 upon initial ligand insertion → (c) flipping out of C1 and deep ligand insertion → (d) wrong hydrogen bond paring between C1 and G12 → (e) flipping out of G12 → (f) good paring between C1 and G12. Multiple flipping of C1 and G12 can be monitored by following the continuous change of the base-to-base distance between C1 and G12 (Fig. 7): whereas the distance was ~5 Å without fluctuation before and after the ligand intercalation, the distance fluctuated during the ligand intercalation from ~15 ns to 90 ns. The continuous changes of more energetic and structural properties are shown and described in Fig. S9 of SI.

4. Discussion & conclusion

The anthracyclines doxorubicin and daunomycin as two of most effective anticancer drugs function by interacting with DNA in different binding modes, among which the insertion between DNA base pairs and formation of an intercalation state might be the most important mode. Thus, the mechanism of the intercalation process is critical for understanding the anticancer activity of the drugs. Based on the crystal structures of small model DNA fragments compounded with anthracyclines [9–12], the rise–insertion mechanism has been implicitly assumed [13]. Yet, this simple mechanism does not explain well the multiple step kinetic behavior observed in the ultra-fast kinetic experiment [15]: a rapid "outside" binding, a weak off-pathway complex, drug intercalation, and conformational adjustment of the DNA–drug complex and an additional conformational rearrangement of the complex. Based on umbrella sampling along the "unbinding" pathway, Mukherjee et al. [16] recently suggested that the "outside" binding is a minor groove binding mode, but the rise–insertion mechanism was still assumed for the later process of intercalation. Clearly, there is an inconsistency between the kinetic model and the rise–insertion mechanism regarding the additional slow conformational rearrangement of the complex because this additional rearrangement is not necessary for the rise–insertion mechanism.

To probe the detailed dynamics of the intercalation process as well as to validate force fields in direct binding simulations, we constructed two unbound DNA systems with two free doxorubicin drugs using two well-studied DNA fragments [10,11,19,20] (d(CG**AT**CG)$_2$ and d(CG**TA**CG)$_2$) and conducted six independent binding simulations (three for each construct). At the end of 100 ns, all 12 drug molecules bound to DNA in three modes: end-binding, mini-groove binding and intercalation modes. Of particular interest are the two intercalation trajectories observed for both DNA fragments. To the best of our knowledge, this is the very first time that the intercalation processes have been observed in MD simulations with explicit solvent. From the two intercalation trajectories, we observed that the outside binding (minor groove binding) was an intermediate state towards the final intercalation state. This is consistent with the structural feature of DNA where the minor groove is accessible for the outside molecules, especially positively charged ones. Minor groove binding is enthalpically driven with little cost of entropy. Our observation on the minor groove binding further confirms the previous proposal [16] that the minor groove binding mode is the "outside" binding in the kinetic model.

More importantly, we observed opening and flipping of the local base pair (A3–T10 for the sequence d(CG**AT**CG)$_2$ and C1–G12 for the sequence d(CG**TA**CG)$_2$) upon the drug intercalation. Although there were some terminal base pairs opened in our DNA-only system, no flipping occurred in the middle base pairs and the DNA fragments maintained the double helix structure well. Therefore, the base-pair flipping of A3–T10 was due to the DNA–drug interactions. While minor groove binding is a downhill transition, there is an activation barrier between the minor groove-bound state and the intercalation state. Chaires et al. [14] estimated this free energy barrier to be 14.9 kcal/mol. The favorable interactions between DNA and the drug might compensate the lost of favorable H-bonds between

