An Atomistic View of DNA Dynamics and Its Interaction with Small Binders: Insights from Molecular Dynamics and Principal Component Analysis

Barbara Fresch and Francoise Remacle

Abstract DNA oligomers are promising building blocks for the development of bottom-up nano-devices and molecular logic machines. To control and exploit their unique capabilities of self-assembling and molecular recognition, a deep understanding of their dynamical properties is essential. We theoretically investigate the dynamics of a DNA dodecamer and its complexes with two common ligands, Hoechst 33258 and the ethidium cation, by means of classical molecular dynamics (MD) simulations and principal component analysis (PCA). We study the structural relation between the flexibility of the double helix and the binding process. The dynamics of a terminal base pair unbinding is also analysed as an example of process that involves multiple energy minima in the underlying free energy landscape.

1 Introduction

Biomolecules are often taken as prototypes of efficient molecular machines as they are able to perform highly complex and specialized functions. They are flexible structures, and their conformational changes are the basis of molecular recognition, allosteric regulation, self-assembly and other capabilities enabling their biological functions. Synthetic molecular machines are commonly built by assembling different molecular units to mimic simple motions typical of macroscopic objects, i.e. the directional mechanical motion of one component relative to another [1–3]. Well-known examples of this approach are rotaxane-based molecular machines [4, 5] in which a "wheel" is moving along an "axis". Even if this perspective allowed invaluable progresses in the understanding and control of motion at the molecular level, the dynamical properties enabling the functioning of many biological molecular machines are much more complex than a simple directional

B. Fresch (✉) · F. Remacle
Department of Chemistry, B6c, University of Liege, 4000 Liege, Belgium

C. Joachim and G. Rapenne (eds.), *Single Molecular Machines and Motors*,
Advances in Atom and Single Molecule Machines, DOI 10.1007/978-3-319-13872-5_2

motion of a component. Functional internal motions may be subtle and usually involve complex correlations between atomic motions.

DNA oligomers in solution, for example, display a rich dynamical behaviour. Despite timescale limitations, classical molecular dynamics (MD) simulations unveil mechanistic details of this motion with an atomic resolution. MD provides information on oligonucleotide flexibility that are difficult, if not impossible, to observe experimentally. Nonetheless, the information contained in a MD trajectory is encoded in a collection of $3N * n$ coordinates if N is the number of atom in the system under study and n the number of configurations recorded during the simulation time. The understanding and the interpretation of such a huge amount of data call for statistical tools that allow extracting most relevant motions and decrease the dimensionality of the problem. Principal component analysis (PCA) [6], also called "essential dynamics" when applied to a molecular trajectory [7], is a statistical analysis technique that allows handling this issue. PCA is based on the definition of a new set of collective coordinates (the principal components (PCs) or "modes") that allow the identification of large-scale concerted motions. The principal components are obtained as the eigenvectors of the atomic displacement covariance matrix. They define directions in the $3N$-dimensional space of atomic coordinates, and the corresponding eigenvalues indicate the extent of the total motion occurring in each direction. It turns out that for many biomolecules, as proteins and nucleic acids, as few as 10–20 modes are sufficient to account for 90 % of the total motion observed during the dynamics. Therefore, PCA allows the reduction of the dimensionality of the dynamics (in relation to data compression applications) and, more importantly for our purposes, it reveals concerted atomic motions that can be related to molecular functions, thus serving as a bridge to connect structural and functional properties of biomolecules.

In this contribution, we analyse the dynamics of a short DNA oligomer (12 base pairs) by all-atom MD simulations and principal components analysis. DNA is well suited to serve as a model to explore relations between structure, dynamics and functions in biologically relevant molecules. On the one side, nucleic acids show exceptional capabilities in molecular recognition and self-assembly. To carry on its biological functions, DNA is transiently deformed, cut and resealed, damaged and repaired, and the strands pulled apart and then re-annealed. Still, a DNA duplex in the canonical B-helix structure is characterized by an overall simple dynamics [8], which on a coarse-grained scale resembles the dynamics of an elastic rod [9, 10]. Nonetheless, from a closer perspective, sequence specific "dynamical motifs" play a significant role in molecular recognition and add complexity to the overall picture [11, 12]. We present a computational study of two different aspects of the dynamics of the DNA oligonucleotide. Firstly, we consider the dynamics involved in the interaction between the DNA dodecamer and two binders: Hoechst 33258 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. [13–15] and references therein). We chose them as prototypes of 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, while ethidium cation intercalates

between base pairs (bps) without strong sequence specificity. We demonstrate that the binding of Hoechst modifies the dynamics of the helix by "locking" the minor groove in a conformation that maximizes the interaction with the ligand. Therefore, the "groove breathing" motion observed in the free DNA oligonucleotide is suppressed upon the binding of Hoechst. The intercalation of ethidium cations induces stretching and untwisting of the double helix greatly enhancing its flexibility [16].

As a second topic, we consider the dynamics of base pair opening in a terminal position of the DNA duplex. Breaking and re-annealing of terminal base pairs are routinely observed in nanosecond MD simulations of short oligonucleotides at room temperature. The opening of a terminal base pair is a conformational transition that brings the dynamics of the system out of the harmonic regime. We will analyse how the underlying energy landscape is related to the principal components extracted from the trajectory.

We first introduce in Sect. 2 the main concepts of PCA applied to a MD trajectory. We then analyse the structure (Sect. 3) and the dynamics (Sect. 4) of the DNA–ligand complexes. Section 5 is devoted to the study of terminal base pair opening. We give the computational details of the MD simulations and of the post-processing of the obtained trajectories in Sect. 6, and we conclude in Sect. 7 summarizing the main points and giving some perspectives on future works.

2 Principal Component Analysis of an MD Trajectory

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.

