


 Cite this: *RSC Adv.*, 2026, 16, 13696

# Ensemble molecular dynamics for predicting binding energies of DNA intercalators

 Anju Choorakottayil Pushkaran and Alya A. Arabi \*

Ensemble molecular dynamics (MD) simulations have been widely used to improve binding energy predictions in protein-ligand systems, but their application to DNA-ligand systems is limited. Following previous studies on DNA-doxorubicin and DNA-proflavine systems, the present study extends the ensemble MD simulations for a total of 16 intercalators spanning different molecular sizes, intercalating scaffolds, flexibilities, and charges. For each DNA-intercalator complex, 25 independent 10 ns MD replicas were simulated using randomized initial velocities. The uncorrected Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) energies significantly overestimated the binding energies and had a poor correlation with the corresponding experimental values (with a correlation coefficient  $R^2$  of 0.45). Including entropy and deformation energy correction terms to the MM/PBSA energies reduced the discrepancy between experimental and predicted binding energies of the cationic intercalators to within ca. 2 kcal mol<sup>-1</sup> (with an  $R^2$  of 0.70). Two additional monocationic intercalators were included for validation. Their predicted binding energies were within 1.2 kcal mol<sup>-1</sup> of the corresponding experimental values. Bootstrap analyses showed that around 10 independent replicas are generally sufficient to reduce the uncertainty in binding energy predictions for cationic intercalators to below 1 kcal mol<sup>-1</sup>, although up to 20 replicas could be needed for more flexible ligands that have relatively high degrees of freedom. Overall, MM/PBSA energies estimated from an ensemble of short MD simulations, and corrected for entropy and deformation, are reliable and computationally efficient for ranking and predicting the binding affinity of the cationic intercalators in this study. This method provides a valuable alternative when experimental binding measurements are impractical or costly.

 Received 2nd January 2026  
 Accepted 14th February 2026

DOI: 10.1039/d6ra00032k

[rsc.li/rsc-advances](https://rsc.li/rsc-advances)

## Introduction

DNA intercalators are planar aromatic molecules that insert between two base pairs of the DNA double helix. They are widely used as anticancer drugs (*e.g.* doxorubicin), antimicrobial agents (*e.g.* proflavine), and molecular probes for DNA detection and imaging (*e.g.* acridine orange).<sup>1</sup> Quantitative assessments of DNA-intercalator interactions are needed as biological activities depend strongly on the binding strengths. Several computational methods can be used to estimate binding energies of biomolecular complexes. For example, alchemical techniques such as free energy perturbation and thermodynamic integration can accurately predict binding energies by sampling the bound and unbound states in explicit solvents.<sup>2,3</sup> However, in practice, these methods are computationally intensive and require extensive sampling and careful convergence checks.<sup>4</sup> End-point free energy methods, such as the MM/PBSA method, are more computationally affordable.<sup>5</sup> The MM/PBSA method combines molecular mechanics energies with continuum solvation models, Poisson-Boltzmann or

Generalized Born, to estimate binding energy from MD simulation trajectories.<sup>6</sup> The advantage of the MM/PBSA method is that it is efficient, but it also has well-documented limitations.<sup>7,8</sup> These include the neglect of certain energetic terms such as entropy and deformation energies,<sup>6,8</sup> and the sensitivity to limited sampling.<sup>9</sup>

High-resolution structures of DNA-intercalator complexes have shown that the intercalation process causes helix distortion.<sup>10</sup> The sugar-phosphate backbone is locally unwound, extended, and the base pairs flanking the intercalator are separated and tilted.<sup>10-12</sup> Therefore, it is important to account for the deformation energy in DNA-intercalator systems. Furthermore, the intercalation typically yields a large favorable enthalpy change due to  $\pi$ - $\pi$  stacking and electrostatic interactions, which is counterbalanced by a substantial loss of entropy.<sup>13</sup> Chaires and co-workers reported that the binding energies of intercalation (on the order of -5 kcal mol<sup>-1</sup> to -10 kcal mol<sup>-1</sup>) are a result of a balance between favorable binding enthalpies and unfavorable entropy penalties.<sup>14-17</sup> Several other studies have also highlighted the importance of accounting for entropy, in addition to enthalpy, when evaluating the binding energies of intercalators binding to DNA.<sup>18-20</sup> Therefore, in DNA-intercalator complexes, both the entropy and deformation correction terms should be added to the predicted

Department of Biochemistry and Molecular Biology, College of Medicine and Health Sciences, United Arab Emirates University, P.O. Box: 15551, Al Ain, United Arab Emirates. E-mail: [alya.arabi@dal.ca](mailto:alya.arabi@dal.ca); [alya.arabi@uaeu.ac.ae](mailto:alya.arabi@uaeu.ac.ae)



MM/PBSA binding energies to reproduce experimentally determined binding energies.

Another major issue with MM/PBSA energies is the sensitivity to limited sampling. Because of the stochastic nature of MD sampling, MM/PBSA binding energies estimated from single MD trajectories are often not reproducible.<sup>9,21</sup> Extending the simulation time to 100 ns may result in a stable complex, however, if restarted with a different starting velocity, it could sample to a different trajectory resulting in a different MM/PBSA energy. It has been reported that ensemble MD simulations, where multiple independent MD simulations are performed using different starting velocities and/or geometries, overcome this challenge of insufficient sampling.<sup>9,21–26</sup> Several studies have shown that an ensemble of shorter simulations often samples broader conformations and yields more reproducible energy estimates than a single long trajectory.<sup>9,21,27</sup> Perez *et al.* quantified this by comparing the results from multiple simulations to a single run of equal total simulation time.<sup>28</sup> Performing independent MD simulations helps to average over MD noise and present the error in the predictions in terms of standard deviations.<sup>22</sup> In another study, the ensemble MD simulation method was reported to accurately reproduce the experimental ranking of ligands, with a Spearman rank coefficient of >0.8,<sup>29</sup> although the absolute energies deviated significantly from the experimental values.

A recent study explored both limitations of the energy correction terms and the sensitivity to limited sampling in two DNA-intercalator complexes, DNA-doxorubicin and DNA-proflavine.<sup>20</sup> It was shown that by averaging results over 25 independent MD simulations of 10 ns each, and by including the entropy and deformation terms, the predicted binding energies were within *ca.* 1.0 kcal mol<sup>-1</sup> of the experimental binding energies for both complexes. The study also demonstrated that single simulations, whether long (100 ns) or short (10 ns), often produced binding energies that were off from experimental values (by up to 5 kcal mol<sup>-1</sup>). However, averaging results across an ensemble of 25 replicas (short 10 ns or long 100 ns simulations) brought the predicted binding energies much closer to the experimental values, within 1 kcal mol<sup>-1</sup>. The study further reported that the average binding energies predicted from ensembles of short *vs.* long simulations resulted in differences of no greater than 1 kcal mol<sup>-1</sup>.<sup>20</sup> However, this discussed study is limited to only two complexes. Validation on a wider and diverse set of intercalators is important to confirm the reliability of the ensemble MD-based MM/PBSA method for DNA-intercalator complexes. Such validation is important for guiding the design and development of novel DNA intercalators for anticancer, antimicrobial, and diagnostic applications, especially for systems where experimental binding data are unavailable or difficult to obtain.

In the present study, the reliability of the ensemble MD-based MM/PBSA method is evaluated for a total of 18 DNA-intercalator complexes involving a diverse set of intercalators with varying molecular sizes (approximately 180 Da to 540 Da), intercalating scaffolds (with two to three fused rings in the scaffold), flexibility (low to high degrees of freedom), and charges (neutral, monocationic, and dicationic). The performance is assessed for both predicting absolute binding energies and ranking relative

energies of DNA-intercalator complexes. Two intercalators, 9-aminoacridine (9-AA)<sup>1+</sup> and amonafide<sup>1+</sup>, were used for further validation purposes. MM/PBSA energy decomposition was performed to partition the total binding energy into individual interaction components. Lastly, the minimum number of replicas required to obtain reliable binding energy estimates is analyzed. To our knowledge, this is the first study to evaluate ensemble-based DNA-intercalator binding energies across such a diverse set of intercalators.

## Computational methods

### Docking of intercalators into DNA

The structure of a double-stranded DNA, 5'-CCATCGCTACC-3', was obtained from the Protein Data Bank (PDB ID: 1HX4).<sup>30</sup> The structure is bound to an intercalator from the phenanthrenes class. This ligand was removed to create an empty intercalation site where other intercalators can be docked. In this study, 18 intercalators known to bind within GC base pairs, and have available experimental binding energy data, were selected. These include 16 intercalators (doxorubicin<sup>1+</sup>, proflavine<sup>1+</sup>, ellipticine<sup>1+</sup>, coralyne<sup>1+</sup>, idarubicin<sup>1+</sup>, mitoxantrone<sup>2+</sup>, sanguinarine<sup>1+</sup>, quinacrine<sup>2+</sup>, acridine<sup>1+</sup>, thionine<sup>1+</sup>, thalidomide, NB506, ethidium<sup>1+</sup>, berberine<sup>1+</sup>, chelerythrine<sup>1+</sup>, and adriamycinone) for evaluating the accuracy and reliability of the method, as well as two additional intercalators (9-AA<sup>1+</sup> and amonafide<sup>1+</sup>) for validation purposes. These 18 intercalators belong to different chemical classes, such as anthracyclines, acridine derivatives, benzophenanthridine and isoquinoline alkaloids, and phenanthridinium and phenothiazinium dyes. They also vary in molecular size, intercalating scaffolds, degrees of freedom, and charges. The inclusion of such a diverse set of intercalators enables validation of the ensemble MD-based method across broad chemical classes. The three-dimensional structures of the selected intercalators was collected from the PubChem database.<sup>31</sup> The structure of all the intercalators selected in this study was protonated at physiological pH using 'FixpKa' within the 'QUACAPAC' program. Different conformers were generated for all the intercalators using the 'OMEGA' tool.<sup>32</sup>

Docking of the intercalators into the intercalation site of the DNA molecule was carried out using 'FRED' in OpenEye 4.2.1.0.<sup>32,33</sup> The 'Make Receptor' utility<sup>32,33</sup> was used to generate a docking grid box covering the full DNA structure including its intercalation site. The dimensions of the box are 41.8 Å × 27.9 Å × 19.5 Å, and its volume is 22 898 Å<sup>3</sup>. Chemgauss4 docking scores are reported, and the binding poses were analyzed using VIDA 4.4.0.4<sup>34</sup> and PyMOL.<sup>35</sup>

### Molecular dynamics simulations

The docked DNA-intercalator complexes obtained from the previous part were used as starting geometries for the MD simulations performed using NAMD 3.0.<sup>36</sup> The ensemble MD simulation protocol adopted in ref. 20 was used in this study. Briefly, binding energies of different DNA-intercalator complexes were calculated using 1-trajectory MD simulation.<sup>37</sup> In this study, energies of the complex, receptor, and ligand were extracted from the



same MD simulation trajectory of each DNA-intercalator complex. However, for estimating the deformation energy, 2-trajectory MD simulations were used,<sup>37</sup> where the receptor energy is extracted from the MD simulation trajectories of DNA alone (native DNA, generated using the builder tool of PyMOL), and from the same DNA structure in the DNA-intercalator complex (deformed).