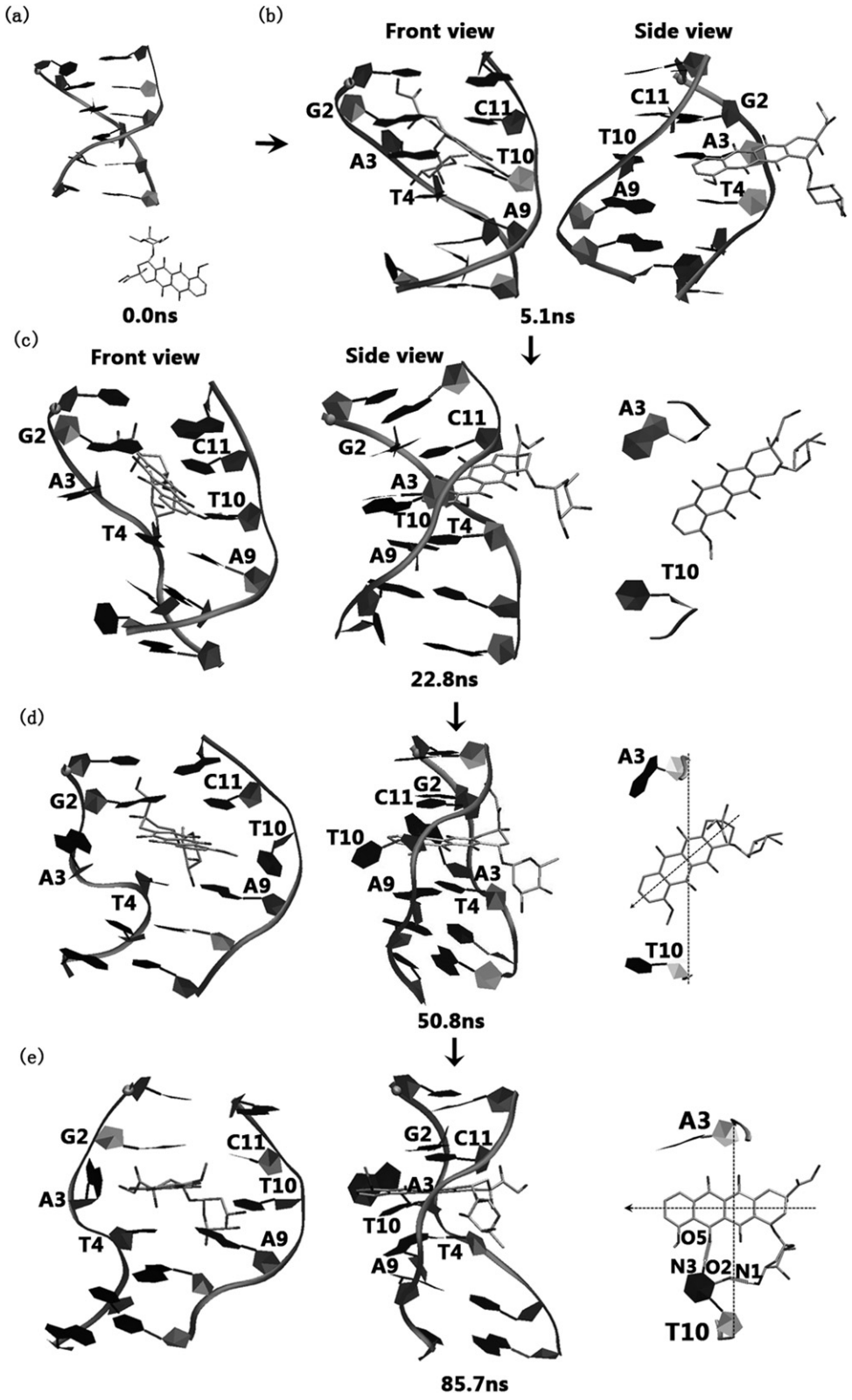

Fig. 5. Representative structures in the intercalation trajectory of sequence d(CG**AT**CG)$_2$.

the base pair and thus facilitates the local opening and flipping of the base pair, in turn the base flipping makes room for the drug insertion. Put together, it is a locally cooperative binding process.

