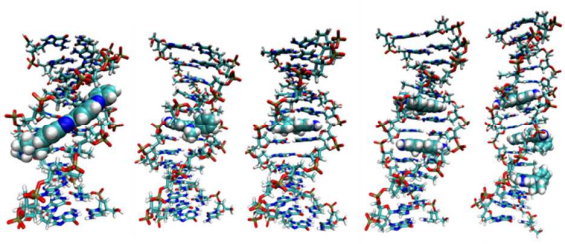
PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Graphical Abstract:



Text:

How drug molecules perturb the conformational freedom of a DNA double helix fragment is investigated by molecular dynamics simulations.

Atomistic account of structural and dynamical changes induced by small binders in the double helix of a short DNA

Barbara Fresch, F. Remacle*

Department of Chemistry, B6c, University of Liege, B4000 Liege, Belgium

Abstract

Nucleic acids are flexible molecules and their dynamical properties play a key role in molecular recognition events. Small binders interacting with DNA fragments induce both structural and dynamical changes in the double helix. We study the dynamics of a DNA dodecamer and of its complexes with Hoechst33258, which is a minor groove binder, and with the ethidium cation, which is an intercalator, by molecular dynamics simulation. The thermodynamics of the DNA-drug interaction is evaluated in connection with the structure and the dynamics of the resulting complexes. We identify and characterize the relevant changes of the configurational distribution of the DNA helix and relate them to the corresponding entropic contributions to the binding free energy. The binder Hoechst locks the breathing motion of the minor groove inducing a reduction of the configurational entropy of the helix that amounts to 20 kcal/mol. On the contrary, intercalations with the ethidium cation enhance the flexibility of the double helix. We show that the balance between the energy required to deform the helix for the intercalation and the gain in configurational entropy is the origin of cooperativity in the binding of a second ethidium and of anti-cooperativity in the binding of a third one. The results of our study provide understanding on the relation between structure, dynamics and energetics in the interaction between DNA fragments and small binders, highlighting the role of dynamical changes and consequent variation of the configurational entropy of the DNA double helix for both types of binders.

***corresponding author: fremacle@ulg.ac.be**

1. Introduction

Understanding the molecular basis of the interaction between DNA and small binders is of fundamental interest for the rational design of new therapeutic agents[1, 2] and developments in DNA-based nano technologies.[3-5] Our current understanding of the molecular recognition between biomolecules and drug-like binders involves a dynamic process where the molecular partners undergo structural and dynamical adaptation to implement binding and unbinding events.[6-11] The central role of molecular flexibility in ligand-receptor interaction is now well recognized,[9, 10, 12] dynamical modifications occurring upon binding have their thermodynamical counterpart in the entropic factors contributing to the overall binding free energy. Even in the computer-aided drug design, where static docking procedures have been applied for decades, approaches to include the changes of configurational entropy upon binding have been recently proposed.[13, 14] All-atom Molecular Dynamics (MD) of nucleic acids is a powerful method to unveil mechanistic details of oligonucleotide flexibility that are difficult, if not impossible, to observe experimentally.[15-18] DNA oligomers in the canonical B-helix structure are characterized by an overall simple dynamics,[7] which on a coarse-grained scale resemble the dynamics of an elastic rod.[19, 20] At a smaller scale, sequence specific dynamical motifs play a significant role in molecular recognition[21-23] and changes in the flexibility of the double helix contribute to allosteric effects and cooperativity in ligand binding.[24] Two broad classes of non-covalent DNA-binding agents have been identified,[25] groove binders and intercalators. Groove binders fit into the DNA minor groove causing no significant perturbation of the DNA structure[26, 27] whereas intercalators bind by inserting a planar aromatic chromophore between adjacent DNA base pairs[28] which requires a deformation of the helical structure.

We focus here on the dynamical changes induced on the double helix of a DNA dodecamer by the interaction with two small binders representative of these two different binding modes: Hoechst33258 and the ethidium cation. These ligands are well known and their complexes with DNA have been studied both experimentally and theoretically (see e.g. refs [29-31] and references therein). We chose them as prototypes of the two different mechanisms of interaction with the DNA double helix: Hoechst is a minor groove binder that binds selectively to A-rich part of the DNA sequence,[7, 30, 32, 33] while the ethidium cation intercalates between base pairs (BPs) without strong sequence specificity.[29, 30] We use MD simulations to access the dynamics of the free oligonucleotide and of its complexes. The ultimate goal is to identify and estimate the thermodynamic contributions corresponding to the structural and dynamical changes induced by the binding. To this end, the computed trajectories are characterized by principal component analysis (PCA)[34, 35] that allows the identification of the most relevant motions. We find that the main collective mode of the free oligonucleotide entails a minor groove breathing motion that is involved in the molecular recognition process between the double helix and the minor groove binder Hoechst. The B-helix does not undergo important structural modifications upon binding with Hoechst, but its dynamics is strongly modified. We identify the changes in the dynamics of the double helix due to the minor groove binding and we show that they contribute to the energetics of the binding process as an entropic loss of about 20 kcal/mol. Contrary to the minor groove binding of Hoechst the intercalation of an ethidium molecule involves a structural deformation of the double helix but does not influence considerably the dynamics of the double helix. We characterize the structural deformations induced

by different binding modes of ethidium and the corresponding energetics. To study multiple intercalations we analyse complexes between the DNA oligonucleotides and one, two and three ethidium molecules. We show that the balance between the energy required to deform the helix and the gain in configurational entropy is the origin of cooperativity in the binding of a second intercalator and of anti-cooperativity in the binding of a third molecule.

Restriction of the motion of the double helix upon ligation with minor groove binders has been theoretically studied in ref[7] and experimentally observed with nuclear magnetic resonance (NMR) techniques,[36] while in ref.[37] cooperative effects are predicted for the intercalation of daunomycin molecules. We strengthen these previous results by connecting explicitly structural and dynamical modifications of the double helix with the thermodynamics of binding and we provide a comparison of the change in the dynamics of the DNA double helix induced by the two main modes of interaction between DNA and drug molecules (i.e. minor groove binding and intercalation). Overall, the results we present here show that dynamical changes of the double helix upon the interaction with small binders play an important role in determining the binding free energy and cooperative effects for multiple bindings.

2. Methods

2.1 Systems

We investigate the dynamics of a free DNA dodecamer (5'- GGTA AATTTAGG – 3', fig. 1a) in the standard B-helix form and of the complexes DNA-Hoechst33258 (1:1) and DNA-ethidium cation with one (1:1), two (1:2) and three (1:3) intercalating molecules. Hoechst is a fluorescent bis-benzimidazole molecule used in several applications in pharmacology (fig. 1b). It interacts with DNA by entering the minor groove with a clear selectivity for AT rich tract of the double helix.[7, 32] In order to identify a good binding site in the AT rich central region of the considered DNA sequence, we performed preliminary unconstrained 6ns equilibrium simulations on two different binding configurations (see fig. S1 of the supplementary material). Molecular mechanics with generalised Born and surface area solvation (MM-GBSA) analysis[38] (see Methods) has been used to evaluate the relative binding free energy of the two binding sites. The more stable complex (fig. 1d) has been used in all the subsequent simulations. The ethidium cation is a simple polycyclic aromatic molecule with short side chains (fig. 1c) and represents a typical DNA intercalator without strong sequence specificity. Intercalators bind to the DNA double helix via the non-covalent stacking interaction with the DNA base pairs.[28, 29, 39] Depending on the binding mode, the localization of the phenyl residue of ethidium cation can be either into the minor groove or into the major groove of the helix. Available crystal structures[40] do not rule out any of the two possibilities and intercalation from the major groove has been observed during MD simulations.[41] Theoretical studies at the quantum level[42-44] on minimal model for intercalation (two or four base pairs) indicate that the side chain of ethidium and even steric constraints with the nucleic acid backbone are of minor importance in the energetics of the binding. We study the (1:1) complex with one molecule of ethidium intercalated on the 5th ApA step of the double helix in two binding configurations: with the phenyl ring into the minor groove (*eth1mg*, fig. 1e) and with the phenyl ring into the major groove (*eth1MG*, fig. 1f). Since this latter complex is energetically favoured (see section 3.2) we maintain the intercalation mode in the major

groove to study the multiple intercalated complexes (1:2) and (1:3), reported in fig. 1g and h, respectively.