For the MD simulations, 18 different DNA-intercalator complexes were prepared using different programs of AMBER-Tools<sup>38</sup> in the AMBER 22 software package.<sup>39</sup> Using the 'ante-chamber' program,<sup>40</sup> AM1-BCC atomic charges<sup>41</sup> and atom types were assigned for all the intercalator structures. Further, the DNA and intercalator structures were parameterized using 'tleap'.<sup>40</sup> The amber leaprc.DNA.OL15 nucleic acid force field<sup>42</sup> and the GAFF2 force field<sup>43</sup> were used for the DNA and intercalators, respectively. The DNA-intercalator complexes were then neutralized by adding sodium counterions and lastly solvated by adding a box of TIP3P water molecules.

Previous studies on DNA-intercalator complexes have shown that average binding energies predicted from an ensemble of short simulations (25 replicas  $\times$  10 ns each) differ by no more than 1 kcal mol<sup>-1</sup> from those obtained using an ensemble of longer simulations (25 replicas  $\times$  100 ns each), with deviations from experimental values being within 1 kcal mol<sup>-1</sup>.<sup>20</sup> Accordingly, in this study, binding energies were estimated using ensemble simulations of 25 replicas, 10 ns each, *i.e.* the total cumulative time, per complex, is 250 ns. Each replica starts with a random starting velocity.

A series of six energy minimization cycles was performed for each of the systems. Initially, a restraint force constant of 500 kcal mol<sup>-1</sup> was applied, and the restraint force constants were gradually reduced by half in each minimization cycle. In the final cycle, the full system undergoes energy minimization without any restraints, allowing the solute and solvent molecules to move freely. Further, the DNA-intercalator systems were gradually heated from 0 K to 300 K for 100 ps. The temperature was controlled by the NVT ensemble. After heating, the DNA-intercalator systems underwent a series of equilibration simulations for a total of 2 ns. During the initial cycles, the heavy atoms in the systems were restrained, and the restraint forces were gradually reduced by half in each cycle. The systems were then equilibrated without any restraint forces for 1.2 ns. Finally, a production run was performed for each of the equilibrated DNA-intercalator systems for 10 ns at a constant pressure and temperature (NPT ensemble). The trajectory data was collected every 20 ps during the production run, for a total of 500 frames. Additionally, to estimate the deformation energy, an ensemble MD simulation was performed on the DNA structure after removing the bound intercalator, using the same protocol.

System stability during the MD simulations was assessed by analyzing, using the 'cpptraj' module of AMBERTools,<sup>44</sup> the RMSD of the heavy atoms in DNA with respect to the starting geometry for each of the replicas.

### Binding energy calculations

Binding energies of DNA-intercalator complexes were estimated using the MMPBSA.py program<sup>37</sup> of AMBER 22. Using the

trajectory data of each simulation, the MM/PBSA energies were estimated.<sup>7,45</sup> For each MD trajectory of 10 ns, 500 frames were considered. The MM/PBSA energy ( $\Delta G_{\text{MM/PBSA}}$ ) of each DNA-intercalator complex was calculated using the following equation:

$$\Delta G_{\text{MM/PBSA}} = G_{\text{DNA-ligand}} - (G_{\text{DNA}} + G_{\text{ligand}}), \quad (1)$$

where  $G_{\text{DNA-ligand}}$ ,  $G_{\text{DNA}}$ , and  $G_{\text{ligand}}$  correspond to Gibbs energies of the DNA-intercalator complex, DNA, and the intercalator, respectively. Gibbs energy ( $G$ ) is given by:

$$G = E_{\text{gas}} + G_{\text{solvation}}, \quad (2)$$

where the gas-phase energy ( $E_{\text{gas}}$ ) includes energies of bond stretching, angle bending between bonded atoms, bond rotation, van der Waals interaction energy, and electrostatic energy. Gibbs energy of solvation ( $G_{\text{solvation}}$ ) includes polar and non-polar contributions of the complex interacting with the surrounding solvent.<sup>7,37</sup>

For an absolute binding energy prediction, entropy and deformation energy correction terms were added to the MM/PBSA energies of the DNA-intercalator complexes. Entropy ( $T\Delta S$ ) was computed using the normal mode program of MMPBSA.py. Given that the normal mode analysis is computationally expensive and it produces unreliable results if a small number of frames are used,<sup>8</sup> 20 equally placed frames were selected from each replica. The deformation energy is the difference in energy between the DNA in its bound form (DNA-intercalator complex) and its unbound (ligand-free DNA) form. The corrected binding energy,  $\Delta G_{\text{corrected}}$ , is given by:

$$\Delta G_{\text{corrected}} = \Delta G_{\text{MM/PBSA}} - T\Delta S + \text{deformation energy} \quad (3)$$

### Statistical analysis

To determine the minimum number of replicas required to reproducibly predict binding energies using the MM/PBSA method, bootstrap analyses were performed. Initially, samples consisting of varying numbers of replicas (from 1 to 25) were generated. For each sample of size ' $n$ ' (*i.e.* ' $n$ ' replicas), 10 000 bootstrap samples were generated by randomly selecting ' $n$ ' binding energy values, with replacement, from the original dataset. The mean binding energies and standard deviations were calculated for each of these bootstrap samples. The estimated standard deviations were plotted as a function of the number of replicas to identify the minimum number needed to reach a standard deviation of maximum 1 kcal mol<sup>-1</sup>.

## Results and discussion

### Docking poses and scores of DNA-intercalator complexes

At physiological pH, doxorubicin, proflavine, ellipticine, coralyne, idarubicin, sanguinarine, acridine, thionine, ethidium, berberine, and chelerythrine exist in their monocationic form, with a +1 charge; mitoxantrone and quinacrine form dications



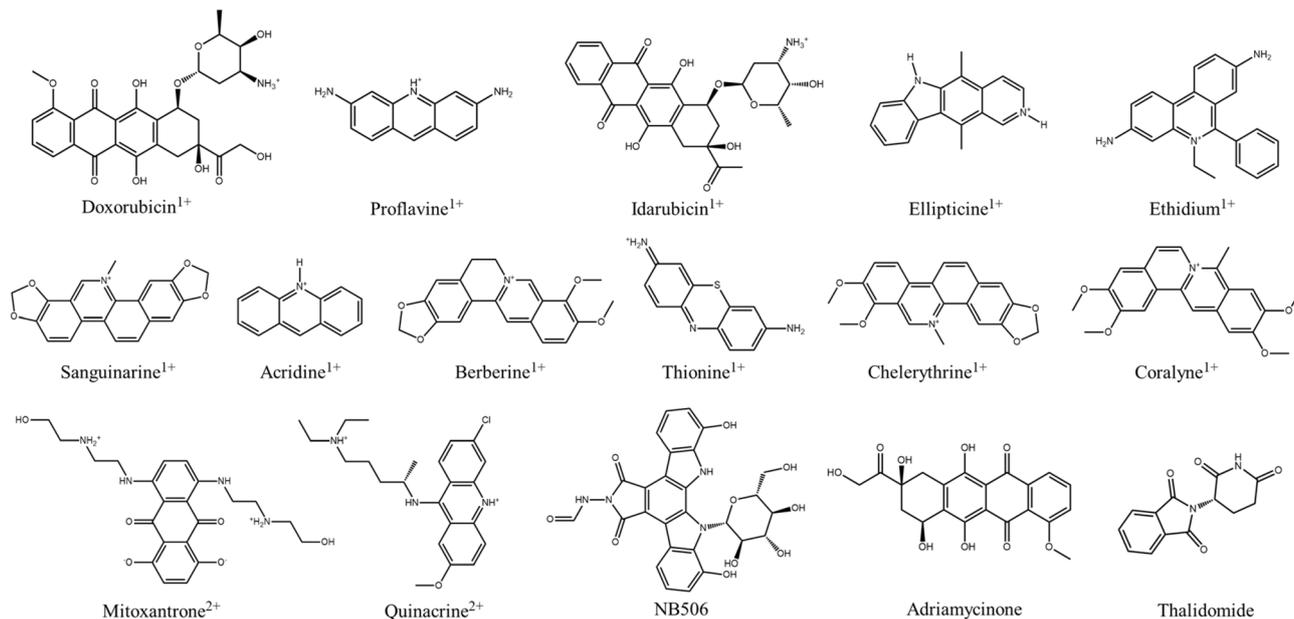


Fig. 1 2D structures of the DNA intercalators and their charges at physiological pH.

with a +2 charge; and thalidomide, NB506, and adriamycinone (also called doxorubicinone) exist in their neutral form. The 2D structures of the intercalators are given in Fig. 1.

The intercalators used in this study vary in the size of the intercalator scaffold, charges, and degrees of freedom. This provides distinct modes of  $\pi$ - $\pi$  stacking, electrostatic interactions, and DNA structural change upon intercalation. Docking of these intercalators was performed at the intercalation site, within two neighboring GC base pairs. Fig. 2 shows the superimposed binding poses of all intercalators at the intercalation site, along with the docking scores of all the DNA-intercalator complexes.

