Biochimie 92 (2010) 1722–1728

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Research paper

Nucleosome stability and accessibility of its DNA to proteins

Peter Prinsen, Helmut Schiessel*

Instituut Lorentz, Universiteit Leiden, P.O. Box 9506, 2300 RA Leiden, The Netherlands

ARTICLE INFO

Article history:
Received 13 April 2010
Accepted 3 August 2010
Available online 13 August 2010

Keywords:
DNA
Nucleosomes
Restriction enzymes

ABSTRACT

In this paper we present a theoretical description of the accessibility of nucleosomal DNA to proteins. We reassess the classical analysis of Polach and Widom (1995) who demonstrated that proteins (in their case restriction enzymes) gain access to buried binding sites inside a nucleosome through spontaneous unwrapping of DNA from the protein spool. We introduce a straightforward nucleosome model the predictions of which show good agreement with experimental data. By fitting the model to the data we obtain the values of two quantities: the adsorption energy to the histone octamer per length of DNA and the extra length that the DNA needs to unwrap beyond the binding site of an enzyme before the enzyme can act as effectively as on bare DNA. Our results indicate that the effective binding energy is surprisingly low which suggests that the nucleosomal parameters are tuned such that two large energies, the DNA bending energy and the pure adsorption energy, nearly cancel. This paper is based on a lecture presented at the summer school "DNA and Chromosomes 2009: Physical and Biological Applications". We follow the lecture as closely as possible which is why we spend more time than usual on issues that are already well-known in the field, and why we discuss some well-known results from a different perspective.

1. Introduction

DNA in eukaryotes is highly compacted through the complexation with cationic protein assemblies; the histone octamers [1]. The compaction mainly serves two functions: to allow the huge amount of DNA (e.g. a total of 2 m of DNA chains per human cell) to fit inside the micron-sized cell nucleus, and to provide an additional layer of control for gene expression by varying the compaction level along the DNA. The latter plays a role in cell differentiation allowing for different cell types, despite them sharing identical genomes.

When we look at the detailed structure of this compacted DNA, called chromatin, the question arises how the DNA can be accessed by transcription factors, polymerases, and other proteins. Most of the DNA seems inaccessible because it is tightly wrapped around millions of protein cylinders. Each of those DNA spools, the so-called nucleosomes, consists of 147 bp (bp) of DNA wrapped in 1 3/4 turns around an octamer of histone proteins and is connected to the next nucleosome by a short stretch of so-called linker DNA of only 10–90 in length. As a result around three quarters of the DNA is wrapped which seems to imply that most of the DNA is inaccessible to DNA binding proteins simply for steric reasons.

From its crystal structure [2] we know that the nucleosome features fourteen regions where the wrapped DNA is in contact with the octamer surface, located where the minor groove of the DNA double helix faces inwards towards the surface of the octamer. In each contact region there are several direct hydrogen bonds as well as positive charges that attract the negatively charged phosphates of the DNA backbone. In order to bind at those sites the DNA has to pay a high price because it needs to bend substantially. To estimate the energy associated with this bending we employ the wormlike chain model, which is very successful in describing DNA elasticity, at least for modestly bent DNA. It is not clear how well this model works for strongly bent DNA inside a nucleosome but we assume it can be employed to calculate a rough estimate of the energy involved. In a nucleosome 127 bp of DNA are bent around the octamer (10 bp at each terminus are essentially unbent [2]). According to the wormlike chain model [1] the energy E_{elastic} associated with this bending is

$$\frac{E_{\text{elastic}}}{k_B T} = \frac{l_P l}{2R_0^2} \tag{1}$$

Here $l_P = 50$ nm is the DNA persistence length, k_B is Boltzmann's constant, T is the temperature, $l \approx 127 \times 0.34\,\text{nm} = 43$ nm is the bent part of the wrapped DNA, and $R_0 \approx 4.3$ nm is the radius of curvature of the centerline of the wrapped DNA (see Fig. 1). This leads to a total bending energy of about 58 k_BT.

In order for the nucleosome to be stable, the pure binding energy of the 14 sites should exceed this 58 k_BT by an energy at least in the order of 1 k_BT per binding site. However, we also expect it to

* Corresponding author. Tel.: +31 71 527 1543; fax: +31 71 527 5511.
E-mail address: schiessel@lorentz.leidenuniv.nl (H. Schiessel).

0300-9084/$ – see front matter © 2010 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.biochi.2010.08.008

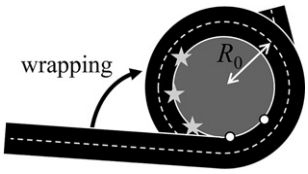

Fig. 1. A partially unwrapped nucleosome with exposed nucleosomal binding sites (stars). The nucleosome can lower its energy by closing those binding sites at the cost of bending the DNA.

be not much more than this because the DNA inside the nucleosome needs to be accessible somehow. If the difference between the pure binding energy and the bending energy is small enough, one can imagine that the nucleosome can make parts of its DNA temporarily accessible through spontaneous unwrapping, as indicated in Fig. 1 which shows a partially unwrapped nucleosome.

Polach and Widom [3] demonstrated that nucleosomes indeed show such opening fluctuations. They studied nucleosome core particles that consist of 147 bp of DNA wrapped around the histone octamer. Since all the DNA is wrapped in that case, one should expect that it would not be accessible to DNA binding proteins if the DNA is too strongly bound. Yet it was found that nucleosomal DNA gets digested when it is exposed to restriction enzymes, a special class of enzymes that cut DNA at specific positions. Moreover, it was observed that the closer the cutting site was located towards the center of the wrapped DNA, the slower this digestion, but the reaction was always possible.

In this paper we give a theoretical estimate of the effective binding energy, i.e. the difference between the pure binding energy and the bending energy, of the DNA to the nucleosome using data from Ref. [3]. Before we introduce a model that allows us to estimate the binding energy (Section 3), we first discuss the relevant experiment [3] and what interesting quantities can be extracted from it.