We note that the final intercalation mode of the d(CG**AT**CG)$_2$ fragment in our simulation is different from the intercalation mode observed in the crystal complex structure in two respects: (1) our mode D had one base (A3) flipping out, while there is no base flipping in the crystal complex structure; (2) whereas our intercalation mode is located on top of the base pair T4–A9, the experimental structure is located above the base pair G2–C11. As to the first

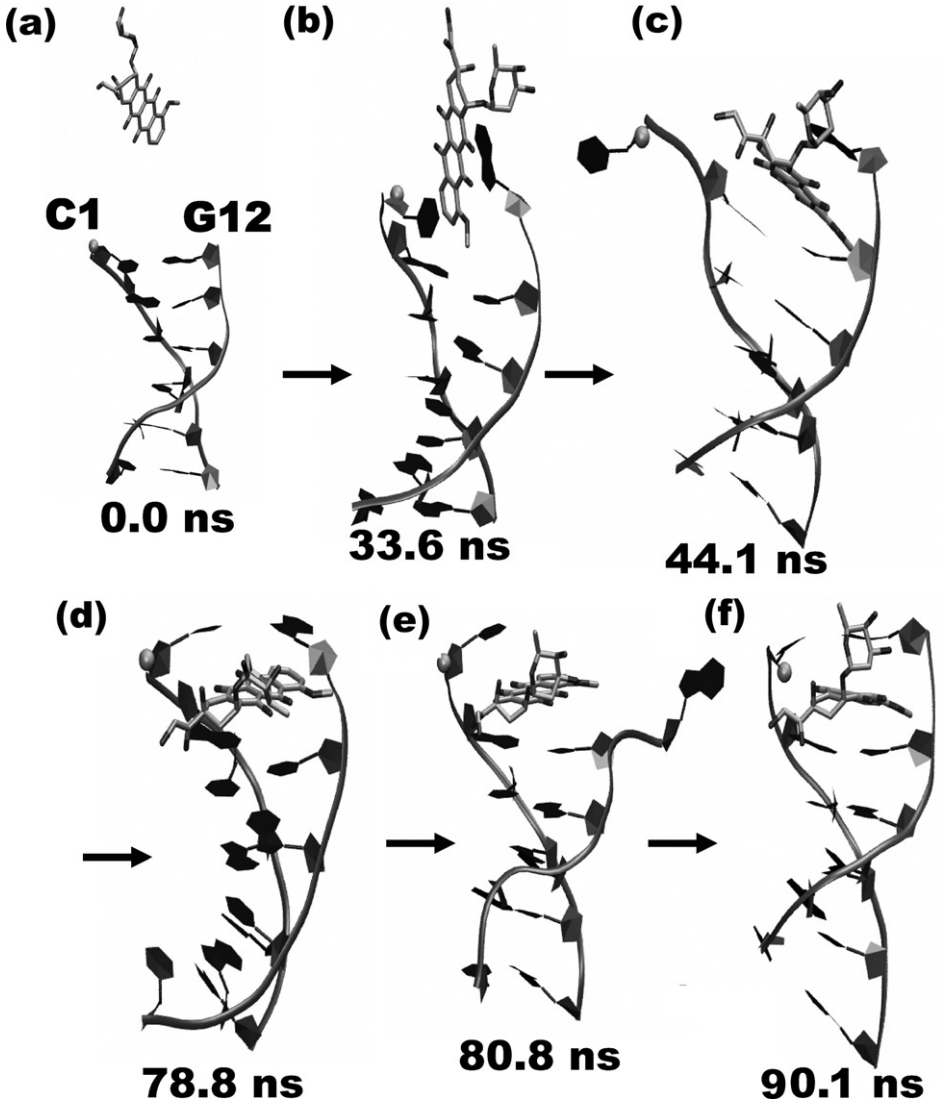

Fig. 6. Representative structures in the intercalation trajectory of sequence d(CG**TA**CG)$_2$.

inconsistency, longer simulations might be required for flipping back of A3 and the formation of base pair A3–T10. As to the second inconsistency, previous experiments [12,40,42] and computationally studies [43–45] show that A–T is also binding site for these anthracyclines with a lower preference over G–C. Therefore, the drug might have multiple binding sites including both A3–T10 and G2–C11 for this DNA fragment. Indeed, we have observed the crystal intercalation mode (G2–C11) in our binding simulations to the closely related d(CG**TA**CG)$_2$ fragment.