A collective coordinate set can be determined as a set of eigenvectors that diagonalizes a second moment matrix. The dynamics of a molecule composed of N atoms is specified by $3N$ Cartesian coordinates $\mathbf{x}(t)$. From a MD trajectory, after the elimination of the overall translational and rotational motion, the variance–covariance matrix of positional fluctuations can be constructed:

$$\mathbf{C} = \left\langle (\mathbf{x}(t) - \langle \mathbf{x} \rangle)(\mathbf{x}(t) - \langle \mathbf{x} \rangle)^T \right\rangle \tag{1}$$

where T indicates the transposed vector and $\langle \ldots \rangle$ denotes the average on the sampled configurations. In practice, an ensemble of $3N$-dimensional coordinate vectors $\mathbf{x}^i$, for $i = 1, \ldots n$, 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 pth and the qth entry of the coordinate vector $\mathbf{x}$ on the ensemble of n realizations, namely

$$\mathbf{C}_{pq} = \frac{1}{n}\sum_{i=1}^{n}\left(x_p^i - \langle x_p\rangle\right)\left(x_q^i - \langle x_q\rangle\right) \tag{2}$$

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 $3N \times 3N$.

A set of eigenvalues and eigenvectors is obtained by solving the standard eigenvalue problem

$$\mathbf{CT} = \Lambda\mathbf{T} \tag{3}$$

with Λ being the diagonal matrix of the eigenvalues and T the transformation matrix whose columns are the eigenvectors of $\mathbf{C}$. The ith eigenvector represents the axis of the ith collective coordinate in the conformational space, while the associate eigenvalue λ_i gives the mean square fluctuation along that axis. Six of the $3N$ eigenvalues are practically zero since the overall translational and rotational motion are removed from the trajectory before the analysis. To obtain $3N$-6 physically meaningful eigenvalues, it is formally necessary to include $3N$-6 configurations in the PCA to ensure completeness of the data set. The motions along these new coordinates are often termed "modes", although they are not necessarily vibrational. In normal mode analysis (NMA), the collective coordinates are defined by the transformation that diagonalizes the Hessian matrix and each (harmonic) mode is associated to its frequency. In principle, one can use the mass-weighted displacement of all the atoms to define the covariance matrix of Eq. (1). This form of PCA is often referred as "quasi-harmonic analysis" since it is equivalent to NMA if the analysed motion is due to fluctuations within a truly harmonic conformational energy surface [17]. However, PCA analysis is more general and can be performed by considering only the coordinates of a chosen subset of atoms. In NMA, higher is the displacement represented by a normal mode, lower is its frequency. Since PCA does not require any assumption on the harmonicity of the motion, the principal components are sorted according to variance rather than frequency. Nonetheless, the largest amplitude components of PCA usually represent the slowest concerted motion represented in the analysed trajectory.

If $\boldsymbol{\eta}_i$ represents the ith eigenvector, or PC, of $\mathbf{C}$, we can define the projection of the original MD trajectory onto the new direction defined by this principal component

$$p_i(t) = \boldsymbol{\eta}_i \cdot (\mathbf{x}(t) - \langle\mathbf{x}\rangle) \tag{4}$$

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 [18, 19], as we will further illustrate in the following. By binning the values of the time-dependent projection $p_i(t)$, one obtains the corresponding probability distribution $P(p_i)$. Notice that the variance of the distribution of each

Fig. 1 Schematic representation of the free energy surface for **a** harmonic, **b** singly hierarchical and **c** multiple hierarchical modes, respectively

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") [20], see Fig. 1a, b. On the other hand, in a sufficiently converged trajectory [21], diffusive dynamics of a principal component leading to bi-modal probability distribution along the trajectory indicates a transition between multiple local energy minima ("multiple hierarchical mode"), see Fig. 1c.

As it is often the case in MD, sampling issues have an important impact on the meaning and on the relevance of the results obtained from the PCA. 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

$$S_{ij} = \boldsymbol{\eta}_i \cdot \boldsymbol{\upsilon}_j \tag{5}$$

Subnanosecond 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 converge remarkably rapidly [22]. The degree of overlap between essential subspaces is often measured as the root mean square inner product (RMSIP) of the two sets of eigenvectors

$$\mathrm{RMSIP}_M = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{M}\left(\boldsymbol{\eta}_i \cdot \boldsymbol{\upsilon}_j\right)^2} \tag{6}$$

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

$$\mathbf{x}_i(t) = p_i(t) \cdot \boldsymbol{\eta}_i + \langle \mathbf{x} \rangle \tag{7}$$

We will use the Cartesian representation of the modes given by Eq. (7) to illustrate the atomic displacement involved in the principal components.

3 Two Case Studies: DNA–Hoechst and DNA–Ethidium Complexes

We perform 15 ns equilibrium simulations of the free DNA dodecamer (5′-GGTA AATTTAGG—3′) in the standard B-helix form and of the complexes DNA–Hoechst 33258 (1:1) and DNA–ethidium cation (1:3), see Fig. 2a. For the details about the simulation protocol, see computational details in Sect. 6.