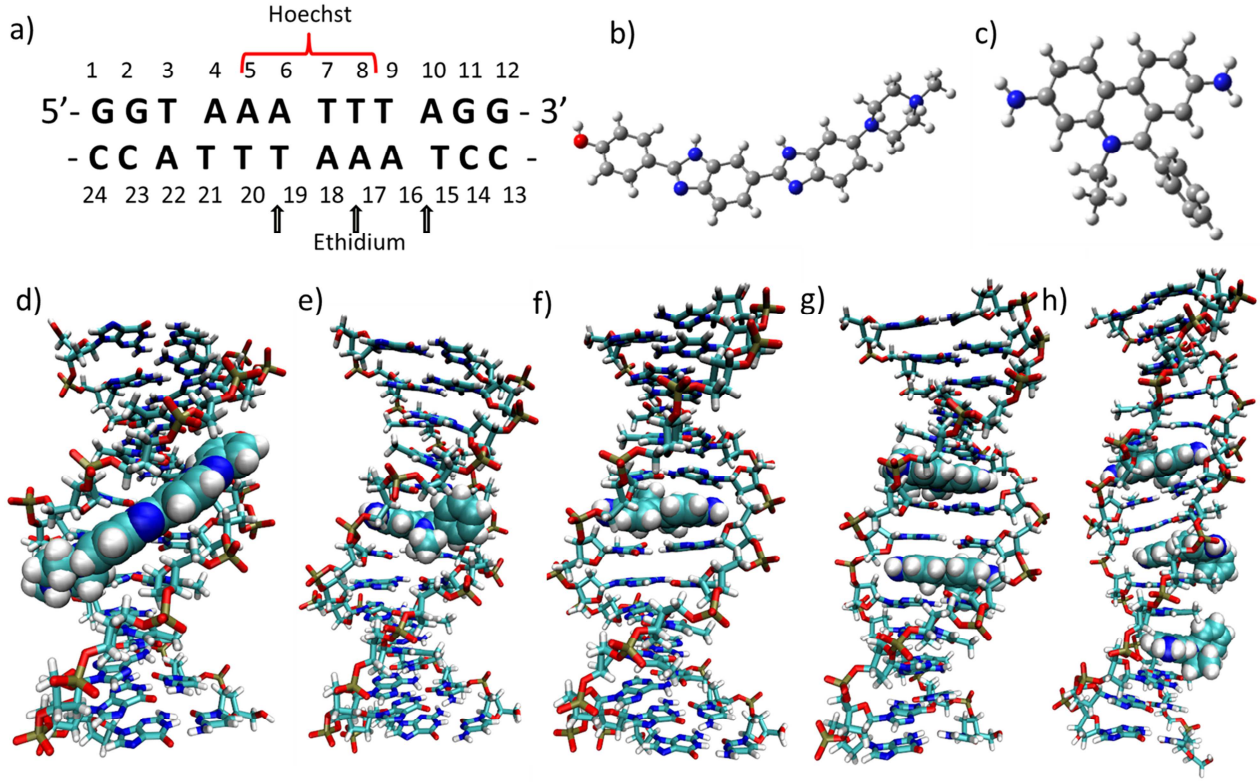


Figure 1: Structural features of the DNA dodecamer and its complexes with Hoechst and ethidium cation. a) sequence of 24 bases in the oligomer with relative numbering. The base pairs (BP) are identified by the number of the nucleotide in the 5'→3' strand. The base pair steps are numbered by the left base pair, i.e. the BP step GpG between the 1st BP (GC) and the 2nd BP (GC) is number 1 and so on. The parenthesis indicates the BPs interacting with Hoechst while the arrows indicate the intercalation sites of the ethidium cation. b) Hoechst molecule. c) ethidium cation molecule. d)-h) Snapshots of low energy structures of the complexes DNA-Hoechst, DNA-ethidium cation (1:1) with the side chains of ethidium into the minor groove (eth1mg) and into the major groove (eth1MG), DNA-ethidium (1:2) and (1:3).

2.2 Simulations

MD simulations of the free DNA dodecamer, DNA-Hoechst and DNA-Ethidium complexes were performed using AMBER 10[45] with the AMBER parm99/bsc0 force field[46] and the TIP3P water model.[47] The force fields for the ligand have been prepared with ANTECHAMBER and corrected according to the parameterization developed in ref[48] for Hoechst. The initial structures were solvated in a cubic periodic box, with a minimum buffer of 10 Å between any DNA or solute atom and the closest box edge. Sodium counterions were added to establish charge neutrality. The resulting system was energy minimized using steepest descent and conjugate gradient methods to relax any residual unfavourable steric interactions introduced during the solvation procedure. First, the system (DNA or DNA complexes) was constrained (500 kcal/mol-Å^2) while the water and counterions were subjected to 10000 cycles of minimization. Then, the full system was allowed to relax during an additional 10000 cycles of unrestrained minimization. The DNA/Hoechst33258/Ethidium cations complexes were restrained (25 kcal/molÅ^2) during a 20 ps, constant volume MD simulation (NVT), during which water and the Na+ atoms were allowed to move freely and the temperature was raised from 0 to 300K using a Langevin temperature control. Next, the system was subjected to 150 ps of constant pressure (NPT) MD to achieve proper density and 5ns (for free DNA and DNA-Hoechst) or 10ns (for DNA-ethidium complexes) of relaxation

before production runs. During MD the long-range electrostatic interactions were treated with the particle-mesh Ewald method[49] using a real-space cut-off distance of $r_{cuttoff}$ = 9 Å. The SHAKE algorithm[50] was used to constrain bond vibrations involving hydrogen atoms, which allowed a time step of 2 fs. We report the analysis of 20 ns of equilibrated dynamics for the free DNA and its complexes, referred to as *dna* and *Hoechst,* for the free dodecamer and the complex with Hoechst. *eth1MG, eth2, eth3* refer to the complexes with 1, 2 and 3 ethidium cations, respectively, with the side chains into the major groove while *eth1mg* denotes the DNA-ethidium (1:1) complex with the side chains of the ligand into the minor groove. The 20ns simulation has been divided into two halves called *A* and *B* of 10 ns each in order to control convergence of parameters and principal components analysis (see Analysis section).

2.3 Analysis

Free energies are estimated by using the MM-GBSA method,[51] in connection with an evaluation of the entropic contribution by normal mode and quasi-harmonic analysis (QHA),[12] as implemented in the MMPBSA.py module[52] of AmberTool12. The MM-GBSA is a post-processing method of evaluating binding free energies of molecules in solution by the analysis of a set of structures collected in the equilibrated portion of a MD trajectory. Solute configurations are sampled from a molecular dynamics simulation using explicit solvent. For each solute configuration, the gas-phase energy is estimated using the same molecular mechanics potential that was used to perform the simulation, but all solvent molecules are ignored, and no cutoffs are used in evaluating the non-bonded interactions. Free energies of solvation are then re-introduced by using a numerical Generalized-Boltzmann (GB) calculation for the electrostatic portion and a surface-area-dependent term (SA) for non-electrostatic contributions to solvation. Extensive tests on 32 molecules chosen as prototypes of nucleic acids and proteins showed that MM-GBSA methodology reproduces solvation free energies of these kinds of systems with a mean error < 1kcal.[53] In practice we sample 100 configurations from the simulations of the free DNA and its complexes. For each ensemble of configurations the estimated free energy can be expressed as

$$G = \langle E_{MM} \rangle + \langle G_{sol} \rangle - T \langle S \rangle = G^{MMGB} - T \langle S \rangle \quad (1)$$

Where the terms on the rhs of the first equality are the molecular mechanics energy without solvent, the solvation free energy as provided by the MM-GBSA and the solute entropy, respectively. The brackets denote the average over the sampled configurations. Binding free energies are then calculated by subtracting the free energies of the unbound receptor and ligand from the free energy of the bound complex

$$\Delta G_{binding} = G^{MMGB}_{complex} - G^{MMGB}_{receptor} - G^{MMGB}_{ligand} - T\Delta S = \Delta G^{MMGB}_{binding} - T\Delta S \quad (2)$$

The entropic contribution related to the reorganization of the solvent upon binding is included in the GBSA methodology. The entropic variation that remains to be evaluated in eq.(2) derives from the loss of rotational and translational degrees of freedom (ΔS_{t+r}) and from changes in the configurational entropy (ΔS_{conf}) of the molecules related to modification of the internal degrees of freedom. The translational and rotational contributions are evaluated according to standard statistical mechanics while the evaluation of the configurational contribution is a problem that can be tackled at different level of approximations.[54] A common route to estimate configurational entropy is based on normal mode analysis and the statistical mechanics of a classical harmonic oscillator. The fundamental limitation of this approach applied to large molecular aggregates is that normal modes are calculated by diagonalization of the Hessian matrix for highly optimized geometries that may bear little similarity to the configurations explored in a room temperature MD simulation. Another common method is based on quasi-harmonic analysis (QHA)[12, 24, 55] which requires the calculation of the eigenvalues ω of the mass-weighted coordinate covariance matrix from the simulation. The quasi-harmonic approach includes some effects of anharmonic terms in the potential, at least to the extent that they influence

the mean square displacement, but it is still limited to motion confined in a quasi-harmonic energy basin. By defining the scaled energy as $\alpha_i = \hbar\omega_i / (kT)$ the entropy is calculated by applying the formula for a quantum mechanical oscillator, as suggested by Andricioaei and Karplus[56]

$$S = k\sum_{i=1}^{3N-6}\left[\frac{\alpha_i}{e^{\alpha_i}-1} - \ln\left(1-e^{-\alpha_i}\right)\right] \quad (3)$$

For a system containing N atoms it is formally necessary to include 3N-6 configurations in the quasi-harmonic analysis to ensure completeness of the data set. We evaluate the entropic contribution by considering the first 5ns of the equilibrated trajectory (2500 configurations). Unless otherwise specified we exclude the terminal G-C base-pairs from the analysis since transient opening of the terminal base pairs are frequent and introduce spurious anharmonic modes in the calculation. The configurational entropy of the system relates to the volume of conformational space accessible to a molecule at a given temperature. Since in a finite-time MD simulation the system might not explore all the accessible region of such a space, the entropy calculated according to eq.(3) depends on the length of the sampling window used to calculate the covariance matrix. In the case in which the system explores a smooth energy landscape during all the simulation time an estimation of the entropy in the limit of infinite sampling can be obtained by fitting its time dependence by the empirical function[12, 24]

$$S = S_\infty - A/t^\alpha \quad (4)$$

where *t* is the simulation time, S_∞ is the asymptotic value of the entropy formally corresponding to a simulation of infinite length while *A* and α are coefficients whose physical meaning remains unclear. We estimate the entropic term by normal mode analysis as well as by the quasi-harmonic approach, extrapolating to the asymptotic value given by (4) when appropriate.