As shown in Fig. 2, all 16 intercalators are bound into the same intercalation site. Positively charged intercalators with large and planar polycyclic aromatic rings had high docking

scores. For example, coralyne<sup>1+</sup> had the highest docking score of  $-13.2$ , followed by doxorubicin<sup>1+</sup> ( $-12.7$ ). Two out of three neutral ligands, thalidomide ( $-8.3$ ) and NB506 ( $-7.5$ ), had the least favorable scores. The docking scores were plotted against their corresponding experimental binding energies (listed in Table 1), as depicted in Fig. 3.

Docking scores showed a weak correlation ( $R^2 = 0.38$ ) with experimental binding energies (see Fig. 3). This result is not unexpected given the approximate nature of docking scores. The largest deviations from the trend line are observed for the neutral intercalators, adriamycinone, thalidomide, and NB506. Adriamycinone had a good docking score of  $-12.0$  but had a relatively weaker experimental binding energy of  $-6.3$  kcal mol<sup>-1</sup>. While docking yielded reasonable binding poses and relative fits, they did not account for DNA conformational flexibility. Therefore, in this study, the docking poses are

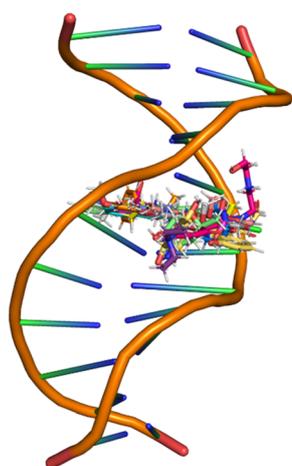


Fig. 2 Superimposed top-ranked docked poses of 16 intercalators within the DNA binding site, along with their docking scores.

| Intercalator                | Docking Score |
|-----------------------------|---------------|
| Coralyne <sup>1+</sup>      | -13.2         |
| Doxorubicin <sup>1+</sup>   | -12.7         |
| Adriamycinone               | -12.0         |
| Chelerythrine <sup>1+</sup> | -11.9         |
| Berberine <sup>1+</sup>     | -11.8         |
| Sanguinarine <sup>1+</sup>  | -11.6         |
| Idarubicin <sup>1+</sup>    | -11.5         |
| Proflavine <sup>1+</sup>    | -11.3         |
| Quinacrine <sup>2+</sup>    | -11.3         |
| Acridine <sup>1+</sup>      | -10.9         |
| Mitoxantrone <sup>2+</sup>  | -10.6         |
| Ellipticine <sup>1+</sup>   | -10.4         |
| Thionine <sup>1+</sup>      | -9.3          |
| Ethidium <sup>1+</sup>      | -9.1          |
| Thalidomide                 | -8.3          |
| NB506                       | -7.5          |

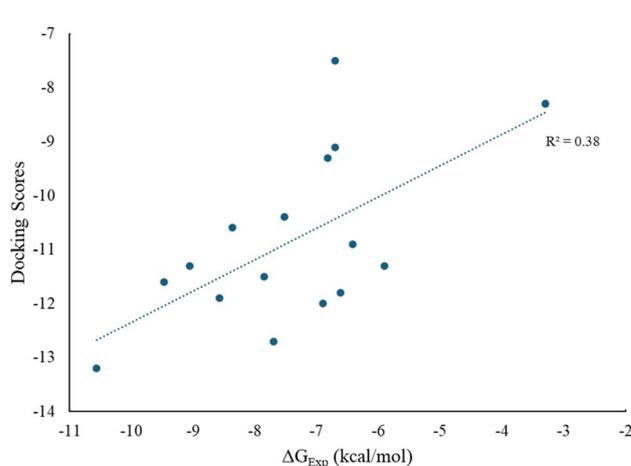


Fig. 3 Docking scores of 16 DNA-intercalator complexes versus their corresponding experimental binding energies,  $\Delta G_{\text{Exp}}$ .



used as starting structures for subsequent ensemble MD simulations, which allow the complexes to relax in explicit solvent and provide the basis for the MM/PBSA binding energy calculations.

### MM/PBSA binding energies

To evaluate the stability of the complexes during the MD simulations, the RMSD values of the DNA heavy atoms relative to the starting structure were analyzed (see Fig. S1 for an illustrative example). The RMSD plots confirmed the stability of the complexes. The average RMSD values were around 2 Å for most of the replicas, with fluctuations around 1 Å. In a few replicas of certain ligands, the average RMSD was around 3 Å, although the fluctuations remained minimal. The ensemble-averaged MM/PBSA binding energies were largely unaffected by the exclusion of replicas with relatively less stable RMSD trends. After excluding the relatively less stable replicas, the evaluated energies deviated from averages computed using all 25 replicas by no more than 0.2 kcal mol<sup>-1</sup> for all complexes except DNA-coralyn<sup>1+</sup>, which had a deviation of 0.5 kcal mol<sup>-1</sup>. Therefore, the energies were estimated using the full set of replicas for all the subsequent analyses. Table 1 lists the average MM/PBSA energies across 25 MD simulation replicas for the 16 complexes, along with the entropy and deformation terms, corrected MM/PBSA binding energies, and experimental values.

In general, the  $\Delta G_{\text{MM/PBSA}}$  increases in magnitude from neutral to monocationic to dicationic intercalators. Consistent with previous findings,<sup>20</sup> despite using ensemble MD simulations, none of the DNA-intercalator systems reproduced the experimental binding energies when the correction terms, for

entropy and deformation, were excluded. For the DNA-intercalator complexes studied, the predicted ensemble average MM/PBSA binding energies are in the range of  $-57.4 \pm 6.7$  kcal mol<sup>-1</sup> (mitoxantrone<sup>2+</sup>) to  $-21.6 \pm 2.6$  kcal mol<sup>-1</sup> (thalidomide), while the experimental values are in the range of  $-10.6$  kcal mol<sup>-1</sup> (coralyne<sup>1+</sup>) to  $-3.3$  kcal mol<sup>-1</sup> (thalidomide). The uncorrected MM/PBSA binding energies are overestimated as they solely account for the highly favorable enthalpic terms driven by the  $\pi$ - $\pi$  stacking of aromatic rings, van der Waals interactions, and favorable electrostatics between the ligand (most of which are positively charged) and the anionic DNA backbone. Experimental thermodynamic studies have also shown that the DNA intercalation process is characterized by a large favorable binding enthalpy, counterbalanced by a large unfavorable entropy.<sup>14-17</sup> It is further observed that the uncorrected MM/PBSA energies have high standard deviations (see Table 1), which reflect high variations in energies among replicas. For example, the percent standard deviations in the binding energies of DNA complexed with NB506, thalidomide, mitoxantrone<sup>2+</sup>, idarubicin<sup>1+</sup> are 19%, 12%, 12%, and 11%, respectively. For the remaining intercalators, this percentage is between 2% and 7%. Overall, the predicted ensemble-averaged MM/PBSA binding energies failed to reproduce the experimental ranking of the compounds ( $R^2 = 0.45$ , see Fig. 4).

### Effect of entropy and deformation corrections on MM/PBSA binding energies

In terms of deformation in DNA-intercalator systems, when an intercalator inserts between the base pairs, the helix unwinds

**Table 1** MM/PBSA binding energies ( $\Delta G_{\text{MM/PBSA}}$ ) of the 16 studied intercalators complexed with DNA, along with the energy correction terms, entropy ( $T\Delta S$ ) and deformation energy, the corrected MM/PBSA binding energies ( $\Delta G_{\text{corrected}}$ ), as well as the experimental binding energies ( $\Delta G_{\text{Exp}}$ ) and their sources. All the listed energies are averaged over 25 replicas of MD simulations per ligand, and are reported with standard deviations in kcal mol<sup>-1</sup>

