A new Monte Carlo based technique to study DNA-ligand interactions

Israel Cabeza de Vaca[1], Fátima Lucas[1,2] and Victor Guallar[1,3,]*

[1] Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, c/Jordi Girona 29, 08034 Barcelona, Spain.

[2] Anaxomics Biotech, Balmes 89, 08008 Barcelona, Spain

[3] Institució Catalana de Recerca I Estudis Avançats, Passeig Lluís Companys 23, 08010 Barcelona, Spain.

KEYWORDS: PELE, DNA, Cisplatin, Markov State Model, Ligand Binding

ABSTRACT

We present a new all-atom Monte Carlo technique capable of performing quick and accurate DNA-ligand conformational sampling. In particular, and using the PELE software as a frame, we have introduced an additional force field, an implicit solvent and an anisotropic network model to effectively map the DNA energy landscape. With these additions, we successfully generated DNA conformations for a test set composed of six DNA fragments of A-DNA and B-DNA. Moreover, trajectories generated for cisplatin and its hydrolysis products identified the best interacting compound and binding site, producing analogous results to microsecond molecular dynamics simulations. Furthermore, a combination of the

Monte Carlo trajectories with Markov State Models produced non-covalent binding free energies in good agreement with the published molecular dynamics results, at a significantly lower computational cost. Overall our approach will allow a quick but accurate sampling of DNA-ligand interactions.

1. INTRODUCTION

Computational drug design efforts focus, to a large degree, on protein-ligand interactions. Nevertheless, there is a great interest in studying DNA-ligand interactions due to its importance, for example, in anticancer therapies.[1] Cisplatin,[2, 3] carboplatin and oxaliplatin complexes[4, 5] are drug examples that bind to DNA causing cross-linking and triggering apoptosis; cisplatin cure rate in early stage testicular cancer is around 85%.[6]

There are three main ways for small molecules to bind DNA: groove binding, intercalation between two base pairs and covalent binding to the bases.[7] Besides, some small molecules can bind to DNA in more than one way. Groove binding is mostly driven by hydrogen bond interactions between ligand and DNA. Intercalation of a small molecule between two adjacent base pairs often requires a planar polyaromatic system that performs pi-pi interactions with DNA bases. Finally, covalent binding is first driven by preferential ligand diffusion and DNA recognition,[8] where positively charged ligands (with large stabilizing Coulomb interactions) dominate,[9] followed by irreversible drug addition. Thus, an all-atom description of the DNA–ligand interaction is a central aspect of drug design.

Different docking algorithms such as AUTODOCK,[10] GLIDE[11] or CDOCKER[12] are used to study binding interactions between drugs and DNA fragments, the main objective being to identify the preferential binding site and to produce a reduced set of ligand orientations. These tools, however, were mostly developed and tested for protein-ligand complexes and the accuracy of DNA-ligand predictions is still questioned.[13] Moreover, docking approaches typically neglect induced fit effects, thus adding bias to the initial conditions (structures). On the other hand, the reduced number of degrees of freedom in DNA (with respect to proteins), together with performance improvement in molecular dynamics (MD) due to GPU acceleration, has been applied to perform microsecond range simulations.[14] These simulations explore whole DNA surface with a ligand and provide the binding energy of different regions in DNA fragments. However, binding free energy calculations are not a trivial task requiring a huge quantity of samples to statistically converge free energies. In this line, the recent introduction of Markov State Model (MSM) techniques[15] applied to long MD trajectories has aided in estimating binding free energies[16] but, in any case, MD sampling remain expensive computationally speaking. Monte Carlo (MC) approaches have become an alternative to reduce the computational time in this kind of problems. Different MC approaches have been developed to sample a large set of nucleic acids conformations based on fixed bond distances and perturbations of the angles and dihedrals such as concerted rotations with flexible bond angles (CRA)[17] or the chain breakage/closure (CBC)[18] algorithms. In particular, CBC algorithm have been applied for DNA flexible docking[19] and to derive atomic resolution data representing the sequence-dependent conformation of DNA duplexes in solution.[20] In addition, CBC has been the basis for further developments of MC sampling techniques for nucleic acids.[21]

The protein energy landscape exploration (PELE) algorithm,[22] in particular, has revealed accurate results in describing protein-ligand interactions,[23-25] at a significant lower computational cost than MD simulations.[26] Currently, it provides protein simulations of biased and non-biased ligand migration, local induced fit, normal mode protein dynamics and protein elastic properties.[27, 28]

In this work, we introduce our recent changes to expand PELE for exploring DNA and DNA-ligand dynamics. First, DNA dynamics are compared with nanosecond MD for six different DNA fragments, achieving excellent agreement. Then, we explored ligand diffusion with cisplatin and its hydrolysis products identifying the binding affinity and mode, in accordance with microsecond MD. Finally, the whole (free) DNA surface exploration was combined with MSM to successfully estimate binding free energies for these three platinum (Pt) compounds.

2. METHODS

The PELE algorithm

PELE is a MC algorithm originally designed to explore the energy landscape between protein and ligands. Each MC step is based on two main parts: perturbation and relaxation (described in detail below). The first is composed by independent perturbations on the ligand and on the receptor. Ligand perturbation consists of a set of random ligand rotations and translations, coupled to quick (steric) local side chain rotamer repositioning (including ligand rotatable bonds), until a non-clashing ligand pose is found. The receptor is perturbed by driving selected atoms using a biased potential to a new position following the direction pointed by a combination of the lowest normal modes. PELE relaxation performs a global minimization using a truncated Newton (TN) algorithm combined with a weak harmonic constraint to partly maintain the new position of the perturbed atoms. In proteins, the relaxation step might also include a more robust side chain prediction using expanded rotamer libraries and the full force field potential energy. Final system conformations are accepted or rejected using a Metropolis criterion. At each accepted step, the binding energy is computed as the internal energy difference between the complex and the free receptor and ligand ($E_{Bind} = E_{AB} - (E_A + E_B)$).