Hoechst is a fluorescent bis-benzimidazole molecule used in several applications in pharmacology. It interacts with DNA by entering the minor groove with a clear selectivity for A–T-rich tract of the double helix [8, 23]. In Fig. 2b, we show the structure of the complex used in the simulations. An H-bond analysis of the obtained trajectory 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, and a weaker hydrogen bond is also established with the nitrogen of the adenine base A6, see Fig. 2a for the numbering convention. The ethidium cation is a simple polycyclic aromatic molecule with short side chains only and represents a typical DNA intercalator without strong sequence specificity. An intercalator binds to the DNA double helix via the non-covalent stacking interaction with the DNA base pairs. Here, we study the DNA–ethidium cation (1:3) complex whose structure is shown in Fig. 2b. Deformations of the DNA double-helix structure are usually described in terms of base pair step helicoidal parameters (*roll*, *tilt*, *twist*, *shift*, *slide* and *rise*) [24, 25]. In Fig. 2c, we report the comparison of the more significant parameters between the free DNA duplex (blue bars), the Hoechst complex (red bars) and the ethidium complex (grey bars), characterizing the different base pair steps (the first bp step is defined by the G1-C24 and the G2-C23 base pairs, the second is between G2-C23 and T3-A22 and so on). The height of the bars in Fig. 2c denotes the average value of the roll, rise and twist parameters during the trajectory, while the error bars denote their standard deviation. Notice that the local flexibility of DNA helix is highly sequence dependent; pyrimidine–purine base pair steps, for example exhibit the largest flexibility and extreme local deformations in B-form DNA [26]. Accordingly, we find that bp steps 3 and 9 (TA) are characterized by a significant value of *roll*, a local helix deformation towards the major groove together with a lower value of the helical twist. It is well known that the binding of Hoechst to the DNA double helix

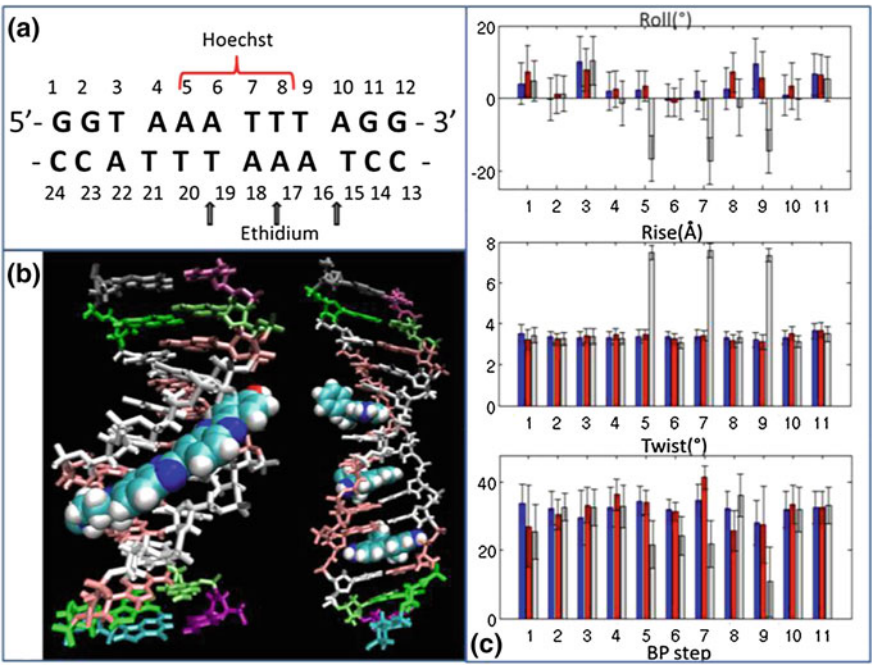

Fig. 2 Structural aspects of the DNA dodecamer and its complexes with Hoechst and ethidium cation. **a** Sequence of 24 bases in the oligomer with relative numbering. The base pairs are identified by the number of the nucleotide in the 5′ → 3′ strand. The base pair steps are numbered as the left base pair, i.e. the bp step between the 1st bp (GC) and the 2nd bp (GC) is number 1 and so on. The *red parenthesis* indicates the bps interacting with Hoechst, while the *arrows* indicate the intercalation sites of ethidium cation. **b** Structures of the two complexes obtained as averages over the MD trajectory. **c** Comparison of structural parameters of the double helix (*Roll*, *Rise and Twist*) between the free DNA (*first bar*, *blue*), Hoechst complex (*second bar*, *red*) and ethidium complex (*third bar*, *grey*). The height of the bars is the average value of the parameter along the trajectory, while error bar represents the standard deviation

does not imply significant structural reorganizations. We observe that the average structural parameters of the double helix in the presence of the minor groove binder remain practically the same as in the free oligomer. Nevertheless, we do find some minor changes in the bp parameters (defining the relative position and orientations of the two bases) within the base pairs directly involved in the binding (5th, 6th, 7th, 8th bps). Specifically, more negative values of *buckle* and *propeller* suggest that the bases are not coplanar within the base pair in order to maximize the favourable interactions with the binder.

Ethidium intercalation strongly distorts the base pair step forming the intercalation site. In our model, the 5th, 7th and 9th bp steps (AA, TT and TA, respectively) are the intercalation sites and they show, as expected, an increased rise (i.e. vertical distance between base pair planes) and lowered twist (expressing the DNA molecule being unwound). Because the distance between the bps in the intercalation sites are

twice the unperturbed distance (about 7.5 against 3.3 Å), the end-to-end distance of the helix of the DNA–ethidium complex amounts to 54.9 Å, which corresponds to an elongation of about 13 Å with respect to the free oligonucleotide.