2. Probing nucleosome accessibility with restriction enzymes

In this section we discuss how the accessibility of a DNA binding site inside the nucleosome was determined experimentally and we discuss the reaction equations associated with the breathing of the nucleosome and the digestion of the DNA. The situation is schematically depicted in Fig. 2(a). Suppose a protein binding site, denoted by the light gray section in Fig. 2, is located somewhere inside the wrapped portion. The protein, denoted by "R", cannot bind in that case. The nucleosome is a dynamic structure however; its DNA unwraps and rewraps spontaneously from both ends and occasionally the DNA unwraps far enough to make the binding site accessible to the protein. We expect that the probability of this happening decreases with the distance from the binding site to the closest terminus of the wrapped portion and is smallest when the binding site is located in the center of the wrapped portion.

To demonstrate and measure this wrapping and unwrapping mechanism Polach and Widom [3] used restriction enzymes. These proteins cut DNA at a specific position in a specific short base sequence. A large number of different kinds of restriction enzymes occur naturally in bacteria and archaea. Their function is to protect the organism against foreign DNA, usually of viral origin. They recognize short sequences in the foreign DNA that do not occur in the organisms own DNA and destroy it by simply cutting at a specific site in the short sequence. Polach and Widom [3] determine the exposure of the cutting site by measuring the rate at which the DNA is degraded. As long as the nucleosome is sufficiently wrapped the restriction enzyme cannot bind due to steric hindrance. When the nucleosome "breathes" spontaneously, i.e. unwraps its DNA far enough beyond the binding site of the enzyme, the enzyme can bind and then either unbind again or cut the DNA at that particular site, see Fig. 2a. By measuring the fraction of uncut DNA as a function of time one can determine the cutting rate. This rate can then be compared to the cutting rate in a solution of bare DNA, i.e. not wrapped around histone octamers, under identical conditions to determine the probability that that particular binding site is accessible to the protein, see Fig. 2b. (Actually, in the experiments the conditions are almost but not exactly identical: since the cutting in the case of bare DNA is much faster than in the case of nucleosomes, the experiments on the bare DNA are performed at restriction enzyme concentrations that are typically two or three orders of magnitude lower than in the case of nucleosomes. However, as we will see later, the cutting rate is proportional to the enzyme concentration. One can then determine the probability that a certain binding site is accessible by comparing the ratios of cutting rate and enzyme concentration for nucleosomes and bare DNA).

We now show how the probability that a certain site is accessible can be determined from the cutting rates of the nucleosomal DNA and the bare DNA. In our derivation of the rate at which the DNA is cut we make three assumptions. The first is that the ratio of the concentration of substrate and the total concentration of enzyme is small so that the concentration of free enzyme can be set equal to the total concentration of enzyme (free and bound). This is the case in the experiments we are considering and it simplifies the theoretical treatment considerably since the rate equations are linear in that case. The second assumption is that the slowest relaxation rate in this system of coupled reaction equations is much slower than the other relaxation rates. This means that in the experiments we see a single relaxation time instead of a combination of two (that might or might not be distinguishable due to measurement errors). Finally, we assume that the equilibrium between open and closed nucleosomes is fast so that the ratio of open and closed nucleosomes is constant during the measurement.

Let us consider first the set of reactions with the bare DNA, Fig. 2b. We denote the bare DNA with "S" (S standing for "site", the site where the enzyme binds), the restriction enzyme with "R", the complex of the enzyme and the DNA by "RS" and the cut DNA with "P" (P stands for "product"). The reaction scheme is then

$$\mathrm{S} + \mathrm{R} \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} \mathrm{RS} \overset{k_{34}}{\rightarrow} \mathrm{P} + \mathrm{R} \tag{2}$$

Here k_{23} and k_{32} denote the forward and backward rates for the binding and unbinding of the restriction enzyme to its target site and k_{34} is the irreversible rate of the cutting of the DNA. For this reaction scheme one can estimate the rate of decrease of the amount of intact, uncut DNA by writing down the rate equations for the concentrations of the different species. In a compact matrix notation the master equation reads as follows:

$$\frac{d}{dt}\begin{pmatrix} c_{\mathrm{S}} \\ c_{\mathrm{RS}} \end{pmatrix} = \begin{pmatrix} -k_{23}c_{\mathrm{R}} & k_{32} \\ k_{23}c_{\mathrm{R}} & -k_{32}-k_{34} \end{pmatrix}\begin{pmatrix} c_{\mathrm{S}} \\ c_{\mathrm{RS}} \end{pmatrix} \tag{3}$$

Here c_{S} is the concentration of sites S, c_{RS} the concentration of bound restriction enzymes and c_{R} the concentration of free, unbound enzymes. These concentrations are functions of time. We assume that the concentration of enzymes is large enough so that this concentration can be considered constant, i.e. $c_{\mathrm{R}} \gg c_{\mathrm{RS}}$ at all times, so we can set $c_{\mathrm{R}}(t) \approx c_{\mathrm{R}}(0) \equiv c_{\mathrm{R,b}}$ (here "b" stands for bare). Eq. (3) is then a set of linear first-order differential equations for c_{S} and c_{RS}. The concentration of the product P is not considered since it is directly related to the concentrations of c_{S} and c_{RS}. Solutions of Eq. (3) are linear combinations of

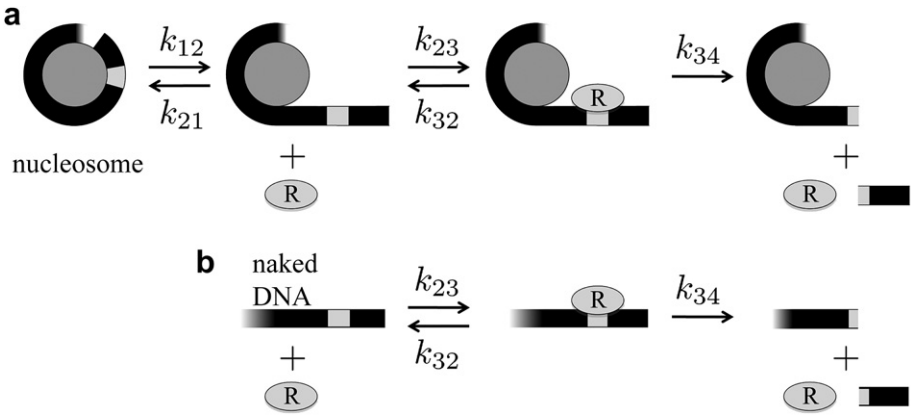

Fig. 2. The setup of Polach and Widom [3]. (a) A fully wrapped nucleosome unwraps spontaneously, thereby exposing the binding site (light gray) for the restriction enzyme R. The enzyme cuts the DNA at this particular site. (b) Same setup in the absence of the histone octamer.