Our simulations provide the first set of drug–DNA atomistic binding simulations for validating newly refined AMBER DNA (parmbsc0) and drug (GAFF) force fields. It is very encouraging that the minor groove binding is observed in all simulations and our second intercalative trajectory reproduces the binding site observed in both NMR [19,20] and X-ray studies [10,11]. These data combined with previous simulation data [16,23–25] support the applicability of the AMBER force fields in studying DNA–drug binding.

Our observations on base flipping may provide an alternative intercalation mechanism to the rise–insertion mechanism. This putative base-flipping mechanism has the following two advantages. First, the DNA deformation energy due to the local base flipping (−20.6 kcal/mol) is less than that of the global stretching

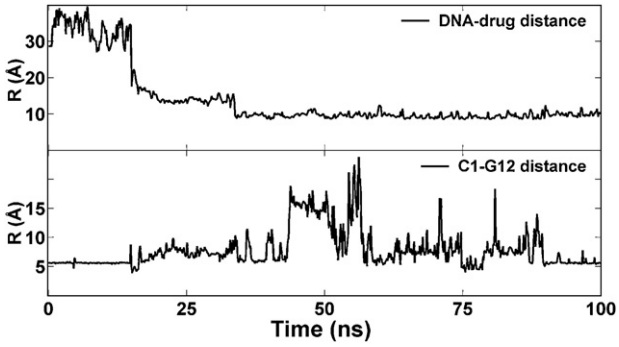

Fig. 7. Continuous change of C1–G12 distance (bottom) upon the intercalation (top) of drug into d(CG**TA**CG)$_2$.

(−23.1 kcal/mol). Second, the DNA deformation penalty is overcome by the favorable binding energy during the intercalation process. In another word, the drug induced flipping lowers the drug insertion free energy barrier. Thus, we suggest that the base pair flipping might be obligatory step or at least an important alternative pathway leading to the intercalation. Of course, the applicability of our intercalation model obtained from limited simulation data with