Ligand perturbation. Ligands are translated and rotated randomly with discrete jumps around +/-25% of a predefined value. This value is selected depending on the exploration type. For global explorations, translation and rotation ranges oscillate around 1-3 Å and 45º. In contrast, standard values are between 0.5-1 Å and 10º for local exploration, where we aim at a ligand-receptor local induced fit sampling. PELE can alternate or change dynamically these parameters defining jump conditions. Users can specify variable translation and rotation parameters using different criteria such as probabilities, root-mean-square deviations (RMSD), ligand solvent accessible area, or (given) separation distances between atoms, residues, etc. Ligand translational direction can be updated at each MC step or kept several steps (before being randomly reassigned) to increase ligand exploration in one region, a particularly useful approach in partially buried binding sites. Moreover, the exploration can be constrained to a cubic, prismatic or spherical box to reduce the conformational search space. To further restrict the exploration space in a dynamical manner, PELE can use a spawning criterion to exchange conformations between independent runs, where users define a criterion, such as a maximum distance between the ligand and an atom selection, interaction energies, etc.

Protein perturbation. This task aims to induce protein global motion by introducing a backbone perturbation following a displacement along one (or a combination) of the lowest normal modes (NM). Normal mode analysis (NMA) approximates the harmonic nature of the global fluctuations through the second derivatives of the potential (the Hessian). In particular, PELE uses an NMA version called Anisotropic Network Model (ANM) [29], an elastic network model based on a simplified coarse grained potential connecting neighbour alpha carbon atoms (within a defined distance cut off). Once the perturbation direction has been chosen, we apply it through an all-atom minimization including a harmonic constraint in the alpha carbons pointing in the NM direction.

Relaxation. The relaxation step is based on a TN minimization using the OPLS all-atom force field [30] and an implicit surface-generalized Born continuum solvent[31] including, at least, all residues involved in the perturbation steps. By default, it adds weak position constraints in the ANM atom nodes with a force constant of 1 kcal/mol. A weak ligand position constraint can be added to maintain the ligand's orientation found during the ligand perturbation. The typical force constant applied to the ligand ranges from 0.1 to 0.5 kcal/mol; a useful procedure when the binding site is narrow and a receptor conformational change is needed to allocate the ligand.

Novel implementations in PELE

Force field and solvent

To study DNA, we have introduced the AMBER parmBSC0 force field,[32, 33] specifically developed for nucleic acids. It has demonstrated the ability to preserve DNA's structure and reproduce feasible fluctuations in microsecond long MD trajectories.[34] We also introduced the Onufriev-Bashford-Case (OBC) implicit solvent[35] which has been developed and tested to reproduce the solvent polar term in macromolecules combined with the ACE non-polar term.[36] A Debye-Hückel term[37] is added to the solvation energy to take into account the ionic strength contribution. Following PELE's implementation, a multiscale[38] non-bonding algorithm is used to speed up atomic pair list generation associated with the non-bonding energy terms. Moreover, cell list optimization provides a quick way to update non-bonding pair lists.

ANM model

Based on the work by Zacharias et al.,[39] the ANM elastic network is generated using the C4' backbone atoms of each DNA base as nodes. Then, an exponential decay model without any cut-off is applied to connect the nodes and calculate the force constants needed for the Hessian matrix (see equation 1 where i and j corresponds to two given nodes). This exponential decay model depends on two constant parameters k_0 and d with values of 1.2 kcal/(mol·Å^2) and 5 Å, respectively, as optimized in a recent study.[39] Eigenvectors produced by the ten smallest normal modes are used to perturb DNA to sample the main global conformations. While PELE can update these ANM eigenvectors after each accepted step, the default behaviour uses the initial ones (updates should be considered if large conformational changes are expected or observed). The perturbation direction can be imposed by the user (a predefined ANM mode) or be based on a weighted average over the eigenvectors generated: after each iteration one random mode is selected with a larger contribution (65% by default), the remaining contribution comes from the average of the other nine eigenvectors.

$$k_{ij} = k_0 e^{-\left(\frac{r_{ij}}{d}\right)^2} \quad (1)$$

Once the direction has been estimated and normalized, final eigenvectors are scaled by a constant factor of 1.5 and placed in the ANM nodes generating the coordinates of a new virtual point. As in proteins, a harmonic constraint with zero equilibrium length is created between each node atom and the virtual point, to be used in the perturbation step. In DNA, however, due to its linear geometry we found a strong dependence on the size and the relaxation force constant; the reduced mobility of short DNA fragments (as opposed to larger ones) imposed increasing slightly the force constant in the harmonic constraint. **Table 1** presents the optimal set of parameters found for each DNA conformation and size.

Table 1. Force constant values (kcal/(mol·Å^2) for the position constraint applied in the PELE global minimization for each representative DNA fragment.

Number of bases	A-DNA	B-DNA
24	3.0	1.5
36	1.5	0.5
48	1.0	0.0

PELE parameters used

In all simulations performed, the ligand perturbation kept the same direction during two consecutive steps to increase the possibility to escape from a local minimum. Translation magnitude was randomly alternated 50% of the time between 1.0 Å and 3.0 Å and rotation angle was generated with a Gaussian distribution around 72º. This set of parameters was chosen to allow a quick DNA surface exploration (using large perturbations) and a local refinement of the binding regions (using small perturbations) in the same run. In ligand diffusion simulations, two position constraints of 10 kcal/mol were added into DNA's end residues (residue 12 and 24) to avoid artefacts with the ligand interaction due to the small fragment size. Ligand movement was restricted to a spherical box of 35 Å and the ionic strength of OBC solvent was set to zero.