| Intercalator                           | $\Delta G_{\text{MM/PBSA}}$<br>(kcal mol <sup>-1</sup> ) | Entropy<br>(kcal mol <sup>-1</sup> ) | Deformation energy<br>(kcal mol <sup>-1</sup> ) | $\Delta G_{\text{corrected}}$<br>(kcal mol <sup>-1</sup> ) | $\Delta G_{\text{Exp}}$<br>(kcal mol <sup>-1</sup> ) | Ref. of $\Delta G_{\text{Exp}}$<br>(kcal mol <sup>-1</sup> ) |
|--|--|--------------------------------------|---|--|--|--|
| <b>Dicationic intercalators</b>        |  |                                      |   |  |  |  |
| Quinacrine <sup>2+</sup>               | $-51.7 \pm 3.8$  | $-21.6 \pm 1.7$                      | $19.7 \pm 3.6$                                  | $-10.4 \pm 3.7$  | -9.1   | 46   |
| Mitoxantrone <sup>2+</sup>             | $-57.4 \pm 6.7$  | $-25.2 \pm 2.8$                      | $20.8 \pm 5.1$                                  | $-11.6 \pm 3.6$  | -8.4   | 47   |
| <b>Monocationic intercalators</b>      |  |                                      |   |  |  |  |
| Coralyne <sup>1+</sup>                 | $-42.3 \pm 0.9$  | $-16.3 \pm 0.7$                      | $15.2 \pm 4.3$                                  | $-10.9 \pm 4.7$  | -10.6  | 48   |
| Sanguinarine <sup>1+</sup>             | $-39.9 \pm 1.3$  | $-15.7 \pm 1.2$                      | $15.9 \pm 2.2$                                  | $-8.9 \pm 2.2$   | -9.5   | 49   |
| Chelerythrine <sup>1+</sup>            | $-40.1 \pm 0.9$  | $-16.0 \pm 0.6$                      | $13.0 \pm 2.7$                                  | $-10.4 \pm 2.9$  | -8.6   | 50   |
| Idarubicin <sup>1+</sup>               | $-42.7 \pm 4.9$  | $-21.2 \pm 0.8$                      | $14.7 \pm 2.1$                                  | $-6.8 \pm 4.4$   | -7.9   | 51   |
| Doxorubicin <sup>1+</sup> <sup>a</sup> | $-50.5 \pm 2.6$  | $-23.5 \pm 1.8$                      | $18.0 \pm 1.4$                                  | $-8.9 \pm 1.6$   | -9.9   | 15   |
| Ellipticine <sup>1+</sup>              | $-39.9 \pm 1.3$  | $-15.4 \pm 0.5$                      | $15.9 \pm 2.2$                                  | $-8.6 \pm 2.5$   | -7.5   | 52   |
| Thionine <sup>1+</sup>                 | $-37.2 \pm 2.4$  | $-15.7 \pm 0.9$                      | $14.9 \pm 2.2$                                  | $-6.6 \pm 3.0$   | -6.8   | 53   |
| Ethidium <sup>1+</sup>                 | $-37.9 \pm 0.7$  | $-16.9 \pm 0.6$                      | $14.5 \pm 2.3$                                  | $-6.6 \pm 2.2$   | -6.7   | 14   |
| Berberine <sup>1+</sup>                | $-37.2 \pm 0.8$  | $-15.7 \pm 0.6$                      | $15.5 \pm 3.0$                                  | $-6.1 \pm 3.0$   | -6.6   | 54   |
| Acridine <sup>1+</sup>                 | $-31.3 \pm 0.7$  | $-12.4 \pm 0.5$                      | $12.9 \pm 2.1$                                  | $-6.1 \pm 2.1$   | -6.4   | 55   |
| Proflavine <sup>1+a</sup>              | $-33.5 \pm 2.4$  | $-15.9 \pm 1.1$                      | $12.0 \pm 2.3$                                  | $-5.6 \pm 1.4$   | -5.9   | 56   |
| <b>Neutral intercalators</b>           |  |                                      |   |  |  |  |
| Adriamycinone                          | $-22.9 \pm 1.4$  | $-16.6 \pm 0.9$                      | $15.9 \pm 2.8$                                  | $9.6 \pm 3.6$  | -6.3   | 15   |
| NB506                                  | $-28.8 \pm 5.5$  | $-20.8 \pm 1.6$                      | $22.2 \pm 4.3$                                  | $14.2 \pm 6.6$   | -6.7   | 57   |
| Thalidomide                            | $-21.6 \pm 2.6$  | $-14.4 \pm 1.0$                      | $12.6 \pm 4.0$                                  | $5.4 \pm 3.6$  | -3.3   | 58   |

<sup>a</sup> The energies are taken from ref. 20.



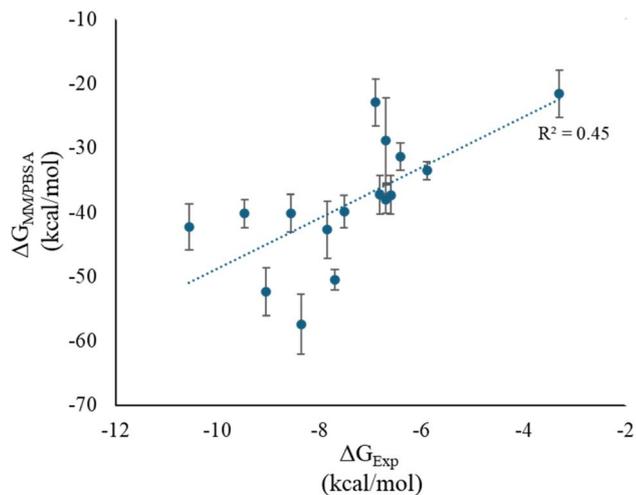


Fig. 4 The uncorrected MM/PBSA ( $\Delta G_{\text{MM/PBSA}}$ ) energies, with the standard deviation bars, of 16 intercalators bound to DNA versus their corresponding experimental binding energies ( $\Delta G_{\text{Exp}}$ ).

and bends. This disrupts the base stacking and sometimes stretches hydrogen bonds, which are energy driven processes.<sup>18,59,60</sup> The deformation energy accounts for the internal energy cost of distorting the receptor from its unbound to its bound conformation. It is a positive energetic penalty that is not part of standard MM/PBSA, unless added explicitly. Jawad *et al.* reported that the deformation energy is an unfavorable component in the binding of doxorubicin to DNA.<sup>18</sup> Similarly in this study, it is observed that the ensemble average deformation energies are positive, ranging from  $12.0 \pm 2.3 \text{ kcal mol}^{-1}$  to  $22.2 \pm 4.3 \text{ kcal mol}^{-1}$  (see Table 1). Including the deformation energy brings the uncorrected MM/PBSA binding energies closer to the experimental values. Table 1 also shows that the deformation energies are relatively independent of the intercalators' charge.

In terms of entropy, the ligand may need to adopt a less favorable conformation to fit in the intercalation site. The entropy term accounts for the loss of degrees of freedom by the ligand and the receptor upon binding. Table 1 lists the estimated entropies for the 16 DNA-intercalator systems. They are largely independent of the intercalator charge. Compared to the uncorrected MM/PBSA energies and the deformation energies, the entropy across the replicas showed minimal fluctuations, as indicated by the smaller standard deviations (see Table 1). The largest ensemble average entropy ( $T\Delta S$ ) was  $-25.2 \pm 2.8 \text{ kcal mol}^{-1}$  (for mitoxantrone<sup>2+</sup>), and the lowest was  $-12.4 \pm 0.5 \text{ kcal mol}^{-1}$  (for acridine<sup>1+</sup>). Large, flexible intercalators with many rotatable bonds, such as mitoxantrone<sup>2+</sup>, tend to incur substantial entropic penalties upon DNA binding due to the restriction of their conformational freedom. Mitoxantrone<sup>2+</sup>, which contains two long flexible side chains, had the highest entropic penalty. This was followed by penalties in the range of  $21\text{--}24 \text{ kcal mol}^{-1}$  for quinacrine<sup>2+</sup>, which contains a long flexible side chain, and for idarubicin<sup>1+</sup>, doxorubicin<sup>1+</sup>, and NB506, each of which contains a substituted heterocyclic ring capable of rotation. Ethidium<sup>1+</sup> also contains a rotatable ring, but since

it is a less flexible benzene ring, the entropic penalty is more moderate ( $16.9 \pm 0.6 \text{ kcal mol}^{-1}$ ). Whereas smaller and relatively rigid molecules, such as acridine, bind mainly *via*  $\pi\text{--}\pi$  stacking and have a smaller entropic penalty. The addition of entropy to the MM/PBSA binding energy further improves the predicted MM/PBSA energy.

Among the monocationic intercalators, the entropy of most of them was within a narrow range of  $-12$  to  $-16 \text{ kcal mol}^{-1}$ , except for idarubicin<sup>1+</sup> and doxorubicin<sup>1+</sup>, which had relatively large entropy values of  $-21.2 \pm 0.8 \text{ kcal mol}^{-1}$  and  $-23.5 \pm 1.8 \text{ kcal mol}^{-1}$ , respectively. Similarly, the range of deformation energies was also  $-12$  to  $-16 \text{ kcal mol}^{-1}$ , except for doxorubicin<sup>1+</sup>, which had a higher deformation energy of  $18.0 \pm 1.4 \text{ kcal mol}^{-1}$ . However, it is also noted that the uncorrected MM/PBSA (*i.e.* enthalpy) values of these two intercalators happen to be large enough to compensate for these exceptionally high energy correction terms, in such a way the corrected MM/PBSA binding energies of these two intercalators remain comparable to experimental values.

It is reported in the literature that MM/PBSA energies estimated from ensemble MD simulations are reasonable for reproducing the ranking of the compounds based on the binding energy, but not for predicting absolute binding energies.<sup>29</sup> However, in the present study, after adding the energy correction terms, the method reproduced not only the ranking of intercalators but also the experimental binding energies. For example, after adding the entropy,  $T\Delta S$  ( $-15.7 \pm 0.9 \text{ kcal mol}^{-1}$ ) and the deformation ( $14.9 \pm 2.2 \text{ kcal mol}^{-1}$ ) energy terms, the corrected MM/PBSA energy (averaged over 25 replicas) of the DNA-thionine<sup>1+</sup> complex was  $-6.6 \pm 3.0 \text{ kcal mol}^{-1}$ . This value aligns well with the experimental binding energy of  $-6.8 \text{ kcal mol}^{-1}$  (see Table 1). For mitoxantrone<sup>2+</sup>, the average corrected MM/PBSA energy ( $-11.6 \text{ kcal mol}^{-1}$ ) shows an over-binding by *ca.*  $3.0 \text{ kcal mol}^{-1}$  compared to the experimental value ( $-8.4 \text{ kcal mol}^{-1}$ ). However, this average is reported with a standard deviation of  $\pm 3.6 \text{ kcal mol}^{-1}$ , *i.e.* the alignment between the computed and the experimental values is still very good. Overall, the corrected MM/PBSA energies for all cationic intercalators, except for mitoxantrone<sup>2+</sup>, differed from the experimental values by no more than  $2 \text{ kcal mol}^{-1}$  (see Table 1). However, the neutral intercalators (adriamycinone, NB506, and thalidomide) were clear outliers, a trend that was already evident in the molecular docking results. These intercalators had repulsive values, with positive corrected MM/PBSA energies of  $9.6 \pm 3.6 \text{ kcal mol}^{-1}$ ,  $14.2 \pm 6.6 \text{ kcal mol}^{-1}$ , and  $5.4 \pm 3.6 \text{ kcal mol}^{-1}$ , respectively (see Table 1). Among these, NB506 was the worst, with an average corrected MM/PBSA energy of  $14.2 \pm 6.6 \text{ kcal mol}^{-1}$ , which is much lower than the experimental value of  $-6.7 \text{ kcal mol}^{-1}$ . Even when considering the individual binding energies from independent simulation trajectories, none of the replicas reproduced the experimentally determined binding energy for any of these three intercalators. Unlike cationic intercalators, neutral intercalators lack charged functional groups and form fewer electrostatic interactions with DNA. Charged groups on the intercalators form specific contacts, such as hydrogen bonds and salt bridges with the neighboring bases,<sup>11,18,61</sup> which also increases the enthalpic