4 From Structure to Dynamics

In this Section, we will explore the dynamical implications of the interaction between the DNA double helix and the binders. To this aim, we first perform a PCA analysis of the free DNA dodecamer. The visual inspection of the 15 ns trajectory of the free DNA reveals that short after 7 ns of simulation, the 1st base pair G1-C24 unbinds: first, the H-bonds between the two bases are broken because of a fluctuation twisting the aromatic rings of the two bases out of the common plane. Then, the guanine unstacks from the neighbour cytosine base and flips and undergoes free rotation in the solvent before establishing some non-canonical H-bond with the oxygen atoms of the sugar–phosphate backbone. Fraying and re-annealing of terminal base pairs are often observed in MD simulations of duplexes at room temperature, and the mechanism of unbinding has been well characterized as part of the melting process of the double helix [27–29]. It is clear that structures generated by the dynamics after the terminal base pair opening belong to a region of the configurational space different from the initial energy basin. In other words, the opening event corresponds to the crossing of some barrier in the underlying energy landscape and allows the system to explore another region of the accessible configurational space. We will consider such a scenario in the next section. Here, we want to focus on the dynamical consequences of the ligand binding; therefore, we retain for the analysis only the first 6 ns of simulation in which all the bps of the double helix are preserved. Moreover, we calculate the principal components by building the covariance matrix of the coordinates of the subset of atoms in the sugar–phosphate backbone. On the one hand, restricting the number of atoms considered in the PCA has the advantage that the storage and diagonalization of the covariance matrix is less demanding. On the other hand, 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 [8, 30]. We found, however, that the main collective modes resulting from the backbone PCA are very similar to those obtained from all-atom analysis [31]. In Table 1, we report the number of principal components accounting for 90 % of the variance, the absolute value of the restricted and of the total variance during the trajectory, and the contribution of the first 2 eigenvectors resulting from the backbone PCA of the DNA dodecamer and the DNA complexes. For the free DNA, 20 PCs account for 90 % of the total variance accumulated during the trajectory. We show in Fig. 3a the projections (see Eq. (4)) of the first two modes (38.7 % of the total variance) as a function of the simulation time: they describe motions that are oscillatory about the mean structure, with no net drift that could suggest a change in the overall

Table 1 Number of principal components representing 90 % of the dynamics total variance, absolute value of the explained and total variance (in $Å^2$) and fraction of variance accounted by the first two principal components

	Number of vectors (90 %)	Explained variance/ total variance	% Explained by the first two PCs
DNA 12-mer	20	308.2/341.2	38.7
DNA–Hoechst(1:1)	20	284.1/315.0	41.0
DNA–ethidium(1:3)	14	507.4/559.8	63.4

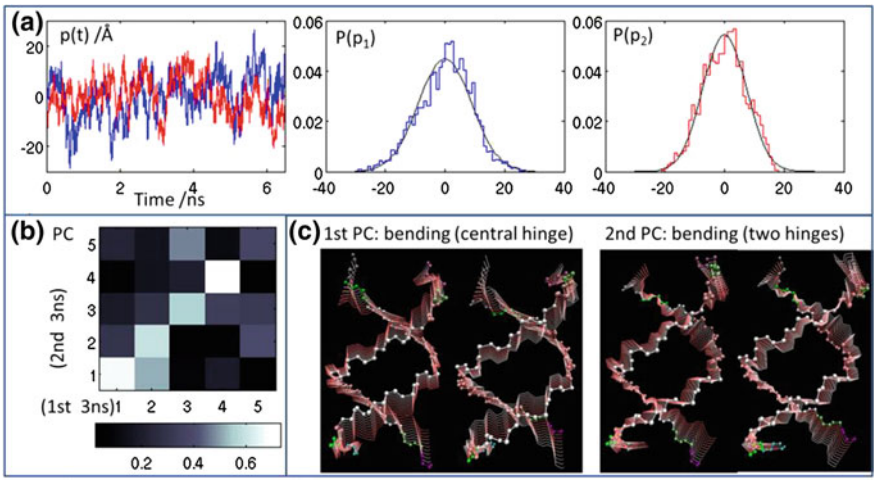

Fig. 3 Principal component analysis of the free DNA dodecamer: **a** projection of the atomic positional fluctuations along the 1st (*blue*) and the 2nd (*red*) principal components as a function of the simulation time and corresponding probability distributions. **b** Matrix of the inner products between the principal components obtained by dividing the trajectory into two halves. **c** Visualization of the atomic motion involved in the 1st and 2nd principal components

conformation of the molecule. The corresponding probability densities, also shown in the figure, are well described by Gaussian distributions expected for motions that explore quasi-harmonic energy basin (see Fig. 1). In Fig. 3b, we report the similarity between the first 5 modes obtained from the first and the second halves of the total trajectory in the form of the matrix of the inner products between them, see Eq. (5). 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. (6) amounts to 0.8. These observations establish the robustness of the PCs resulting from the total trajectory. The first two modes are shown in Fig. 3c: they both correspond to bending modes of the duplex as also results from a normal modes analysis of the DNA double helix [10, 32] and resemble the normal motions of a simple elastic rod. What is more interesting for our purpose is to notice that the first principal component entails a sort of "groove

breathing" motion in the central portion of the double helix, with the backbone of the two strands getting far apart which creates a good condition for the possible approach of a binder molecule.