To understand the dynamical changes of the DNA double helix upon binding of the small ligands we analyse the computed trajectories in terms of Principal Components (PCs). Principal Component Analysis (PCA) is a statistical tool that extracts large-scale motions occurring in the MD trajectory, revealing the structure underlying the atomic fluctuations. It defines a new coordinate system through a linear transformation of the atomic coordinates. Such collective coordinates are then used to define a low-dimensional subspace (often called "essential subspace") in which a significant part of the molecular motion is expected to take place. The collective coordinate set is determined as a set of eigenvectors that diagonalizes the variance-covariance matrix of positional atomic fluctuations, after the elimination of the overall translational and rotational motion. In practice, an ensemble of 3N-dimensional coordinate vectors $\mathbf{x}^i$, for i=1,..n, is generated by recording, at a discrete time interval, the configurations generated during the trajectory. Each element $\mathbf{C}_{pq}$ of the covariance matrix is then calculated as the covariance between the p-th and the q-th entry of the coordinate vector **x** on the ensemble of n realizations, namely

$$\mathbf{C}_{pq} = \frac{1}{n}\sum_{i=1}^{n}\left(x_p^i - \langle x_p\rangle\right)\left(x_q^i - \langle x_q\rangle\right) \quad (5)$$

where x_p^i and x_q^i are the elements of $\mathbf{x}^i$, while $\langle x_p\rangle$ and $\langle x_q\rangle$ denote their average over the ensemble of n realizations. The result is a time independent matrix of dimension 3Nx3N. A set of eigenvalues and eigenvectors is obtained by solving the standard eigenvalue problem $\mathbf{CT} = \Lambda\mathbf{T}$ with Λ being the diagonal matrix of the eigenvalues and **T** the transformation matrix whose columns are the eigenvectors of **C**. The i-th eigenvector represents the axis of the i-th collective coordinate in the conformational space, while the associate eigenvalue λ_i gives the mean square fluctuation along that axis. To obtain 3N-6 physically

meaningful eigenvalues, it is formally necessary to include at least 3N-6 configurations in the principal component analysis to ensure completeness of the data set. If $\boldsymbol{\eta}_i$ represents the i-th eigenvector, or Principal Component (PC), of **C**, we can define the projection of the original MD trajectory onto the new direction defined by this principal component

$$p_i(t) = \boldsymbol{\eta}_i \cdot \left(\mathbf{x}(t) - \langle \mathbf{x} \rangle\right) \tag{6}$$

The time evolution of the projections gives indication on the characteristic of the motion along the PC and contributes to the characterization of the underlying energy landscape.[57, 58] By binning the values of the time dependent projection $p_i(t)$, one obtains the corresponding probability distribution, $P(p_i)$. The variance of the distribution of each projection is the variance of the atomic fluctuation along the i-th principal component, i.e. the corresponding eigenvalue, $\lambda_i = \left\langle p_i(t)^2 \right\rangle$. A Gaussian distribution of the projection p_i corresponds to sampling around one single structure and indicates that the explored region of energy landscape is well described by a quadratic potential ("harmonic modes") or it is at least characterized by a harmonic envelope ("quasi-harmonic or singly-hierarchical modes").[59] The ability of the analysis to extract meaningful information on important functional motions and features of the energy landscape depends on the statistical relevance of the configurational subspace sampled within the simulation. To evaluate the reliability of the principal components it is good practice to divide the simulation in two halves and compare the corresponding modes. Two sets of eigenvectors $\boldsymbol{\eta}_i$ and $\boldsymbol{\upsilon}_j$ can be compared with each other by taking their inner product

$$S_{ij} = \boldsymbol{\eta}_i \cdot \boldsymbol{\upsilon}_j \tag{7}$$

Sub-nanosecond MD simulations of protein dynamics showed that even if the individual principal components extracted from different portions of the dynamics correlate poorly, the subspace spanned by the major principal components converges remarkably rapidly.[60] The degree of overlap between essential subspaces is often measured as the root mean square inner product (RMSIP) of the two sets of eigenvectors

$$\mathrm{RMSIP}_M = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{M}\left(\boldsymbol{\eta}_i \cdot \boldsymbol{\upsilon}_j\right)^2} \tag{8}$$

where M is the dimension of the subspaces. In order to visualise and interpret the motion defined by a PC it is useful to project it back to Cartesian coordinates as follows

$$\mathbf{x}_i(t) = p_i(t) \cdot \boldsymbol{\eta}_i + \langle \mathbf{x} \rangle \tag{9}$$

We will use the Cartesian representation of the modes given by eq.(9) to illustrate the atomic displacement involved in the principal components. PCA analysis was carried on as implemented in the PCAsuite software.[61] The convergence of the first five PCs of the free oligonucleotide has been verified by comparing the results obtained by using different portions and different lengths (5 ns, 10 ns and 20 ns) of the total 20 ns trajectory.

Deformations of the DNA double helix structure will be described in terms of the standard six base pair parameters (*shear, buckle, stretch, propeller, stagger, opening)* and six dimer step parameters (*roll*, *tilt*, *twist*, *shift*, *slide*, and *rise).* [62, 63] The minor groove width at the level of the i-th base pair is defined as the distances between the P atoms in different strands separated by 3 base pairs, i.e., the distance between P '(i – 2)···P(i + 2) across the minor groove with three intervening base pairs. Here P(i) and P'(i) are the 5'-phosphates of the complementary nucleotides that comprise base pair i, with the prime used to denote the nonsequence (complementary) strand. For example, the computed minor groove widths at the 5th base-pair step

corresponds to the distances between P'(3)···P(7). All the analyses were performed by the different moduli of the AmberTools12 suite of programs.

3. Results and Discussion

3.1 Minor-Groove Binding

We start by considering the binding of Hoechst to the minor groove of the DNA double helix. The structure of the DNA-Hoechst complex is showed in fig.1d. The H-bond analysis of the 20ns trajectory of the DNA-Hoechst complex reveals that the interaction occurs mainly through the formation of hydrogen bonds between the N-donors of the Hoechst and the free oxygen of the thymine bases T19-20-7, a weaker hydrogen bond is also established with the nitrogen of the adenine base A6, (see Figure 1 for the numbering convention). The end-to-end distance of the double helix (calculated as the distance between the –OH termination of the guanine and cytosine residues at the 5' termini) in the DNA-Hoechst complex (39.7Å±2.3) is practically the same as for the free dodecamer (39.2Å± 2.6). In figure 2 we report relevant parameters describing the structure of the double helix in the free oligonucleotide as well as in the complex with Hoechst. The height of the bars denotes the average value of the selected parameters while the error bars denote their variability quantified as standard deviation calculated along the trajectory. The first two bars (blue and green) of Figure 2 refer to the two halves of the 20ns DNA simulation: all the parameters show consistent values indicating convergence of the results and a good sampling of the helix conformations. The local flexibility of the DNA helix is sequence dependent, pyrimidine-purine base pair steps exhibiting the largest flexibility and deformability in the B-form DNA.[64] Accordingly, we find that BP steps 3 and 9 (TpA) are characterized by a significant positive value of *roll,* that indicates a local bending of the helix, together with a lower value of the local *twist*. Binding of Hoechst to the DNA double helix does not imply significant structural reorganisations as previously reported by both experimental[32] and theoretical[7] studies. The analysis of the structural descriptors of the complex (brown bars in Figure2) shows that binding with Hoechst slightly affects the parameters describing the base pairs directly involved in the interaction (5^{th}, 6^{th}, 7^{th}, 8^{th} BPs). Specifically, higher average values of *buckle* and more negative values of *propeller* deformations indicate that the bases are not co-planar within the base pair in order to maximize the favourable interactions with the binder. However, the base pairs remain relatively free to move as indicated by the fluctuations of these parameters about their average value. The only effect we observe on the helical descriptors is a slight over-twist of the BP steps ApA5 and ApA7, together with the unwinding of the TpT8 base pair step probably due to the steric disturbance of the piperazinyl ring of Hoechst.

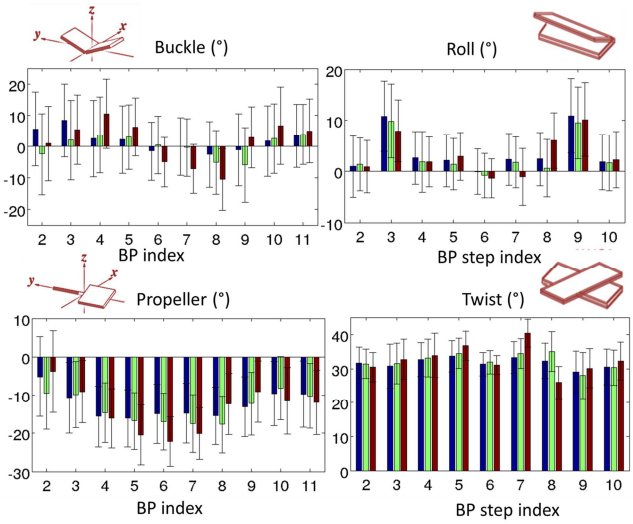


Figure 2: Selected parameters (in degrees) describing the conformation of the base pairs and of helix. The height of the bars denotes the parameter average value while the error bars denote their variability quantified as standard deviation calculated along the trajectory. The first and the second bars (blue and green) refer to first and the second halves of the simulation of the free DNA oligonucleotide. The third bar (red) refers the the simulation of the DNA-Hoechst complex.