contribution. In fact, experimental studies have shown that removing a charged amino group from a classic intercalator, ethidium bromide, significantly reduced the binding affinity by at least 100-fold, and also flipped the binding energy from enthalpy-driven to entropy-driven.<sup>62</sup> Furthermore, among the outliers, adriamycinone is also an aglycone form of doxorubicin. Therefore, it is possible that the removal of the sugar moiety not only reduces electrostatic interactions but can also alter the mode of binding. Experimental studies by Tartakoff *et al.* reported that adriamycinone, an aglycone of doxorubicin, does not bind to the DNA through intercalation despite retaining a planar aromatic ring structure.<sup>63</sup> In the case of thalidomide, although experimentally measured binding energies have been reported for its binding as a possible DNA intercalator, certain other studies report that it binds potentially to DNA grooves.<sup>64</sup> For NB506, this molecule has been reported to preferentially bind to triplex DNA among 12 different nucleic acid structures and sequences tested.<sup>65</sup> In addition, it has been reported that the binding energy of NB506 to DNA is significantly less favorable than that of the classic intercalator, doxorubicin.<sup>57</sup> Therefore, the disagreement between MM/PBSA-predicted and experimental binding energies for the three neutral ligands may be because of the incorrect assumptions about their binding mode in solution. Furthermore, the experimental binding energies of the intercalators compiled from the literature were obtained using different techniques, such as spectrophotometric titration,<sup>46–48,50,51,54,56,66,67</sup> isothermal titration calorimetry,<sup>14,49,53</sup> and equilibrium binding assays.<sup>52,55,57,58</sup> Deviations by up to *ca.* 2.0 kcal mol<sup>-1</sup> were observed in the experimental values of certain DNA-intercalator complexes.<sup>68,69</sup> This is due to variations in the methods or experimental conditions such as DNA sequence, buffer composition, pH, and ionic strength. Such differences may also influence the observed correlations with computational predictions. Lastly, beyond these listed reasons, having more accurate predicted energies for the cationic intercalators than for neutral ones, may be attributed to the more consistent treatment of coulombic interactions relative to other energy terms in the MM/PBSA method (the energy decomposition analysis will be discussed at the end of this section).

Although adding entropy and deformation energy corrections improved the agreement between corrected MM/PBSA energies and experimental values, the overall correlation was reduced, with an  $R^2$  value of 0.32 (see Fig. 5) (compared to  $R^2$  of 0.45 in Fig. 4). This is even lower than the correlation obtained with docking scores. However, when the three outliers (NB506, thalidomide, and adriamycinone, all of which are neutral) were excluded from the analysis, the  $R^2$  value improved to 0.70. Corrected MM/PBSA binding energies ( $R^2 = 0.70$ ) outperformed both uncorrected MM/PBSA energies and docking scores in ranking cationic ligands ( $R^2 = 0.34$  and  $R^2 = 0.30$ , respectively). Furthermore, since the experimental reference values were obtained from different methods, the correlation analysis was repeated with experimental binding energies measured using exclusively isothermal titration calorimetry. This further improved the correlation, ( $R^2 = 0.85$ ). To validate the predictive capability implied by this correlation (*i.e.*  $R^2 = 0.70$  when

experimental methods are not filtered), binding energies were estimated for two additional charged DNA-intercalator complexes, DNA-9AA<sup>1+</sup> and DNA-amonafide<sup>1+</sup>. The corrected MM/PBSA binding energies of these two complexes, obtained from ensemble MD simulations of 25 replicas, were  $-5.2 \pm 2.6$  kcal mol<sup>-1</sup> and  $-6.5 \pm 4.1$  kcal mol<sup>-1</sup>, respectively. Using these values in the regression relationship obtained from the set of charged intercalators ( $y = 1.27x + 1.68$ ,  $R^2 = 0.70$  from Fig. 5), the experimental binding energies are predicted to be  $-5.4 \pm 2.0$  kcal mol<sup>-1</sup> and  $-6.4 \pm 3.2$  kcal mol<sup>-1</sup> for DNA-9AA<sup>1+</sup> and DNA-amonafide<sup>1+</sup>, respectively. These values are well aligned with the actual experimental binding energies,  $-6.6$  kcal mol<sup>-1</sup> for 9AA<sup>1+</sup> and  $-6.9$  kcal mol<sup>-1</sup> for amonafide<sup>1+</sup>. This agreement indicates that the regression relationship can be used to predict the binding energies of additional intercalators. However, in this case, the corrected MM/PBSA energies alone perform equally well, as they are not only reliable for ranking but also reasonably accurate in absolute terms. Overall, these results demonstrate that corrected MM/PBSA energies from ensemble MD simulations reliably predict the binding energies of cationic intercalators. Such predictive capability is particularly valuable in cases where new intercalators are difficult or costly to synthesize, have limited stability, or require challenging or resource-intensive experimental assays to determine their binding affinities. In these situations, this established regression relationship enables the simulated MM/PBSA binding energies to be translated into estimated experimental values, thereby offering a practical and efficient alternative for preliminary screening and prioritization of candidate molecules.

Furthermore, MM/PBSA energy decomposition was performed to partition the total binding energy into individual interaction components. This analysis identifies the relative contributions of van der Waals (VDWAALS) and electrostatic (EEL) interactions, as well as polar (EPB) and non-polar (ENPOLAR) solvation terms (see Table 2). This analysis reveals

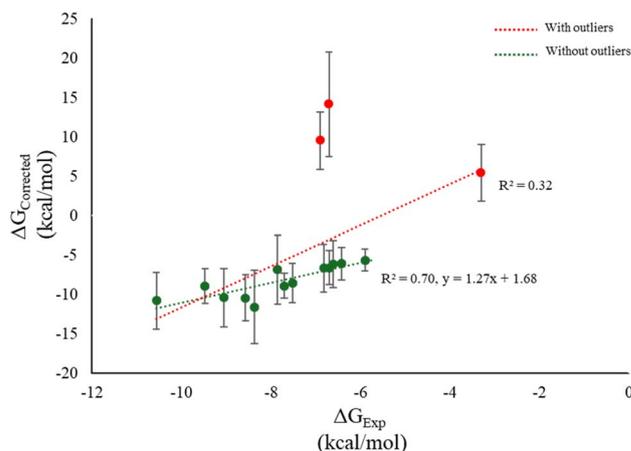


Fig. 5 Corrected MM/PBSA ( $\Delta G_{\text{Corrected}}$ ) energies with standard deviation bars of intercalators bound to DNA versus their corresponding experimental binding energies ( $\Delta G_{\text{Exp}}$ ). The red trendline is for all the 16 DNA-intercalator systems and the green trendline is after excluding the three outliers shown in red.



**Table 2** Components of the MM/PBSA energy for the intercalation of different ligands into DNA. The components include van der Waals interactions (VDWAALS), electrostatic interactions (EEL), polar solvation (EPB), and non-polar solvation (ENPOLAR). The energy terms are reported as an average from 25 replicas of 10 ns MD simulations, along with the standard deviation, in kcal mol<sup>-1</sup>

| Intercalator                           | VDWAALS (kcal mol <sup>-1</sup> ) | EEL (kcal mol <sup>-1</sup> ) | EPB (kcal mol <sup>-1</sup> ) | ENPOLAR (kcal mol <sup>-1</sup> ) |
|--|-----------------------------------|-------------------------------|-------------------------------|-----------------------------------|
| Quinacrine <sup>2+</sup>               | -56.8 ± 3.7                       | -1034.1 ± 26.9                | 1042.1 ± 26.3                 | -4.3 ± 0.2                        |
| Mitoxantrone <sup>2+</sup>             | -60.2 ± 5.7                       | -1054.1 ± 35.3                | 1061.5 ± 34.4                 | -4.2 ± 0.2                        |
| Coralyne <sup>1+</sup>                 | -51.3 ± 1.7                       | -509.9 ± 5.6                  | 522.6 ± 7.5                   | -3.6 ± 0.1                        |
| Sanguinarine <sup>1+</sup>             | -47.2 ± 1.1                       | -514.9 ± 4.5                  | 525.3 ± 5.0                   | -3.3 ± 0.1                        |
| Chelerythrine <sup>1+</sup>            | -47.8 ± 0.9                       | -510.5 ± 2.6                  | 521.6 ± 3.0                   | -3.4 ± 0.1                        |
| Idarubicin <sup>1+</sup>               | -60.3 ± 4.1                       | -525.6 ± 19.7                 | 549.6 ± 16.9                  | -4.1 ± 0.1                        |
| Doxorubicin <sup>1+</sup> <sup>a</sup> | -68.2 ± 2.9                       | -554.3 ± 12.0                 | 577.6 ± 11.6                  | -4.8 ± 0.2                        |
| Ellipticine <sup>1+</sup>              | -44.4 ± 0.4                       | -511.2 ± 3.9                  | 518.8 ± 2.6                   | -2.9 ± 0.0                        |
| Thionine <sup>1+</sup>                 | -38.9 ± 0.3                       | -510.1 ± 8.5                  | 514.4 ± 6.3                   | -2.7 ± 0.0                        |
| Ethidium <sup>1+</sup>                 | -42.3 ± 0.2                       | -499.1 ± 1.5                  | 506.5 ± 1.2                   | -3.1 ± 0.1                        |
| Berberine <sup>1+</sup>                | -45.3 ± 1.0                       | -500.6 ± 2.5                  | 511.9 ± 3.2                   | -3.3 ± 0.1                        |
| Acridine <sup>1+</sup>                 | -34.4 ± 0.3                       | -515.9 ± 4.0                  | 521.6 ± 3.4                   | -2.5 ± 0.0                        |
| Proflavine <sup>1+</sup>               | -36.2 ± 0.8                       | -512.9 ± 5.9                  | 518.4 ± 4.0                   | -2.7 ± 0.1                        |
| Adriamycinone                          | -47.4 ± 1.4                       | -5.2 ± 2.3                    | 32.9 ± 2.2                    | -3.3 ± 0.1                        |
| NB506                                  | -55.7 ± 2.7                       | -32.0 ± 10.8                  | 63.2 ± 10.9                   | -4.2 ± 0.2                        |
| Thalidomide                            | -33.8 ± 2.5                       | -14.3 ± 1.9                   | 29.1 ± 1.8                    | -2.6 ± 0.1                        |

<sup>a</sup> The energies are taken from ref. 20.

differences in interaction patterns among ligands with different charges and chemical structures.