We then consider the essential dynamics of the two complexes: DNA–Hoechst (1:1) and DNA–ethidium cation (1:3). As for the free DNA, we analyse a portion of the trajectory (of about 6 ns) in which no base pair opening is observed and for which the time evolution of the principal components are compatible with the motion in a quasi-harmonic energy basin. As one can see from Table 1, the interaction with Hoechst does not entail a significant change in the overall observed variance, while the structure of the DNA–ethidium complex is definitely more flexible and floppy than the free oligonucleotide, the total variance being significantly higher. In Fig. 4a, we show the total root mean squared fluctuation amplitude of the backbone, specifying the contribution coming from the different nucleotides; the upper part of the graph shows the displacement of the atoms in the nucleotides 1–12, while the lower part shows the displacement of the complementary bases 13–24 (see numbering in Fig. 2a). As a common feature of the free DNA oligonucleotide and both complexes, the terminal portions of the double helix are characterized by higher mobility, while the central part (bps from 5 to 8) is more rigid. In comparison with the free oligonucleotide, all the base pairs in the DNA–ethidium complex experience larger fluctuations during the trajectory, while the presence of the Hoechst molecule reduces the total displacement of the central base pairs but only slightly. However, the total root mean squared displacement along the trajectory does not contain information about the relative direction of the atomic motion. To analyse the changes on the backbone dynamics induced by the presence of the ligands, we look at the main principal components of the dynamics of the DNA–ligand complexes. The analysis of the matrix of the inner products between the principal components of the free DNA and of the DNA–ligand complexes showed in Fig. 4b shows that the PCs of the complexes preserve an overall correlation with the PCs of the free oligonucleotide, especially in the case of Hoechst binding. Nonetheless, by analysing the modes of the DNA–Hoechst complex (see Fig. 4c), we find that the breathing of the groove which was part of the first PC of the free oligonucleotide (Fig. 3c) is completely absent in the first five PCs of the complex. The first mode of the DNA–Hoechst complex is still a bending motion around a central hinge (which explains the correlation with the 1st PC of DNA in Fig. 4b), but the backbone of the two strands moves concurrently in the same direction (which explains the correlation with the 2st PC of DNA in Fig. 4b). In the second mode, the central part of the helix is practically blocked, and the motion is concentrated on the terminal arms. The change in the groove width is thus suppressed by the presence of the Hoechst molecule. This can be intuitively understood since the groove is "locked" in the conformation that maximizes the favourable interaction between the DNA and the binder. On the contrary, the first two principal modes of the DNA–ethidium complex are qualitatively very similar to the first two modes of the free oligonucleotide (compare Fig. 4c with Fig. 3d). Nonetheless, the correlation between the modes of the complex and the free DNA is lower in this case, due to the structural deformation of the helix upon intercalation described in the previous section.

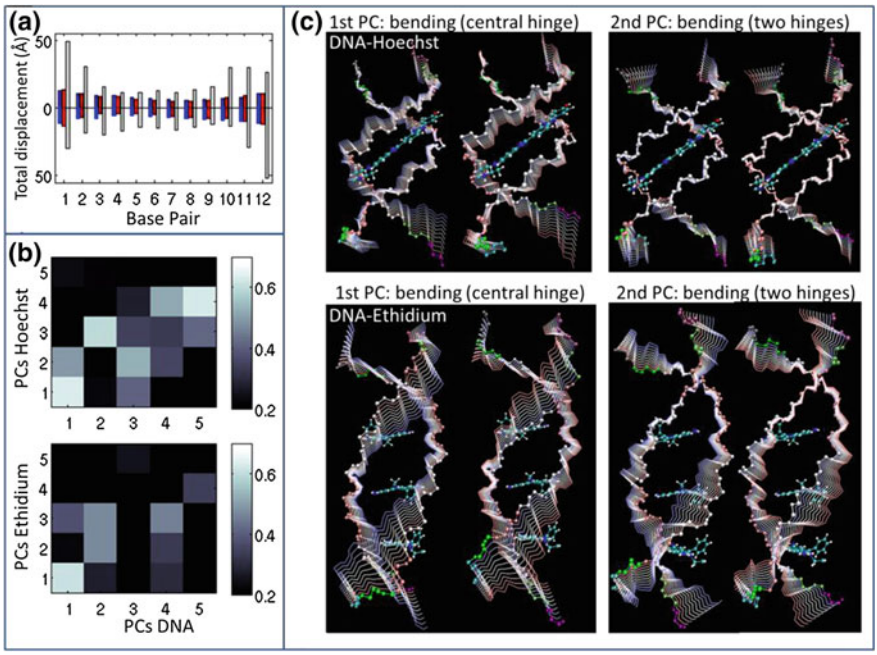

Fig. 4 Principal component analysis of the DNA complexes with Hoechst (1:1) and ethidium cation (1:3): **a** total root mean square displacement of the backbone atoms grouped by nucleotide. The upper part of the graph shows data for nucleotides 1–12 and the lower part for nucleotides 13–24, the numbering convention is given in Fig. 2a. The *bars* refer to free DNA (*blue*), Hoechst complex (*red*) and ethidium complex (*grey*). **b** Matrixes of the inner products between the first five principal components of the free oligonucleotide and of the complexes. **c** Visualization of the atomic motion involved in the 1st and 2nd modes of the DNA–Hoechst and the DNA–ethidium complexes

5 Exploring the Conformational Space: Opening of a Terminal Base Pair

The essential dynamics of the short oligonucleotide presented in the previous section shows that, on relatively short timescales, the dynamics is dominated by fluctuations within a local minimum (that can be approximated well by a system's local normal modes). However, since the energy landscape is characterized by multiple minima, on longer timescales, we can expect that the system crosses some barrier and starts to explore other regions of phase space. When this is the case, the large fluctuations are dominated by a largely anharmonic diffusion between multiple wells. To illustrate this point, we will now consider the whole 15 ns trajectory of the DNA 12-mer. As previously mentioned, an opening of the terminal GC-1 base pair is observed after 7 ns. The fraying of the terminal base pair is energetically costly because it involves the rupture of the H-bonds of the base pair and the