We calculate the binding free energy for the Hoechst complex by applying the MM-GBSA scheme described in the Methods section, eq.(2). The contributions to the GBSA free energy of binding are reported in Table 1. The interpretation of the experimental thermodynamic binding data of minor-groove binders usually assumes that the contributions to the binding free energy arising from conformational changes are negligible compared to other forces driving ligand-DNA complexation.[25, 27] This assumption is motivated by the observation that the structures of the double helix and of the ligand are not considerably distorted as observed from x-ray crystallographic studies. Thus, the binding of a ligand to the minor groove of DNA is treated as a rigid-body association. This view is supported by the successful application of the single trajectory (ST) protocol to evaluate the binding energy from molecular dynamics simulations.[65] Within the ST approach, the MD procedure is performed only for the ligand–DNA complex, followed by the extraction of the trajectories of the free molecules, which is equivalent to the assumption that the conformations of the DNA and the ligand in the bound and unbound states are similar. In practice, taking all structures from a single simulation cancels the noise resulting from sampling inconsistencies and the error inherent in force-field and implicit solvation energies, and it is often preferred.[66] For the DNA-Hoechst complex, we implement both the single trajectory (ST) and the multiple trajectory (MT) protocols. In the multiple trajectory approach the MM-GBSA energy contributions of the receptor and the ligand in the unbound state are evaluated on trajectories obtained from separate simulations. The two approaches give very similar results for the DNA-Hoechst complex (see Table 1). The dominant terms of the MM-GBSA binding free energy analysis are Van der Waals and electrostatic interactions, as expected for a positively charged binder establishing H-bond interactions with the DNA bases. The estimation of the entropic contribution by normal mode analysis and quasi-harmonic analysis (QHA), both calculated with a ST protocol, shows a remarkable agreement. The result of the MM-GBSA/ normal mode analysis in the single trajectory approach gives a binding free energy of -34.5±6 kcal/mol and comparable values are obtained by considering the MM-GBSA energy difference calculated with the multiple trajectory approach or by using the QHA estimation of the entropy

calculated with the ST protocol. This value is considerably higher than the experimental binding energy, evaluated between -7kcal/mol and -12kcal/mol depending on the binding site and the measurement technique.[32, 33] We then estimate the configurational entropy of the receptor (i.e. the DNA double helix) from the trajectory of the free oligonucleotide. It is 20 kcal/mol higher than the value estimated by using the dynamics of the double helix in the bound state. By using the free oligonucleotide as receptor we calculate a binding free energy of -11.3 kcal/mol in excellent agreement with the experimental values. The extra entropic contribution deriving from considering the dynamics of the free oligonucleotide rather than the dynamics of the bounded receptor suggests that ligation modifies the dynamics of the double helix and this is what we analyse in details below.

To investigate the dynamical effects of the binding we analyse the trajectories of the free oligonucleotide and of the Hoechst complex in terms of Principal Components (PCs) of the motion of the sugar-phosphate backbone and of the whole double helix. A backbone-only analysis is helpful to disentangle different components of the molecular recognition process: variation in the sugar-phosphate backbone affects the minor groove width producing specific movements essential to clamp the ligands.[7, 67] We found, however, that the main collective modes resulting from the backbone PCA are very similar to those obtained from all-atom analysis.[12] In order to assess the robustness of the analysis we first consider the dependence of the PCs obtained from the covariance matrix of the backbone atoms on the simulation length. In Fig.3a we report the similarity between the first 5 modes obtained from the first (*dnaA*) and the second halves (*dnaB*) of the total 20ns trajectory in the form of the matrix of the inner products between them, see eq. (7). The higher value of the products clustered along the diagonal shows the good convergence of the main principal components. The overlap of the essential subspaces spanned by the first five PCs of the two different portions of the trajectory calculated as the RMSIP defined in eq.(8) amounts to 0.9. The motions described by the eigenvectors are oscillatory about the mean structure, with no net drift that could suggest a change in the overall conformation of the molecule. The corresponding probability densities, shown in figure 3b for the fist eigenvector, are well described by Gaussian distributions expected for motions that explore quasi-harmonic energy basin. These observations establish the robustness of the PCs resulting from the total 20 ns trajectory of the free oligonucleotide. The first two modes resulting from the analysis of the backbone atoms and from the all-atom PCA are shown in Fig 3c: both modes correspond to bending of the duplex as also results from a normal modes analysis of the DNA double helix[20, 68] and resemble the normal motions of a simple elastic rod.

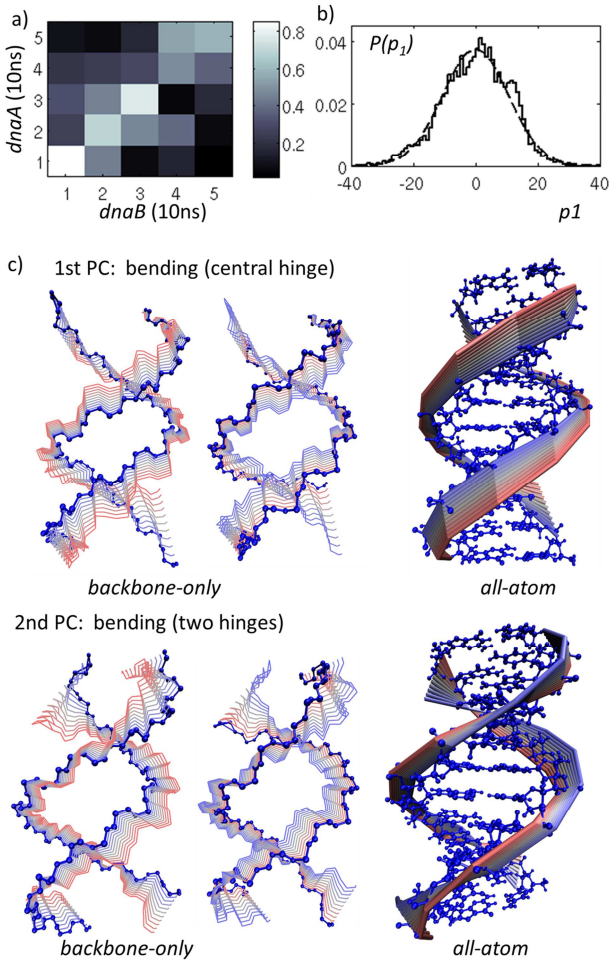


Figure 3: Principal Component Analysis of the free DNA dodecamer: a) Matrix of the inner products between the principal components, eq.(7), obtained by dividing the trajectory into two halves. b) probability distribution of the projection of the atomic positional fluctuations along the 1st principal component, eq.(6). c) Visualization of the atomic motion obtained from eq. (9) involved in the 1st and 2nd principal components of the only-backbone and all-atom PCA. For the backbone modes, the two limit structures together with the traces of the intermediate configurations of the backbone are shown. For comparison, in the representation of the all-atom PCs the trace of the backbone motion involved in the modes is reported. The changing colours of the traces denote different snapshots during the oscillation.

The first principal component entails a "groove breathing" motion in the central portion of the double helix (see figure 1c and the animation available in supplementary materials), with the backbone of the two strands getting far apart during the bending. This motion favours the recognition of a minor groove binder like Hoechst[7] that has to reach the central region of the double-helix in order to establish the H-bond with the DNA bases. The analysis of the principal components on the 20ns trajectory of the DNA-Hoechst complex reveals the presence of a large-amplitude anharmonic motion (see figure SI-2 in the supplementary material) that is due to the transient fraying of the terminal base-pair GC1. To avoid the introduction of spurious effects we refer to the principal components obtained from the first 5ns of the simulation, i.e. prior to the un-pairing of the terminal base-pair. The total variance due to the backbone motion registered during 5ns of the dynamics of the Hoechst complex is smaller that the variance describing the dynamics of the free oligonucleotide during the same time (a summary on the variance data obtained from PCA and the overlap between essential subspaces are reported in Table SI-1 of supplementary material). The question whether individual PCs calculated on relatively short simulations are meaningful, especially with reference to functional motions in protein, has been subject of debates.[60, 69, 70] For the DNA-based systems considered here, the convergence analysis performed in the free oligonucleotide trajectory shows that the first five PCs obtained by analysing a 5ns trajectory are essentially the same obtained by using the whole 20 ns trajectory. Moreover, the elimination of clearly anharmonic modes corresponding to a barrier crossing and exploration of new energy basins (see also the

discussion in Figure SI-2) insures that the motions represented by the PCs are the modes corresponding to a well-defined and quasi-harmonic energy basin that is sufficiently sampled during 5ns of dynamics. The similarity matrix between the first five principal components of the dynamics of the backbone in the free DNA oligonucleotide and in the complexes with Hoechst is reported in Figure 4a and suggests that the first PC of DNA-Hoechst backbone dynamics is analogous to the second PC of the free oligonucleotide. The visual inspection of the modes (see Figure 4b and the mode animation in SI) confirms that in the DNA-Hoechst complex the first and the second modes are switched: the 1st PC is the bending around two hinges while the 2nd PC corresponds to the bending around the central hinge. However, despite the high overlap of the two vectors, the 2nd mode of the Hoechst complex is quite different from the 1st mode of the free oligonucleotide for what concerns the motion of the central region of the double-helix. As one can noticed from the representation of the 2nd PC in figure 4b, the central part of the helix remains rigid while the analogous mode in the free oligonucleotide contains the groove breathing motion. The change in the groove width is thus suppressed by the presence of the Hoechst molecule. The groove is locked in the conformation that maximizes the favourable interaction between the DNA and the binder. The suppression of the groove breathing dynamics is reflected in the minor groove width calculated during the 20 ns simulation in the central part of the double helix reported in figure 4c. The average width of the groove decreases for all the central base pairs (from BP 4 to 8) in the Hoechst complex and the width variation (quantified as twice the standard deviation along the trajectory and represented by the error bars) is greatly reduced especially for the base pairs directly interacting with the binder. The configurational entropy of the double-helix in the free DNA oligonucleotides and in the Hoechst complex for different simulation length are reported in figure 4d. We excluded from the analysis the four GC base pairs located at the termini of the double helix since they bring an-harmonic contributions due to reversible openings observed along the trajectory. The difference in configurational entropy of the double helix is independent from the simulation length and equals 20kcal/mol, as we previously estimated in the first 5ns of the trajectory. By fitting the entropy with eq. (4) we find an asymptotic value of 482.2kcal/mol in the Hoechst complex and 502.6kcal/mol for the free oligonucleotide, which again correspond to a negative variation of 20kcal/mol in the configurational entropy of the DNA double-helix. We can thus conclude that the binding with Hoechst restricts the flexibility of the double helix especially by locking the breathing motion of the minor groove. This dynamical modification is reflected in an entropic penalty for the binding that amounts to 20 kcal/mol and it is thus critical to recover a binding energy compatible with the experimental value.