In all 16 DNA-intercalator complexes, VDWAALS is a consistently large and favorable (highly negative) contributor to the binding (see Table 2). The van der Waals term captures the  $\pi$ - $\pi$  stacking interactions of the planar aromatic intercalators with DNA base pairs, which is a key feature of the intercalation mechanism.<sup>11,18,70</sup> For example, the van der Waals term was reported to be the major stabilizing factor for the intercalation of doxorubicin<sup>1+</sup> within DNA.<sup>18-20</sup> This term correlates with the size and planarity of the intercalators.<sup>71</sup> In other words, intercalators with large, planar, and rigid polyaromatic scaffolds exhibit more favorable van der Waals energies due to strong stacking interactions with neighboring DNA bases. For example, molecules with a large aromatic surface area, *e.g.* doxorubicin<sup>1+</sup> (-68.2 ± 2.9 kcal mol<sup>-1</sup>), mitoxantrone<sup>2+</sup> (-60.2 ± 5.7 kcal mol<sup>-1</sup>), and idarubicin<sup>1+</sup> (-60.3 ± 4.1 kcal mol<sup>-1</sup>), had the largest VDWAALS contribution. Whereas, smaller or less planar intercalators, such as acridine<sup>1+</sup> (-34.4 ± 0.3 kcal mol<sup>-1</sup>) and proflavine<sup>1+</sup> (-36.2 ± 0.8 kcal mol<sup>-1</sup>), had smaller VDWAALS contribution. The flexibility or rigidity of the intercalator structure also influenced fluctuations in the VDWAALS term across replicas of each complex. Rigid and smaller intercalators, such as ethidium<sup>1+</sup> and thionine<sup>1+</sup>, had relatively small standard deviations in the VDWAALS term ( $\pm 0.2$  and  $\pm 0.3$  kcal mol<sup>-1</sup>, respectively). On the other hand, flexible intercalators, such as mitoxantrone<sup>2+</sup> and idarubicin<sup>1+</sup>, which have multiple rotatable bonds, had larger fluctuations in VDWAALS term (*ca.*  $\pm 4$  and  $\pm 6$  kcal mol<sup>-1</sup>, see Table 2).

The EEL and EPB are strongly dependent on the charge of the ligand. Compared to the neutral intercalators, many of the cationic intercalators had large attractive coulombic interactions, particularly with the negatively charged phosphate backbone of DNA (see Table 2). Dicationic intercalators, such as mitoxantrone<sup>2+</sup> and quinacrine<sup>2+</sup>, had the largest EEL (-1054.1 ± 35.3 kcal mol<sup>-1</sup> and -1034.1 ± 26.9 kcal mol<sup>-1</sup>, respectively) and

EPB (1061.5 ± 34.4 kcal mol<sup>-1</sup> and 1042.1 ± 26.3, respectively) contributions. Monocationic intercalators had approximately half of these contributions. As expected, the EEL and EPB values for the neutral intercalators were relatively very low (below 32.0 ± 10.8 for EEL and 63.2 ± 10.9 kcal mol<sup>-1</sup> for EPB). The electrostatic interaction energy in the gas phase (listed as EEL in Table 2) is largely counterbalanced by the polar solvation energy. Therefore, the net electrostatic contribution (difference between EEL and EPB) provides a relatively small contribution to the binding of intercalators to the DNA. This observation is consistent with findings from previous studies.<sup>18-20</sup> It is worth noting that both EEL and EPB show large standard deviations, especially for dicationic intercalators. For example, mitoxantrone<sup>2+</sup> has a standard deviation of  $\pm 35.3$  and  $\pm 34.4$  for EEL and EPB, respectively. Small fluctuations in either term can lead to large errors in their combined contribution. For example, if EEL is -500 kcal mol<sup>-1</sup> and EPB is 500 kcal mol<sup>-1</sup>, their net contribution is close to zero. However, a small fluctuation of only 1 to 2% in either term (*i.e.*  $\pm 5$  to  $\pm 10$  kcal mol<sup>-1</sup>) can change the net value by several kcal mol<sup>-1</sup>. This relates to the poor reproducibility of MM/PBSA energy estimated from single trajectories. The ENPOLAR term accounts for the hydrophobic effects. All 16 intercalators showed a favorable ENPOLAR term, ranging from roughly -2.7 ± 0.0 to -4.8 ± 0.2 kcal mol<sup>-1</sup> in the ensemble averages. Overall, the cationic intercalators had relatively low uncertainty in the electrostatic (<4%) and polar solvation (<3%) terms, while neutral intercalators had significantly higher uncertainties in the electrostatic terms (EEL and EPB), with fluctuations exceeding 30% of the predicted value for NB506.

### Optimization of replica number in ensemble MD simulation for intercalators

In a previous study on the DNA-doxorubicin and DNA-proflavine systems, it was reported that using eight replicas of 10 ns each is sufficient to reproduce the absolute intercalation energies.<sup>20</sup> In this



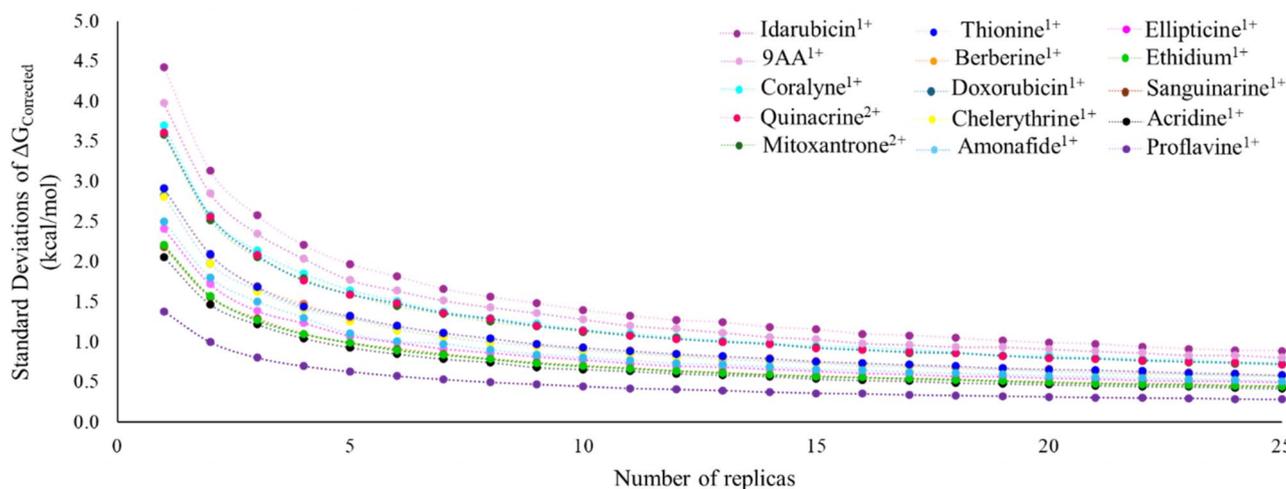


Fig. 6 Standard deviations of the corrected MM/PBSA energies (in kcal mol<sup>-1</sup>) with respect to the number of replicas of 10 ns MD simulations for the cationic DNA-intercalator systems.

section, Fig. 6 shows the bootstrap analyses for corrected MM/PBSA binding energies from ensemble simulations of the cationic DNA-intercalator complexes (including 9-AA<sup>1+</sup> and amonafide<sup>1+</sup>).

Having a cutoff error of *ca.* 1 kcal mol<sup>-1</sup> has been reported as the gold standard in advanced free energy calculations.<sup>4</sup> From Fig. 6, targeting a 1 kcal mol<sup>-1</sup> uncertainty appears to be a practical compromise between precision and computational cost. The bootstrap analyses in Fig. 6 show that sampling with an increased number of replicas decreased the standard deviation in the corrected MM/PBSA energies, until a plateau is reached. In other words, after a certain optimal number of replicas, the energy is effectively converged and running more replicas will have only negligible improvements in precision. The minimum number of replicas required to predict binding energies varies from system to another. Proflavine required only 2 replicas to fall below an accepted standard deviation threshold of 1 kcal mol<sup>-1</sup>, while, on the other extreme, 20 replicas were needed for the idarubicin<sup>1+</sup> system to meet the same threshold. The DNA-intercalator systems that required more than ten replicas were coralyne<sup>1+</sup>, idarubicin<sup>1+</sup>, quinacrine<sup>2+</sup>, and mitoxantrone<sup>2+</sup>. The standard deviation of the corrected MM/PBSA energies for these systems was greater than 3.5 kcal mol<sup>-1</sup>. For sanguinarine<sup>1+</sup>, acridine<sup>1+</sup>, and ethidium<sup>1+</sup>, five replicas were required. Overall, most complexes with cationic intercalators (11 systems) had reached *ca.* 1 kcal mol<sup>-1</sup> uncertainty using up to 10 independent replicas.

## Conclusions

This study shows that ensemble MD simulations with entropy and deformation energy corrections improve the reproducibility and accuracy of MM/PBSA binding energy predictions for DNA-intercalator complexes. Ensemble MD simulations efficiently reproduced the experimental ranking of the cationic intercalators. They also accurately reproduced the experimental binding energies within *ca.* 1 to 2 kcal mol<sup>-1</sup>. Although the overall correlation between corrected MM/PBSA and experimental binding energies yielded an *R*<sup>2</sup> of 0.32, excluding the

neutral intercalators showed a better agreement with an *R*<sup>2</sup> of 0.70. This demonstrates that the protocol used in this study is more reliable for the selected charged intercalators. This linear regression was used to predict experimental binding energies of two additional cationic ligands that were not included in the original set of intercalators, and the predicted energies were within *ca.* 1.0 kcal mol<sup>-1</sup> of the actual values. This supports the reliable use of the method, when combined with entropy and deformation energy corrections, for predicting binding energies of cationic DNA intercalators, although the results remain sensitive to starting geometries and initial velocities in MD simulations. Energy decomposition analyses highlighted the major contributing energy terms for the different types of intercalators. In addition, bootstrap analyses showed that ten independent MD replicas per system are sufficient to achieve binding energy uncertainties below 1 kcal mol<sup>-1</sup> for most cationic systems, although up to 20 replicas were required for a few cases.