Molecular dynamics simulations

MD simulations have been performed using Amber12 package[40] with the parmBSC0 force field. Explicit solvent simulations have been set up using a truncated octahedral water box with TIP3P water model[41] where the distance between the solute unit and the box edges was set to 12 Å, and the systems were neutralized adding Na^+ ions. The equilibration protocol consisted of initial solvent minimization, followed by a global minimization and 200 picoseconds heating the system from 0 to 300 K using a weak-coupling algorithm with constant pressure. The time step used has been 0.5 femtoseconds in the equilibration and production runs, with the SHAKE algorithm[42, 43] constraining hydrogen bond lengths. Non-bonding interactions have been evaluated using a cutoff of 9 Å with the Particle-Mesh-Ewald (PME)[44] method to compute long-range electrostatic interactions. Constant pressure and temperature (NPT ensemble) has been applied to the system using a Berendsen barostat and thermostat.[45] Each simulation consisted of 200 nanoseconds. Analogous implicit solvent simulations have been carried out using AMBER parmBSC0 force field and the OBC implicit solvent.

DNA conformational analysis

MD simulations with explicit and implicit solvent were carried out on each structure to provide a reference DNA set of conformations. Six independent PELE trajectories (each running in one single computing core) were simulated for each system. Then, these trajectories were joined in one trajectory removing the first 50 frames of each one, considered part of the equilibration. DNA conformations have been analysed and compared using the root mean square fluctuation (RMSF), principal component analysis (PCA) and DNA base pair step parameters. PRODY[46] library was used to compute RMSF and PCAs. 3DNA[47] software was used to calculate the rise, roll, twist, slide, shift and tilt DNA bases topological parameters.

Absolute binding free energies

To test the ligand–DNA capabilities of PELE, we have selected three platinum compounds: cisplatin, CPT, neutrally charged; the first hydrolysis product, CPT1, with formal charge +1; and the second product, CPT2, with formal charge +2. Force field ligand parameters and charges were extracted from Lucas et al. [8]. We started each PELE simulation from six different ligand positions 20 Å away from the DNA fragment. The same B-DNA fragment with 24 bases employed in a previous study [8] with protein data bank (PDB) entry 2K0V[48] was used. The structure has the same sequence of the damaged DNA sequence 3LPV[49] with cisplatin cross-linked in the G6-G7 base pair (pdb entry 3LPV).[49]

Binding free energies were estimated using MSM with the software package EMMA.[50, 51] MSM defines states and uses the transition between them to describe equilibrium properties. PELE simulation frames were aligned to a reference structure using the DNA atoms P, C2 and C4'. MSM was constructed following the next steps[51]: 1) extract cartesian coordinates of the central Pt atom of cisplatin molecules; 2) generate 300 clusters using the K-means algorithm; 3) assign each snapshot to a (clustered) microstate using Voronoi discretisation; 4) check connectivity between microstates to determine the largest set of them; 5) assure that the implied timescales become constant after a certain lag time (τ); stationary distribution of the microstates is computed using $\pi = \pi T_{ij}$ where T_{ij} corresponds to the transition matrix between microstates. After this analysis, the stationary distribution corresponds to the eigenvector with eigenvalue of the transition matrix equal to one. Potential mean force (PMF) profile is then computed using the Boltzmann inversion of the stationary distribution, $G_i = -k_B T \log \pi_i$, and the binding free energy through $\Delta G_0 = -k_B T \log {}^{v_b}/_{v_0} - \Delta w$, where k_B is the Boltzmann constant, T = 300 K, v_0=1661 Å^3 (1 M ligand concentration), v_b is the PMF bound volume and Δw corresponds to the difference between the minimum (bound state) and the bulk average (unbound state) values in the PMF profile.

3. RESULTS

DNA conformational analyses

To test the DNA conformational sampling obtained by the ANM model, A-DNA and B-DNA fragments with 24, 36 and 48 bases were generated using NAB tool[52] as a test set. The initial 24 bases sequence was taken from the PDB[53] entry 2K0V[48] corresponding to d((CCTCTGGTCTCC)·d(GGAGACCAGAGG)). Sequences with 36 bases and 48 bases were generated replicating the 24 bases sequence motif (see supporting information Figure S1 for a picture of the final models and the sequences for the 36 and 48 bases fragments).

Analyses of the RMSF, PCA and bases topological parameters show that PELE explores an equivalent conformational space as the (200 nanosecond) MD simulations (Figure 1). For the RMSF analysis, we obtained around 5000 frames from each MD and PELE trajectory where the initial structure was chosen as reference for both trajectories. Figure 1A shows the RMSF for the B-DNA fragment with 24 bases, where points represents the RMSF of atoms P, C2 and C4', respectively (in each individual residue), and residues are arranged in ascending order from 1 to 24. Initial, medium and final parts of the plot (nodes 1-6, 30-42 and 66-72 belong to terminal residues in each strand) correspond to the 5' and 3'-ends of the double strands. As expected in movements derived from the lowest normal modes, the RMSF plot shows higher fluctuations in the DNA for MD than PELE, since these fluctuations correspond to higher frequency modes; all other bases present an excellent agreement. [54, 55] RMSF plots showed the same agreement for the other five systems studied (see supporting information Figure S2).