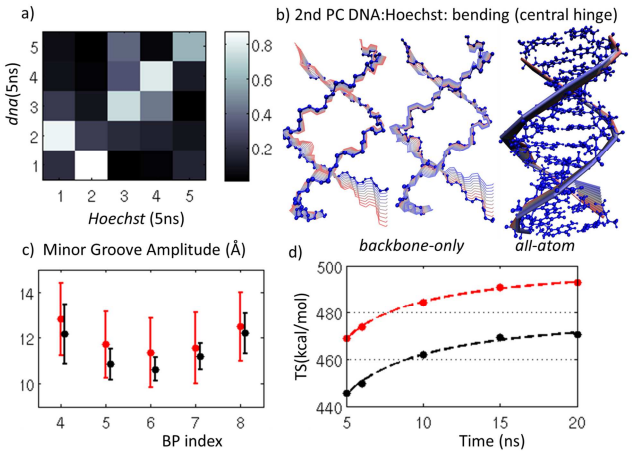


Figure 4: Principal Component Analysis of the DNA-Hoechst complex: a) Matrix of the inner products between the principal components of the backbone motion of the complex and the free oligomer. b) Visualization of the atomic motion obtained from eq. (9) involved in the 2nd principal components of the only-backbone and all-atom PCA in the Hoechst complex, see legend of figure 2. c) average value and fluctuation of minor groove width calculated during the 20ns simulation in the central part of the double helix for the Hoechst complex (black) and the free oligomer (red). d) QHA entropy as a function of the simulation length and fitting (dashes) with eq.(4).

Table 1: MMGBSA analysis of the DNA:Hoechst complex in the single and multiple trajectory protocol. Decomposition of $\Delta G_{binding}$ in the MM-GBSA scheme: Van der Waals (VDWAALS), electrostatic (EEL), polar (EGB) and non-polar (ESURF) contributions to the solvation free energy, total gas phase (DELTA G gas) and solvation (DELTA G solv) binding energy, and resulting MM-GBSA binding energy (DELTA TOTAL). Estimation of the translational, rotational and configurational entropy by normal mode and quasi-harmonic analysis from the single trajectory and QHA estimation by considering the free DNA as receptor. All the values are in kcal/mol.

Energy Components	**Average/Std. Err. (STP)**	**Average/Std. Err. (MTP)**
VDWAALS	-64.63 ± 0.24	-63.7 ± 2.2
EEL	-573.7 ± 1.2	-586.2 ± 13
EGB	590.6 ± 1.1	603.9 ± 12
ESURF	-6.95 ± 0.01	-7.07 ± 0.1
DELTA G gas	-638.3 ± 1.3	-648.6 ± 13
DELTA G solv	583.6 ± 1.1	596.9 ± 12
DELTA TOTAL	**-54.73 ± 0.32**	**-51.7 ± 6**
Entropy Term TΔS	**Normal Mode/QHA**	**QHA**
Translational	-13.1 ± 0/ -13.1	-13.06
Rotational	-11.25 ± 0.02/ -11.26	-11.2
Vibrational	4.1 ± 1.6/ 2.36	-16.1
DELTA S total	**-20.2 ± 1.6/ -21.9**	**-40.4**
$\Delta G_{binding}$	**-34.5 ± 6/ -32.7**	**-11.3**

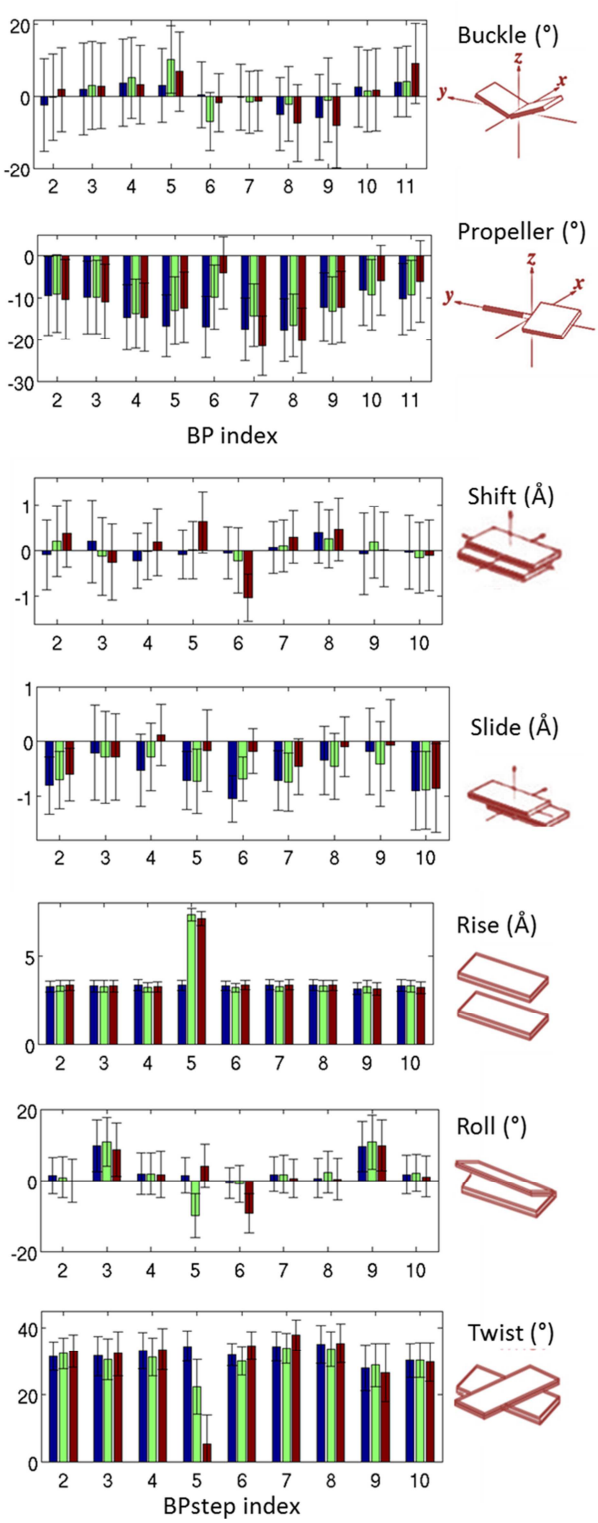
3.2 Intercalation of one ethidium cation

In the complex DNA-ethidium (1:1) the chosen intercalation site is the ApA5 base pair step. We compare two intercalation modes: in *eth1mg* the phenyl ring of the ligand is into the minor groove while in *eth1MG* the phenyl ring points toward the major grove of the helix (figure 1e and f). Unlike the binding to the minor groove discussed above, it is clear that intercalation requires structural change in the double helix. The first step in the binding process of each intercalator is the drug-induced creation of the intercalation site.[71-74] The intercalation site differs from the undisturbed DNA by a doubling of the base–base stacking distance followed by unwinding and conformational changes of the sugar phosphate backbone. Because of the formation of the intercalation site, a MM-GBSA estimation of the binding free energy in a single trajectory protocol overestimates the binding free energy since it does not take into account the energy spent to deform the double helix. The binding energies obtained for the two complexes are reported in Table 2 while the complete results of the analysis, including the different MM-GBSA contributions to the total energy of each species are reported in Table2 of the SI. In the first column of Table 2 we report the binding energy calculated with the ST protocol that ignores the energy required to form the intercalation site. By comparing the MM-GBSA energy of the receptor calculated within the ST protocol and the energy of the free DNA oligonucleotide we calculate the energetic penalty associated to the creation of the intercalation site that is

$$\Delta G_{deformation}^{MMGB} = G_{receptor}^{MMGB,(ST)}\left(eth1\right) - G^{MMGB}\left(DNA\right). \quad (10)$$

Rather unexpectedly, we found that the energies required to create the intercalation site for the two binding modes are significantly different, the energy required for *eth1mg* (23.8 kcal/mol) being more than twice the energy required for the *eth1MG* complex (9.9 kcal/mol). By adding the deformation energy, eq. (10), to the ST binding energy we obtain the MM-GBSA estimation of the binding energy. The intercalation mode of *eth1MG* is energetically favoured mainly because it requires less energy to deform the double helix. The estimation of the entropic contribution does not change this result since the QHA estimation of the entropy of the two complexes gives comparable value (equal within 1 kcal/mole, see Table 2). In comparison to the minor groove binding of Hoechst, the intercalation of a single molecule of ethidium does not modify much the configurational dynamics of the double helix, we observe only a slight increase of configurational entropy upon binding that suggests an enhanced flexibility of the helix with respect to the unbound state. The total solute entropy variation accompanying the binding is still negative due to the lost of global translational and rotational degree of freedom. By taking into account the different deformation energies, we find that the binding configuration with the phenyl ring into the minor groove is less favoured than the binding energy of the ethidium with the side chain directed toward the major groove. The total binding energy of -4.7kcal/mol we obtain for eth1mg compares well with the value calculated with different theoretical methodologies[29] for the same binding configuration. For the binding in the eth1MG the calculated binding energy amounts to -9.6 kcal/mol, both values compare fairly well with the experimental estimation of -7kcal/mol.[75]

In order to understand the structural origin of the different energies required to create the intercalation site in the two binding configurations we compare, in figure 5, the parameters describing the average structure of the double helix in the two complexes *eth1MG* (second bar) and



eth1mg (third bar). The values characterizing the free oligonucleotide helix are also reported for reference (first bar).