## Author contributions

Anju Choorakottayil Pushkaran: data curation, writing – original draft, writing – review & editing, visualization, investigation, validation, formal analysis. Alya A. Arabi: conceptualization, data curation, writing – original draft, writing – review & editing, visualization, investigation, validation, formal analysis, resources, project administration, supervision, funding acquisition.

## Conflicts of interest

The authors declare that they have no competing interest.

## Data availability

The sources for DNA and intercalator structures, as well as the details of the software used in this study, are provided in the 'Computational methods' section.



Supplementary information (SI): the RMSD plots for 25 independent MD simulation replicas of the DNA-thionine<sup>1+</sup> complex. See DOI: <https://doi.org/10.1039/d6ra00032k>.

## Acknowledgements

This study is supported by the UAEU Strategic Research Program via the Zayed Bin Sultan Al Nahyan Center for Health Sciences (code: G00003650), and by the College of Medicine and Health Sciences Research Grant (code: 12M325). We would like to acknowledge OpenEye Scientific (<https://www.eyesopen.com/>) for providing us with the academic version of their software.

## References

- 1 M. Godziba and S. Ciesielski, Natural DNA Intercalators as Promising Therapeutics for Cancer and Infectious Diseases, *Curr. Cancer Drug Targets*, 2020, **20**, 19–32.
- 2 S. Wan, A. P. Bhati and P. V. Coveney, Comparison of Equilibrium and Nonequilibrium Approaches for Relative Binding Free Energy Predictions, *J. Chem. Theory Comput.*, 2023, **19**, 7846–7860.
- 3 A. P. Bhati, S. Wan and P. V. Coveney, Equilibrium and Non-equilibrium Ensemble Methods for Accurate, Precise and Reproducible Absolute Binding Free Energy Calculations, *J. Chem. Theory Comput.*, 2025, **21**(1), 440–462.
- 4 V. Gapsys, A. Yildirim, M. Aldeghi, Y. Khalak, D. van der Spoel and B. L. de Groot, Accurate Absolute Free Energies for Ligand–Protein Binding Based on Non-equilibrium Approaches, *Commun. Chem.*, 2021, **4**, 61.
- 5 S. Genheden and U. Ryde, The MM/PBSA and MM/GBSA Methods to Estimate Ligand-binding Affinities, *Expert Opin. Drug Discovery*, 2015, **10**, 449–461.
- 6 T. Tuccinardi, What is the Current Value of MM/PBSA and MM/GBSA Methods in Drug Discovery?, *Expert Opin. Drug Discovery*, 2021, **16**, 1233–1237.
- 7 T. Hou, J. Wang, Y. Li and W. Wang, Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations, *J. Chem. Inf. Model.*, 2010, **51**, 69–82.
- 8 H. Sun, L. Duan, F. Chen, H. Liu, Z. Wang, P. Pan, F. Zhu, J. Z. H. Zhang and T. Hou, Assessing the Performance of MM/PBSA and MM/GBSA Methods. 7. Entropy Effects on the Performance of End-point Binding Free Energy Calculation Approaches, *Phys. Chem. Chem. Phys.*, 2018, **20**, 14450–14460.
- 9 B. Knapp, L. Ospina and C. M. Deane, Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas, *J. Chem. Theory Comput.*, 2018, **14**, 6127–6138.
- 10 C. A. Frederick, L. D. Williams, G. Ughetto, G. A. Van Der Marel, J. H. Van Boom, A. Rich and A. H.-J. Wang, *Worldwide Protein Data Bank*, 1989.
- 11 C. Pérez-Arnaiz, N. Busto, J. M. Leal and B. García, New Insights into the Mechanism of the DNA/Doxorubicin Interaction, *J. Phys. Chem. B*, 2014, **118**, 1288–1295.
- 12 A. Soni, P. Khurana, T. Singh and B. Jayaram, A DNA Intercalation Methodology for an Efficient Prediction of Ligand Binding Pose and Energetics, *Bioinformatics*, 2017, **33**, 1488–1496.
- 13 K. J. Breslauer, D. P. Remeta, W. Y. Chou, R. Ferrante, J. Curry, D. Zaunczkowski, J. G. Snyder and L. A. Marky, Enthalpy-Entropy Compensations in Drug-DNA Binding Studies, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 8922–8926.
- 14 J. Ren, T. C. Jenkins and J. B. Chaires, Energetics of DNA Intercalation Reactions, *Biochemistry*, 2000, **39**, 8439–8447.
- 15 J. B. Chaires, S. Satyanarayana, D. Suh, I. Fokt, T. Przewloka and W. Priebe, Parsing the Free Energy of Anthracycline Antibiotic Binding to DNA, *Biochemistry*, 1996, **35**, 2047–2053.
- 16 J. B. Chaires, Energetics of Drug-DNA Interactions, *Biopolymers*, 1997, **44**, 201–215.
- 17 J. B. Chaires, A Thermodynamic Signature for Drug–DNA Binding Mode, *Arch. Biochem. Biophys.*, 2006, **453**, 26–31.
- 18 B. Jawad, L. Poudel, R. Podgornik, N. F. Steinmetz and W.-Y. Ching, Molecular Mechanism and Binding Free Energy of Doxorubicin Intercalation in DNA, *Phys. Chem. Chem. Phys.*, 2019, **21**, 3877–3893.
- 19 B. Jawad, L. Poudel, R. Podgornik and W.-Y. Ching, Thermodynamic Dissection of the Intercalation Binding Process of Doxorubicin to dsDNA with Implications of Ionic and Solvent Effects, *J. Phys. Chem. B*, 2020, **124**, 7803–7818.
- 20 A. C. Pushkaran and A. A. Arabi, Accurate Prediction of DNA-intercalator Binding Energies: Ensemble of Short or Long Molecular Dynamics Simulations?, *Int. J. Biol. Macromol.*, 2025, **306**, 141408.
- 21 B. Knapp, L. Ospina and C. M. Deane, Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas, *J. Chem. Theory Comput.*, 2018, **14**, 6127–6138.
- 22 S. Wan, A. P. Bhati, A. D. Wade and P. V. Coveney, Ensemble-Based Approaches Ensure Reliability and Reproducibility, *J. Chem. Inf. Model.*, 2023, **63**, 6959–6963.
- 23 S. Wan, B. Knapp, D. W. Wright, C. M. Deane and P. V. Coveney, Rapid, Precise, and Reproducible Prediction of Peptide–MHC Binding Affinities from Molecular Dynamics That Correlate Well with Experiment, *J. Chem. Theory Comput.*, 2015, **11**, 3346–3356.
- 24 S. K. Sadiq, D. W. Wright, O. A. Kenway and P. V. Coveney, Accurate Ensemble Molecular Dynamics Binding Free Energy Ranking of Multidrug-Resistant HIV-1 Proteases, *J. Chem. Inf. Model.*, 2010, **50**, 890–905.
- 25 S. Yan, J. M. Peck, M. Ilgu, M. Nilsen-Hamilton and M. H. Lamm, Sampling Performance of Multiple Independent Molecular Dynamics Simulations of an RNA Aptamer, *ACS Omega*, 2020, **5**, 20187–20201.
- 26 A. A. Arabi and A. C. Pushkaran, Artificial Intelligence Models Trained on Data From Molecular Dynamics Trajectories, *BMC Artif. Intell.*, 2026, **2**, 3.
- 27 M. Adler and P. Beroza, Improved Ligand Binding Energies Derived from Molecular Dynamics: Replicate Sampling Enhances the Search of Conformational Space, *J. Chem. Inf. Model.*, 2013, **53**, 2065–2072.