$$\mathbf{c}^i(t) = \mathbf{c}^i e^{\lambda^i t} = \begin{pmatrix} c_1^i \\ c_2^i \end{pmatrix} e^{\lambda^i t} \quad i = +,- \tag{4}$$

with $\mathbf{c}^+ = (\lambda^+ + k_{32} + k_{34}, k_{23}c_{R,b})^T$ and $\mathbf{c}^- = (\lambda^- + k_{32} + k_{34}, k_{23}c_{R,b})^T$ the eigenvectors of the 2×2 matrix in Eq. (3) and

$$\lambda^{\pm} = \frac{1}{2}\Big(\pm \sqrt{(k_{32} + k_{34} + k_{23}c_{R,b})^2 - 4k_{23}c_{R,b}k_{34}} - k_{32} - k_{34} - k_{23}c_{R,b}\Big) \tag{5}$$

the corresponding eigenvalues, where the plus sign in $\pm$ should be used for λ^+ and the minus sign for λ^-. $-\lambda^+$ and $-\lambda^-$ are called the relaxation rates or decay rates of the components proportional to $\mathbf{c}^+$ and $\mathbf{c}^-$, respectively, as can be seen from Eq. (4).

Generally, the initial concentrations, $c_S(0)$ and $c_{RS}(0)$, are not known. However, we want to measure a single decay rate in the experiments so we assume that $|\lambda^-| \gg |\lambda^+|$. This is the case if and only if $(k_{32} + k_{34} + k_{23}c_{R,b})^2 \gg k_{23}c_{R,b}k_{34}$. Then after a very short time $\approx 1/|\lambda^-|$ only the component with the smaller decay rate, $-\lambda^+$, survives whereas the faster mode has died out. The experimentally determined rate constant k_{bare} that controls the decay of the bare DNA is then simply

$$k_{bare} = -\lambda^+ \approx \frac{k_{23}c_{R,b}k_{34}}{k_{32} + k_{34} + k_{23}c_{R,b}} \tag{6}$$

If we assume, for simplicity, that $c_{RS}(0) = 0$ (which holds if we add the restriction enzyme to the DNA solution at $t = 0$) then a sufficient condition for $c_{RS} \ll c_{R,b}$ is $k_{23}c_S(0) \ll \lambda^+ - \lambda^- \approx -\lambda^- \approx k_{32} + k_{34} + k_{23}c_{R,b}$.

We now determine the corresponding rate constant for the cutting of the nucleosomal DNA. The reaction scheme, Fig. 2a, is as follows:

$$N + R \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} S + R \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} RS \overset{k_{34}}{\rightarrow} P + R \tag{7}$$

As before, we assume that $c_R \gg c_{RS}$ so that $c_R(t) \approx c_R(0) \equiv c_{R,n}$ ("n" stands for nucleosome) in which case we have three linear first-order differential equations for c_N, c_S and c_{RS}. We also assume that the first reaction in (7), the equilibrium between the open and closed nucleosome, is fast compared to the other ones, namely the binding and unbinding of the restriction enzyme and the cutting of the DNA by the restriction enzyme. One can show that this is the case if $k_{21} \gg k_{32} + k_{34} + k_{23}c_{R,n}$ (this follows for example from Appendix A, specifically Eq. (28) together with the assumption $|\lambda_1| \gg |\lambda_2|$). Then the most negative eigenvalue (which corresponds to the fastest rate) of the matrix pertaining to the three differential equations (analogous to the matrix in Eq. (3) above) is $\lambda_1 \approx -(k_{12} + k_{21})$ (see Appendix A). After a short time $\approx 1/|\lambda_1|$ the ratio c_N/c_S is approximately constant and equal to k_{21}/k_{12} and we can simplify the master equation to

$$\frac{d}{dt}\begin{pmatrix} c_N + c_S \\ c_{RS} \end{pmatrix} = \begin{pmatrix} -k_{23}c_{R,n}p_{open} & k_{32} \\ k_{23}c_{R,n}p_{open} & -k_{32} - k_{34} \end{pmatrix}\begin{pmatrix} c_N + c_S \\ c_{RS} \end{pmatrix} \tag{8}$$

where

$$p_{open} \equiv \frac{c_S}{c_N + c_S} \approx \frac{k_{12}}{k_{12} + k_{21}}. \tag{9}$$

The quantity p_{open} has a simple meaning: it is the probability to find the binding site open or, in other words, it is the fraction of time the binding site is open. Eq. (8) corresponds to the following reaction scheme

$$D + R \underset{k_{32}}{\overset{k_{23}p_{open}}{\rightleftharpoons}} RS \overset{k_{34}}{\rightarrow} P + R \tag{10}$$

where D represents the intact DNA, i.e. N and S lumped together. This reaction scheme is the same as that for the bare DNA case, Eq. (2), with c_S replaced by $c_D = c_N + c_S$ and k_{23} by $k_{23}p_{open}$. There is an additional factor p_{open} because the restriction site is only available in a fraction p_{open} of the DNA molecules.

Analogous to the case of bare DNA, we assume that the smaller of the two eigenvalues (in absolute value), λ_3, is much smaller than the other one, λ_2, which is the case if and only if $(k_{32} + k_{34} + k_{23}c_{R,n}p_{open})^2 \gg k_{23}c_{R,n}p_{open}k_{34}$. Then for times larger than $\approx 1/|\lambda_2|$ D will decay with a single decay rate

$$k_{nucl} \approx \frac{k_{23}c_{R,n}p_{open}k_{34}}{k_{32} + k_{34} + k_{23}c_{R,n}p_{open}} \tag{11}$$

If we assume again, for simplicity, that $c_{RS}(0) = 0$ (restriction enzyme added to the DNA solution at $t = 0$) and if we also assume that c_N and c_S are in equilibrium at $t = 0$, in other words $k_{12}c_N(0) = k_{21}c_S(0)$, then a sufficient condition for $c_{RS} \ll c_{R,n}$ is $k_{23}c_S(0) = k_{23}p_{open}(c_S(0) + c_N(0)) \ll k_{32} + k_{34} + k_{23}c_{R,n}p_{open}$.