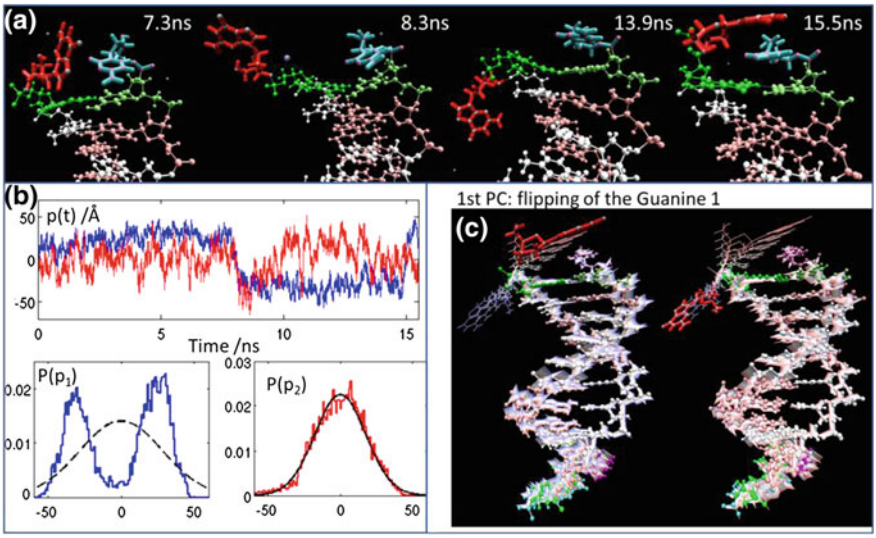

Fig. 5 Principal component analysis of the whole 15 ns trajectory of the free DNA dodecamer showing terminal base pair opening: **a** mechanistic details of the opening of the terminal base pair GC-1. **b** Projection of the atomic positional fluctuations along the 1st (*blue*) and the 2nd (*red*) principal components as a function of the simulation time and corresponding probability distributions. **c** Visualization of the atomic motion involved in the 1st principal component

un-staking of at least one base from the next one. Nonetheless, the increased motional freedom that follows the fraying makes the peeling process entropically driven. The stepwise ("zipper-like") process starting with the fraying/peeling of one of the two double-helix ends and propagating along the dodecamer has been identified as the main unfolding route in the thermal melting of DNA oligonucleotides [29, 33]. In the condition of our simulation (room temperature and absence of external forces), the base pair rupture process is of course reversible and biased towards the re-annealing. The mechanism of the opening we observed is illustrated in Fig. 5a by snapshots of the dynamics taken at different times. After the rupture of the canonical H-bonds, due to a fluctuation which rotates the two bases out-of-plane (snapshot at 7.3 ns in Fig. 5a), the guanine G1 explores different conformations, while the cytosine C24 conserves the inter-strand staking interaction with the next base C23. At 8.3 ns, G1 is completely unfolded before establishing non-canonical H-bonds with the atoms of the backbone and thus residing for about 5 ns in this "peeled conformation" (representative snapshot at 13.9 ns). Short before the end of the simulations, the guanine leaves this kinetic intermediate and goes back on the top of the helix. However, it does not find straightly the optimal Watson–Crick pairing; instead, the structure is again trapped in a local minima characterized by the stacking between G1 and C24 (snapshot at 15.5 ns). It is clear that, because of the opening process described above, the system visits a portion of the energy landscape during the simulation time that is not well approximated by a harmonic potential. Let us see how this is reflected by the essential dynamics resulting from a

PCA analysis of the whole trajectory. To analyse the base opening motion, we do not confine the analysis to the backbone atoms, since it is clear by simple visual inspection that the reaction coordinate is mainly composed of the coordinates of the G1 base atoms. Accordingly, we perform an all-atom PCA. Figure 5b shows the projection along the first two principal components of the dynamics and the corresponding probability distributions. The first PC, which account for the 30 % of the total displacement, shows very clearly the transition between two distinct conformations suggesting the crossing of a barrier in an underlying energy profile characterized by multiple minima (as the one schematically shown in Fig. 1c). The corresponding probability distribution is bi-modal, which is a typical signature of a multiple hierarchical mode. The time dependence of the second mode also shows the signature of the transition around 8 ns, but the overall distribution is well represented by a Gaussian. The Gaussian probability distribution suggests that, along the direction identified by the 2nd PC, the system remains within a single energy basin. Moreover, by plotting the projection of the higher modes as a function of the projection of the first mode, one can verify that the first PC is statistically un-correlated from the others. This means that it represents a good reaction coordinate for the base pair opening process. The motion described by the first PC is shown in Fig. 1c and clearly corresponds to the flipping of the G1 base. Given a longer trajectory which presents multiple binding/unbinding events, thermodynamic properties can be derived as ensemble averages and one can use the principal components to obtain a map of the free energy landscape [19, 34].

6 Computational Details

MD simulations of the DNA, DNA–Hoechst 33258 complex and DNA–ethidium cation (1:3) were performed using AMBER 10.0 with the AMBER ff99bsc0 force field for nucleic acids and the TIP3P water model. The force fields for the ligand have been prepared with ANTECHAMBER and corrected according to the parameterization developed in Ref. [35] for Hoechst and in Ref. [13] for the ethidium cation. 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 counter-ions were added to establish charge neutrality. The resulting system was energy minimized using steepest decent and conjugate gradient methods to relieve any residual unfavourable steric interactions introduced during the solvation procedure. First, the systems (DNA and DNA complexes) were restrained (500 kcal/mol-Å^2), while the water and counter-ions were subjected to 10,000 cycles of minimization. Then, the full system was allowed to relax during an additional 10,000 cycles of unrestrained minimization. The DNA oligonucleotide and Hoechst 33258/ethidium cations systems were restrained (25 kcal/mol Å^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 300 K using a Langevin temperature control. Next, the system was subjected to

150 ps of constant pressure (NPT) MD to achieve proper density and 1 ns of relaxation before production runs. During MD, the long-range electrostatic interactions were treated with the particle mesh Ewald method [36] using a real-space cut-off distance of $r_{\text{cut-off}} = 9$ Å. The SHAKE algorithm [37] was used to constrain bond vibrations involving hydrogen atoms, which allowed a time step of 2 fs. Frames are recorded every 1,000 steps (2 ps).

Analysis of the trajectory to extract the helical parameter of the double helix was performed with the cpptraj module of AmberTools12. PCA analysis was carried on as implemented in the PCA suite software [38].