Fluctuation analyses of the bases' topological parameters allow us to evaluate the structural integrity along the simulations. We have focused on the parameters: roll, rise, twist, slide, shift and tilt (see 3DNA [47] for a detailed explanation). Figure 1B shows a comparison between MD trajectories with explicit and implicit solvent and PELE for the B-DNA fragment with 24 bases, where the reference value corresponds to the initial structure generated with NAB tools. Overall, the agreement between PELE and MD is excellent. Roll is the only one that showed significant differences between the reference and the simulations; after a few MD and MC steps, DNA's ends were slightly collapsed reducing the chain length and the average roll value. In production runs, one might choose to use a weak constraint on the ends if emulating the effects of a larger

DNA chain, or to avoid DNA-ligand overestimated interactions (see Methods). In any case, changes in 5′ and 3′ DNA ends do not significantly affect the minor and major groove size and are not important for the binding process. In the remaining DNA fragments studied, PELE and MD sampling produced a similar parameter agreement (see supporting information Figure S5).

PCA was used to extract the most important motions from the conformational sampling trajectories. We used the inner product over the first ten principal components (PCs) and projections over the first two PCs to compare both simulation methods. Figure 1C shows the inner product matrix in a colour map with good overlapping between the lowest 4-5 modes, similar to the one we could obtain with two different force fields. As usual, PCs were sorted in decreasing maximum variance order. Thus, the most significant modes are the lowest ones because they represent the highest contribution to the variance of the fluctuations. All six DNA fragments studied showed similar correlations for the inner product matrix diagonal (see supporting information Figure S3). As expected from applying a simple ANM approximation, in some instances the (variance) ordering from MD and PELE trajectories is shifted.

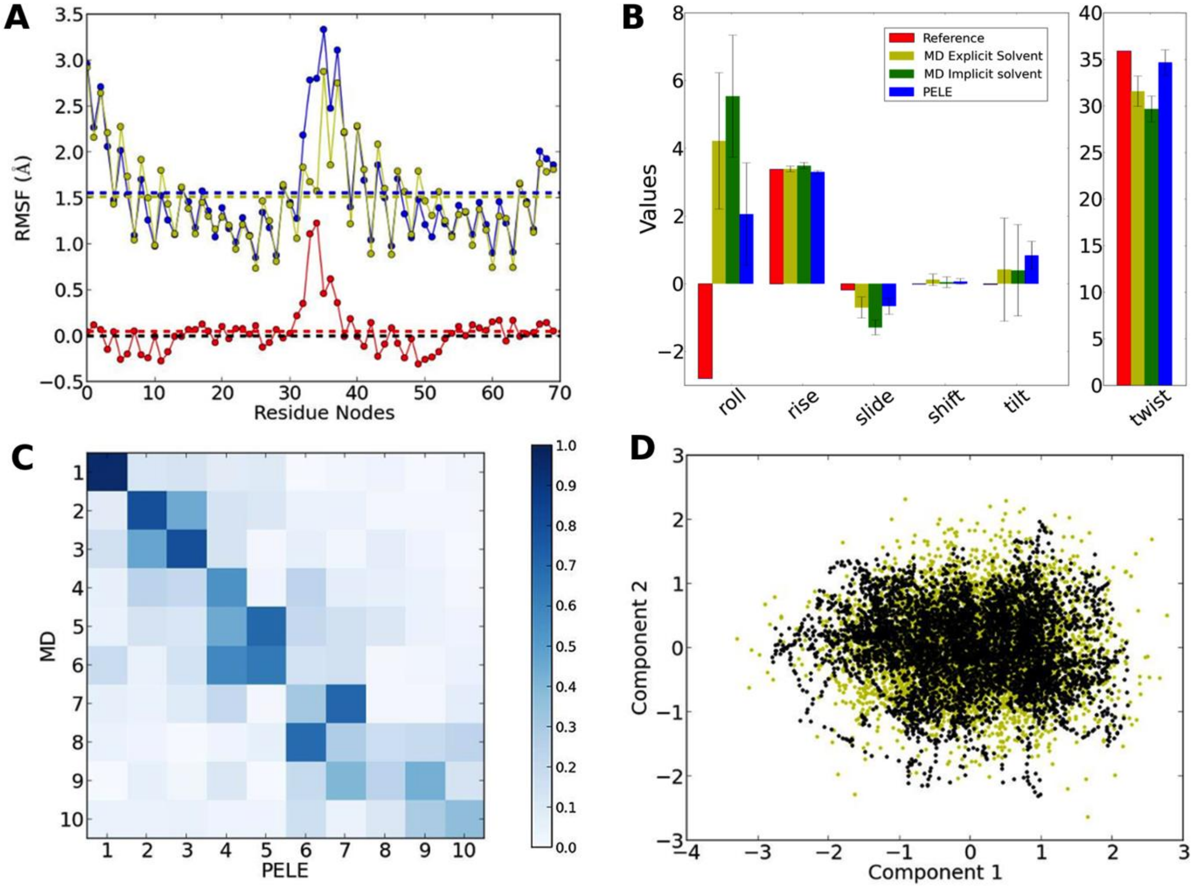

Figure 1. B-DNA with 24 bases analysis. Panel A shows the RMSF results for MD (blue), PELE (yellow) and the difference between them (red). Each point represents the RMSF of atoms P, C2 and C4', respectively (in each individual residue), and residues are arranged in ascending order from 1 to 24. Panel B, base pair step parameters comparison between DNA canonical structure and MD (explicit and implicit solvent) and PELE. Panel C corresponds to the normalized cross correlation matrix for the first ten PCs. Panel D is the 2D projection plot of the first two PCs for MD (yellow) and PELE (black).

The first two PCs projections contain the most significant fluctuation information of each trajectory; Figure 1D shows the projections for the 24 B-DNA fragment simulation for the first two eigenvectors. Clearly, PELE and MD explore the same area showing good agreement between their conformations. The other five DNA systems PCs projection plots show similar good correlation (see supporting information Figure S4).