One can determine p_{open} from k_{nucl}. However, one needs to know the values of the rate constants k_{23}, k_{32} and k_{34}. If $k_{32} + k_{34} \gg k_{23}c_{R,b}, k_{23}c_{R,n}p_{open}$ matters simplify considerably. In that case

$$\frac{k_{\mathrm{nucl}}}{c_{\mathrm{R,n}}} \approx p_{\mathrm{open}} \frac{k_{\mathrm{bare}}}{c_{\mathrm{R,b}}} \tag{12}$$

In other words, comparing the two rates for DNA cutting, that for nucleosomes, k_{nucl}, and that for bare DNA, k_{bare}, we can easily determine p_{open} without explicitly having to determine the other rate constants. Note that we can write $k_{\mathrm{nucl}} \approx k_{23} c_{\mathrm{R,n}} p_{\mathrm{open}} p_{\mathrm{cut}}$ where $p_{\mathrm{cut}} \equiv k_{34}/(k_{32}+k_{34})$ is the probability that the enzyme cuts the DNA when it is bound to it (as opposed to unbinding). In other words, the enzyme binding determines the rate of digestion of the DNA. To summarize, the restrictions on the rate constants and concentrations that lead to Eq. (12) are:

- R1: $k_{32}+k_{34} \gg k_{23} c_{\mathrm{S}}$
- R2: $k_{32}+k_{34} \gg k_{23} c_{\mathrm{R,n}} p_{\mathrm{open}}$
- R3: $k_{32}+k_{34} \gg k_{23} c_{\mathrm{R,b}}$
- R4: $k_{21} \gg k_{32}+k_{34}+k_{23} c_{\mathrm{R,n}}$

In the experiments of Polach and Widom [3] $K_{\mathrm{m}} \equiv (k_{32}+k_{34})/k_{23} \gg c_{\mathrm{S}}$ for both the experiments with bare DNA and the ones with nucleosomes, which implies that R1, and thus the assumption $c_{\mathrm{RS}} \ll c_{\mathrm{R,n}}, c_{\mathrm{R,b}}$, is justified. In the same paper it is claimed that $k_{21} \geq 10^5\ \mathrm{s}^{-1}$. However, subsequent measurements showed that $k_{21} \approx 20-90\ \mathrm{s}^{-1}$ [4]. Together with $k_{23} \lesssim 10^8\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$ [5] this implies that $k_{21}/k_{23} \geq 200-900\ \mathrm{nM} \gg K_{\mathrm{m}} + c_{\mathrm{R,n}}$ ($c_{\mathrm{R,n}} \leq 100$ nM and $K_{\mathrm{m}} \approx 1$–10 nM, see [3]). This means that indeed $k_{21} \gg k_{32}+k_{34}+k_{23} c_{\mathrm{R,n}}$ (R4). Finally, since $p_{\mathrm{open}} \leq 10^{-2}$, as we will see shortly, and since $c_{\mathrm{R,b}}$ is typically about two to three orders of magnitude smaller than $c_{\mathrm{R,n}}$, we have $K_{\mathrm{m}} \gg c_{\mathrm{R,b}}, c_{\mathrm{R,n}} p_{\mathrm{open}}$ so $k_{32}+k_{34} \gg k_{23} c_{\mathrm{R,b}}, k_{23} c_{\mathrm{R,n}} p_{\mathrm{open}}$ (R2 and R3). All the assumptions that lead to Eq. (12) are thus fulfilled.

Fig. 3 shows the results of the experiment. We plot p_{open} as a function of x_b, which is the position along the DNA of the binding site of the respective restriction enzyme (in bp). Experiments have been performed for positions close to the entrance, which is at $x_b = 1$ bp, up to close to the middle of the wrapped portion, which is at $x_b = 74$ bp. Note that the accessibility is greatly reduced for binding sites anywhere in the nucleosomal DNA as compared to bare DNA, even for binding sites close to the terminus of the wrapped portion. Moreover, the data points lie roughly along a straight line in the log–linear plot. This suggests that the probability decays exponentially from the termini towards the middle of the wrapped portion.

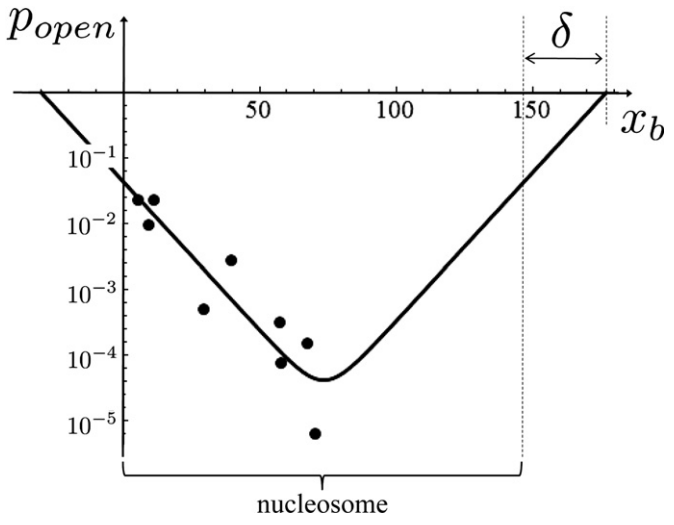

Fig. 3. Probability p_{open} for a binding site to be open. The position of the binding site, x_b, is given in base pairs. The termini of the wrapped portion are at $x_b = 1$ bp and at $x_b = 147$ bp. The data are taken from the restriction enzyme analysis [3]. Shown is also the theoretical curve, Eq. (14).

Note that it seems that we have more restrictions on the rate constants than Polach and Widom even though we arrive at the same expression for the rate of decrease of total nucleosome concentration as they do (Eq. (12)). However, they make some assumptions that were not made explicit in their paper. We discuss this further in Appendix B.

Also note that Eq. (10) is a different effective reaction scheme from that found in Ref. [6]. In that paper it is claimed that the right side of Eq. (7) follows the classical Michaelis–Menten relationship. This is not true in general however, since S is constantly replenished from the stock of closed nucleosomes, N. Their assumption is valid if $-(k_{32}+k_{34})$ is the largest eigenvalue, in absolute value, but this is not the case here as can easily be seen from estimates of the reaction constants. So, contrary to their claims, they do put restrictions on the reaction rates (the steady state assumption in the Michaelis–Menten equation implies additional restrictions).