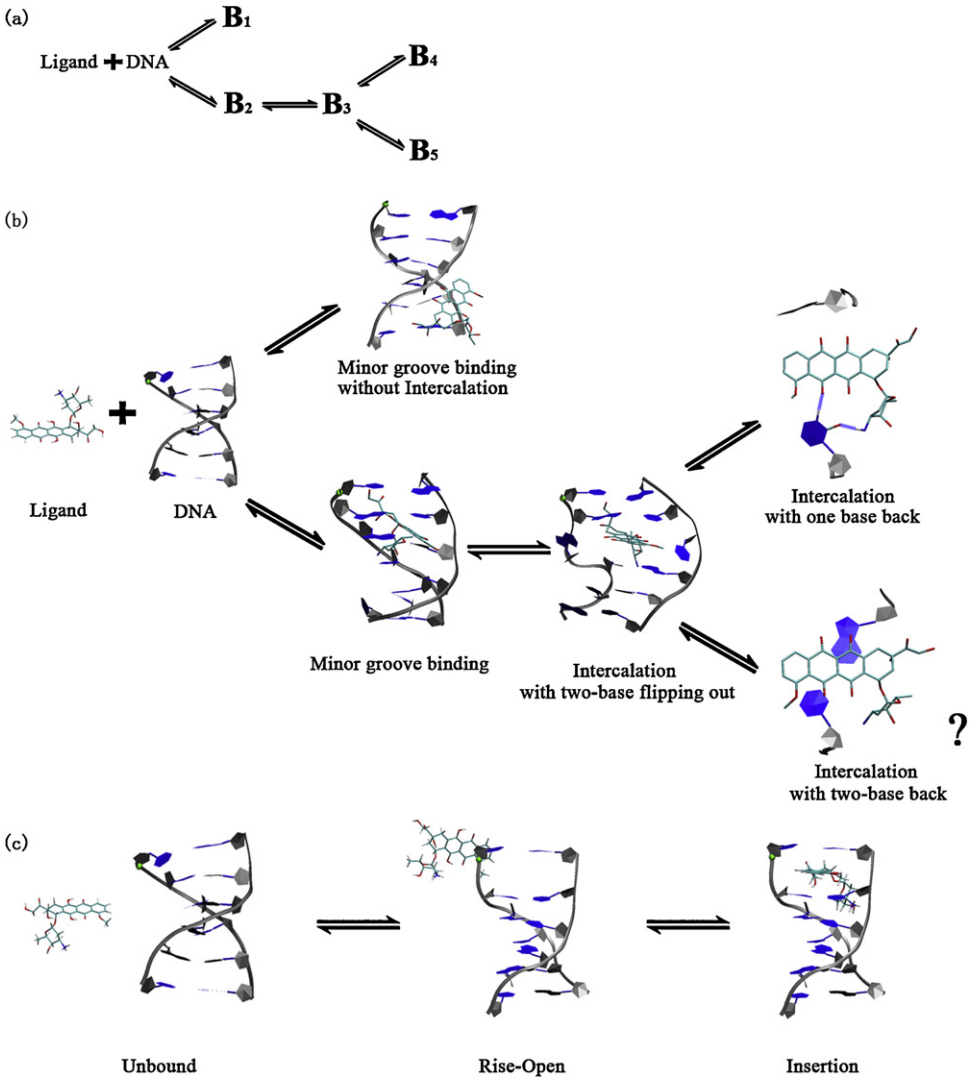

Fig. 8. Experimental kinetic model (a), our proposed flipping–intercalation mechanism (b) and the previously assumed rise–insertion mechanism (c). The ? state is extrapolated from our simulations.

the short DNA fragments to long DNA chains remains to be established. Our structural data clearly shows the terminal fraying effect of the short DNA fragments, which may affect the DNA stability and thus DNA–drug interactions. Further simulations as well as experiments with longer fragments in the future are required to clarify this point.

Nonetheless, our binding simulations may provide structural insights into the experimental kinetic models (Fig. 8a and b). The five step kinetic model proposed by Rizzo et al. [15] includes a parallel arrangement of step 1 (one off-pathway weak bound step) and 2 (on-pathway weak bound step), followed by on-pathway intercalation step 3, followed by another parallel arrangement of step 4 and 5 (either conformational rearrangement of the drug–DNA complex or redistribution of bound drug to preferred sites without dissociation). Corresponding to our simulation, since not all minor groove binding in our simulation led to the intercalation state, we suggest that step 1 might be off-pathway minor groove bindings. Our direct binding simulations confirmed that step 2 is on-pathway minor groove binding. The step 3 may correspond to our binding mode C with an opening and flipping of base, while mode D with flipping back of one base is consistent with step 4. In addition, the crystal intercalation mode with two bases flipping back might be step 5. In contrast, the assumed rise–insertion mechanism is not completely consistent with the experimental kinetic model.

In addition, our drug-induced base flipping may also shed light on the drug's anti-cancer mechanism. Because of the high energy cost for breaking H-bonds in spontaneous base opening for DNA and RNA [46,47], base flipping is usually activated by strong protein–DNA interactions for DNA strand separation and unwinding during DNA replication or transcription [48–50]. Therefore, the drug-induced base flipping would disrupt the normal protein–DNA/RNA interactions for their normal function, which may lead to the death of cancer cells.