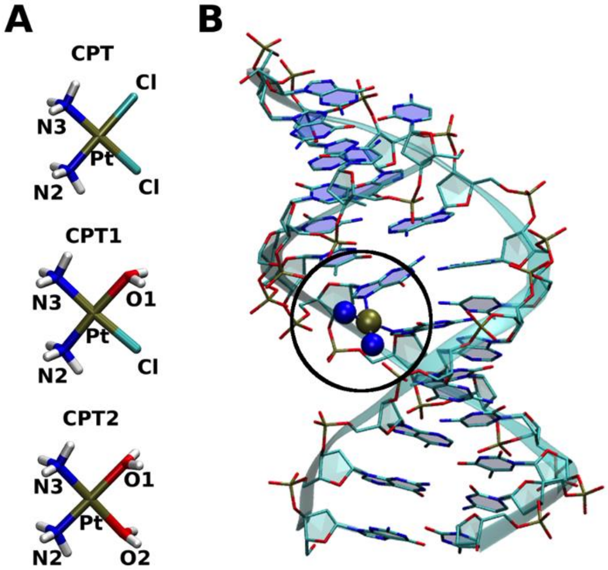

Figure 2. Panel A, view of cisplatin compounds CPT-parent drug, CPT1-mono-aquo and CPT2-di-aquo. Panel B, representation of PDB ID 3PLV structure showing the cross-linked cisplatin in the binding site (black circle).

DNA-ligand exploration

To test the DNA-ligand conformational sampling capabilities, we have studied the interaction with CPT and its aqua derivatives CPT1 and CPT2, with an overall molecular charges of 0, +1 and +2, respectively. Figure 2A shows the atomic representation of the three compounds and Figure 2B the crystallographic DNA cisplatin bound structure with the binding site highlighted. PELE's binding energy (Figure 3A**)** with respect to the binding site distance clearly indicates that

CPT spends most of the time in the bulk with very few visits to the binding site, where the ligand shows low interaction energies. In fact, other DNA regions (mainly the loose ends) presented better interaction energies than the binding site. Introducing a positive charge into the ligand (by replacing a chlorine substituent by a water molecule) clearly produces a significant shift in the ligand exploration. CPT1 is now able to identify the binding site showing more favourable interaction energies than the neutral variant. Following this trend CPT2 improved the binding site recognition, producing a clear precursor structure for the covalent addition. Figure 3B shows the ligand distribution around the DNA fragment during the simulations for the three ligands, showing excellent agreement with the results found in the previous microsecond MD study (see supporting information Figure S6). In line with the interaction energy profiles, CPT shows low ligand concentration close to the DNA fragment, CPT1 presents a larger exploration, and CPT2 has the highest cluster concentration around the DNA molecule and especially in the binding region. The ligand structure adopted in both MD and PELE in the binding site is shown in Figure 4, indicating (besides the agreement in both methods) a clear covalent addition precursor. To further quantify the equivalence of these ligand distributions between PELE and MD, we performed the MSM analyses and computed the absolute binding free energies.

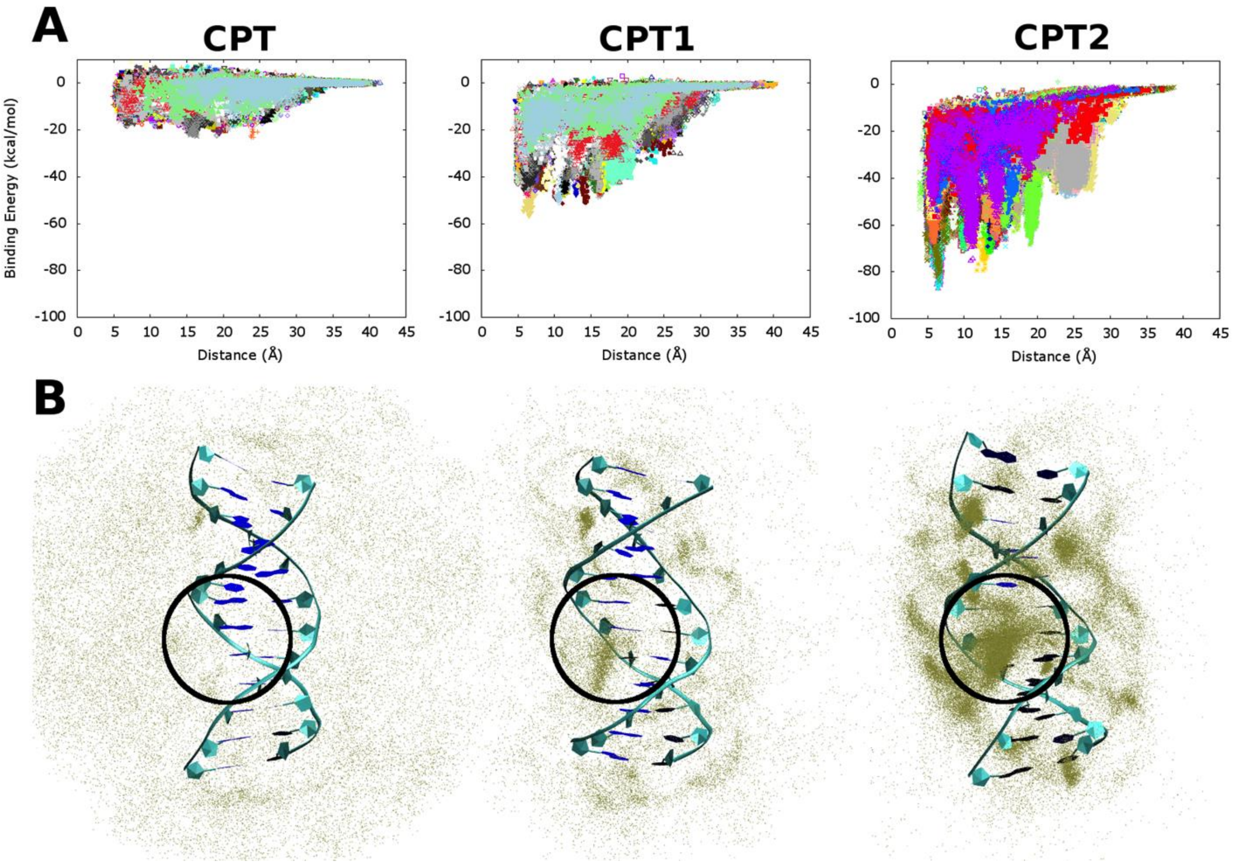

Figure 3. Panel A, PELE binding energy profiles for CPT, CPT1 and CPT2. Distance is measured from the Pt atom position to the binding site N7 atom from guanine 6. Panel B shows Pt atom position for each PELE trajectory frame. Black circles show cisplatin binding site.