3. Model for nucleosome breathing

The experiments of Polach and Widom [3] demonstrate that DNA binding proteins can reach their target sites within nucleosomal DNA because of spontaneous opening fluctuations of the nucleosome. In this section we extract the energetics involved in this so-called *site exposure mechanism* from the data shown in Fig. 3. For this we need to relate the experimentally measured quantity p_{open} to the adsorption energy per base pair on the nucleosome $f_{\mathrm{crit}}a$. Here f_{crit} is the critical force that would be needed to unpeel the DNA from the histone octamer and $a \approx 0.34$ nm is the distance between base pairs. We introduce the dimensionless parameter $q \equiv f_{\mathrm{crit}} a/k_B T$ and we assume that the unwrapping state of the nucleosome only depends on the number of unwrapped base pairs at each end of the DNA. We number the base pairs of the DNA that can be adsorbed on the histone octamer from $x = 1$ to $x = L = 147$. The unwrapping state of a nucleosome is then characterized by the section $\{x_L, \cdots, x_R\}$ that is still wrapped, where $1 \leq x_L \leq x_R \leq L$. The dimensionless complexation energy of a fully wrapped nucleosome is $-qL$ whereas a partially unwrapped nucleosome has a lower complexation energy $-q(x_R - x_L + 1)$. We sum over all possible states weighted with the corresponding Boltzmann factor to get the partition function

$$Z = \sum_{x_L=1}^{L} \sum_{x_R=x_L}^{L} e^{q(x_R - x_L + 1)} \approx \frac{e^{q(L+2)}}{q^2} \tag{13}$$

where we used the fact that $f_{\mathrm{crit}} = O(1\mathrm{pN})$ and $k_B T \approx 4.1$ pN nm at room temperature so $q \ll 1$ and $qL \gg 1$.

We now suppose that there is a restriction site between base pairs x_b and $x_b + 1$ with $1 \leq x_b \leq L-1$. Note that many restriction enzymes do not cut simply between two base pairs but produce overhangs (short single-stranded sections). In that case we define the restriction site as exactly between the cuts in the two single strands. We are interested in the probability that the restriction site is accessible to the restriction enzyme. We assume that in order for the restriction enzyme to bind it is not sufficient that the DNA is unwrapped up to the restriction site but that δ extra base pairs of DNA have to be unwrapped. This is schematically depicted in Fig. 4. The probability that the restriction site is accessible is then

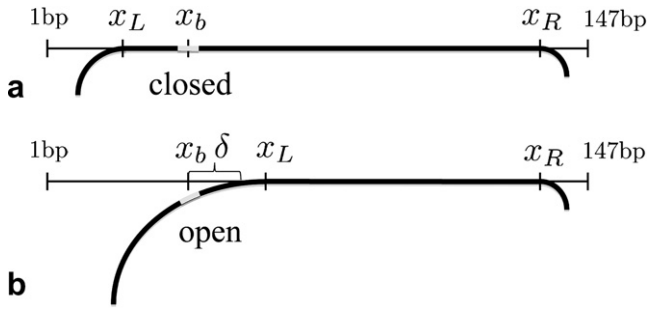

Fig. 4. Schematic depiction of different wrapping states of the nucleosome. In case (a) the binding site (light gray) is closed and thus not accessible to the restriction enzyme. Case (b) shows a situation in which the left end of the nucleosome is unwrapped to a position x_L with $x_L \geq x_b + \delta$. That means the binding site is open.

$$p_{\text{open}} = \frac{1}{Z}\left(\sum_{x_L = x_b+\delta+1}^{L} \sum_{x_R = x_L}^{L} e^{q(x_R - x_L + 1)} + \sum_{x_L = 1}^{x_b - \delta} \times \sum_{x_R = x_L}^{x_b - \delta} e^{q(x_R - x_L + 1)} \right) \approx e^{-q\delta}\left(e^{-qx_b} + e^{-q(L - x_b)} \right) \tag{14}$$

Since $L \gg 1$ we can treat x_b as a continuous variable. A least-square fit of Eq. (14) to the data in Fig. 3 leads to the curve depicted in that same figure and does indeed show a reasonable agreement. The optimal fit parameters (± one standard deviation) are $q = 0.104 \pm 0.016$ and $\delta = 30 \pm 12$ bp. The latter value suggests that a substantial amount of DNA needs to be unwrapped before the restriction enzyme can cut as efficiently as on a bare DNA substrate. The former gives a critical unwrapping force $f_{\text{crit}} = 1.3 \pm 0.2$ pN or, written differently, $f_{\text{crit}} = 0.31 \pm 0.05\ k_BT/\text{nm}$. The net adsorption energy of the total amount of DNA is $E_{\text{net}} = f_{\text{crit}} 50\ \text{nm} = 15 \pm 2\ k_BT$.

4. Discussion

We have introduced a model that allows us to determine from experiments two important parameters that characterize the successful binding and cutting of restriction enzymes on a nucleosomal binding site: the effective adsorption energy of the DNA on the histone octamer per unit length of DNA (or critical force) f_{crit} and the extra length of DNA δ that needs to be unwrapped before the restriction enzyme cuts as effectively as on bare DNA. We found δ to be quite substantial, namely $\delta = 30 \pm 12$ bp. It would be interesting to perform similar experiments but with a nucleosome complexed to longer stretches of DNA and to study whether the cutting efficiency is reduced even for binding sites outside the wrapped part but within a distance δ of the nucleosomal entrance points. The theoretical curve in Fig. 3 predicts p_{open} for those cases.

The absorption energy of the DNA on the histone octamer was found to be $f_{\text{crit}} = 0.31 \pm 0.05\ k_BT/\text{nm}$ which amounts to a net adsorption energy of $E_{\text{net}} \approx 15\ k_BT$. We mentioned in the introduction that the adsorption energy per binding site should not be too much larger than k_BT to allow for breathing but also not smaller than k_BT to have well-defined binding sites. Interestingly the average net binding energy per site is around $15\ k_BT/14 \approx 1\ k_BT$, i.e. at the lower boundary of the expected range. This is a surprisingly small number, especially taking into account the fact that we calculated above, in Eq. (1), that the elastic energy E_{elastic} is about 4 times larger. This suggests that nature has tuned the pure adsorption energy E_{ads} such that its value is close to E_{elastic}, namely

$$E_{\text{net}} = E_{\text{ads}} - E_{\text{elastic}} \approx 15 k_BT \tag{15}$$

with $E_{\text{ads}} \approx 75 k_BT$ and $E_{\text{elastic}} \approx 60 k_BT$.