Figure 5: Selected parameters describing the conformation of the base pairs and of helix in the free oligonucleotide (blue), eth1MG (green) and eth1mg (brown). The height of the bars denotes the average value of the selected parameters while the error bars denote their variability quantified as standard deviation calculated along the trajectory.

The average step *rise* of the intercalation site ApA5 increases from 3.34 Å ±0.31 in the free oligonucleotide to 7.33 Å ±0.36 and 7.10 Å ± 0.39 in *eth1MG* and *eth1mg*, respectively. This is a local deformation since the *rise* parameters of the neighbouring BP steps are unaffected by the intercalation (Figure 5e) in good agreement with previous results obtained for other small intercalators.[29, 37] The end-to-end distance of the helix in the complex is also enhanced by about 4 Å with respect to the free helix. The formation of the intercalation site involves a *roll* deformation in the corresponding BP step,[71] we found that the intercalated BP step is characterized by a negative *roll* in *eth1MG* and a positive *roll* in *eth1mg* indicating the opening of the BP step in the direction of the side chains of ethidium (fig 5f). This is accompanied by a further opening of the intercalation site due to opposite values of *buckle* in the two BPs directly interacting with the intercalator (see panel a of figure 5). On the other hand the staking interaction with the ligand reduces the value of *propeller* of the same BPs. Intercalation entails significant untwisting of the intercalation site and increased *twist* fluctuations, with a more marked untwisting for the binding mode *eth1mg* (from 34.5°±4.5 of the free ApA5 to 23.2±7.6 of the intercalated ApA5 in *eth1MG* and 5.3±8.4 in *eth1mg*). Also the values of *shift* and *slide* (fig. 5c-d) show that the configuration of the double helix, especially in its central part around the intercalation site, is more disturbed when the side chains of ethidium are directed into the minor groove. The more prominent deformation of the double helix caused by the presence of the ethidium with the phenyl group into the minor groove is the origin of the higher energy

required to create the intercalation site and of the consequent decreasing of the binding energy in *eth1mg* with respect to *eth1MG*.

Table 2: Analysis of the binding free energy in the intercalated complexes. The energetics refers to the single intercalation process, i.e. the unbound state of the (1:n) complex is defined as the (1:n-1) complex, for n=1,2,3. $\Delta G_{binding}^{MMGB,(ST)}$ **is the MM-GBSA free energy of binding obtained in a single trajectory approach while** $\Delta G_{deformation}^{MMGB}$ **denotes the difference in free energy between the deformed helix and the free oligonucleotide. The sum of these two terms give the MM-GBSA binding energy. The configurational entropic contribution to the binding energy is evaluated by QHA. All the values are in kcal/mol.**

	$\Delta G_{binding}^{MMGB,(ST)}$	$\Delta G_{deformation}^{MMGB}$	$\Delta G_{binding}^{MMGB}$	$T\Delta S_{conf}$ (*)	$T\Delta S_{TOT}$	$\Delta G_{binding}$
eth1mg	-41.7	23.8	-17.9	9.8	-13.2	-4.7
eth1MG	-33.9	9.9	-24.0	8.6	-14.4	-9.6
eth2	-33.3	17.7	-15.6	21.0	-2.1	-13.5
eth3	-34.2	25.7	-8.5	19.5	-3.6	-4.9

(*) all the entropic term are evaluated by QHA over 5ns of simulation and they refer to the change in entropy due to the last intercalation event. In the evaluation, the terminal G-C base pairs have been excluded to avoid anharmonic contributions resulting from base-pair opening events.

3.3 Multiple Intercalations

According to the results discussed in the previous section, we study multiple intercalations by inserting the ligand with the side chains directed toward the major groove of the helix. The second intercalation into the position TpT7 (*eth2*, see fig 1g) produces the same effects described above at the level of the involved base pairs. The *rise* of this second intercalation site also increases to 7.2 Å and this is reflected by the end-to-end distance that amounts to 47.1 Å ±2.5 (7.9 Å longer than the free dodecamer). A significant unwinding of the central part of the helix is observed: in addition to the un-twisting of the TpT7 BP step that accommodate the ligand, the intercalation of the second ethidium induces a further untwisting of the first intercalation site (ApA5) and of the ApT6 step located between the two intercalation sites. Probably to compensate the unwinding of the central part of the helix a significant increment on the *twist* of the steps TpA3 and 9, which are the second next BP step from the intercalation sites, is observed (the plot of the parameters for the multiple intercalated complexes are available in figure SI-3). As noticed above the TpA pyrimidine-purine base pair steps in position 3 and 9 represent the most flexible point of our sequence. Even if relatively distant from the intercalation sites, their relative conformational freedom compensates the unwinding of the central part of the helix. The third molecule of ethidium is intercalated in TpA9 reducing the twist of this base pare step to 12.6°±9.7. The average end-to-end distance of the DNA-ethidium complex (1:3) amounts to 54.9Å±2.7, meaning that the intercalation of the third ethidium cation in position 9 causes a higher elongation than the previous intercalations in position 5 and 7. Since the *rise* of the TpA9 after intercalation is comparable to the rise of ApA5 and TpT7, the increased elongation is due to the important unwinding of the base pair step TpA9.

In order to compare the dynamics of the double helix in the free oligonucleotide and in the intercalated complexes we investigate the principal components of the helix dynamics in the

trajectory of the free DNA and those of the (1:1), (1:2) and (1:3) complexes. Intercalations with ethidium decisively enhance the flexibility of the backbone and of the entire double helix, as it is shown by the higher total variance registered during the trajectory of the bounded helixes (see Table SI1). The principal component analysis of the helix motion in the (1:1) complex dynamics shows that the principal modes are qualitatively the same as those of the free oligonucleotide, while the similarity is gradually lost by increasing the number of intercalating molecules (the similarity matrices between the intercalated complexes and the free DNA are reported in figure SI-4). In Figure 6a the first mode is represented for all the ethidium complexes: the motion of the backbone includes the breathing dynamics of the minor groove already observed for the free oligonucleotide. The average minor groove width (figure 6b) is increased upon intercalation for the base pairs directly interacting with the binder and also for next-neighbour base-pairs while the amplitudes of the minor groove width fluctuations are not affected by the first two intercalation at the BP steps 5 and 7. As it is suggested by the end-to-end distance of the complex, the third intercalation in the more flexible TpA9 step involves structural and dynamical changes that are more pronounced than for the two intercalations at ApA5 and TpT7. Upon the third intercalation the amplitude of the minor groove width fluctuation around the base-pairs 7 and 8 increases significantly, suggesting that all the bottom half of the double helix (i.e. BPs from 6 to 12) becomes more flexible. To quantify the motional freedom gained by the double helix upon multiple intercalations we calculate the atomic fluctuations along all the PCs of the backbone dynamics and we show in figure 6c the total backbone fluctuations for each base pair and for each complex. The first intercalation of ethidium in position 5 does not change significantly the flexibility of the double helix. In the complex 1:2 we observe that the backbone around the central base-pairs, especially AT6 and TA7 that are in between the two intercalators, becomes more floppy. The third intercalation between TA9 and AT10 considerably changes the dynamical freedom of the entire bottom half of the double helix. This result clearly shows that, even if the structural modifications due to intercalation have a local character,[29, 37] the influence on the dynamical properties of the double helix propagates through several base pairs and the dynamics of a consistent part of the system is perturbed. The increased motional freedom of the helix correlates with an increment of its configurational entropy. In figure 6d we report the configurational entropy of the double helix in the three ethidium complexes as a function of the simulation time and we compare it with the one of the free oligonucleotide. The results obtained for the (1:2) and (1:3) complexes have to be taken with caution since the first principal component is clearly anharmonic (see fig 4 in SI) and thus inconsistent with the quasi-harmonic assumption of eq.(3). Nonetheless, the calculated entropies indicate a clear qualitative trend: a positive configurational entropy variation of the double helix accompanies each intercalation event, this variation is of the order of 5kcal/mol for the first two intercalations at ApA5 and TpT7. The entropy variation of the double helix following the third intercalation is up to four times larger, which is consistent with an overall higher flexibility of a large part of the helix. The long range dynamical modifications can play a role in the cooperativity found for the intercalation of other small intercalating molecules[37] since an enhanced flexibility of the helix may favour the formation of an extra intercalation site. In order to study the energetic of multiple intercalations we applied the MM-GBSA method to the ethidium complexes. We focus on the energetic of the single intercalation, meaning that the unbound state of the (1:n) complex is defined as the (1:n-1) complex, for n=1,2,3. The results are summarized in Table 2. By neglecting the deformation energy, i.e. in the STP approach, the MM-GBSA binding energy is practically the same for all the intercalation events and it is around -34 kcal/mol. The evaluation of the deformation energy follows eq.(10): to define the