Supporting information

The movies of the two intercalation trajectories, the detailed analyses of the simulations and the AMBER GAFF force field of doxorubicin and the PDB files from Fig. 3 are included in the supporting materials.

Acknowledgments

This project was funded by the National Science Foundation of China to HL (Grant 30870474). Usage of AMBER and VMD graphics packages is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2012.05.006.

References

[1] R. Martínez, L. Chacón-García, Current Medicinal Chemistry 12 (2005) 127–151.
[2] K.M. Tewey, T.C. Rowe, L. Yang, B.D. Halligan, L.F. Liu, Science 226 (1984) 466–468.
[3] M. Waring, Nature 219 (1968) 1320–1321.
[4] J.B. Lepecq, C. Paoletti, Journal of Molecular Biology 27 (1967) 87–89.
[5] I. Haq, P. Lincoln, S.D.C.B. Norden, B.Z. Chowdhry, J.B. Chaires, Journal of the American Chemical Society 117 (1995) 4788–4796.
[6] J.B. Chaires, N.D. Dattagupta, D.M. Crothers, Biochemistry Biokhimiia 22 (1983) 284–292.
[7] K. Chen, N. Gresh, B. Pullman, Journal of Biomolecular Structure and Dynamics 3 (1985) 445–466.
[8] J.B. Chaires, N. Dattagupta, D.M. Crothers, Biochemistry Biokhimiia 21 (1982) 3933–3940.
[9] G.J. Quigley, A.H.J. Wang, G. Ughetto, G.V.D. Marel, J.H.V. Boom, A. Rich, Proceedings of the National Academy of Sciences of the United States of America 77 (1980) 7204–7208.
[10] A.H. Wang, G. Ughetto, G.J. Quigley, A. Rich, Biochemistry Biokhimiia 26 (1987) 1152–1163.
[11] C.A. Frederick, L.D. Williams, G. Ughetto, G.A. van der Marel, J.H. van Boom, A. Rich, A.H. Wang, Biochemistry Biokhimiia 29 (1990) 2538–2549.
[12] C.M. Nunn, L. Van Meervelt, S. Zhang, M.H. Moore, O. Kennard, Journal of Molecular Biology 222 (1991) 167–177.
[13] M. Prabhakaftan, S.C. Harvey, Biopolymers 27 (1988) 1239–1248.
[14] J.B. Chaires, N. Dattagupta, D.M. Crothers, Biochemistry Biokhimiia 24 (1985) 260–267.
[15] V. Rizzo, N. Sacchi, M. Menozzi, Biochemistry Biokhimiia 28 (1989) 274–282.
[16] A. Mukherjee, R. Lavery, B. Bagchi, J.T. Hynes, Journal of the American Chemical Society 130 (2008) 9747–9755.
[17] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T.E.r. Cheatham, C.A. Laughton, M. Orozco, Biophysical Journal 92 (2007) 3817–3829.
[18] C.A. Andac, A.M. Miandji, U. Hornemann, N. Noyanalpan, International Journal of Biological Macromolecules 48 (2011) 531–539.
[19] R. Bortolini, S. Mazzini, R. Mondelli, E. Ragg, C. Ulbricht, S. Vioglio, S. Penco, Applied Magnetic Resonance 7 (1994) 71–87.
[20] S. Mazzini, R. Mondelli, E. Ragg, Journal of Chemical Society, Perkin Transactions 2 (1998) 1983–1991.
[21] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Journal of Chemical Physics 79 (1983) 926–935.
[22] J. Aqvist, Journal of Physical Chemistry 94 (1990) 8021–8024.
[23] R. Lavery, K. Zakrzewska, D. Beveridge, T.C. Bishop, D.A. Case, T.E.I. Cheatham, S. Dixit, B. Jayaram, F. Lankas, C. Laughton, J.H. Maddocks, A. Michon, R. Osman, M. Orozco, A. Perez, T. Singh, N. Spackova, J. Sponer, Nucleic Acids Research 38 (2010) 299–313.