- 28 J. J. Perez, M. S. Tomas and J. Rubio-Martinez, Assessment of the Sampling Performance of Multiple-Copy Dynamics versus a Unique Trajectory, *J. Chem. Inf. Model.*, 2016, **56**, 1950–1962.
- 29 D. W. Wright, B. A. Hall, O. A. Kenway, S. Jha and P. V. Coveney, Computing Clinically Relevant Binding Free Energies of HIV-1 Protease Inhibitors, *J. Chem. Theory Comput.*, 2014, **10**, 1228–1241.
- 30 C. H. Lin, X. Huang, A. Kolbanovskii, B. E. Hingerty, S. Amin, S. Broyde, N. E. Geacintov and D. J. Patel, Molecular Topology of Polycyclic Aromatic Carcinogens Determines DNA Adduct Conformation: A Link to Tumorigenic Activity, *J. Mol. Biol.*, 2001, **306**, 1059–1080.
- 31 S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang and E. E. Bolton, PubChem 2025 update, *Nucleic Acids Res.*, 2024, **53**, D1516–D1525.
- 32 P. C. D. Hawkins, A. G. Skillman, G. L. Warren, B. A. Ellingson and M. T. Stahl, Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database, *J. Chem. Inf. Model.*, 2010, **50**, 572–584.
- 33 M. McGann, FRED and HYBRID Docking Performance on Standardized Datasets, *J. Comput.-Aided Mol. Des.*, 2012, **26**, 897–906.
- 34 VIDA 4.4.0.4, *OpenEye, Cadence Molecular Sciences*, Santa Fe, NM, <https://www.eyesopen.com>.
- 35 *The PyMOL Molecular Graphics System, Version 3.0*, Schrödinger, LLC.
- 36 J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kalé, K. Schulten, C. Chipot and E. Tajkhorshid, Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD, *J. Chem. Phys.*, 2020, **153**, 044130.
- 37 B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, MMPBSA.py: An Efficient Program for End-State Free Energy Calculations, *J. Chem. Theory Comput.*, 2012, **8**, 3314–3321.
- 38 D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, N. Forouzes, T. J. Giese, A. W. Götz, H. Gohlke, S. Izadi, K. Kasavajhala, M. C. Kaymak, E. King, T. Kurtzman, T.-S. Lee, P. Li, J. Liu, T. Luchko, R. Luo, M. Manathunga, M. R. Machado, H. M. Nguyen, K. A. O'Hearn, A. V. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, A. Risheh, S. Schott-Verdugo, A. Shajan, J. Swails, J. Wang, H. Wei, X. Wu, Y. Wu, S. Zhang, S. Zhao, Q. Zhu, T. E. Cheatham, D. R. Roe, A. Roitberg, C. Simmerling, D. M. York, M. C. Nagan and K. M. Merz, AmberTools, *J. Chem. Inf. Model.*, 2023, **63**, 6183–6191.
- 39 D. A. Case, H. M. Aktulga, K. Belfon, I. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, *et al.*, *Amber 2022*, University of California, San Francisco, 2022.
- 40 H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, J. T. Berryman, S. R. Brozell, F. S. Carvahol, D. S. Cerutti, T. E. Cheatham III, G. A. Cisneros, V. W. D. Cruzeiro, T. A. Darden, N. Forouzes, M. Ghazimirsaeed, G. Giambaşu, T. Giese, M. K. Gilson, H. Gohlke, A. W. Goetz, J. Harris, Z. Huang, S. Izadi, S. A. Izmailov, K. Kasavajhala, M. C. Kaymak, I. Kolossv'ary, A. Kovalenko, T. Kurtzman, T. S. Lee, P. Li, Z. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, M. Manathunga, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, A. Rahnamoun, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, J. Wang, X. Wu, Y. Wu, Y. Xiong, Y. Xue, D. M. York, C. Zhao, Q. Zhu, D. A. Case and P. A. Kollman, *Amber 2022*, University of California, San Francisco, 2022.
- 41 A. Jakalian, D. B. Jack and C. I. Bayly, Fast, Efficient Generation of High-quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation, *J. Comput. Chem.*, 2002, **23**, 1623–1641.
- 42 M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo and P. Jurečka, Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA, *J. Chem. Theory Comput.*, 2015, **11**, 5723–5736.
- 43 X. He, V. H. Man, W. Yang, T.-S. Lee and J. Wang, A Fast and High-Quality Charge Model for the Next Generation General AMBER Force Field, *J. Chem. Phys.*, 2020, **153**, 114502.
- 44 D. R. Roe and T. E. Cheatham, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
- 45 P. Su, C. Tsai, S. Mehboob, K. E. Hevener and M. E. Johnson, Comparison of Radii Sets, Entropy, QM Methods, and Sampling on MM-PBSA, MM-GBSA, and QM/MM-GBSA Ligand Binding Energies of *F. tularensis* Enoyl-ACP Reductase FabI, *J. Comput. Chem.*, 2015, **36**, 1859–1873.
- 46 M. Hossain and G. Suresh Kumar, DNA Intercalation of Methylene Blue and Quinacrine: New Insights into Base and Sequence Specificity from Structural and Thermodynamic Studies with Polynucleotides, *Mol. Biosyst.*, 2009, **5**, 1311.
- 47 S. Dogra, P. Awasthi, M. Nair and R. Barthwal, Interaction of Anticancer Drug Mitoxantrone with DNA Hexamer Sequence D-(ctcgag)<sub>2</sub> by Absorption, Fluorescence and Circular Dichroism Spectroscopy, *J. Photochem. Photobiol., B*, 2013, **123**, 48–54.
- 48 K. Bhadra, M. Maiti and G. S. Kumar, Thermodynamics of the Binding of Cytotoxic Protoberberine Molecule Coralyne to Deoxyribonucleic Acids, *Biochim. Biophys. Acta, Gen. Subj.*, 2008, **1780**, 298–306.
- 49 M. Hossain and G. Suresh Kumar, Thermodynamic Profiles of the DNA Binding of Benzophenanthridines Sanguinarine



- and Ethidium: A Comparative Study with Sequence Specific Polynucleotides, *J. Chem. Thermodyn.*, 2010, **42**, 1273–1280.
- 50 P. Basu and G. Suresh Kumar, Elucidation of the DNA Binding Specificity of the Natural Plant Alkaloid Chelerythrine: A Biophysical Approach, *J. Photochem. Photobiol., B*, 2014, **138**, 282–294.
- 51 C. Ozluer and H. E. S. Kara, In Vitro DNA Binding Studies of Anticancer Drug Idarubicin Using Spectroscopic Techniques, *J. Photochem. Photobiol., B*, 2014, **138**, 36–42.
- 52 G. Dodin, M. Schwaller, J. Aubard and C. Paoletti, Binding of Ellipticine Base and Ellipticinium Cation to Calf-thymus DNA: A Thermodynamic and Kinetic Study, *Eur. J. Biochem.*, 1988, **176**, 371–376.
- 53 P. Paul, M. Hossain and G. Suresh Kumar, Calorimetric and Thermal Analysis Studies on the Binding of Phenothiazinium Dye Thionine with DNA Polynucleotides, *J. Chem. Thermodyn.*, 2011, **43**, 1036–1043.
- 54 X.-L. Li, Y.-J. Hu, H. Wang, B.-Q. Yu and H.-L. Yue, Molecular Spectroscopy Evidence of Berberine Binding to DNA: Comparative Binding and Thermodynamic Profile of Intercalation, *Biomacromolecules*, 2012, **13**, 873–880.
- 55 S. Ichimura, M. Zama, H. Fujita and T. Ito, The Nature of Strong Binding between Acridine Orange and Deoxyribonucleic Acid As Revealed by Equilibrium Dialysis and Thermal Renaturation, *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.*, 1969, **190**, 116–125.
- 56 A. Basu and G. Suresh Kumar, Thermodynamic Characterization of Proflavine–DNA Binding through Microcalorimetric Studies, *J. Chem. Thermodyn.*, 2015, **87**, 1–7.
- 57 C. Carrasco, H. Vezin, W. D. Wilson, J. Ren and J. B. Chaires, DNA Binding Properties of the Indolocarbazole Antitumor Drug NB-506, *Anti-Cancer Drug Des.*, 2001, **16**, 99–107.
- 58 H. P. Koch and M. J. Czejka, Evidence for the Intercalation of Thalidomide into DNA: Clue to the Molecular Mechanism of Thalidomide Teratogenicity?, *Z. Naturforsch., C: J. Biosci.*, 1986, **41**, 1057–1061.
- 59 H. M. Berman and P. R. Young, The Interaction of Intercalating Drugs with Nucleic Acids, *Annu. Rev. Biophys. Bioeng.*, 1981, **10**, 87–114.
- 60 A. S. Biebricher, I. Heller, R. F. H. Roijmans, T. P. Hoekstra, E. J. G. Peterman and G. J. L. Wuite, The Impact of DNA Intercalators on DNA and DNA-processing Enzymes Elucidated through Force-Dependent Binding Kinetics, *Nat. Commun.*, 2015, **6**, 7304.
- 61 E. E. Brossard and S. A. Corcelli, Mechanism of Daunomycin Intercalation into DNA from Enhanced Sampling Simulations, *J. Phys. Chem. Lett.*, 2024, **15**, 5770–5778.
- 62 N. C. Garbett, N. B. Hammond and D. E. Graves, Influence of the Amino Substituents in the Interaction of Ethidium Bromide with DNA, *Biophys. J.*, 2004, **87**, 3974–3981.
- 63 S. S. Tartakoff, J. M. Finan, E. J. Curtis, H. M. Anchukaitis, D. J. Couture and S. Glazier, Investigations into the DNA-Binding Mode of Doxorubicinone, *Org. Biomol. Chem.*, 2019, **17**, 1992–1998.
- 64 S. Yasmeen, F. A. Qais, M. Rana, A. Islam and Rahisuddin, Binding and Thermodynamic Study of Thalidomide with Calf Thymus DNA: Spectroscopic and Computational Approaches, *Int. J. Biol. Macromol.*, 2022, **207**, 644–655.
- 65 J. Ren, C. Bailly and J. B. Chaires, Nb-506, an Indolocarbazole Topoisomerase I Inhibitor, Binds Preferentially to Triplex DNA, *FEBS Lett.*, 2000, **470**, 355–359.
- 66 J. M. Crenshaw, D. E. Graves and W. A. Denny, Interactions of Acridine Antitumor Agents with DNA: Binding Energies and Groove Preferences, *Biochemistry*, 1995, **34**, 13682–13687.
- 67 L. Xie, Y. Xu, F. Wang, J. Liu, X. Qian and J. Cui, Synthesis of New Amonafide Analogues via Coupling Reaction and Their Cytotoxic Evaluation and DNA-Binding Studies, *Bioorg. Med. Chem.*, 2009, **17**, 804–810.
- 68 X.-L. Li, Y.-J. Hu, H. Wang, B.-Q. Yu and H.-L. Yue, Molecular Spectroscopy Evidence of Berberine Binding to DNA: Comparative Binding and Thermodynamic Profile of Intercalation, *Biomacromolecules*, 2012, **13**(3), 873–880.
- 69 D. Bhowmik, M. Hossain, F. Buzzetti, R. D'Auria, P. Lombardi and G. S. Kumar, Biophysical Studies on the Effect of the 13 Position Substitution of the Anticancer Alkaloid Berberine on Its DNA Binding, *J. Phys. Chem. B*, 2012, **116**(7), 2314–2324.
- 70 F. Zunino, A. Di Marco, A. Zaccara and R. A. Gambetta, The Interaction of Daunorubicin and Doxorubicin with DNA and Chromatin, *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.*, 1980, **607**, 206–214.
- 71 D. Řeha, M. Kabeláč, F. Ryjáček, J. Šponer, J. E. Šponer, M. Elstner, S. Suhai and P. Hobza, Intercalators. 1. Nature of Stacking Interactions between Intercalators (Ethidium, Daunomycin, Ellipticine, and 4',6-Diaminide-2-phenylindole) and DNA Base Pairs. Ab Initio Quantum Chemical, Density Functional Theory, and Empirical Potential Study, *J. Am. Chem. Soc.*, 2002, **124**, 3366–3376.