deformation energy involved in the second intercalation, the energy difference in the rhs of eq. (10) refers to the energy of the complex (1:1) extracted as receptor from the trajectory of the complex (1:2) that has to be compared with the energy of the complex (1:1) in its own trajectory. The same scheme is applied for the third intercalator. The MM-GBSA energy required to create the intercalation site increases monotonically with the number of intercalators already present in the helix. We find 9.9 kcal/mol for the first intercalation site, 17.7 kcal/mol for the second intercalation site and 25.7 kcal/mol for the third intercalation site. Accordingly, the MM-GBSA free energy of binding without the inclusion of the solute entropy decreases for subsequent intercalation events with no sign of cooperative effect. The entropic contribution coming from the solute entropy has been estimated by QHA by subtracting the entropy of the (1:n-1) complex and the ligand from the entropy of the (1:n) complex for n=1,2,3 and the result are reported in Table2. The entropy variation due to the loss of global translational and rotational degrees of freedom upon binding is constant for all the intercalations and amounts to about -23kcal/mol. Each intercalation entails a gain in configurational entropy. For the first intercalation it amounts to $T\Delta S_{conf} = 8.6$ kcal/mol and this value raises to $T\Delta S_{conf} \simeq 20$ kcal/mol for the second and the third intercalations. The balance between the MM-GBSA free energy of binding and the entropy variation due to the change in the solute dynamics gives a binding free energy, eq. (2), that depends on the number of molecules already intercalated in the helix. We calculate $\Delta G_{binding} = -9.6$ kcal/mol for the first intercalation at the ApA5 step, $\Delta G_{binding} = -13.5$ kcal/mol for the second intercalation at TpT7 and $\Delta G_{binding} = -4.9$ kcal/mol for the third intercalation at TpA9. We thus find a cooperative effect for the second binding, which results to be more favourable than the first one as already found for other intercalators.[37] Cooperativity arises mainly from the gain in conformational freedom and the consequent gain in configurational entropy of the (1:2) complex with respect to the (1:1) complex. Despite the gain in configurational entropy, the third intercalation is disfavoured because of the higher energetic cost required for the creation of the intercalation site.

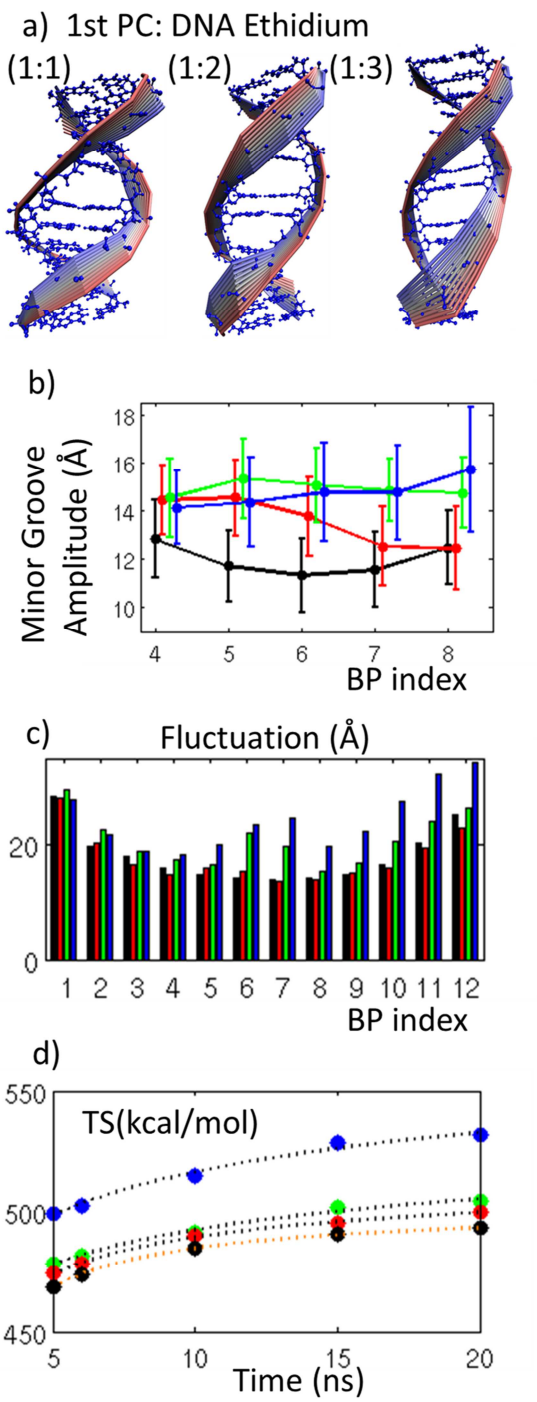


Figure 6: a) Visualization of the atomic motion obtained from eq. (9) involved in the 1st principal component of the all-atom PCA in the intercalated complexes. b) average value and fluctuation of minor groove width calculated during the 20ns simulation in the central part of the double helix for the free oligomer (black), eth1mg (red), eth2 (green) and eth3 (blue). c) Total root mean square atomic fluctuations grouped by base pairs. d) QHA entropy of the double helix in the free oligonucleotide and in the intercalated complexes as a function of the simulation length and fitting with eq.(4).

4. Conclusions

We analysed the structural and dynamical modifications occurring in the double helix of a DNA dodecamer upon binding with small ligands. As prototype of minor groove binder, we consider the Hoechst molecule, while we used the ethidium cation to study intercalation. Changes in the dynamics of the double helix entail an entropic contribution in the thermodynamics of binding that cannot be ignored. In particular, we find that binding with Hoechst causes the minor groove to be locked in a configuration that is narrower than in the free oligonucleotide and the suppression of the groove

breathing motion involved in the molecular recognition of the drug molecule. The consequent negative variation in the configurational entropy of the double helix contributes 20 kcal/mol in the binding free energy of the complex and it has to be taken into account in order to obtain binding energies comparable with the experimental values. On the contrary, intercalation enhances the flexibility of the helical scaffold. We studied two binding configurations of the ethidium cation, finding that the orientation of the ligand with the side chains directed toward the major groove requires fewer changes in the helical structure and it is thus slightly favoured over the configuration in which the side-chains are into the minor groove. The energetics of subsequent intercalations depends on the number of previously intercalated molecules. Structural deformations upon intercalation are localized but dynamical changes propagate through the base-pairs and affect the flexibility of the entire double helix. Due to the balance between the energy required to create the intercalation site and the gain in configurational entropy of the solute upon intercalation, we find a cooperative effect that favours the intercalation of the second ethidium molecule but disfavours the intercalation of a third molecule. The study of multiple intercalated complexes considered here points out that the thermodynamics of multiple bindings is a complex problem. Further investigations are required to elucidate a number of interesting issues as, for example, whether the dynamical changes are site-dependent, whether the favoured binding mode changes depending on the binding mode of the previously intercalated molecules, possible site specificity arising after the first intercalation and the role of the total length of the oligonucleotide. Overall, our results highlight the important influence of dynamical changes in the DNA double-helix on the thermodynamics of molecular recognition process, for both minor groove binding agents and intercalating drugs. Principal Components Analysis reveals that the main changes in the dynamics of the double helix upon binding occur in the slower, collective motions of the DNA double helix. Since these global modes characterize the DNA double helix dynamics rather independently of its specific sequence, we are confident that the main results we reported are also robust with respect to changes in the sequence of the dodecamer.

Acknowledgments: We gratefully thank the NANOFORCE ARC project of the University of Liege. BF acknowledges a post-doctoral fellowship from the University of Liège. FR acknowledges support from the Fonds National de la Recherche Scientifique (Belgium).

Electronic Supplementary Information (ESI) available: preliminary studies on the binding site of Hoechst, analysis of anharmonic principal component due to base pair opening in the trajectory of the DNA-Hoechst complex, table of variances and overlap of essential subspaces obtained by PCA, table of the MM-GBSA energy components for all the considered systems, comparison of base pair and base pair step parameters in DNA-ethidium complexes, further details on the PCA of DNA-ethidium complexes. Movies of the 1st and 2nd PCs of the free oligonucleotide, DNA-Hoechst complex and DNA-ethidium complexes. See DOI: 10.1039/b000000x/