[24] S. Cosconati, L. Marinelli, R. Trotta, A. Virno, S. De Tito, R. Romagnoli, B. Pagano, V. Limongelli, C. Giancola, P.G. Baraldi, L. Mayol, E. Novellino, A. Randazzo, Journal of the American Chemical Society 132 (2010) 6425–6433.
[25] E. Fadrna, N.a. Spackova, J. Sarzynska, J. Koca, M. Orozco, T.E. Cheatham III, T. Kulinski, J. Sponer, Journal of Chemical Theory Computations 5 (2009) 2514–2530.
[26] C.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman, Journal of Physical Chemistry 97 (1993) 10269–10280.
[27] J.M. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Journal of Computational Chemistry 25 (2004) 1157–1174.
[28] D.A. Case, T.E.I. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, Journal of Computational Chemistry 26 (2005) 1668–1688.
[29] J. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Journal of Chemical Physics 23 (1977) 327–341.
[30] U. Essmann, L. Perera, M.L. Berkowitz, T.A. Darden, H. Lee, L.G. Pedersen, Journal of Chemical Physics 103 (1995) 8577–8593.
[31] P. Procacci, B.J. Berne, Molecular Physics 83 (1994) 255–272.
[32] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Journal of Chemical Physics 81 (1984) 3684–3690.
[33] S.W. Chiu, M. Clark, S. Subramaniam, E. Jakobsson, Journal of Computational Chemistry 21 (2000) 121–131.
[34] X. Daura, K. Gademann, B. Jaun, D. Seebach, W.F. van Gunsteren, A.E. Mark, Angewandte Chemie International Edition 38 (1999) 236–240.
[35] R. Lavery, H. Sklenar, Journal of Biomolecular Structure and Dynamics 6 (1989) 655.
[36] D. Bhattacharya, M. Bansal, Journal of Biomolecular Structure and Dynamics 6 (1988) 93–104.
[37] R.E. Dickerson, EMBO Journal 8 (1989) 1.
[38] X.J. Lu, W.K. Olson, Nucleic Acids Research 31 (2003) 5108–5121.
[39] W. Humphrey, A. Dalke, K. Schulten, Journal of Molecular Graphics 14 (1996) 33–38.
[40] 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.I. Cheatham, Accounts of Chemical Research 33 (2000) 889–897.
[41] J. Kongsted, P. Soderhjelm, U. Ryde, Journal of Computer-Aided Molecular Design 23 (2009) 395–409.
[42] J.B. Chaires, Biochemistry Biokhimiia 22 (1983) 4204–4211.
[43] D. Řeha, M. Kabelác, F. Ryjácek, J. Sponer, J.E. Sponer, M. Elstner, S. Suhai, P. Hobza, H. Pavel, Journal of the American Chemical Society 124 (2002) 3366–3376.
[44] G. Barone, C.F. Guerra, N. Gambino, A. Silvestri, A. Lauria, A.M. Almerico, F.M. Bickelhaupt, Journal of Biomolecular Structure and Dynamics 26 (2008) 115–129.
[45] D.J. Cashman, G.E. Kellogg, Journal of the American Chemical Society 47 (2004) 1360–1374.
[46] M. Gueron, J.L. Leroy, Methods in Enzymology 261 (1995) 383–413.
[47] C.J. Chen, I.M. Russu, Biophysical Journal 87 (2004) 2511–2545.
[48] X. Cheng, R.J. Roberts, Nucleic Acids Research 29 (2001) 3784–3795.
[49] I.G. Lyakhov, P.N. Hengen, D. Rubens, T.D. Schneider, Nucleic Acids Research 29 (2001) 4892–4900.
[50] R.J. Roberts, Cell 82 (1995) 9–12.