Being so dynamic, however, might come at a cost: the nucleosome might not be very stable and easily fall apart. This is especially the case if a protein binds at a DNA binding site that is located deep inside the nucleosome. Once the protein is bound, the nucleosome cannot rewrap but might easily unwrap completely and disintegrate. Another example is the case when the nucleosome is under tension which can easily happen inside the nucleus where many motor proteins are at work all the time. Experiments [7] where nucleosomes have been unwrapped completely in micromanipulation setups suggest that the last turn is much more stable against unwrapping than the above analysis suggests. We interpreted this as a consequence of the two-turn geometry of the nucleosome by arguing that the turns feel an effective repulsion [8]. This repulsion would help the first turn to unwrap more easily leading to the substantial breathing observed in Ref. [3]. The breathing data, Fig. 3, might indeed suggest that p_{open} is significantly reduced for the most inner data point. However, the neighboring data points show no reduction of p_{open}, even though the DNA might have to be unwrapped more than one turn due to the large value of the extra length δ. We have fitted a model with two adsorption energies, one for the first 3/4 turn and one for the last turn, to the data in Fig. 3 but the fit is not very different from the single adsorption energy model. The estimated errors in the parameters are much larger in the latter case from which we conclude that the error in the measurements are too large to justify using the more complicated model.

As mentioned earlier, it is unclear whether the wrapping and unwrapping of the nucleosome can be described by a first-order rate process. In fact, the DNA is adsorbed on the histone octamer at 14 equally spaced contact points. The range of the adsorption potential is presumably small whereas the DNA has to stretch completely to recover the bending energy. This means that there is an energy barrier between the adsorped and desorped state at each contact point. In that case the wrapping should be described as an equilibrium process between DNA adsorped at all points, desorped at one point, desorped at two points and so on. One can show that if the desorption rate is much smaller than the adsorption rate then the differences between the relaxation rates are much smaller than their absolute values which means that it is very hard experimentally to distinguish between a single relaxation rate and a combination of relaxation rates. We will come back to this point in a future paper.

All this is unimportant if the equilibration between different adsorption states is fast compared to the enzyme binding and DNA cutting. In that case only the equilibrium distribution of adsorption states is relevant and one can speak of a probability that the restriction site is open without having to refer to an opening and closing rate.

Note that in our model for nucleosome breathing we assume that there is a constant binding energy per base pair even though in reality there is an adsorption site every 10 bp [2]. Furthermore, we treat the cutting position x_b as a continuous variable even though it is not. These assumptions simplify the computations considerably and the only difference is that the fit in Fig. 3 would otherwise consist of horizontal lines with vertical jumps every 10 bp. The spread in the experimental data is too big to determine where these jumps occur.

The current paper focused on equilibrium properties of nucleosomal breathing. In the meantime data on the dynamics of breathing have become available. Those data are extracted from experiments that use FRET (fluorescence resonance energy transfer) in which two fluorescent dyes are placed at strategic positions allowing direct monitoring of the breathing dynamics [4,9–16]. We are currently extending our model to explain the dynamics of

breathing as observed in those FRET experiments (see also Ref. [17]).

Acknowledgments

We thank Jonathan Widom, Marc Emanuel and Nima Hamedani Radja for stimulating discussions. This work is part of the research programme of the 'Stichting voor Fundamenteel Onderzoek der Materie (FOM)', which is financially supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)'.

Appendix A

If the concentration of free enzyme c_R is approximately constant and equal to the initial concentration $c_{R,n}$ then the rate equations corresponding to the reaction scheme in Eq. (7) are

$$\frac{d}{dt}\begin{pmatrix} c_N \\ c_S \\ c_{RS} \end{pmatrix} = \begin{pmatrix} -k_{12} & k_{21} & 0 \\ k_{12} & -k_{21}-k_{23}c_{R,n} & k_{32} \\ 0 & k_{23}c_{R,n} & -k_{32}-k_{34} \end{pmatrix}\begin{pmatrix} c_N \\ c_S \\ c_{RS} \end{pmatrix} \tag{16}$$

The rate constants, which are equal to minus the eigenvalues λ_i, $i = 1,2,3$, can be calculated from the characteristic equation of the matrix

$$\lambda_i^3 + a_2\lambda_i^2 + a_1\lambda_i + a_0 = 0 \tag{17}$$

where the coefficients, which are all positive, are given by

$$a_2 = k_{12} + k_{21} + k_{32} + k_{34} + k_{23}c_{R,n}, \tag{18}$$

$$a_1 = (k_{12} + k_{21})(k_{32} + k_{34}) + k_{23}c_{R,n}(k_{12} + k_{34}) \tag{19}$$

and

$$a_0 = k_{12}k_{23}c_{R,n}k_{34}. \tag{20}$$

We now assume that

$$k_{23}c_{R,n}(k_{12} + k_{34}) \ll (k_{12} + k_{21})(k_{32} + k_{34}), \tag{21}$$

which is equivalent to $k_{12}k_{23}c_{R,n} \ll (k_{12} + k_{21})(k_{32} + k_{34})$ and $k_{23}c_{R,n}k_{34} \ll (k_{12} + k_{21})(k_{32} + k_{34})$, since all reaction rates and concentrations are positive. The smallest eigenvalue (in absolute value), λ_3, is found by equating the last two terms in Eq. (17)

$$\lambda_3 \approx -\frac{a_0}{a_1} \approx -\frac{k_{12}k_{23}c_{R,n}k_{34}}{(k_{12} + k_{21})(k_{32} + k_{34})}. \tag{22}$$

It is easy to show that for $\lambda = \lambda_3$ the first two terms in Eq. (17) are indeed much smaller than the last two:

$$\left|\frac{\lambda_3^3}{a_1\lambda_3}\right| \approx \frac{k_{12}}{(k_{12} + k_{21})}\frac{k_{34}}{(k_{32} + k_{34})}\frac{k_{12}k_{23}c_{R,n}}{(k_{12} + k_{21})(k_{32} + k_{34})} \frac{k_{23}c_{R,n}k_{34}}{(k_{12} + k_{21})(k_{32} + k_{34})} \ll 1 \tag{23}$$

where we have used Eq. (21) and the fact that $0 < (a/a + b) < 1$ if a and b are positive, and

$$\left|\frac{a_2\lambda_3^2}{a_1\lambda_3}\right| \approx \frac{k_{12}}{(k_{12} + k_{21})}\frac{k_{23}c_{R,n}k_{34}}{(k_{12} + k_{21})(k_{32} + k_{34})} + \frac{k_{34}}{(k_{32} + k_{34})}\frac{k_{12}k_{23}c_{R,n}}{(k_{12} + k_{21})(k_{32} + k_{34})} + \frac{k_{23}c_{R,n}k_{34}}{(k_{12} + k_{21})(k_{32} + k_{34})}\frac{k_{12}k_{23}c_{R,n}}{(k_{12} + k_{21})(k_{32} + k_{34})} \ll 1. \tag{24}$$