References

1. R. L. M. van Montfort and P. Workman, *Trends in Biotechnology*, 2009, **27**, 315-328.
2. R. Galindo-Murillo, L. Ruiz-Azuara, R. Moreno-Esparza and F. Cortes-Guzman, *Physical Chemistry Chemical Physics*, 2012, **14**, 15539-15546.
3. S. N. Syed, H. Schulze, D. Macdonald, J. Crain, A. R. Mount and T. T. Bachmann, *Journal of the American Chemical Society*, 2013, **135**, 5399-5407.
4. A. A. Greschner, K. E. Bujold and H. F. Sleiman, *Journal of the American Chemical Society*, 2013, **135**, 11283-11288.
5. G. Brancolini, A. Migliore, S. Corni, M. Fuentes-Cabrera, F. J. Luque and R. Di Felice, *ACS Nano*, 2013, **7**, 9396-9406.
6. L. Zidek, M. V. Novotny and M. J. Stone, *Nat Struct Mol Biol*, 1999, **6**, 1118-1121.
7. C. E. Bostock-Smith, S. A. Harris, C. A. Laughton and M. S. Searle, *Nucleic Acids Research*, 2001, **29**, 693-702.
8. S. K. Sinha and S. Bandyopadhyay, *The Journal of Chemical Physics*, 2011, **135**, -.
9. A. Mukherjee, *The Journal of Physical Chemistry Letters*, 2011, **2**, 3021-3026.
10. M. Kolář, T. s. Kubař and P. Hobza, *The Journal of Physical Chemistry B*, 2010, **114**, 13446-13454.
11. J. Dolenc, R. Baron, C. Oostenbrink, J. Koller and W. F. van Gunsteren, *Biophysical Journal*, 2006, **91**, 1460-1470.
12. S. A. Harris and C. A. Laughton, *Journal of Physics: Condensed Matter*, 2007, **19**, 076103.
13. R. Rohs, I. Bloch, H. Sklenar and Z. Shakked, *Nucleic Acids Research*, 2005, **33**, 7048-7057.
14. G. Schneider, *Nat Rev Drug Discov*, 2010, **9**, 273-276.
15. H. Ma, C. Wan, A. Wu and A. H. Zewail, *Proceedings of the National Academy of Sciences*, 2007, **104**, 712-716.
16. A. Pérez, F. Lankas, F. J. Luque and M. Orozco, *Nucleic Acids Research*, 2008, **36**, 2379-2394.
17. A. Pérez, F. J. Luque and M. Orozco, *Journal of the American Chemical Society*, 2007, **129**, 14739-14745.
18. M. F. Hagan, A. R. Dinner, D. Chandler and A. K. Chakraborty, *Proceedings of the National Academy of Sciences*, 2003, **100**, 13922-13927.
19. A. Pérez, J. R. Blas, M. Rueda, J. M. López-Bes, X. de la Cruz and M. Orozco, *Journal of Chemical Theory and Computation*, 2005, **1**, 790-800.
20. A. Matsumoto and N. Go, *The Journal of Chemical Physics*, 1999, **110**, 11070-11075.
21. W. K. Olson, A. A. Gorin, X.-J. Lu, L. M. Hock and V. B. Zhurkin, *Proceedings of the National Academy of Sciences*, 1998, **95**, 11163-11168.
22. J. R. Bothe, K. Lowenhaupt and H. M. Al-Hashimi, *Journal of the American Chemical Society*, 2011, **133**, 2016-2018.
23. F. Lankaš, J. Šponer, J. Langowski and T. E. Cheatham Iii, *Biophysical Journal*, 2003, **85**, 2872-2883.
24. S. A. Harris, E. Gavathiotis, M. S. Searle, M. Orozco and C. A. Laughton, *Journal of the American Chemical Society*, 2001, **123**, 12658-12663.
25. J. B. Chaires, *Biopolymers*, 1997, **44**, 201-215.
26. W. Treesuwan, K. Wittayanarakul, N. G. Anthony, G. Huchet, H. Alniss, S. Hannongbua, A. I. Khalaf, C. J. Suckling, J. A. Parkinson and S. P. Mackay, *Physical Chemistry Chemical Physics*, 2009, **11**, 10682-10693.
27. V. V. Kostjukov, A. A. H. Santiago, F. R. Rodriguez, S. R. Castilla, J. A. Parkinson and M. P. Evstigneev, *Physical Chemistry Chemical Physics*, 2012, **14**, 5588-5600.
28. A. Rescifina, C. Zagni, M. G. Varrica, V. Pistarà and A. Corsaro, *European Journal of Medicinal Chemistry*, 2014, **74**, 95-115.
29. T. Kubař, M. Hanus, F. Ryjáček and P. Hobza, *Chemistry – A European Journal*, 2006, **12**, 280-290.

30. L. Strekowski and B. Wilson, *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 2007, **623**, 3-13.
31. I. Haq, *Archives of Biochemistry and Biophysics*, 2002, **403**, 1-15.
32. I. Haq, J. E. Ladbury, B. Z. Chowdhry, T. C. Jenkins and J. B. Chaires, *Journal of Molecular Biology*, 1997, **271**, 244-257.
33. A. L. Drobyshev, A. S. Zasedatelev, G. M. Yershov and A. D. Mirzabekov, *Nucleic Acids Research*, 1999, **27**, 4100-4105.
34. S. Hayward and B. Groot, in *Molecular Modeling of Proteins*, ed. A. Kukol, Humana Press, 2008, vol. 443, pp. 89-106.
35. A. Amadei, A. B. M. Linssen and H. J. C. Berendsen, *Proteins: Structure, Function, and Bioinformatics*, 1993, **17**, 412-425.
36. G. L. Olsen, E. A. Louie, G. P. Drobny and S. T. Sigurdsson, *Nucleic Acids Research*, 2003, **31**, 5084-5089.
37. M. Trieb, C. Rauch, F. R. Wibowo, B. Wellenzohn and K. R. Liedl, *Nucleic Acids Research*, 2004, **32**, 4696-4703.
38. P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Accounts of Chemical Research*, 2000, **33**, 889-897.
39. T. Konakahara, H. Komatsu, N. Sakai and B. Gold, *Journal of Computer Chemistry, Japan*, 2007, **6**, 27-32.
40. S. C. Jain, C.-c. Tsai and H. M. Sobell, *Journal of Molecular Biology*, 1977, **114**, 317-331.
41. R. R. Monaco, *Journal of Nucleic Acids*, 2010, **2010**, 4.
42. K. M. Langner, P. Kedzierski, W. A. Sokalski and J. Leszczynski, *The Journal of Physical Chemistry B*, 2006, **110**, 9720-9727.
43. K. M. Langner, T. Janowski, R. W. Góra, P. Dziekoński, W. A. Sokalski and P. Pulay, *Journal of Chemical Theory and Computation*, 2011, **7**, 2600-2609.
44. J. Cerny and P. Hobza, *Physical Chemistry Chemical Physics*, 2007, **9**, 5291-5303.
45. *D.A. Case; T.A. Darden; T.E. Cheatham, III; C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C. Walker, W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, and P.A. Kollman (2008), AMBER 10, University of California, San Francisco.*
46. A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham Iii, C. A. Laughton and M. Orozco, *Biophysical Journal*, 2007, **92**, 3817-3829.
47. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926-935.
48. K. E. Furse, B. A. Lindquist and S. A. Corcelli, *The Journal of Physical Chemistry B*, 2008, **112**, 3231-3239.
49. T. Darden, D. York and L. Pedersen, *The Journal of Chemical Physics*, 1993, **98**, 10089-10092.
50. S. Miyamoto and P. A. Kollman, *Journal of Computational Chemistry*, 1992, **13**, 952-962.
51. V. Tsui and D. A. Case, *Journal of the American Chemical Society*, 2000, **122**, 2489-2498.
52. B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *Journal of Chemical Theory and Computation*, 2012, **8**, 3314-3321.
53. B. Jayaram, D. Sprous and D. L. Beveridge, *The Journal of Physical Chemistry B*, 1998, **102**, 9571-9576.
54. J. Carlsson and J. Aqvist, *Physical Chemistry Chemical Physics*, 2006, **8**, 5385-5395.
55. S. A. Harris, Z. A. Sands and C. A. Laughton, *Biophysical Journal*, 2005, **88**, 1684-1691.
56. I. Andricioaei and M. Karplus, *The Journal of Chemical Physics*, 2001, **115**, 6289-6292.
57. A. L. Tournier and J. C. Smith, *Physical Review Letters*, 2003, **91**, 208106.
58. G. G. Maisuradze, A. Liwo and H. A. Scheraga, *Journal of Molecular Biology*, 2009, **385**, 312-329.

59. A. Kitao, S. Hayward and N. Go, *Proteins: Structure, Function, and Bioinformatics*, 1998, **33**, 496-517.
60. A. Amadei, M. A. Ceruso and A. Di Nola, *Proteins: Structure, Function, and Bioinformatics*, 1999, **36**, 419-424.
61. T. Meyer, C. Ferrer-Costa, A. Pérez, M. Rueda, A. Bidon-Chanal, F. J. Luque, C. A. Laughton and M. Orozco, *Journal of Chemical Theory and Computation*, 2006, **2**, 251-258.
62. R. E. B. Dickerson, M.; Calladine, C. R.; Diekmann, S.; Hunter, W. N.; O. Kennard; von Kitzing, E.; Lavery, R.; Nelson, H. C. M.; Olson, W. K.; Saenger, W.; Shakked, Z.; Sklenar, H.; Soumpasis, D. M.; Tung, C. S.; Wang, A. H. J.; Zhurkin, *The EMBO journal*, 1989, **8**, 1-4.
63. X. J. Lu and W. K. Olson, *Nucleic Acids Research*, 2003, **31**, 5108-5121.
64. D. L. Beveridge, S. B. Dixit, G. Barreiro and K. M. Thayer, *Biopolymers*, 2004, **73**, 380-403.
65. H. Wang and C. A. Laughton, *Methods*, 2007, **42**, 196-203.
66. J. M. J. Swanson, R. H. Henchman and J. A. McCammon, *Biophysical Journal*, 2004, **86**, 67-74.
67. C. Laughton and B. Luisi, *Journal of Molecular Biology*, 1999, **288**, 953-963.
68. A. Matsumoto and W. K. Olson, *Biophysical Journal*, 2002, **83**, 22-41.
69. M. A. Balsera, W. Wriggers, Y. Oono and K. Schulten, *The Journal of Physical Chemistry*, 1996, **100**, 2567-2572.
70. O. F. Lange and H. Grubmüller, *The Journal of Physical Chemistry B*, 2006, **110**, 22842-22852.
71. A. Mukherjee, R. Lavery, B. Bagchi and J. T. Hynes, *Journal of the American Chemical Society*, 2008, **130**, 9747-9755.
72. M. Wilhelm, A. Mukherjee, B. Bouvier, K. Zakrzewska, J. T. Hynes and R. Lavery, *Journal of the American Chemical Society*, 2012, **134**, 8588-8596.
73. W. D. Sasikala and A. Mukherjee, *Physical Chemistry Chemical Physics*, 2013, **15**, 6446-6455.
74. W. D. Sasikala and A. Mukherjee, *The Journal of Physical Chemistry B*, 2012, **116**, 12208-12212.
75. N. C. Garbett, N. B. Hammond and D. E. Graves, *Biophysical Journal*, 2004, **87**, 3974-3981.