Eur. Phys. J. E (2013) **36**: 21
DOI 10.1140/epje/i2013-13021-4

THE EUROPEAN PHYSICAL JOURNAL E

Regular Article

Rigid-body molecular dynamics of DNA inside a nucleosome

Arman Fathizadeh[1], Azim Berdy Besya[2], Mohammad Reza Ejtehadi[3,a], and Helmut Schiessel[4,b]

[1] Institute for Nanoscience and Nanotechnology, Sharif University of Technology, P.O. Box 14588-89694, Tehran, Iran
[2] Engineering Department AIRIC, Km. 15, Karaj special road, P.O. Box: 13445-55, Tehran, Iran
[3] Physics Department, Sharif University of Technology, P.O. Box 11155-9161, Tehran, Iran
[4] Instituut-Lorentz for Theoretical Physics, P.O.Box 9506, 2300 RA Leiden, The Netherlands

Received 17 October 2012 and Received in final form 13 February 2013
Published online: 13 March 2013 –

Abstract. The majority of eukaryotic DNA, about three quarter, is wrapped around histone proteins forming so-called nucleosomes. To study nucleosomal DNA we introduce a coarse-grained molecular dynamics model based on sequence-dependent harmonic rigid base pair step parameters of DNA and nucleosomal binding sites. Mixed parametrization based on all-atom molecular dynamics and crystallographic data of protein-DNA structures is used for the base pair step parameters. The binding site parameters are adjusted by experimental B-factor values of the nucleosome crystal structure. The model is then used to determine the energy cost for placing a twist defect into the nucleosomal DNA which allows us to use Kramers theory to calculate nucleosome sliding caused by such defects. It is shown that the twist defect scenario together with the sequence-dependent elasticity of DNA can explain the slow time scales observed for nucleosome mobility along DNA. With this method we also show how the twist defect mechanism leads to a higher mobility of DNA in the presence of *sin* mutations near the dyad axis. Finally, by performing simulations on 5s rDNA, 601, and telomeric base pair sequences, it is demonstrated that the current model is a powerful tool to predict nucleosome positioning.

1 Introduction

The DNA of eukaryotic cells has typically macroscopic lengths but needs to fit inside micron-sized cell nuclei. To achieve this, eukaryotic DNA is hierarchically folded with the help of proteins into the chromatin complex [1]. On the first level, DNA is wrapped around millions of protein cylinders resulting in a string of so-called nucleosomes. The core of each nucleosome is a cylinder composed of an octamer of histone proteins and it is wrapped by a 147 base pairs (bp) long DNA stretch. The complexes are connected via short stretches of unbound DNA, the linker DNA. According to the nucleosome crystal structure [2], DNA is bound to the octamer at 14 binding sites where the minor groove of the DNA double helix faces the cylinder defining a left-handed superhelical wrapping path of one and three quarter turns. The higher-order structures beyond the nucleosome are still highly debated [3] and are not discussed here any further.

The tight wrapping of nucleosomal DNA around the histone octamers raises the question of how other DNA binding proteins can bind to their target base pair sequences. If such a sequence happens to be located inside a wrapped portion, steric exclusion would hinder its access. As it turns out, however, spontaneous fluctuations of the nucleosomes provide transient access. There are at least two mechanisms: nucleosome breathing or site exposure [4–13], where the DNA partially unwraps from the protein core, and nucleosome sliding [14–19] where the protein cylinder moves as a whole along the DNA. Nucleosome breathing occurs spontaneously from either end of the wrapped DNA portion as the result of the sequential opening of binding sites. Experimentalists detect nucleosome breathing either by measuring protein accessibility to nucleosomes [4–8] or by employing fluorescence resonance energy transfer between dyes strategically placed on the nucleosomal DNA and/or histones [9–12].

Nucleosome sliding results from defects that are spontaneously formed on the nucleosome. Experimental evidence [18,19] points toward small 1 bp twist defects where an extra or a missing bp is located between two neighboring binding sites [20–22]. Such a defect can be accommodated inside the nucleosome by the over- or understretching and over- or under-twisting of the corresponding DNA section. Twist defects are produced at either end of the wrapped portion. If such a defect manages to diffuse through the nucleosome and to fall off at the other end, it causes a 1 bp step of the nucleosome along the DNA.

[a] e-mail: ejtehadi@sharif.edu
[b] e-mail: schiessel@lorentz.leidenuniv.nl

The cost of a twist defect had been estimated to be about $9k_BT$ [20], where T denotes the absolute temperature and k_B the Boltzmann constant. This means that defects are rare and especially the occurrence (and interaction) of more than one defect on a nucleosome is of no practical interest. By assuming 14 identical binding sites offering 13 identical defect locations, it was concluded in ref. [20] that only in one of 14 cases a twist defect manages to cross through a nucleosome and hence to cause a sliding event.

Earlier models of the nucleosome (like in refs. [20, 23–26]) assumed the DNA molecule to be mechanically homogeneous along its contour. Newer models also account for bp-step specific mechanical properties that have a pronounced effect on the DNA bending energetics as DNA is severely bent in the complex. Morozov *et al.* [27] include in their model sequence-specific DNA mechanics but account for the histone-DNA interaction only through a very simplified potential. That potential is at a minimum when the DNA forms an ideal superhelix with a radius and pitch inferred from the crystal structure [2] and varies quadratically with the deviation from that ideal geometry. The model allows to predict the positioning of nucleosomes on various sequences. On the other hand, the model of Becker and Everaers [28] starts from nucleosome crystal structures and infers from the DNA conformation forces and torques that act on the molecule. This DNA nanomechanics approach detects not only the strong impact of the binding sites on DNA deformations but also a smaller one that stems from the histone tails. It successfully predicts the sequence specific locations of twist defects in the crystal structures. Whereas both approaches [27, 28] account for the DNA mechanics in a similar way, their modeling of the DNA-histone interaction are completely different. The former approach is completely unspecific in just building in a potential to bend the DNA into an ideal superhelix whereas the latter approach forces the DNA into the very specific conformation found in the crystal structure.

In this paper we introduce a model that attempts to come to a more physical description of the energetics of the DNA binding sites. A schematic representation of the model is depicted in fig. 1. The details of the model are explained in the following section. Our model constrains the DNA more than in ref. [27] but less than in ref. [28]. Giving a more realistic representation of the DNA-histone interaction allows us to study a wider range of phenomena observed for nucleosomes, *e.g.* thermal fluctuations inside the crystal and nucleosome sliding through twist defects. In sect. 3 we present the results of various simulations that we performed with our model. After tuning the model such that it reproduces the experimental values of the B-factors for the backbone phosphates, we measure the total energy of the system with twist defects at different positions and compare our findings with the preferred locations of twist defects in two crystallographic structures. As the next step nucleosome sliding is investigated based on twist defect diffusion. We then show how *sin* mutations near the dyad axis have a strong impact on nucleosome sliding by altering the dynamics of the twist defect diffusion. Finally, the effects of sequence-dependent elasticity of DNA on nucleosome positioning and sliding is studied for the 5S rDNA, the 601, and four telomeric sequences.

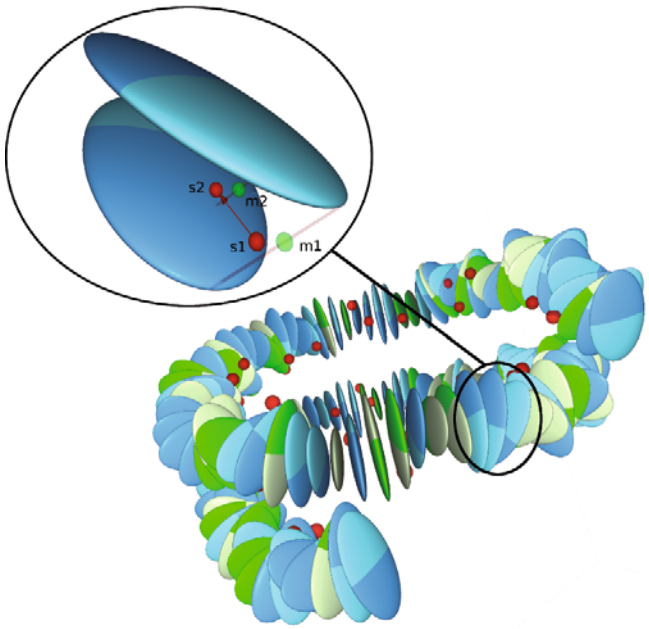

Fig. 1. (Color online) Schematic representation of the model. Main figure: Every base is shown with a different color. Blue ellipsoids represent A-T and green ellipsoids represent G-C base pairs. The example shown here corresponds to the palindromic DNA sequence of NCP147. Red spheres are s1 springs. s2 springs are not shown to avoid complexity. Inset: two adjacent base pairs and the location of binding sites. The red points represent the binding site fixed in space and the green point represents the location of the phosphate halfway between two base pairs. The direction of the red arrow that connects s1 to s2, was determined from a simulation where the s2 springs were absent (see text for details).

2 Model

2.1 Interaction potential between base pairs

We consider a system of successive rigid objects where each object represents a base pair. Each base pair interacts with its adjacent neighbors via a harmonic potential

$$U = \frac{1}{2}(\boldsymbol{\psi} - \boldsymbol{\psi}_0)^T \cdot \mathbf{K} \cdot (\boldsymbol{\psi} - \boldsymbol{\psi}_0), \qquad (1)$$

where $\boldsymbol{\psi}$ is a vector with 6 components, specifying relative orientation and separation of the adjacent base pairs. The rotational (Twist, Tilt, and Roll) and translational (Shift, Slide and Rise) parameters are defined at a reference midframe via CEHS representation [29]. $\boldsymbol{\psi}_0$ represents the equilibrium configuration of base pairs and $\mathbf{K}$ is a 6×6 stiffness matrix. Both are sequence-dependent. Having 4 different bases results in 16 different possible combinations for two successive base pairs which are reduced to 10 after considering the symmetries.

There are different methods to obtain 10 possible sets of $\boldsymbol{\psi}_0$ and $\mathbf{K}$. Olson *et al.* [30] report the parameters by analyzing the crystal structures of naked B-DNA and protein-DNA configurations. In a different approach, Lankas *et al.* [31] used all-atom molecular dynamics simulations to find parameters for the harmonic potential between base pairs (see ref. [32] for a more recent molecular dynamics parametrization). Becker *et al.* [33] combine both methods and define a mixed parametrization method. They show that by using stiffness matrices from all-atom molecular dynamics [31] and equilibrium parameters from protein-DNA crystals [30], a more precise potential for DNA modeling is achieved (MP parametrization). In the current study we use the MP parametrization for the interaction of the base pairs.

2.2 Interaction at binding sites

The interaction between the histone octamer and the DNA is only considered at the 14 groups of binding sites. To mimic the configuration at the binding sites in the crystal structure, every binding site in our model consists of 3 phosphate groups of DNA connected to the octamer. We model these bonds as harmonic springs

$$U_{\rm bind} = \frac{1}{2} k_{\rm bind} (\Delta r)^2 - E_{\rm bind}. \qquad (2)$$

There are five unknown parameters to be determined per bond: the stiffness of the spring, $k_{\rm bind}$, the depth of the binding energy, $E_{\rm bind}$, and three for the coordinates of the site. To find the locations of the binding sites we use the crystal structure of NCP147 [34]. This structure is perfect (no twist defects in the structure) and contains a palindromic sequence of DNA. We find bonded phosphate groups from the local minima of the B-factor curve of the phosphorous atoms in both DNA strands and obtain their coordinates from the pdb file as well. The B-factor of an atom is a measure of the fluctuations of its position. Phosphates bound to the protein core are expected to show less fluctuations and can thus be identified by a smaller value of the B-factor (its precise definition is given later below). These points are shown schematically in fig. 1 by s1. The nucleosomal DNA in our model must be connected to these points. Since the rigid bp model does not contain anything except the bp's themselves, we need to define the locations of our virtual phosphates. As a good approximation to the geometry of real DNA double helices, we place the actual locations of the phosphates in a given conformation at the midpoint of a line which connects the ends of two adjacent base pairs as shown in fig. 1 (green sphere m1). We denote the coordinate of the center of mass of the i-th ellipsoid by $\hat{x}^i$ and its orientation with respect to a reference frame by its local frame $\{\hat{d}_1^i, \hat{d}_2^i, \hat{d}_3^i\}$, where $\hat{d}_3$ is normal to the bp plane and $\hat{d}_2$ points along the major axis of the ellipsoids. Then the position vector $\hat{x}_{m1}$ of m1 in the reference frame is obtained as

$$\hat{x}_{m1} = \frac{1}{2}\left(\hat{x}^i + \hat{x}^{i+1} + \frac{w}{2}(\hat{d}_2^i + \hat{d}_2^{i+1})\right), \qquad (3)$$

where the indexes i and $i+1$ correspond to the bp's involved in a binding site and $w \approx 18\,\text{Å}$ is the bp width. According to this definition, Δr can be written as $\Delta r = |\hat{x}_{m1} - \hat{x}_{s1}|$ with $\hat{x}_{s1}$ as the position vector of point s1 in the reference frame.

According to the equipartition theorem, the stiffness of each binding site follows from the fluctuations at that point

$$k_{\rm bind} = \frac{3 B_f k_B T}{8\pi^2}, \quad \text{with} \quad B_f = \frac{8}{3}\pi^2 \langle (\Delta r)^2 \rangle, \qquad (4)$$

where B_f is the crystallographic B-factor and T is the temperature of the crystal structure, here 103 K. This gives us a hint of the order of magnitude for the spring constant but further adjustments were needed (see below). Within the accuracy of the crystal structure data the binding sites reflect the two-fold symmetry of the core particle very closely. In our model we impose perfect symmetry by averaging the coordinates and stiffnesses of the corresponding binding sites.

The structure obtained in this way has still some degrees of freedom showing fluctuations that are too large compared to the experiment, especially a sideway rolling motion of the DNA. In reality, hydrogen bonds between the phosphates and the proteins control the direction of the bonded groups. We need therefore an extra constraint at every binding site. In fig. 1, s1 shows the location of the phosphorous atom obtained from the symmetrized NCP147 structure. To restrict the direction of the bp's at the binding sites, we add for each binding site another spring at position s2. This extra spring acts on the point m2 between two bp's involved in a binding site. The coordinates of point m2 can be found in a similar way as m1 via eq. (4) by using $(w - d_s)$ instead of w with $d_s = |\hat{x}_{s2} - \hat{x}_{s1}|$. These two springs must be located in a reference direction $(\hat{x}_{s2} - \hat{x}_{s1})$. To find this direction, we first simulate the system without the extra springs. The average direction of the $\hat{d}_2$ axes of two bp's at that site is chosen to be the reference direction. The stiffnesses of both springs are set equal and the distance d_s between them is kept as a free parameter. This free parameter is found through fitting to the B-factor curve of the NCP147 as explained in the next section. The parameters obtained from this procedure are then used for different DNA sequences.

The only parameters that are left to be determined are the depths of the binding sites, $E_{\rm bind}$. In most of our simulations bonded bp's are not allowed to detach; the values of $E_{\rm bind}$ are then not important. Only in sect. 3.3 these values are needed to estimate the energy landscape felt by a twist defect that moves around the nucleosome. In that section we explain how we extracted the values of the binding energies of the 14 binding sites from experiments.

2.3 Coarse-grain molecular dynamics simulations

All of our simulations are done by considering rigid-body dynamics for the base pairs and integrating the equations of motion using a symplectic algorithm [35]. It is assumed

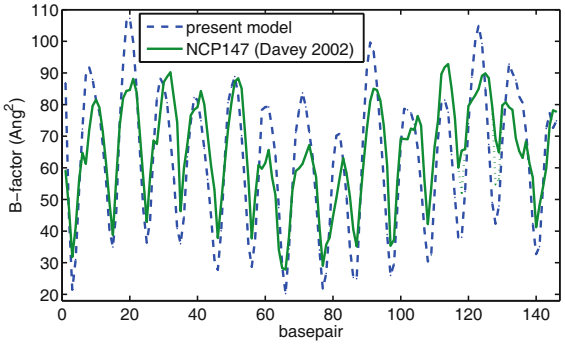

Fig. 2. (Color online) Comparison of the B-factor curves between the experiment [2] and the present model.

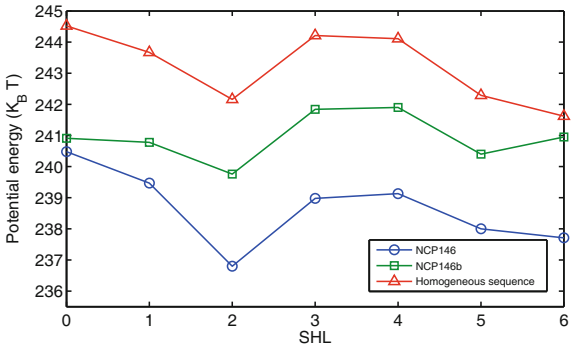

Fig. 3. (Color online) Total elastic energy of the system for twist defect located on different superhelical locations for NCP146, NCP146b, and for a homogenous DNA chain.

that the mass of each base pair is distributed uniformly in a geometry given in [36] to find the corresponding moment of inertia. Forces are calculated directly from the partial derivatives of the interaction potential. For torque calculation we use a numerical method based on virtual work exerting small virtual rotations about three mutual axes on every bp at each time step (see appendix A for more details). Simulations are performed in a canonical ensemble with the Nose-Hoover Chain thermostat. For every simulation run the system is given enough time for equilibration. Our coarse-grained model allows us to increase the time steps in our simulations up to about 10 fs which is bigger than its value in ordinary all-atom MD. This together with the radical reduction in the number of degrees of freedom in the rigid base pair model leads to a very efficient simulation of nucleosomal DNA. For instance, a 1 ns simulation for 147 bp nucleosomal DNA on a 2.4 GHz CPU takes about 2 hours. The presented data in this study are obtained from several MD simulations with total simulation time of about 2.5 microseconds. Due to the complexity and the size of the nucleosome structure this required simulation time is hardly achievable with all-atom molecular dynamics simulations. For example, 240 cores benchmarked in an all-atom nucleosome model consisting of 158000 atoms at between 8-24 ns/day [37].

3 Results and discussion

3.1 B-factor

For the first simulation we tried to reproduce the B-factor curve of the phosphates for the NCP147 structure. A NVT simulation at 103 K was performed and after equilibration of the system samples were taken from the location of the m1-points (fig. 1) of the model to calculate the B-factor for both strands of DNA. After doing some adjustments on the stiffness of the springs and location of the second spring we were able to reproduce the B-factor curve with a reasonable accuracy as it is shown in fig. 2. The obtained distance between two springs at each site was chosen to be 3 Å everywhere.

3.2 Preferred location of twist defect

We used our model to study the energy differences of the system when a twist defect is located on different superhelical locations (SHL's), for two different sequences, the NCP146 and the NCP146b structure. The twist defect in NCP146 is located between SHL+0.5 and SHL+2.5 and in NCP146b located between SHL+3.5 and SHL+5.5 [28,34]. The effect of histone tails is absent in our model and it might affect the location of the twist defects but it has been shown that even without considering this effect, rigid base pair models are able to predict twist defect location in crystal structures [28]. Figure 3 shows the results of our simulations for these sequences. Every point of these curves belongs to a separate simulation with a missing base pair at the corresponding SHL. After removing a base pair, the system is relaxed at the crystal temperature and the time-averaged potential energy of the system is measured. For NCP146 the energy curve shows an absolute minimum at the 2nd SHL corresponding to the experimental location. For NCP146b the curve also shows an absolute minimum at the 2nd SHL; in addition a local minimum appears at 5th SHL. So the preferred locations of the twist defects correspond to minima (in one case only a local one) of the elastic energy of our model.

The same calculations were performed for homogeneous DNA where its parameters were obtained by averaging over the equilibrium values and stiffness matrices of all possible bp steps, see fig. 3. Remarkably all three curves show more or less the same trends. This finding suggests that the locations and stiffnesses of the binding sites play a more important role than the DNA sequence itself in determining the preferred twist defect location. We make use of this finding in the next section where we study twist defect diffusion using homogenous DNA instead of performing a huge number of simulations for every nucleosome position of a given bp sequence.

We can measure the absolute cost of twist defects at different SHL's by comparing the energies of the nucleosome with twist to that of a perfect 147 bp complex. To account for the fact that the latter structure contains one bp more, we multiply the potential energy of the perfect

147 bp model by a factor of 146/147. From this we find on average a $4.5k_BT$ cost for placing a twist defect into a nucleosome. Remarkably, this number is substantially cheaper than what was estimated before [20,28]. Thus according to our model twist defects occur more likely. As we have a more accurate description of the microscopic structure of the nucleosome, we believe that this smaller value might be closer to the real value. As we shall see later, other effects, unaccounted for in ref. [20], compensate for this discrepancy such that our new prediction of nucleosome repositioning finds numbers for the overall dynamics similar to the former reference. The by far most important new effect comes from the fact that binding sites have different strengths as explained in the next section.

3.3 Twist defect diffusion

As mentioned in the introduction, nucleosome sliding is very likely caused by the diffusion of twist defects around the histone octamer [20,21]. A twist defect enters spontaneously from either end of the wrapped DNA portion. The twist defect can then diffuse inside the nucleosome. The defect moves when a thermal fluctuation causes the opening of a binding site next to the defect, smearing it out over two adjacent SHL's. This transition state is unstable and the DNA forms again a new bond to the octamer. If this bond is created in a way to localize the twist defect at the next SHL, the defect has moved by one step corresponding to 10 bp. To find the energy landscape that the defect experiences as it moves through the nucleosome, we need to determine the energies of a nucleosome with a defect at every SHL and every transient state in between. For twist defects at different SHL's we just need to repeat what we have done in the previous section, but now at room temperature. However, for the transition states we need to know the values of the binding energies. In a recent experiment, Hall *et al.* [38] unzipped DNA into a nucleosome and measured dwell times, showing that the strength of different binding sites is not the same. This result can be used to estimate the strength of every binding site. If we suppose that the binding energy of each site has an exponential relation with its corresponding dwell time and $t_{\mathrm{dwell}} = Ce^{E_{\mathrm{bind}}/k_BT}$ with C being a proportionality constant, then we have

$$E^i_{\mathrm{bind}} = k_BT(\log t^i_1 + \log t^i_2 - 2\log C), \tag{5}$$

where E^i_{bind} is the binding energy at the i-th binding site, t^i_1 and t^i_2 are the dwell times for the two peaks that were observed for each binding site in the experiment.

In the following C is chosen such that the overall binding energy agrees with the value determined from an experiment where nucleosomes were unwrapped by an external force [39]. A theoretical study [26] based on that experiment allows to estimate the average energy for breaking each binding site

$$E_{\mathrm{average}} = (f_0 + f_1)d + \frac{1}{14}E_{\mathrm{el}}. \tag{6}$$

Table 1. Estimated energy of each binding site.

Binding site SHL	Energy (k_BT)
6.5	9.2
5.5	11.2
4.5	12.8
3.5	9.2
2.5	10.1
1.5	14.4
0.5	16.2

Here f_0 is the smaller force needed for the unwrapping of the first turn, f_1 is the extra force for the unwrapping of the last turn (reflecting the electrostatic repulsion between the two DNA turns), d is the average length of DNA between binding sites, about 3.4 nm, and E_{el} is the total elastic energy of the DNA bent around the nucleosome. By minimizing the energy of the system we find $E_{\mathrm{el}} \approx 70k_BT$ for homogeneous DNA. This value is much smaller than elastic energies reported in the previous pdb based model of ref. [28] where the DNA conformation was forced to be close to that of the crystal structures. Putting this together with $f_0 = 0.7k_BT/\mathrm{nm}$ and $f_1 = 1.4k_BT/\mathrm{nm}$ [26] we obtain $E_{\mathrm{average}} \approx 12k_BT$. Using eq. (5) with C chosen such that the average binding energy amounts to $12k_BT$ leads to the values shown in table 1. Note that for the outmost binding sites no peaks in the dwell time could be observed, reflecting the weakness of those sites. We assume here that their energies are equal to the one of the second weakest position that is at SHL3.5. We checked that as long as the outer binding sites are weak, the results presented in this paper do not depend on that particular choice.

With these numbers at hand we determine now the height of barriers felt by a diffusing twist defect. The energy cost for bond breaking is given in table 1 and the gain in elastic energy by stretching the defect over two neighboring locations is obtained by a simulation where the corresponding binding site is turned off. Figure 4 shows the energy landscape that the twist defect feels while it goes through the nucleosomal DNA. In this plot every point is a separate simulation of the system at room temperature to determine the time-averaged potential energy. Integer numbers correspond to cases where the twist defect is localized at the corresponding SHL and half-integers describe the transition states. The change in energy at these transition states is the sum of the energy cost for breaking the corresponding binding site (taken from table 1) and the gain in the elastic energy of the DNA.

Kramers' rate theory predicts that the rate with which a twist defect enters the nucleosome is given by

$$k_{\mathrm{enter}} = \nu_0\, e^{\frac{-U_b}{k_BT}}, \tag{7}$$

where ν_0 is the attempt frequency and U_b is the barrier height that the twist defect encounters when entering the nucleosome. The attempt frequency can be roughly estimated by realizing that a stretch of about 10 bp length

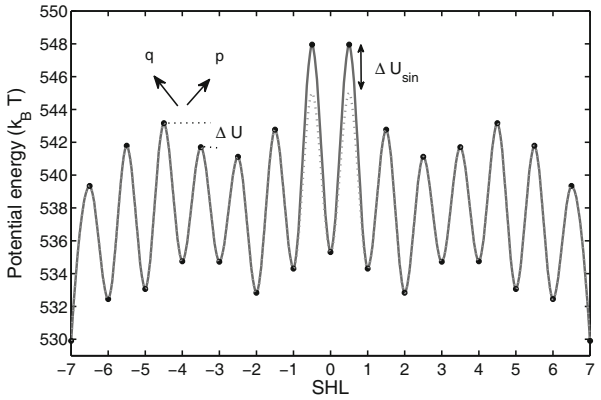

Fig. 4. Energy landscape for a twist defect moving through a nucleosome. At the third position the probabilities for the twist defect to go to the right, p, and to the left, q, are indicated together with ΔU, the energy difference between the barriers. *Sin*-mutations are accounted for by weakening the middle binding sites by an amount ΔU_{sin}.

needs to make a fraction of corkscrew motion when the defect goes over a barrier. Considering that stretch as a cylinder, the rotational friction constant is given by $\zeta_{\text{eff}} = (2\pi/10\,\text{bp})4\pi\eta R^2 L$, where $L = 10\,\text{bp} \approx 34\,\text{Å}$, $R \approx w/2 \approx 9\,\text{Å}$, $\eta = 10^{-3}\,\text{Pa s}$ (the viscosity of water) and $(2\pi/10\,\text{bp})$ corresponds to translating 10 bp to a full turn, 2π. With this we estimate the attempt frequency to be $1.6 \times 10^{10}\,\text{s}^{-1}$ [20]. Inserting into eq. (7) the height of the initial barrier in the energy landscape of fig. 4, we obtain $k_{\text{enter}} = 1.2 \times 10^6\,\text{s}^{-1}$.

Only those twist defects that pass through all of the barriers lead to nucleosome repositioning but not those that fall off at the same side where they have been created. We calculate now that crossing probability for twist defects. At each location a twist defect has different rates for going forward or backward. Suppose at a given location there is an energy difference ΔU between the corresponding transition states (see fig. 4). As a result the escape rate to the right, r, and to the left, s, are not equal but obey $s = re^{-\Delta U/k_B T}$. Therefore the probabilities of the twist defect to go to the right, p, and to the left, q, are given by

$$p = \frac{r}{r+s} \quad \text{and} \quad q = \frac{s}{r+s}\,. \tag{8}$$

Finding the probabilities for all 13 sites and considering both ends as absorbing boundaries we can construct the transition matrix for a twist defect which enters from SHL-7 and exits from SHL+7 as

$$\begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ q_{-6} & 0 & p_{-6} & & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & q_{-5} & 0 & & p_{-5} & & \cdot & \cdot & & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & q_{+6} & & 0 & & p_{+6} \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & & 1 \end{pmatrix} \tag{9}$$

where the p_i's and q_i's are the probabilities for the twist defect at the i-th SHL to go to the right and to the left.

Table 2. Probabilities for the twist defect to go to the right for each SHL.

p_{-6}	p_{-5}	p_{-4}	p_{-3}	p_{-2}	p_{-1}	p_0
0.076	0.320	0.707	0.837	0.062	0.007	0.5

These values are given in table 2 for the first 7 locations. All the other values can be found by the symmetry ($q_i = p_{-i}$ and $p_i = 1 - q_i$).

By removing the columns and rows that correspond to the absorbing boundaries, we obtain a 13×13 submatrix T. The fundamental matrix of this transition is $Q = (I - T)^{-1}$ and the probability p^* for a twist defect to go from the SHL-7 to the SHL+7 is given by $(QR)_{1,2}$ where R is a 13×2 submatrix of the transition matrix with $R_{1,1} = q_{-6}$ and $R_{13,2} = p_{+6}$ and other components equal to zero.

This calculation leads to $p^* \approx 1/9000$. Using this probability and accounting for the fact that there are two types of twist defects (extra or missing base pair) entering the nucleosome from two sides, we find the diffusion coefficient for the sliding of the nucleosome along the DNA

$$D = 2k_{\text{enter}}p^* \approx 240\,\text{bp}^2/\text{s}. \tag{10}$$

This value is much higher than the experimental value where repositioning is observed on the time scale of an hour. However, so far we have not considered the effect of the DNA sequence on repositioning rates. Remarkably, a number of the same order of magnitude, $D \approx 580\,\text{bp}^2/\text{s}$, has been found in in ref. [20] (also there before accounting for the bp sequence effect). Even though the cost for twist defects had been estimated to be much more expensive, the crossing probability was much higher (as equally strong binding sites were assumed), overall compensating the differences. Before we consider sequence effects, we discuss first how the transition probability for twist defect can be affected by mutations.

3.4 Effect of sin mutations on the probability of twist defect transfer

It has been shown that some mutations in the histone proteins, called *sin*-mutations, enhance the mobility of nucleosomes along DNA [40]. One group of *sin*-mutations occurs near the dyad axis affecting binding sites at SHL±0.5. This effect makes nucleosome sliding about 4 times faster [40]. In our model we account for *sin*-mutations by weakening these binding sites by an amount ΔU_{sin} (see fig. 4). We then repeat our calculations for the twist defect passage probability with the new values for the binding energies of the most inner binding sites. In fig. 5 the ratio of the transition probability with and without mutations is shown *versus* the amount of change in the binding energy of the sites at SHL±0.5. Our model predicts a 4 times higher transition rate —and thus a four times higher mobility— when the inner binding sites are weakened by just 13%. This pronounced sensitivity is

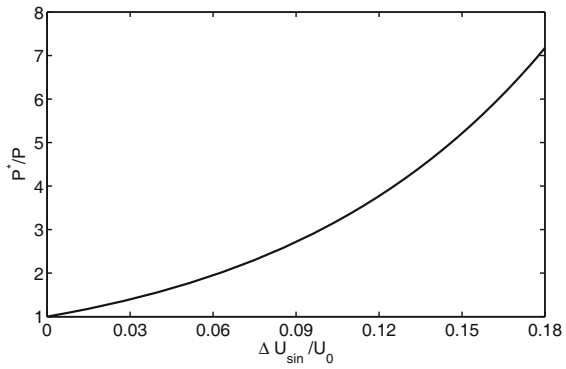

Fig. 5. Ratio of probabilities for passing of the twist defect through the nucleosome with and without *sin* mutations *versus* change in the strength of the binding sites at SHL±0.5. U_0 is the binding strength of those sites in the absence of mutations.

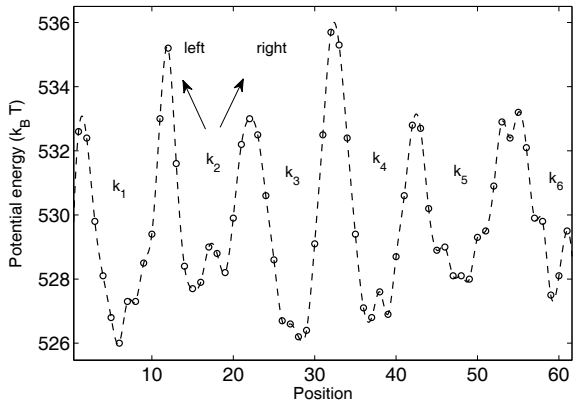

Fig. 6. Potential energy felt by the nucleosome at different locations around the 5S rDNA positioning sequence.

caused by the fact that many defects are reflected at the most inner sites as these are the strongest in the nucleosome, see fig. 4.

3.5 Effect of sequence-dependent elasticity of DNA on nucleosome positioning and repositioning

To study the sequence-dependent properties of nucleosomal DNA, we start with a 207 base pairs long sequence containing the 5S rDNA positioning sequence [14]. This sequence offers the nucleosome 61 different possible positions. We performed separate simulations for the 147 wrapped bp for each position. When the nucleosome is moved by one bp, we need to rotate it also by one tenth of a full turn. In fig. 6 average elastic energies are shown for all 61 positions. This curve shows several interesting features. The absolute minimum of the curve occurs at position 6 which is in agreement with Morozov *et al.* [27]. Also this position is very close to the positioning site found in the experiment [14]. Sequence-dependent elasticity of DNA also predicts the experimentally observed 10 base pair distance between the preferred nucleosome locations reflecting the optimal direction of bending of the DNA on the octamer.

Table 3. Rates of going over barriers for 5S rDNA for different positions (s^{-1}), see fig. 6.

Rate	Left	Right
k_1	0.091	0.254
k_2	0.251	0.122
k_3	0.173	0.259
k_4	0.229	0.201
k_5	0.133	0.065
k_6	0.109	-

We use now this finding, together with what we obtained for the diffusion of a single twist defect in eq. (10), to estimate the effective diffusion constant of nucleosome sliding along such a piece of DNA. Using the (approximate) 10 bp periodicity, we can estimate the effective diffusion constant as follows:

$$D_{\text{eff}} = \frac{100\,\text{bp}^2}{2T_{\text{eff}}}\,. \tag{11}$$

Here T_{eff} is the time typically spent in each minimum. The inverse of this time, $k = T_{\text{eff}}^{-1}$, follows from Kramer's rate theory to be

$$k = \frac{\sqrt{U''(x_{\min})U''(x_{\max})}}{2\pi\zeta} e^{-A/k_BT}. \tag{12}$$

In this equation $U''(x_{\max})$ and $U''(x_{\min})$ are the curvatures of the elastic energy at the top of the barriers and at the bottom of the minima, A is the barrier height, and $\zeta = k_BT/D$ with $D \approx 240\,\text{bp}^2/\text{s}$, see eq. (10). By fitting parabolic functions to every minimum and maximum and measuring barrier heights at the different positions, the rate values at different locations can be determined. These values are shown in table 3. According to this table, the average rate is about $0.172\,\text{s}^{-1}$. Putting this value into eqs. (11) and (12) we arrive at a diffusion constant of $8.6\,\text{bp}^2/\text{s}$ which is close to the experimental value $\approx 1\,\text{bp}^2/\text{s}$. We conclude that the twist defect scenario together with sequence-dependent elasticity explains the slow time scale of the nucleosome sliding phenomenon.

We repeated the same simulation for the 601 positioning sequence. From all 135 different possible positions, we only simulated between positions 28 to 72 that contain position 50, the position with the highest affinity. According to fig. 7 our model shows a 10 bp separation between preferred locations. In comparison with the 5S rDNA sequence, the absolute minimum of this curve, at position 40 in our model, has a lower elastic energy than each of the 6 positions of the 5S rDNA sequence. This is compatible with the fact that the 601 sequence has a higher affinity for nucleosomes [41]. In addition, we find here substantially higher barriers between the preferred locations. This suggests a lower mobility in the vicinity of the 601 positioning sequence as compared to 5S rDNA. Our calculations also show higher barriers for the 601 sequence in comparison

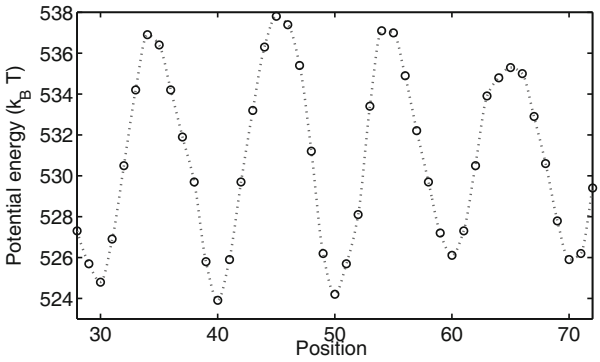

Fig. 7. Potential energy for the 601 positioning sequence. Only positions 28 to 72 are shown.

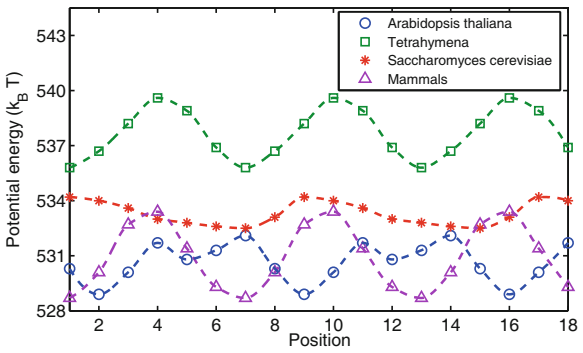

Fig. 8. (Color online) Potential energy of the nucleosome as a function of the position for four different telomeric sequences.

Table 4. Effective diffusion coefficient for different telomeric sequences.

Sequence	D_{eff} (bp^2/s)
Arabidopsis thaliana (GGGTTTA)	45.1
Mammals (GGGTTA)	14.5
Saccharomyces cerevisiae (GGTGTGTG)	122.3
Tetrahymena (GGGGTT)	31.5

with [27]. Doing the same calculation as before we obtain here an effective diffusion constant $D_{\text{eff}} \approx 0.03\,\text{bp}^2/\text{s}$, which is 2 orders of magnitude smaller than that for 5S rDNA.

We finally studied the sequence-dependent elasticity for telomeric sequences. Telomeric sequences are short tandem repeats of DNA (usually 6 to 8 base pairs long) and are found at the chromosome ends. It has been shown that these sequences do not have a strong affinity for nucleosomes and that nucleosome mobility is enhanced [19,42]. Here we simulated nucleosomal DNA for 4 different telomeric sequences (see table 4). Comparing the absolute minima of the elastic energy curves in fig. 8 with each other and with the 5S rDNA and 601 sequences we find for the lowest possible energy of each sequence:

$$U_{601} < U_{5\text{S rDNA}} < U_{thaliana} \approx U_{\text{mammals}} < U_{cerevisiae} < U_{Tetrahymena}. \quad (13)$$

The above qualitative result shows a good agreement with *in vitro* affinity measurements of these sequences [41,42].

From fig. 8 one can also see that for telomeric sequences one has smaller barriers between different positions in comparison to the 601 and 5S rDNA positioning sequences. This explains the high mobility of nucleosomes on telomeric sequences. In table 4 we present the effective diffusion constant for these four telomeric sequences. They are between 1.7 and 14 times larger than the one we found for the 5S rDNA sequence.

4 Conclusion

Here we introduced a rigid base pair molecular dynamics model for nucleosomal DNA including interactions to the histone octamer via 14 binding sites. Our model is capable to reproduce experimental results found for the nucleosome crystal structure: fluctuations of the DNA phosphates and the preferred locations of one base pair twist defects. More importantly, we demonstrated that the model is also very successful in predicting various effects occurring for room temperature: twist defect diffusion through nucleosomes causing their repositioning, the role of *sin* mutations in affecting nucleosome mobility and that of the underlying base pair sequence in slowing down the kinetics. Our coarse-grained model allows thus to simulate such processes on biologically relevant time scales which would be otherwise inaccessible.

In conclusion, our results suggest that the elasticity of DNA, even in harmonic approximation, is able to explain a wide range of phenomena that has been observed for nucleosomes.

We thank Prof. Ralf Everaers and Dr. Behrouz Eslami-Mossallam for helpful discussions.

Appendix A. Force and torque calculations for the elastic energy of base pairs

Suppose that we have two neighboring bp's interacting with each other and that the conformation of the second bp is given by the rotation matrix R and the location vector $\mathbf{x}$ in the local coordinates of the first bp. The three components of the force acting on the second bp in the local Cartesian coordinates of the first bp can be written as

$$F_k = -\frac{\partial U(R,x)}{\partial x_k}, \qquad k = 1,2,3, \quad (A.1)$$

where the x_k's are the components of the Cartesian coordinates, the F_k's are the force components in the corresponding direction, and $U(R,\mathbf{x})$ is defined in eq. (1). By

using the symmetry of the stiffness matrix one finds the forces as

$$\mathbf{F}_k = -\frac{\partial \psi_i}{\partial x_k} K_{ij}(\psi_j - \psi_{0j}) \text{ summation over } i,j = 1,\ldots,6, \tag{A.2}$$

where ψ=[Twist,Tilt,Roll,Shift,Slide,Rise] is the relative conformation of the base pairs and K, ψ_0, and ψ are defined in the midframe coordinate system [29]. The three rotational components of ψ do not depend on the relative translation of the bp's [29]. So the partial derivative of these components with respect to x_k is equal to zero

$$\frac{\partial \psi_i}{\partial x_k} = 0, \qquad i = 1,2,3. \tag{A.3}$$

On the other hand translational parameters can be found from the midframe coordinate Rm [29]:

$$\psi_i = Rm_{i-3j} x_j, \qquad i = 4,5,6. \tag{A.4}$$

Knowing that the midframe orientation does not depend on the x_k's, partial derivatives of Shift, Slide, and Rise with respect to x_k can be easily obtained as

$$\frac{\partial \psi_i}{\partial x_k} = Rm_{i-3j}\frac{\partial x_j}{\partial x_k} = Rm_{i-3j}\delta_{jk} = Rm_{i-3k}, \quad i = 4,5,6. \tag{A.5}$$

By substituting eqs. (A.3) and (A.5) in (A.2) we have

$$\mathbf{F}_k = -Rm_{i-3k} K_{ij}(\psi_j - \psi_{0j}), \quad i = 4,\ldots,6, \qquad j = 1,\ldots,6. \tag{A.6}$$

To obtain the acting torque on the second bp we have used the virtual work method numerically. Virtual rotations are applied in three orthogonal directions and the difference between the potential energies before and after a small virtual rotation is measured. From the ratio of the difference in the potential energies and h, the amount of the virtual rotation, the value of torque at the corresponding direction can be found

$$M_i = -\frac{U(h\epsilon_i \cdot R, \mathbf{x}) - U(R, \mathbf{x})}{h}, \qquad i = 1,2,3. \tag{A.7}$$

Here M_i is the component of torque in the corresponding direction, and ϵ_i is the asymmetric third-order tensor. We choose here $h = 0.00001$. To reduce computational costs, force and torque on the first bp are calculated with the equilibrium conditions.

References

1. H. Schiessel, J. Phys.: Condens. Matter **15**, R699 (2003).
2. K. Luger, A.W. Mäder, S.K. Richmond, D.F. Sargent, T.J. Richmond, Nature (London) **389**, 251 (1997).
3. K. Maeshima, S. Hihara, M. Eltsov, Curr. Opin. Cell Biol. **22**, 291 (2010).
4. K.J. Polach, J. Widom, J. Mol. Biol. **254**, 130 (1995).
5. J.D. Anderson, J. Widom, J. Mol. Biol. **296**, 979 (2000).
6. J.D. Anderson, J. Widom, Mol. Cell. Biol. **21**, 3830 (2001).
7. J.D. Anderson, P.T. Lowary, J. Widom, J. Mol. Biol. **307**, 977 (2001).
8. J.D. Anderson, A. Thaström, J. Widom, Mol. Cell. Biol. **22**, 7147 (2002).
9. G. Li, M. Levitus, C. Bustamante, J. Widom, Nat. Struct. Mol. Biol. **12**, 46 (2005).
10. L. Kelbauskas, N. Chan, R. Bash, J. Yodh, N. Woodbury, D. Lohr, Biochemistry **46**, 2239 (2007).
11. A. Gansen, A. Valeri, F. Hauger, S. Felekyan, S. Kalinin, K. Toth, J. Langowski, C.A.M. Seidel, Proc. Natl. Acad. Sci. U.S.A. **106**, 15308 (2009).
12. W.J.A. Koopmans, R. Buning, T. Schmidt, J. van Noort, Biophys. J. **97**, 195 (2009).
13. P. Prinsen, H. Schiessel, Biochimie **92**, 1722 (2010).
14. S. Pennings, G. Meersseman, E.M. Bradbury, J. Mol. Biol. **220**, 101 (1991).
15. G. Meersseman, A. Pennings, E.M. Bradbury, EMBO J. **11**, 2951 (1994).
16. S. Pennings, G. Meersseman, E.M. Bradbury, Proc. Natl. Acad. Sci. U.S.A. **91**, 10275 (1994).
17. A. Flaus, T.J. Richmond, J. Mol. Biol. **275**, 427 (1998).
18. J.M. Gottesfeld, J.M. Belitsky, C. Melander, P.B. Dervan, K. Luger, J. Mol. Biol. **321**, 249 (2002).
19. S. Pisano, E. Marchioni, A. Galati, R. Mechelli, M. Savino, S. Cacchione, J. Mol. Biol. **369**, 1153 (2007).
20. I.M. Kulic, H. Schiessel, Phys. Rev. Lett. **91**, 148103 (2003).
21. F. Mohammad-Rafiee, I.M. Kulic, H. Schiessel, J. Mol. Biol. **344**, 47 (2004).
22. H. Schiessel, Eur. Phys. J. E **19**, 251 (2006).
23. E.M. Mateescu, C. Jeppesen, P. Pincus, Europhys. Lett. **46**, 493 (1999).
24. K.K. Kunze, R.R. Netz, Phys. Rev. Lett. **85**, 4389 (2000).
25. H. Schiessel, J. Widom, R.F. Bruinsma, W.M. Gelbart, Phys. Rev. Lett. **86**, 4414 (2001).
26. I.M. Kulic, H. Schiessel, Phys. Rev. Lett. **92**, 228101 (2004).
27. A.V. Morozov, K. Fortney, D.A. Gaykalova, V.M. Studitsky, J. Widom, E.D. Siggia, Nucl. Acids Res. **37**, 4707 (2009).
28. N.B. Becker, R. Everaers, Structure **17**, 579 (2009).
29. W.K. Olson, M. Bansal, S.K. Burley, R.E. Dickerson, M. Gerstein, J. Mol. Biol. **313**, 229 (2001).
30. W.K. Olson, A. Gorin, X. Lu, L. Hock, V. Zhurkin, Proc. Natl. Acad. Sci. **95**, 11163 (1998).
31. F. Lankas, P. Sponer, J. Langowski, T.E. Cheatham, Biophys. J. **85**, 2872 (2003).
32. R. Lavery, K. Zakrzewska, D. Beveridge, T.C. Bishop, D.A. Case, T. Cheatham, S. Dixit, B. Jayaram, F. Lankas, C. Laughton *et al.*, Nucleic Acids Res. **38**, 299 (2010).
33. N. Becker, L. Wolff, R. Everaers, Nucleic Acid Res. **34**, 5638 (2006).
34. C.A. Davey, D.F. Sargent, K. Luger, A.W. Maeder, T.J. Richmond, J. Mol. Biol. **319**, 1097 (2002).
35. H. Kamberaj, R.J. Low, M.P. Neal, J. Chem. Phys. **122**, 1906216 (2005).
36. B. Mergell, M.R. Ejtehadi, R. Everaers, Phys. Rev. E **68**, 021911 (2003).
37. M. Biswas, J. Langowski, T.C. Bishop, WIREs Comput. Mol. Sci. (2013).

38. M.A. Hall, A. Shundrovsky, L. Bai, R.M. Fulbright, J.T. Lis, M.D. Wang, Nature Struc. Mol. Biol. **16**, 124 (2009).
39. B.D. Brower-Toland, C.L. Smith, R.C. Yeh, J.T. Lis, C.L. Peterson, M.D. Wang, Proc. Natl. Acad. Sci. **99**, 1960 (2002).
40. A. Flaus, C. Rencurel, H. Ferreira, N. Wiechens, T. Owen-Hughes, EMBO J. **23**, 343 (2004).
41. P.T. Lowary, J. Widom, J. Mol. Biol. **276**, 19 (1998).
42. S. Cacchione, M.A. Cerone, M. Savino, FEBS Lett. **400**, 37 (1997).