We now show that λ_3 is indeed the smallest eigenvalue (in absolute value). The three solutions to Eq. (17) obey the relations

$$\lambda_1\lambda_2\lambda_3 = -a_0, \tag{25}$$

$$\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 = a_1 \tag{26}$$

and

$$\lambda_1 + \lambda_2 + \lambda_3 = -a_2. \tag{27}$$

Eqs. (22) and (25) show that $\lambda_1\lambda_2 \approx a_1$ which leads to the conclusion that λ_1 and λ_2 have the same sign (since a_1 is positive) and, together with Eq. (26), that $|\lambda_1\lambda_3 + \lambda_2\lambda_3| \ll \lambda_1\lambda_2$. Combining Eqs. (21) and (22) leads to $k_{12}^{-1} + k_{34}^{-1} \ll -\lambda_3^{-1}$ or, equivalently, $-\lambda_3 \ll k_{12}$ and $-\lambda_3 \ll k_{34}$ which implies $-\lambda_3 \ll a_2$. Combining this result with Eq. (27) leads to $\lambda_1 + \lambda_2 \approx -a_2$ so λ_1 and λ_2 are both negative. The equation $|\lambda_1\lambda_3 + \lambda_2\lambda_3| \ll \lambda_1\lambda_2$ can then be rewritten as $-\lambda_1^{-1} - \lambda_2^{-1} \ll -\lambda_3^{-1}$ or, in other words, all eigenvalues are negative and $|\lambda_1| \gg |\lambda_3|$ and $|\lambda_2| \gg |\lambda_3|$. We remark that Eq. (21) implies that $a_2 \approx k_{12} + k_{21} + k_{32} + k_{34}$ which, together with $\lambda_1\lambda_2 \approx a_1$ and $\lambda_1 + \lambda_2 \approx -a_2$, leads to the conclusion that $\lambda_1 \approx -(k_{12} + k_{21})$ and $\lambda_2 \approx -(k_{32} + k_{34})$ and that either can be the largest.

The (exact) eigenvectors corresponding to the three eigenvalues are

$$v_i = \begin{pmatrix} (\lambda_i + k_{21} + k_{23}c_{R,n})(\lambda_i + k_{32} + k_{34}) - k_{23}c_{R,n}k_{32} \\ k_{12}(\lambda_i + k_{32} + k_{34}) \\ k_{12}k_{23}c_{R,n} \end{pmatrix} \tag{28}$$

for $i = 1, 2, 3$. The solution to Eq. (16) is a linear combination of these eigenvectors

$$\begin{pmatrix} c_N \\ c_S \\ c_{RS} \end{pmatrix} = \sum_{i=1}^{3} \alpha_i v_i e^{\lambda_i t}. \tag{29}$$

We assume that $c_{RS}(0) = 0$ in which case $\alpha_1 + \alpha_2 + \alpha_3 = 0$ and $k_{12}\sum_{i=1}^{3}\alpha_i\lambda_i = c_S(0)$. We also assume that c_N and c_S are in equilibrium at $t = 0$, in other words, $k_{12}c_N(0) = k_{21}c_S(0)$. Then $k_{12}\sum_{i=1}^{3}\alpha_i\lambda_i^2 = -(k_{32} + k_{34} + k_{23}c_{R,n})c_S(0)$. The coefficient α_1 is given by $\alpha_1(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3) = (\lambda_1 + k_{12} + k_{21})c_S(0)/k_{12}$ and the other ones are found by permuting the indices. The maximum concentration of complex c_{RS} is $k_{12}k_{23}c_{R,n}\sum_{i=1}^{3}\alpha_i\exp(\lambda_i\tau)$ where τ is the solution of $\sum_{i=1}^{3}\alpha_i\lambda_i\exp(\lambda_i\tau) = 0$. We then have the exact expression

$$\frac{c_{RS}}{c_{R,n}} \geq k_{23}c_S(0)\left[\frac{k_{12} + k_{21} + \lambda_3}{\lambda_2(\lambda_1 - \lambda_3)}e^{\lambda_3\tau} - \frac{k_{12} + k_{21} + \lambda_1}{\lambda_2(\lambda_1 - \lambda_3)}e^{\lambda_1\tau}\right] \tag{30}$$

If we now use the fact that $k_{12} + k_{21} + \lambda_3 < k_{12} + k_{21}$, $|\lambda_3| \ll |\lambda_1|$, $0 < \exp(\lambda_i\tau) < 1$ and $|k_{12} + k_{21} + \lambda_1| \ll k_{12} + k_{21}$ we find

$$\frac{c_{RS}}{c_{R,n}} \lesssim \frac{k_{23}c_S(0)(k_{12} + k_{21})}{\lambda_1\lambda_2} \approx \frac{k_{23}c_S(0)}{k_{32} + k_{34}}. \tag{31}$$

This means that $c_{RS} \ll c_{R,n}$ if $k_{23}c_S(0) = k_{23}p_{open}(c_S(0) + c_N(0)) \ll k_{32} + k_{34}$.