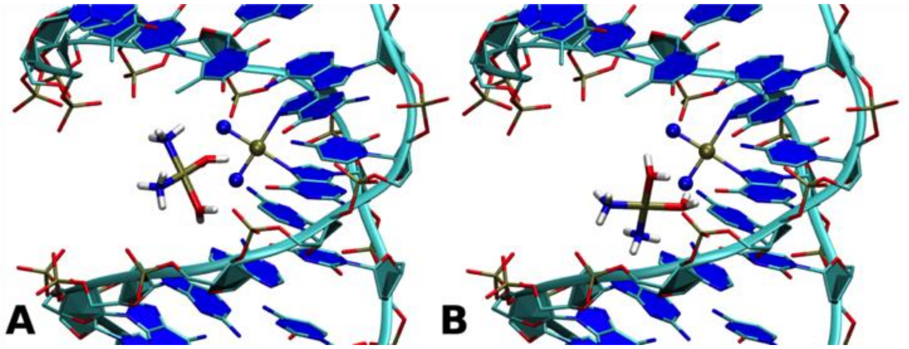

Figure 4. Representative cisplatin di aqua orientations for the binding site cluster. Panel A and B correspond to MD and our MC approach, respectively.

Binding free energies of cisplatin compounds

In a previous study,[8] we applied MSM to microsecond MD trajectories generated for the three cisplatin compounds: CPT, CPT1 and CPT2. Results were compared with steered molecular dynamics (SMD) and Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) techniques to determine the similarity between different methods. Here, we have applied MSM to our Monte Carlo simulations generated in the previous section to study the ability to reproduce the ligand affinities. To this end, we have generated the 2D PMF profiles obtained through MSM (see supporting information S7) and estimated binding free energies following the procedure described in the Methods section.

Binding free energies are summarized and compared with the previous published results in Table 2. PELE's results, -0.7±0.2, -2.0±0.5 and -2.8±1.0 kcal/mol for CPT, CPT1 and CPT2, respectively, are in good agreement with those obtained in microsecond MD simulations. The binding free energy for CPT showed the maximum difference (0.7 kcal/mol) between PELE and MD. This difference comes from the ligand conformational sampling: in MD, CPT was able to find more weak local minima around the DNA structure, particularly in the minor groove, increasing the binding free energy. PELE, due to the implicit solvent model and the ligand perturbation, was not able to keep the ligand in these clusters long enough to converge the results. Nevertheless, PELE clearly discriminates ligand affinity and provides quantitative absolute binding energies for ligands with significant binding energy.

Table 2. Absolute Binding free energies (kcal/mol) comparison for CPT, CPT1 and CPT2 drugs. MD results have been extracted from.[8]

		CPT	CPT1	CPT2
	MSM	-1.4	-2.1	-2.8
MD	SMD	-1.6	-2.6	-2.8
	MMPBSA	-2.4	-3.3	-3.8
PELE	MSM	-0.7	-2.0	-2.8

4. CONCLUSIONS

The PELE algorithm is today a well-established Monte Carlo method for studying protein-ligand interactions, with a good compromise between speed and accuracy. Here, we have presented the expansion of the program to allow its usage to study DNA-ligand interactions. To this aim, several modifications including an additional implicit solvent, ANM model and a force field have been implemented. Altogether, with these additions PELE is now able to reproduce conformations and ligand distributions obtained at microsecond scale by MD. In particular, we demonstrated its ability to explore similar DNA conformations obtained with MD for different A-DNA and B-DNA fragments of various sizes. The comparison between DNA structures using RMSF, PCA (inner product and projections) and the base pair step DNA parameters for both methods confirmed the similarity of the conformational exploration. Certainly, our Monte Carlo based approach has limitations, such as a reduced set of normal modes, their approximate nature or the lack of time evolution, that will not make it the best tool for an exhaustive dynamical DNA exploration. Nevertheless, it produces a great quick conformational search to be coupled with ligand dynamics.

The potential of PELE in exploring the DNA-ligand conformational space was tested against recent non-biased microsecond molecular dynamic simulations for cisplatin and two of its aqua derivatives. The well-defined trend (difference) observed for the three ligands (Figure S6), together with the extensive MD simulation data,[8] makes of this system a nice test set. Clearly PELE is capable of reproducing the non-covalent DNA-ligand interactions for the three systems. As expected, differences are only observed for the (very) weak-binding CPT compound, for which the ligand perturbation step leads to a reduced population of the DNA surface (minor groove occupation seen in MD not observed in PELE). As discussed earlier, however, CPT most likely does not bind to DNA[8]; true binders produce quantitative absolute free energies.

Importantly, besides ligand (space) distribution, ranking and absolute free energies, the correct orientation of the pre-covalent bound compound is observed (Figure 4). This is an important feature since obtaining receptor-ligand induced fit orientations is a crucial aspect in drug development projects, from which to design new compounds. Such information is quickly obtained, within 1-2 CPU hours in a commodity cluster (~16 cores), with PELE. Further computation of absolute binding free energies, for instance by using MSM, is not a trivial task, requiring approximately 128 cores for 24 hours. Nevertheless, this still constitutes an improvement over the 1.5 microseconds simulation necessary to reach convergence in molecular dynamics. Moreover, since each core performs an independent simulation, the method scales linearly with computational resources. This speed up in time and scalability opens the door for *in silico* accurate screening of DNA binders using affordable resources and simulation time.