7 Conclusions

Conformational freedom of molecules is related to the underlying free energy landscape, which for biological molecule is high dimensional and likely characterized by many local minima accessible at room temperature. Some concerted motions are important biologically as they describe the manner in which the biomolecule is particularly flexible. PCA analysis extracts from a MD trajectory a set of collective structural changes of the molecule, i.e. principal components of the dynamics, which contribute significantly to the observed atomic displacement. It is well known that a large part of the overall motion of the system during a simulation can be described in terms of only a few principal components. These large displacement motions dominate the range of molecular configurations explored during thermal agitation, and they are often related to the capability of the molecule to perform biological relevant functions. We have illustrated these concepts by considering the dynamics of a DNA dodecamer in solution. Molecular recognition between the oligonucleotide and binders is related to a groove breathing mode that facilitates the approach of the ligand to the base pairs of the double helix. We observed that the binding of Hoechst in the central part of the dodecamer suppresses this mode by "locking" the minor groove in a configuration that maximises the interaction between the DNA and the drug. The intercalation of three ethidium cations significantly distorts the B-helix structure but does not suppress the breathing mode of the helix backbone. The principal components of the dynamics of a terminal base pair opening are also analysed. We showed that this motion cannot be described within a quasi-harmonic approximation since the system explores multiple minima in the underlying energy landscape. The first principal component, describing the flipping of one terminal base, is a good reaction coordinate for the opening process. The decomposition of MD trajectories in term of principal components is a powerful method to unveil the atomic details of complex dynamical processes involving the DNA double helix. Further work in this direction may improve our understanding of important processes, such as the mechanism of unfolding under the action of external forces as probed my modern single-molecule manipulation experiments, or the effects of specific drugs on the stability and on the elastic properties of DNA.

Acknowledgements This work was supported by the NANOFORCE ULg ARC project. BF thanks ULg for a post-doctoral fellowship. FR is a Director of research from Fonds National de la Recherche Scientifique, Belgium. We thank A.S. Duwez (ULg) for fruitful discussions about related experimental aspects.

References

1. Kay, E.R., Leigh, D.A., Zerbetto, F.: Synthetic molecular motors and mechanical machines. Angew. Chemi. Int. Ed. **46**, 72–191 (2007). doi:10.1002/anie.200504313
2. Joachim, C., Rapenne, G.: Molecule concept nanocars: chassis, wheels, and motors? ACS Nano **7**, 11–14 (2013). doi:10.1021/nn3058246
3. Klok, M., Boyle, N., Pryce, M.T., Meetsma, A., Browne, W.R., Feringa, B.L.: MHz unidirectional rotation of molecular rotary motors. J. Am. Chem. Soc. **130**, 10484–10485 (2008)
4. Berna, J., Leigh, D.A., Lubomska, M., Mendoza, S.M., Perez, E.M., Rudolf, P., Teobaldi, G., Zerbetto, F.: Macroscopic transport by synthetic molecular machines. Nat. Mater. **4**, 704–710 (2005). http://www.nature.com/nmat/journal/v4/n9/suppinfo/nmat1455_S1.html
5. Lewandowski, B., De Bo, G., Ward, J.W., Papmeyer, M., Kuschel, S., Aldegunde, M.J., Gramlich, P.M.E., Heckmann, D., Goldup, S.M., D'Souza, D.M., Fernandes, A.E., Leigh, D. A.: Sequence-specific peptide synthesis by an artificial small-molecule machine. Science **339**, 189–193 (2013)
6. Hayward, S., Groot, B.: In: Kukol, A. (ed.) Molecular Modeling of Proteins. Methods Molecular Biology™, Ch. 5, vol. 443, pp. 89–106 Humana Press (2008)
7. Amadei, A., Linssen, A.B.M., Berendsen, H.J.C.: Essential dynamics of proteins. Proteins: Struct. Funct. Bioinf. **17**, 412–425 (1993). doi:10.1002/prot.340170408
8. Bostock-Smith, C.E., Harris, S.A., Laughton, C.A., Searle, M.S.: Induced fit DNA recognition by a minor groove binding analogue of Hoechst 33258: fluctuations in DNA A tract structure investigated by NMR and molecular dynamics simulations. Nucleic Acids Res. **29**, 693–702 (2001). doi:10.1093/nar/29.3.693
9. Pérez, A., Blas, J.R., Rueda, M., López-Bes, J.M., de la Cruz, X., Orozco, M.: Exploring the essential dynamics of B-DNA. J. Chem. Theory Comput. **1**, 790–800 (2005). doi:10.1021/ct050051s
10. Matsumoto, A., Go, N.: Dynamic properties of double-stranded DNA by normal mode analysis. J. Chem. Phys. **110**, 11070–11075 (1999)
11. Harris, S.A., Gavathiotis, E., Searle, M.S., Orozco, M., Laughton, C.A.: Cooperativity in drug—DNA Recognition: a molecular dynamics study. J. Am. Chem. Soc. **123**, 12658–12663 (2001). doi:10.1021/ja016233n
12. Bothe, J.R., Lowenhaupt, K., Al-Hashimi, H.M.: Sequence-specific B-DNA flexibility modulates Z-DNA formation. J. Am. Chem. Soc. **133**, 2016–2018 (2011). doi:10.1021/ja1073068
13. Kubař, T., Hanus, M., Ryjáček, F., Hobza, P.: Binding of cationic and neutral phenanthridine intercalators to a DNA oligomer is controlled by dispersion energy: quantum chemical calculations and molecular mechanics simulations. Chem. Eur. J. **12**, 280–290 (2006). doi:10.1002/chem.200500725
14. Strekowski, L., Wilson, B.: Noncovalent interactions with DNA: an overview. Mutat. Res. Fundam. Mol. Mech. Mutagen. **623**, 3–13 (2007). 10.1016/j.mrfmmm.2007.03.008
15. Haq, I.: Thermodynamics of drug–DNA interactions. Arch. Biochem. Biophys. **403**, 1–15 (2002). doi:10.1016/S0003-9861(02)00202-3
16. Fresch, B., Remacle, F.: Atomistic account of structural and dynamical changes induced by small binders in the double helix of a short DNA. Phys. Chem. Chem. Phys. **16**, 14070–14082 (2014). doi:10.1039/C4CP01561D

17. Hayward, S., Kitao, A., Gō, N.: Harmonic and anharmonic aspects in the dynamics of BPTI: a normal mode analysis and principal component analysis. Protein Sci. **3**, 936–943 (1994). doi:10.1002/pro.5560030608
18. Tournier, A.L., Smith, J.C.: Principal components of the protein dynamical transition. Phys. Rev. Lett. **91**, 208106 (2003)
19. Maisuradze, G.G., Liwo, A., Scheraga, H.A.: Principal component analysis for protein folding dynamics. J. Mol. Biol. **385**, 312–329 (2009). doi:10.1016/j.jmb.2008.10.018
20. Kitao, A., Hayward, S., Go, N.: Energy landscape of a native protein: Jumping-among-minima model. Proteins Struct. Funct. Bioinf. **33**, 496–517 (1998). doi:10.1002/(SICI)1097-0134(19981201)33:4<496::AID-PROT4>3.0.CO;2-1
21. Hess, B.: Similarities between principal components of protein dynamics and random diffusion. Phys. Rev. E **62**, 8438–8448 (2000)
22. Amadei, A., Ceruso, M.A., Di Nola, A.: On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. Proteins Struct. Funct. Bioinf. **36**, 419–424 (1999). doi:10.1002/(SICI)1097-0134(19990901)36:4<419::AID-PROT5>3.0.CO;2-U
23. Haq, I., Ladbury, J.E., Chowdhry, B.Z., Jenkins, T.C., Chaires, J.B.: Specific binding of hoechst 33258 to the d(CGCAAATTTGCG)2 duplex: calorimetric and spectroscopic studies. J. Mol. Biol. **271**, 244–257 (1997). doi:10.1006/jmbi.1997.1170
24. Dickerson, R.E.B.M., Calladine, C.R., Diekmann, S., Hunter, W.N., Kennard, O., 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, V.B.: Definitions and nomenclature of nucleic acid structure parameters. EMBO J. **8**, 1–14 (1989)
25. Lu, X.J., Olson, W.K.: 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res. **31**, 5108–5121 (2003). doi:10.1093/nar/gkg680
26. Beveridge, D.L., Dixit, S.B., Barreiro, G., Thayer, K.M.: Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. Biopolymers **73**, 380 (2004)
27. Hagan, M.F., Dinner, A.R., Chandler, D., Chakraborty, A.K.: Atomistic understanding of kinetic pathways for single base-pair binding and unbinding in DNA. Proc. Natl. Acad. Sci. **100**, 13922–13927 (2003). doi:10.1073/pnas.2036378100
28. Kannan, S., Zacharias, M.: Simulation of DNA double-strand dissociation and formation during replica-exchange molecular dynamics simulations. Phys. Chem. Chem. Phys. **11**, 10589–10595 (2009). doi:10.1039/B910792B
29. Perez, A., Orozco, M.: Real-time atomistic description of DNA unfolding. Angew. Chem. Int. Ed. **49**, 4805–4808 (2010). doi:10.1002/anie.201000593
30. Laughton, C., Luisi, B.: The mechanics of minor groove width variation in DNA, and its implications for the accommodation of ligands. J. Mol. Biol. **288**, 953–963 (1999). doi:10.1006/jmbi.1999.2733
31. Harris, S.A., Laughton, C.A.: A simple physical description of DNA dynamics: quasi-harmonic analysis as a route to the configurational entropy. J. Phys.: Condens. Matter **19**, 076103 (2007)
32. Matsumoto, A., Olson, W.K.: Sequence-dependent motions of DNA: a normal mode analysis at the base-pair level. Biophys. J. **83**, 22–41 (2002). doi:10.1016/S0006-3495(02)75147-3
33. Wong, K.-Y., Pettitt, B.M.: The pathway of oligomeric DNA melting investigated by molecular dynamics simulations. Biophys. J. **95**, 5618–5626 (2008). doi:10.1529/biophysj.108.141010
34. Papaleo, E., Mereghetti, P., Fantucci, P., Grandori, R., De Gioia, L.: Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: the myoglobin case. J. Mol. Graph. Model. **27**, 889–899 (2009). doi:10.1016/j.jmgm.2009.01.006
35. Furse, K.E., Lindquist, B.A., Corcelli, S.A.: Solvation dynamics of hoechst 33258 in water: an equilibrium and nonequilibrium molecular dynamics study. J. Phys. Chem. B **112**, 3231–3239 (2008). doi:10.1021/jp711100f

36. Darden, T., York, D., Pedersen, L.: Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. J. Chem. Phys. **98**, 10089–10092 (1993)
37. Miyamoto, S., Kollman, P.A.: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem. **13**, 952–962 (1992). doi:10.1002/jcc.540130805
38. Meyer, T., Ferrer-Costa, C., Pérez, A., Rueda, M., Bidon-Chanal, A., Luque, F.J., Laughton, C.A., Orozco, M.: Essential dynamics: a tool for efficient trajectory compression and management. J. Chem. Theory Comput. **2**, 251–258 (2006). doi:10.1021/ct050285b

http://www.springer.com/978-3-319-13871-8

Single Molecular Machines and Motors
Proceedings of the 1st International Symposium on Single Molecular Machines and Motors, Toulouse 19-20 June 2013
Joachim, C.; Rapenne, G. (Eds.)
2015, VIII, 196 p. 122 illus., 85 illus. in color., Hardcover
ISBN: 978-3-319-13871-8