Appendix B

Here we report the assumptions made by Polach and Widom in their paper, some of which were not made explicit. They mention the following restrictions on the rate constants: $k_{21} \gg k_{23}c_{R,n}$ and $k_{23}c_S \ll k_{32} + k_{34}$ ($k_{21} \gg k_{12}$ ($k_{eq}^{conf} \ll 1$ in their notation) is not necessary to arrive at Eq. (12)). However, they also make the steady state approximation for c_S and c_{RS} which means $|dc_S/dt| \ll k_{12}c_N + k_{32}c_{RS} \approx k_{21}c_S + k_{23}c_{R,n}c_S$ and $|dc_{RS}/dt| \ll (k_{32} + k_{34})c_{RS} \approx k_{23}c_{R,n}c_S$. If there is one single decay rate during the measurement (which is

the assumption) then $|c_N^{-1}dc_N/dt| \approx |c_S^{-1}dc_S/dt| \approx |c_{RS}^{-1}dc_{RS}/dt| \approx k_{nucl} \approx k_{23}c_{R,n}p_{open}p_{cut}$. The steady state assumptions then imply $k_{21} + k_{23}c_R \gg k_{23}c_{R,n}p_{open}p_{cut}$ and $k_{32} + k_{34} \gg k_{23}c_{R,n}p_{open}p_{cut}$. There is one additional assumption that is made which is not immediately obvious. From the reaction equations we have the exact relation $d(c_N + c_S)/dt = -(k_{34}c_{RS} + dc_{RS}/dt)$. In the derivation of k_{nucl} in Ref. [3] it is assumed (implicitly) that $d(c_N + c_S)/dt = -k_{34}c_{RS}$, in other words, that $k_{34}c_{RS} \gg dc_{RS}/dt$ which is the case if (and only if) $k_{32} + k_{34} \gg k_{23}c_{R,n}p_{open}$ (for bare DNA we find in a similar way $k_{32} + k_{34} \gg k_{23}c_{R,b}$). Finally, the restriction $k_{21} \gg k_{23}c_{R,n}$ can be replaced by the less restrictive one $k_{12} + k_{21} \gg k_{23}c_{R,n}p_{cut}$. This means that Polach and Widom's restrictions are really

- P1: $k_{32} + k_{34} \gg k_{23}c_S$
- P2: $k_{32} + k_{34} \gg k_{23}c_{R,n}p_{open}$
- P3: $k_{32} + k_{34} \gg k_{23}c_{R,b}$
- P4: $k_{12} + k_{21} \gg k_{23}c_{R,n}p_{cut}$
- P5: $k_{21} + k_{23}c_{R,n} \gg k_{23}c_{R,n}p_{open}p_{cut}$

where P4 and P5 can be replaced by the more restrictive $k_{21} \gg k_{12} + k_{23}c_{R,n}$. This means the only restrictions Polach and Widom do not mention in their paper are P2 and P3.

We also have restrictions P1, P2 and P3 but we have the restriction $k_{21} \gg k_{32} + k_{34} + k_{23}c_{R,n}$ instead of P4 and P5. Our restriction implies P4 and P5 but the reverse is not necessarily true. It turns out however that our assumption that $-(k_{12} + k_{21})$ is the largest eigenvalue (in absolute value) is too strong; it can also be the second largest eigenvalue. This is shown in Appendix A where it is also shown that the only assumptions one needs to derive $k_{nucl} \approx k_{23}c_{R,n}p_{open}p_{cut}$ are P1, P2 and P4; P5 is not needed (and P3 is needed for the bare DNA case).

References

[1] H. Schiessel, The physics of chromatin, J. Phys. Condens. Matter 15 (2003) R699–R774.

[2] K. Luger, A.W. Mäder, R.K. Richmond, D.F. Sargent, T.J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution, Nature 389 (1997) 251–260.

[3] K.J. Polach, J. Widom, Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation, J. Mol. Biol. 254 (1995) 130–149.

[4] G. Li, M. Levitus, C. Bustamante, J. Widom, Rapid spontaneous accessibility of nucleosomal DNA, Nat. Struct. Mol. Biol. 12 (2005) 46–53.

[5] A. Fersht, Structure and Mechanism in Protein Science. W.H. Freeman and Company, New York, 1999.

[6] C. Wu, A. Travers, A 'one-pot' assay for the accessibility of DNA in a nucleosome core particle, Nucl. Acids Res. 32 (2004) e122.

[7] B.D. Brower-Toland, C.L. Smith, R.C. Yeh, J.T. Lis, C.L. Peterson, M.D. Wang, Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA, Proc. Natl. Acad. Sci. USA 99 (2002) 1960–1965.

[8] I.M. Kulić, H. Schiessel, DNA spools under tension, Phys. Rev. Lett. 92 (2004) 228101.

[9] M. Tomschik, H. Zheng, K. van Holde, J. Zlatanova, S.H. Leuba, Fast, long-range, reversible conformational fluctuations in nucleosomes revealed by single-pair fluorescence resonance energy transfer, Proc. Natl. Acad. Sci. USA 102 (2005) 3278–3283.

[10] M. Tomschik, K. van Holde, J. Zlatanova, Nucleosome dynamics as studied by single-pair fluorescence resonance energy transfer: a reevaluation, J. Fluoresc. 19 (2009) 53–62.

[11] W.J.A. Koopmans, A. Brehm, C. Logie, T. Schmidt, J. van Noort, Single-pair FRET microscopy reveals mononucleosome dynamics, J. Fluoresc. 17 (2007) 785–795.

[12] W.J.A. Koopmans, R. Buning, T. Schmidt, J. van Noort, spFRET using alternating excitation and FCS reveals progressive DNA unwrapping in nucleosomes, Biophys. J. 97 (2009) 195–204.

[13] A. Gansen, F. Hauger, K. Tóth, J. Langowski, Single-pair fluorescence resonance energy transfer of nucleosomes in free diffusion: optimizing stability and resolution of subpopulations, Anal. Biochem. 368 (2007) 193–204.

[14] A. Gansen, K. Tóth, N. Schwarz, J. Langowski, Structural variability of nucleosomes detected by single-pair Förster resonance energy transfer: histone acetylation, sequence variation, and salt effects, J. Phys. Chem. B 113 (2009) 2604–2613.

[15] A. Gansen, A. Valeri, F. Hauger, S. Felekyan, S. Kalinin, K. Tóth, J. Langowski, C.A.M. Seidel, Nucleosome disassembly intermediates characterized by single-molecule FRET, Proc. Natl. Acad. Sci. USA 106 (2009) 15308–15313.

[16] L. Kelbauskas, J. Sun, N. Woodbury, D. Lohr, Nucleosome stability and dynamics vary significantly when viewed by internal versus terminal labels, Biochem. 47 (2008) 9627–9635.

[17] W. Möbius, R.A. Neher, U. Gerland, Kinetic accessibility of buried DNA sites in nucleosomes, Phys. Rev. Lett. 97 (2006) 208102.