In summary, we introduced here the expansion of our Monte Carlo code PELE to study DNA-ligand interactions. By introducing last generation force field and solvent models, together with specific DNA sampling algorithms, we are capable of quickly and accurately explore the DNA-ligand induced fit space for the non-covalent recognition process.

ASSOCIATED CONTENT

Supporting Information.

The following files are available free of charge on the ACS Publications website at DOI. Picture of the six DNA fragments tested. Analysis and comparison (RMSF, principal component projections and cross correlation map of principal components, base pair step DNA parameters) with MD for the six DNA fragments studied. Visual comparison between MC and MD for the cisplatin distributions. PMF and lagtime generated with the MC trajectories. (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: victor.guallar@bsc.es

Notes

These authors declare no competing financial interest.

ACKNOWLEDGMENT

The authors thank the Barcelona Supercomputing Center for computational resources. This work was supported by grants from the European Research Council—2009-Adg25027-PELE European project and the Spanish Ministry of Economy and Competitiveness CTQ2013-48287 and “Juan de la Cierva” to F.L.

ABBREVIATIONS

MD, molecular dynamics; PELE, protein energy landscape exploration; NM, normal modes; MNA, normal mode analysis; ANM, anisotropic network model; TN, truncated Newton; RMSD, root-mean-square deviations; RMSF, root-mean-square fluctuations; PCA, principal component analysis; OBC, Onufriev-Bashford-Case; MSM, Markov state models; MMPBSA, Molecular Mechanics Poisson-Boltzmann Surface Area; SMD, steered molecular dynamics; CPT, cisplatin; CPT1, mono aquo cisplatin; CPT2, di aquo cisplatin.

REFERENCES

1. Palchaudhuri, R.; Hergenrother, P. J., DNA as a target for anticancer compounds: methods to determine the mode of binding and the mechanism of action. *Current opinion in biotechnology* **2007,** *18*, 497-503.
2. Rosenberg, B.; Van Camp, L.; Krigas, T., Inhibition of cell division in Escherichia coli by electrolysis products from a platinum electrode. *Nature* **1965,** *205*, 698-699.
3. Rosenberg, B.; Vancamp, L., Platinum compounds: a new class of potent antitumour agents. *Nature* **1969,** *222*, 385-386.
4. Weiss, R. B.; Christian, M. C., New cisplatin analogues in development. *Drugs* **1993,** *46*, 360-377.
5. Alderden, R. A.; Hall, M. D.; Hambley, T. W., The discovery and development of cisplatin. *Journal of chemical education* **2006,** *83*, 728.
6. Einhorn, L. H., Treatment of testicular cancer: a new and improved model. *Journal of clinical oncology* **1990,** *8*, 1777-1781.
7. Hurley, L. H., DNA and its associated processes as targets for cancer therapy. *Nature Reviews Cancer* **2002,** *2*, 188-200.
8. Lucas, M. F.; de Vaca, I. C.; Takahashi, R.; Rubio-Martínez, J.; Guallar, V., Atomic level rendering of DNA-drug encounter. *Biophysical journal* **2014,** *106*, 421-429.
9. Hannon, M. J., Supramolecular DNA recognition. *Chemical Society Reviews* **2007,** *36*, 280-295.
10. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry* **2009,** *30*, 2785-2791.
11. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry* **2004,** *47*, 1739-1749.
12. Wu, G.; Robertson, D. H.; Brooks, C. L.; Vieth, M., Detailed analysis of grid- based molecular docking: A case study of CDOCKER—A CHARMm- based MD docking algorithm. *Journal of computational chemistry* **2003,** *24*, 1549-1562.
13. Gilad, Y.; Senderowitz, H., Docking studies on DNA intercalators. *Journal of Chemical Information and Modeling* **2013,** *54*, 96-107.

14. Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *Journal of chemical theory and computation* **2012,** *8*, 1542-1555.

15. Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S., Progress and challenges in the automated construction of Markov state models for full protein systems. *The Journal of chemical physics* **2009,** *131*, 124101.

16. Buch, I.; Giorgino, T.; De Fabritiis, G., Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **2011,** *108*, 10184-10189.

17. Ulmschneider, J. P.; Jorgensen, W. L., Monte Carlo backbone sampling for nucleic acids using concerted rotations including variable bond angles. *The Journal of Physical Chemistry B* **2004,** *108*, 16883-16892.

18. Sklenar, H.; Wüstner, D.; Rohs, R., Using internal and collective variables in Monte Carlo simulations of nucleic acid structures: chain breakage/closure algorithm and associated Jacobians. *Journal of computational chemistry* **2006,** *27*, 309-315.

19. Rohs, R.; Bloch, I.; Sklenar, H.; Shakked, Z., Molecular flexibility in ab initio drug docking to DNA: binding-site and binding-mode transitions in all-atom Monte Carlo simulations. *Nucleic acids research* **2005,** *33*, 7048-7057.

20. Zhang, X.; Machado, A. C. D.; Ding, Y.; Chen, Y.; Lu, Y.; Duan, Y.; Tham, K. W.; Chen, L.; Rohs, R.; Qin, P. Z., Conformations of p53 response elements in solution deduced using site-directed spin labeling and Monte Carlo sampling. *Nucleic acids research* **2014,** *42*, 2789-2797.

21. Minary, P.; Levitt, M., Conformational optimization with natural degrees of freedom: A novel stochastic chain closure algorithm. *Journal of Computational Biology* **2010,** *17*, 993-1010.

22. Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V., PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput.* **2005,** *1*, 1304–1311.

23. Kotev, M.; Lecina, D.; Tarrago, T.; Giralt, E.; Guallar, V., Unveiling prolyl oligopeptidase ligand migration by comprehensive computational techniques. *Biophys J* **2015,** *108*, 116-25.

24. Hosseini, A.; Espona-Fiedler, M.; Soto-Cerrato, V.; Quesada, R.; Pérez-Tomás, R.; Guallar, V., Molecular interactions of prodiginines with the BH3 domain of anti-apoptotic Bcl-2 family members. *PloS one* **2013,** *8*.

25. Lucas, M. F.; Guallar, V., An Atomistic View on Human Hemoglobin Carbon Monoxide Migration Processes. *Biophysical journal* **2012,** *102*, 887-896.

26. Cossins, B. P.; Hosseini, A.; Guallar, V., Exploration of protein conformational change with PELE and meta-dynamics. *Journal of Chemical Theory and Computation* **2012,** *8*, 959-965.

27. Madadkar-Sobhani, A.; Guallar, V., PELE web server: atomistic study of biomolecular systems at your fingertips. *Nucleic acids research* **2013,** *41*, W322-W328.

28. Giannotti, M. I.; Cabeza de Vaca, I.; Artés, J. M.; Sanz, F.; Guallar, V.; Gorostiza, P., Direct Measurement of the Nanomechanical Stability of a Redox Protein Active Site and Its Dependence upon Metal Binding. *Journal of Physical Chemistry B* **2015**.

29. Doruker, P.; Atilgan, A. R.; Bahar, I., Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to α- amylase inhibitor. *Proteins: Structure, Function, and Bioinformatics* **2000,** *40*, 512-524.

30. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **1996,** *118*, 11225-11236.
31. Gallicchio, E.; Zhang, L. Y.; Levy, R. M., The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. *Journal of computational chemistry* **2002,** *23*, 517-529.
32. Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M., Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophysical journal* **2007,** *92*, 3817-3829.
33. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A., Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry* **1995,** *16*, 1357-1377.
34. Pérez, A.; Luque, F. J.; Orozco, M., Dynamics of B-DNA on the microsecond time scale. *Journal of the American Chemical Society* **2007,** *129*, 14739-14745.
35. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large- scale conformational changes with a modified generalized born model. *Proteins: Structure, Function & Bioinformatics* **2004,** *55*, 383-394.
36. Calimet, N.; Schaefer, M.; Simonson, T., Protein molecular dynamics with the generalized Born/ACE solvent model. *Proteins: Structure, Function, and Bioinformatics* **2001,** *45*, 144-158.
37. Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A., Solvation free energies of peptides: Comparison of approximate continuum solvation models with accurate solution of the Poisson-Boltzmann equation. *The Journal of Physical Chemistry B* **1997,** *101*, 1190-1197.
38. Zhu, K.; Shirts, M. R.; Friesner, R. A.; Jacobson, M. P., Multiscale optimization of a truncated Newton minimization algorithm and application to proteins and protein-ligand complexes. *Journal of Chemical Theory and Computation* **2007,** *3*, 640-648.
39. Setny, P.; Zacharias, M., Elastic Network Models of Nucleic Acids Flexibility. *Journal of Chemical Theory and Computation* **2013,** *9*, 5460-5470.
40. Case, D.; Darden, T.; Cheatham III, T. E.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K., AMBER 12. *University of California, San Francisco* **2012,** *1*.
41. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983,** *79*, 926-935.
42. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977,** *23*, 327-341.
43. Miyamoto, S.; Kollman, P. A., Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **1992,** *13*, 952-962.
44. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics* **1993,** *98*, 10089-10092.
45. Berendsen, H. J.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J., Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **1984,** *81*, 3684-3690.
46. Bakan, A.; Meireles, L. M.; Bahar, I., ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **2011,** *27*, 1575-1577.

47. Lu, X. J.; Olson, W. K., 3DNA: a software package for the analysis, rebuilding and visualization of three- dimensional nucleic acid structures. *Nucleic acids research* **2003,** *31*, 5108-5121.
48. Bhattacharyya, D.; Ramachandran, S.; Sharma, S.; Pathmasiri, W.; King, C. L.; Baskerville-Abraham, I.; Boysen, G.; Swenberg, J. A.; Campbell, S. L.; Dokholyan, N. V.; Chaney, S. G., Flanking bases influence the nature of DNA distortion by platinum 1,2-intrastrand (GG) cross-links. *PLoS One* **2011,** *6*, e23582.
49. Todd, R. C.; Lippard, S. J., Structure of duplex DNA containing the cisplatin 1,2-{Pt(NH3)2}2+-d(GpG) cross-link at 1.77 A resolution. *Journal of inorganic biochemistry* **2010,** *104*, 902-8.
50. Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F., Markov models of molecular kinetics: Generation and validation. *The Journal of chemical physics* **2011,** *134*, 174105.
51. Senne, M.; Trendelkamp-Schroer, B.; Mey, A. S.; Schütte, C.; Noé, F., EMMA: A software package for Markov model building and analysis. *Journal of Chemical Theory and Computation* **2012,** *8*, 2223-2238.
52. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005,** *26*, 1668-1688.
53. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The protein data bank. *Nucleic acids research* **2000,** *28*, 235-242.
54. Zgarbová, M.; Otyepka, M.; Sponer, J.; Lankas, F.; Jurečka, P., Base pair fraying in molecular dynamics simulations of DNA and RNA. *Journal of Chemical Theory and Computation* **2014,** *10*, 3177-3189.
55. Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H., Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophysical Journal* **2005,** *89*, 3721-3740.

Table of Contents graphic: