The physics of chromatin

2003 J. Phys.: Condens. Matter 15 R699

(http://iopscience.iop.org/0953-8984/15/19/203)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **15** (2003) R699–R774 PII: S0953-8984(03)55677-6

TOPICAL REVIEW

The physics of chromatin

Helmut Schiessel

Max-Planck-Institut für Polymerforschung, Theory Group, PO Box 3148, D-55021 Mainz, Germany

Received 21 March 2003
Published 6 May 2003
Online at stacks.iop.org/JPhysCM/15/R699

Abstract

Recent progress has been made in the understanding of the physical properties of chromatin—the dense complex of DNA and histone proteins that occupies the nuclei of plant and animal cells. Here I will focus on the two lowest levels of the hierarchy of DNA folding into the chromatin complex. (i) The nucleosome, the chromatin repeating unit consisting of a globular aggregate of eight histone proteins with the DNA wrapped around it: its overcharging, the DNA unwrapping transition, the 'sliding' of the octamer along the DNA. (ii) The 30 nm chromatin fibre, the necklace-like structure of nucleosomes connected via linker DNA: its geometry, its mechanical properties under stretching and its response to changing ionic conditions. I will stress that chromatin combines two seemingly contradictory features: (1) high compaction of DNA within the nuclear envelope and, at the same time, (2) accessibility to genes, promoter regions and gene regulatory sequences.

Contents

1. Introduction

Higher developed organisms face the problem of storing and retrieving a huge amount of genetic information—and this in each cell separately. For instance, the human genome corresponds to 3 billion base pairs (bp) of the DNA double helix, two copies of which make up 2 m of DNA chains that have to be stored within the tiny micrometre-sized nucleus of each cell [1]. These 2 m are composed of 46 shorter DNA pieces, each of which, if not condensed, would form a swollen coil of roughly 100 μm diameter—clearly much too large to fit into the nucleus. On the other hand, the densely packed genome would form a ball of just $\sim$2 μm diameter due to the huge aspect ratio of contour length (2 m) versus diameter (20 Å) of the DNA material. Hence, the DNA indeed fits into the nucleus but a suitable compaction mechanism is required.

DNA is a highly charged macromolecule carrying two negative elementary charges per 3.4 Å. In the presence of multivalent counterions DNA condenses into dense, often toroidal, aggregates [2], resembling the DNA packaged in viral capsids [3–5]. In viruses the DNA has just to be stored whereas such a simple method of DNA compaction cannot work for the long DNA chains in eucaryotic cells (cells of fungi, plants and animals) where many portions of the DNA have to be accessible to a large number of proteins (gene regulatory proteins, transcription factors, RNA polymerases etc). Therefore the substrate that these proteins interact with must be much more versatile to allow access to certain regions of the DNA and to hide (i.e. to silence) other parts. That way each cell can regulate the expression of its genes separately according to its state in the cell cycle, the amount of nutrients present, etc. Furthermore, the differentiation of cells into the various types that make up a multicellular organism relies to a large extent on the way the DNA, which is identical in all the cells, is packaged.

The substrate that combines all these features is chromatin, a complex of DNA and so-called *histone proteins*. In 1974 it was realized that the fundamental unit of chromatin is the *nucleosome* [6, 7]: roughly 200 bp of DNA are associated with one globular octameric aggregate of eight histone proteins consisting of two molecules each of the four core histones H2A, H2B, H3 and H4. A stretch of 147 bp DNA is wrapped in a 1-and-3/4 left-handed superhelical turn around the octamer and is connected via a stretch of linker DNA to the next such protein spool. Each octamer, together with the wrapped DNA, forms a *nucleosome core particle* with a radius of $\sim$5 nm and a height of $\sim$6 nm which carries a large negative electric charge [8, 9].

While the structure of individual core particles is now documented in great detail mainly on the basis of high-resolution x-ray analyses [10, 11], much less is known about the higher-order structures to which they give rise. When the fibre is swollen—as this is the case for low ionic strength—it has the appearance of 'beads-on-a-string'. It is sometimes referred to as the '10 nm fibre' since its 'beads' have $\sim$10 nm diameter [12]. With increasing salt concentration, heading towards physiological conditions (roughly 100 mM), the fibre becomes denser and thicker, attaining a diameter of $\sim$30 nm [13]. Longstanding controversy surrounds the structure of this so-called 30 nm *fibre* [14–17]. In the *solenoid models* [12, 18, 19] it is assumed that the chain of nucleosomes forms a helical structure with the linker DNA bent in between, whereas the *zig-zag* or *crossed linker models* [20–24] postulate straight linkers that connect nucleosomes that are located on opposite sites of the fibre. The higher-order folding of the 30 nm fibre into structures on scales up to micrometres is yet to be elucidated. In figure 1 I sketch the steps of the DNA folding starting with a DNA chain of length $\sim$1 cm and $\sim 10^6$ octamers and ending up at the highly condensed *chromosome*. This highly condensed structure occurs before cell division and contains the chain and its copy neatly packaged for distribution into the two daughter cells. The size of the chromosome is $\sim$10 000 times smaller than the contour length of the original chain.

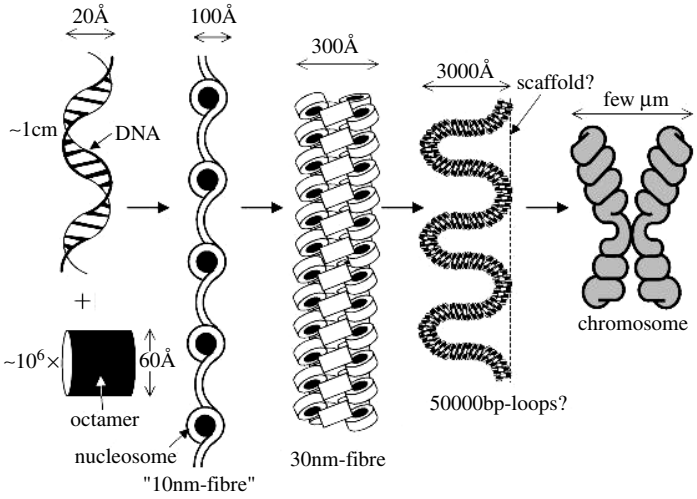

Figure 1. Steps of the DNA compaction into chromatin. The DNA molecule of length ∼1 cm is compacted with the help of 10^6 histone octamers leading to a 10 000-fold reduction of its original length (see text for details).

It is instructive to draw a comparison between the structure and function of chromatin and that of a daily-life example: the library. As the nucleus stores a long one-dimensional string of bp, so the library contains a huge one-dimensional string of letters, the text written down in all of its books. A book like [1] contains ∼10 km of text, a library with 10 000 books stores roughly 100 000 km of text! How can the user find and retrieve the little piece of information of interest? The way this is handled is that the text is folded in a hierarchical fashion in lines, pages, books and shelves. This makes it relatively easy, with the help of a few markers, to find the corresponding text passage. Furthermore, all the text is stored in a dense fashion but the book of interest can be taken out of the shelve and opened at the appropriate page without perturbing the rest of the library. Apparently, the result of this hierarchical structure is a relatively high efficiency in storing a huge amount of information in a relatively small space and, at the same time, having high accessibility to it.

The similarities between hierarchies in the library and in chromatin are pretty obvious. What is less obvious, and in many respects is still an open question, is how the dense chromatin structure can be opened locally to allow access to its genes. As mentioned above for one nucleosomal repeat length, typically 200, 147 bp are wrapped around the octamer, i.e., roughly 75% of the DNA chain is tightly associated with the histone aggregates. It is known that many essential proteins that interact with DNA do not have access to DNA when it is wrapped (reviewed in [25]). Moreover, even the unwrapped sections, the linker DNA, are somewhat buried inside the dense 30 nm fibre. Therefore, it is necessary for the cell to have mechanisms at hand to open (unfold) the fibre and then, somehow, to unwrap the DNA from the protein spools or to temporarily remove them from the DNA piece of interest.

This leads to the problem of how the chromatin structure changes its shape with time. As this structure involves length scales of many orders of magnitude (from ångströms to micrometres) then its dynamical processes take place in a wide range of timescales, beginning with fluctuations of the nucleosome structure in the micro- to millisecond timescale [26] up to large scale variations in the condensation degree of the chromosome that follows the cell cycle with a typical period of hours to days [1].

It has been shown through competitive protein binding to nucleosomal DNA [27, 28] that thermal fluctuations might lead to partial unwrapping of the DNA from the nucleosome. This mechanism provides intermittent access to nucleosomal DNA. Not only unwrapping but even '*sliding*' of nucleosomes along DNA seems to be facilitated by thermal fluctuations, as has been demonstrated in well-defined *in vitro* experiments [29–31]. Whereas this kind of repositioning dynamics is quite slow—even at elevated temperatures the timescale is of the order of minutes to hours—nucleosome repositioning also appears to be of great importance *in vivo* where it is aided by *chromatin-remodelling complexes*, large multi-protein complexes that use energy by burning ATP (reviewed in [7, 32–35]). These complexes might catalyse and direct the displacement of nucleosomes out of regions where direct access to DNA has to be granted (like promoter regions of transcriptionally active genes).

Based on a range of experiments it has also been speculated that RNA polymerase, the protein complex that transcribes (copies) genes, can itself act 'through' nucleosomes without having to disrupt the structure completely [36–40]. An appealing picture is the idea that the polymerase gets around the nucleosome in a loop. Such a mechanism is especially important since genes are typically distributed over a DNA portion of the length of hundreds to ten thousands of bp, which means that there are a few up to thousands of nucleosomes with which a polymerase has somehow to deal with during transcription.

Concerning the unfolding of the 30 nm fibre, I have already mentioned above that a decrease in the ionic strength leads to a swelling of the fibre. Obviously, *in vivo* such a global swelling is not possible within the tiny space available within the nucleus. Hence only the unfolding of local regions should be expected as the above given analogy of a library suggests. Experimental observations seem to indicate that the swelling degree can indeed be tuned locally by the acetylation of the lysine-rich (i.e. cationic) *tails* of the eight core histones that appear to be long flexible polyelectrolyte chains [41] that extend out of the globular part. Furthermore, transcriptionally active chromatin portions show a depletion in the *linker histone* H1, a cationic protein that is believed to act close to the entry–exit region of the DNA at the nucleosome. As long as H1 is present the fibre is relatively dense and the individual nucleosomes are inured to thermal fluctuations (no unwrapping, sliding or transcription through the octamer). If H1 is missing the chromatin fibre appears to be much more open and the nucleosomes become 'transparent' and mobile due to thermal fluctuations [31].

Undoubtedly chromatin lies at the heart of many essential biological processes, ranging from gene expression to cell division. Most of these processes are controlled by a huge number of specific proteins. Their investigation clearly belongs to the realm of biologists and is beyond the scope of a physicist. Nevertheless, many insights gained in this field were achieved through *in vitro* experiments under relatively well-controlled conditions, in some cases essentially only involving DNA and histone proteins. Furthermore, new physical methods like micromanipulation experiments allow us to gain access to certain physical properties of chromatin. The purpose of this paper is to discuss some of these results and, especially, to review the physical theories they gave rise to.

The interest of physicists in single nucleosomes was mainly sparked by the above mentioned fact that the core particle carries a large negative net charge. This is due to the fact that much more negatively charged DNA is wrapped around the cationic octamer than necessary for its neutralization. Beginning around 1998, considerable activity started among several research groups to explain on which physical facts *overcharging* is based (recently reviewed in [42]), a phenomenon not only occurring in chromatin but also in DNA–lipid complexes, multilayer adsorption of polyelectrolytes, etc. To gain insight into this the nucleosome was translated into different types of toy models, usually consisting of one charged chain and an oppositely charged sphere and the amount of wrapping was calculated [43–55] (cf also the

related studies [56–71]). It turned out that overcharging is a fairly robust phenomenon that occurs even if certain mechanisms are neglected (like counterion release upon adsorption). It was also shown that multi-sphere complexation can lead to *undercharged* systems [49, 51].

The relevance of these toy models for 'real' nucleosomes might be questionable, especially since the binding sites of the DNA to the octamer are relatively specific [10, 11] and since under physiological conditions (100 mM) the screening length is fairly short—10 nm, i.e., half the diameter of the DNA double helix. Nevertheless, these models might give some insight into the unwrapping transition that takes place when the DNA, which is fairly rigid on the nucleosome length scale, unwraps from the nucleosome due to a decrease of the adsorption energy. This can be achieved by a change in the salt concentration and has indeed been observed experimentally [8, 72]. A simple approach to this problem, comparing adsorption energies between a 'sticky' spool and a semiflexible chain and its bending energy, has been given in [73] and led to the prediction of an *unwrapping transition* in an all-or-nothing fashion, i.e., the cylinder is fully wrapped by the chain or not wrapped at all. A more complex picture has been found in [46, 47] where the unwrapping has been calculated for complexes with short chains ('nucleosome core particles'). Here the chain showed different degrees of wrapping (dependent on the ionic strength) which seems to agree fairly well with the corresponding experimental observations on core particles [8, 72]. It was shown in [55, 74] that besides the wrapped and unwrapped structures there is for longer chains the realm of open *multi-loop* complexes (so-called *rosettes*) that have now also been observed in computer simulations [70].

The nucleosome repositioning, mentioned above, is another single-nucleosome problem that has been investigated theoretically [75–77]. It is suggested that the nucleosome 'sliding' is based on the formation of loops at the ends of the nucleosome which then diffuse as defects around the spool, leading to the repositioning. This process somewhat resembles polymer reptation in the confining tube of the surrounding medium (a polymer network or melt), cf [78–80]. A different 'channel' for repositioning might be a corkscrew motion of the DNA helix around the octamer induced by twist defects which has now also been studied theoretically [77].

On the level of the 30 nm fibre, recent progress has been made in the visualization of these fibres via electron cryomicroscopy [23]. The micrographs reveal, for lower ionic strength, structures that resemble the crossed linker model mentioned above. However, for increasing ionic strength the fibres become so dense that their structure still remains obscure. An alternative approach was achieved via the micromanipulation of single 30 nm fibres [81–83]. The stretching of the fibres showed interesting mechanical properties, namely a very low stretching modulus for small tension, a force plateau around 5 pN and, at much higher tensions, sawtooth-type patterns.

These experimental results led to a revival of interest in 30 nm fibre models. Theoretical studies [24, 84, 85] as well as computer simulations [86, 87] attempted to explain the mechanical properties of the fibre. What all these models have in common is that they assume straight linkers in accordance with experimental observations, at least for lower ionic strength [23]. The low stretching modulus of the fibres is then attributed to the bending and twisting of the linker DNA which is induced by the externally applied tension. Indeed, all the models find relatively good agreement with the experimental data. The 5 pN force plateau is usually interpreted as a reversible condensation–decondensation transition of the fibre which can be attributed to a small attractive interaction between the nucleosomes [24, 86]. This is also in good agreement with recent studies on single nucleosome core particles where such a weak attraction has been observed and attributed to a tail-bridging effect [88, 89]. Finally, the sawtooth pattern which leads to a non-reversible lengthening of the fibre probably reflects the 'evaporation' of histones from the DNA [83, 90].

The geometry of the crossed-linker models suggest that the density of nucleosomes in the fibre depends to a large extend on the entry–exit angle of the DNA at the nucleosomes. It is known that in the presence of the linker histone the entering and exiting strands are forced together in a 'stem' region [23]. A recent study [91] investigated how the electrostatic repulsion between the two strands dictates this entry–exit angle. In particular, it was shown how this angle can be controlled *in vitro* via a change in the salt concentration. It also has been speculated that, via biochemical mechanisms that control the charges in the entry–exit region (like the acetylation of certain histone tails), the cell can locally induce a swelling of the fibre [17].

Before discussing in the following all the above mentioned issues concerning the physics of chromatin, I note that there are also important studies using a bottom-down approach by studying the physical properties of whole chromosomes that have been extracted from nuclei preparing for cell division. Via micropipette manipulation of these mitotic chromosomes it has been demonstrated that they are extremely deformable by an externally applied tension [92–94] and that a change in ionic strength induces a hypercondensation or decondensation of the chromosome [95]. Meiotic and mitotic chromosomes were compared to simple polymer systems like brushes and gels [96]. A problem with this bottom-down approach is still the lack of knowledge of the chromosome structure at this level and of the proteins that cause them. I will therefore dispense with a discussion on this subject.

This review is organized as follows. Section 2 gives a discussion of single nucleosome problems. After providing some experimental facts on the structure of the nucleosome (section 2.1), I discuss simple model systems (section 2.2), the unwrapping transition (section 2.3) and the nucleosome repositioning (section 2.4). Section 3 features the next level of folding, the 30 nm fibre. I briefly review some of the proposed models (section 3.1) and then give a systematic account of the crossed-linker model (sections 3.2 and 3.3) and its mechanical response to stretching, bending and twisting (section 3.4). Then fibre swelling (section 3.5) is discussed. Section 4 gives a conclusion and outlook.

2. Single nucleosome

2.1. Experimental facts on the core particle

The structure of the nucleosome core particle is known in exquisite detail from x-ray crystallography: the octamer in absence of the DNA was resolved at 3.1 Å resolution [97] and the crystal structure of the complete core particle at 2.8 Å resolution [10] and recently at 1.9 Å [11]. The octamer is composed of two molecules each of the four core histone proteins H2A, H2B, H3 and H4. Each histone protein has in common a similar central domain composed of three α-helices connected by two loops. These central domains associate pairwise in the form of a characteristic 'handshake' motif which leads to crescent-shaped heterodimers, namely H3 with H4 and H2A with H2B[1]. These four 'histone-fold' dimers are put together in such a way that they form a cylinder with $\sim$65 Å diameter and $\sim$60 Å height. With grooves, ridges and relatively specific binding sites they define the wrapping path of the DNA, a left-handed helical ramp of 1 and 3/4 turns, 147 bp length and a $\sim$28 Å pitch. In fact, the dimers themselves are placed in the octamer in such a way that they follow this solenoid: they form a tetrapartite, left-handed superhelix, a spiral of the four heterodimers $(\text{H2A–H2B})^1$, $(\text{H3–H4})^1$, $(\text{H3–H4})^2$ and $(\text{H2A–H2B})^2$. This aggregate has a two-fold axis of symmetry (the dyad axis)

[1] At physiological conditions stable oligomeric aggregates of the core histones are the H3–H4 tetramer (an aggregate of two H3 and two H4 proteins) and the H2A–H2B dimer. The octamer is stable if it is associated with DNA or at higher ionic strengths.

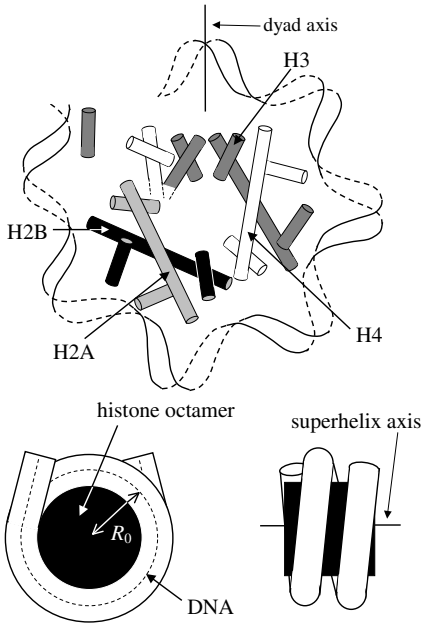

Figure 2. Schematic views of the nucleosome core particle. Top: upper half of the 8 core histones and the nucleosomal DNA. Bottom: a simplified model, where the octamer is replaced by a cylinder and the DNA by a WLC. Also indicated are the dyad axis and the DNA superhelix axis.

that goes through the $(H3–H4)_2$ tetramer apex and is perpendicular to the superhelix axis. A schematic view of the nucleosome core particle is given in figure 2.

There are 14 regions where the wrapped DNA contacts the octamer surface, documented in great detail in [10] and [11]. These regions are located where the minor grooves of the right-handed DNA double helix face inwards towards the surface of the octamer. Each crescent-shaped heterodimer has three contact points, two at its tips and one in the middle, altogether making up 12 of the 14 contacts. Furthermore, at each end of the wrapped section (the termini of the superhelix) there is a helical extension of the nearby H3 histone making contact with a minor groove. At each contact region there are several direct hydrogen bonds between histone proteins and the DNA sugar–phosphate backbone [10] as well as bridging water molecules [11]. Furthermore, there is also always a (cationic) arginine side chain extending into the DNA minor groove. The free energies of binding at each sticking point are different which can be concluded from the fact that for each binding site there is a different number of hydrogen bonds located at different positions; this is also reflected in the fluctuations of the DNA phosphate groups in the nucleosome crystal [10]. However, a reliable quantitative estimate of these energies is still missing.

An indirect method of estimating the adsorption energies at the sticking points is based on studies of competitive protein binding to nucleosomal DNA [27, 28, 98]. Many proteins are not able to bind to DNA when it is wrapped on the histone spool due to steric hindrance from the octamer surface. However, thermal fluctuations temporarily expose portions of the nucleosomal DNA via the unwrapping from either end of the superhelix. It was demonstrated that sites which are cut by certain restriction enzymes showed—compared to naked DNA—an increased resistance to digestion by these enzymes when they are associated with the octamer. Furthermore, the further these sites are from the termini of the superhelix the less frequently

they become exposed for cutting. The rate for digestion is reduced by roughly a factor of 10^{-2} for sites at the superhelical termini and $\sim 10^{-4}$–10^{-5} for sites close to the centre of the nucleosomal DNA portion (i.e., close to the dyad axis). From these findings one can estimate that the adsorption energy per sticking point is of order $\sim$1.5–2 k_BT.

It is important to note that the 1.5–2 k_BT does not represent the *pure* adsorption energy but instead the *net* gain in energy which is left after the DNA has been bent around the octamer to make contact with the sticking point. A rough estimate of this deformation energy can be given by describing the DNA as a semiflexible chain with a persistence length l_P of $\sim$500 Å [99]. Then the elastic energy [100] required to bend the 127 bp of DNA around the octamer (10 bp at each terminus are essentially straight [10]) is given by

$$E_{elastic}/k_BT = l_P l/2R_0^2. \tag{1}$$

Here l is the bent part of the wrapped DNA ($\sim 127 \times 3.4$ Å $= 431$ Å) and R_0 is the radius of curvature of the centreline of the wrapped DNA (cf figure 2) which is roughly 43 Å [10]. Hence, the bending energy is of order 58 k_BT. This number, however, has to be accepted with caution. First of all it is not clear if equation (1) is still a good approximation for such strong curvatures. Then it is known that the DNA does not bend uniformly around the octamer [10] and also the DNA might show modified elastic properties due to its contacts with the octamer. Nevertheless, using this number one is led to the conclusion that the bending energy per ten bp, i.e., per sticking site, is of order $60\ k_BT/14 \sim 4\ k_BT$.

Together with the observation that the net gain per sticking point is $\sim 2\ k_BT$ this means that the pure adsorption energy is on average $\sim 6\ k_BT$ per binding site. Note that the huge pure adsorption energy of $\sim 6\ k_BT \times 14 \sim 85\ k_BT$ per nucleosome is cancelled to a large extent by the $\sim 58\ k_BT$ from the DNA bending, a fact that has important consequences for the unwrapping transition discussed in section 2.3 and in particular, for the nucleosome repositioning reviewed in section 2.4.

Of great importance are also flexible, irregular tail regions of the core histones which make up $\sim$28% of their sequences [41]. Each histone has a highly positively charged, flexible tail (which is the N-end of the polypeptide chain [1]) that extends out of the nucleosome structure. Some of them exit between the two turns of the wrapped DNA, others on the top or bottom of the octameric cylinder. These N-tails are extremely basic due to a large amount of lysine and arginine aminoacids (aas). They are sites of post-translational modification and are crucial for chromatin regulation. In particular, the tails have a strong influence on the structure of the 30 nm chromatin fibre, as I will discuss in more detail in section 3. X-ray scattering data on core particles [88] suggest that the tails are adsorbed on the complex for small ionic strengths and extended at high salt concentrations (cf figure 8 in [88]). If the octamer is associated with a longer DNA piece the N-tails desorb at higher ionic strength.

Finally, let me mention the amount of charge found on the nucleosome core particle [8]. The histone octamer contains 220 basic side chains (arginine and lysine). From these are about 103 are located in the flexible histone tails mentioned above. The rest, 117 residues, are in the globular part of the octamer, of which 31 are exposed to the solvent, the rest being involved in intra- and interprotein ionic interactions. On the other hand, one has 147 bp of DNA wrapped around the octamer, each contributing two phosphate groups. Hence there are 294 negative charges from the DNA versus 220 positive charges of the octamer (or even less, 134 if the charges buried inside the octamer are not counted), i.e., the nucleosomal complex is overcharged by the DNA. At first sight this is a surprising fact and indeed it led to the development of simplified toy models containing charged spheres and oppositely charged chains which I will discuss in the following section.

2.2. Polyelectrolyte-charged sphere complexes as model systems for the nucleosome

The complexation of (semi-) flexible polyelectrolytes and oppositely charged macro-ions is an important ingredient in biological processes. For instance, the non-specific part of the interaction between DNA and proteins is governed by electrostatics [101]. In fact, the nucleosome is an example of this kind of interaction. A number of experimental [102–105] and theoretical [43, 106–109] studies have demonstrated that the complexation of *highly* charged macro-ions (e.g. DNA) is governed by an unusual electrostatics mechanism: *counterion release*. The free energy of complexation is then dominated by the entropy increase of the released counterions that had been condensed before complexation. This electrostatic contribution to the free energy has to compete with the energy cost of deforming one or both macromolecules to bring them in close contact.

In the next section I discuss how the counterion release leads to an overcharged sphere–chain complex which might be considered as a simplified model system of the nucleosome. I also consider the case of a polyelectrolyte chain placed in a solution of oppositely charged spheres (section 2.2.2). Both sections follow closely the treatments given in [43, 51]. Afterwards I review studies where it was found that overcharging also occurs in weakly charged systems due to 'standard' electrostatics (section 2.2.3). Section 2.2.4 is devoted to physiological conditions (strong screening) and also to the question of whether such toy models can 'explain' the large net charge of nucleosomes.

2.2.1. Single-sphere complex (highly charged case). Consider a single sphere of radius R_0 with its charge eZ homogeneously smeared out over the surface (the 'octamer') and a flexible rod with a charge per unit length $-e/b$, persistence length l_P, contour length $L \gg R_0$ and radius r (the 'DNA chain'). Both are placed in a salt solution characterized by a Bjerrum length $l_B \equiv e^2/\epsilon k_B T$ (ϵ: dielectric constant of the solvent; in water $\epsilon = 80$ and $l_B = 7$ Å at room temperature) and a Debye screening length $\kappa^{-1} = (8\pi c_s l_B)^{-1/2}$. Furthermore the solvent is treated as a continuum. Clearly, such a model neglects all the intricate features of the nucleosome discussed in section 2.1. It is, however, indispensable to start from such a simple model system to identify general features that occur in polyelectrolyte–macro-ion complexes.

I will focus in this section on salt concentrations c_s that are sufficiently small such that κ^{-1} is large compared to the sphere radius, $\kappa R_0 \ll 1$. The persistence length is assumed to be large compared with R_0. The chain is here highly charged which means that the so-called *Manning parameter* $\xi \equiv l_B/b$ is required to be much larger than one. In this case $(1-\xi^{-1})L/b \simeq L/b$ counterions are condensed on the chain reducing the effective line charge density to the value $-e/l_B$ [110, 111]. The entropic cost to 'confine' a counterion close to the chain is $\Omega k_B T$ with $\Omega = 2\ln(4\xi\kappa^{-1}/r)$ [51, 112]. This leads to the following entropic electrostatic charging free energy of the isolated chain in the salt solution:

$$\frac{F_{chain}(L)}{k_B T} \simeq \frac{\Omega L}{b} \tag{2}$$

On the other hand the corresponding electrostatic charging free energy of the spherical macro-ion of charge Z is given by

$$\frac{F_{sphere}(Z)}{k_B T} \simeq \begin{cases} \dfrac{l_B Z^2}{2R_0} & \text{for } |Z| < Z_{max} \\ |Z|\tilde{\Omega}(Z) & \text{for } |Z| \gg Z_{max} \end{cases} \tag{3}$$

where $\tilde{\Omega}(Z) = 2\ln(|Z|l_B\kappa^{-1}/R_0^2)$ [110] and $Z_{max} \approx \tilde{\Omega}R_0/l_B$ (see below). For weakly charged spheres ($|Z| < Z_{max}$) F_{sphere} is the usual electrostatic charging energy. In the highly

charged case, $|Z| \gg Z_{max}$, most of the counterions are localized close to the sphere with an entropic cost $\tilde{\Omega}(Z)k_BT$ per counterion leading to equation (3) for $|Z| \gg Z_{max}$. Only the small fraction Z_{max}/Z of counterions is still free, leading to an effective sphere charge Z_{max}. The value of Z_{max} follows from the balance of electrostatic charging energy $l_B Z_{max}^2/2R_0$ and counterion entropy $-\tilde{\Omega}(Z)Z_{max}$ [113].

The total free energy of the sphere–chain complex can be determined as follows. Assume that a length l of the chain has been wrapped around the sphere. Divide the sphere–chain complex in two parts: the sphere with the wrapped part of length l of the chain and the remaining chain of length $L - l$. The first part, which I will refer to as the 'complex', carries a net charge $Z(l) = Z - l/b$. The electrostatic free energy $F_{compl}(l)$ of the complex is then estimated to be equal to $F_{sphere}(Z(l))$ (neglecting higher-order multipole contributions). There is a special length $l_{iso} = bZ$, the *isoelectric wrapping length*, at which $Z(l_{iso}) = 0$. The usual principle of charge neutrality would lead one to expect that the total free energy is minimized at this point.

The total free energy of the sphere–chain complex is approximately given by the following terms [51]:

$$F_1(l) = F_{compl}(l) + F_{chain}(L - l) + F_{compl-chain}(l) + E_{elastic}(l). \tag{4}$$

The first two terms have already been discussed. The third term is the electrostatic free energy of the interaction between the complex and the remainder of the chain which is of the order

$$\frac{F_{compl-chain}(l)}{k_BT} \simeq Z^*(l)\ln(\kappa R_0) \tag{5}$$

where $Z^*(l)$ is the *effective* charge of the complex (the smaller value of $Z(l)$ and Z_{max}). The final term in equation (4) describes the elastic energy of the wrapped portion of the chain that has a typical curvature $1/R_0$ and is given already above in equation (1).

Following [51] the two cases $|Z(l)| < Z_{max}$ and $|Z(l)| > Z_{max}$ have to be treated separately. The first case applies for wrapping lengths l between $l_{min} = l_{iso} - bZ_{max}$ and $l_{max} = l_{iso} + bZ_{max}$. The free energy (4) takes then the form

$$\frac{F_1(l)}{k_BT} \simeq \frac{l_B}{2R_0}\left(Z - \frac{l}{b}\right)^2 + \frac{Cl}{b} + \text{constant} \tag{6}$$

where $C = l_P b/2R_0^2 - \ln(\kappa R_0) - \Omega$. For the second case, when $|Z(l)| > Z_{max}$, one finds

$$\frac{F_1(l)}{k_BT} \simeq \frac{B^{\mp}l}{b} + \text{constant} \tag{7}$$

with $B^{\mp} = l_P b/2R_0^2 - \Omega \mp \tilde{\Omega}$. The '$-$' sign refers to the case $l < l_{min}$ (i.e. $Z(l) > Z_{max}$) when for every segment b of adsorbed length a negative counterion of the sphere and a positive counterion of the chain are released while the '+' sign refers to the case $l > l_{max}$ (equivalently, $Z(l) < -Z_{max}$) when for every adsorbed segment a positive counterion is transferred from the chain to the sphere leading to a change $k_B(\Omega - \tilde{\Omega})$ of its entropy. The three different cases are depicted in figure 3.

Using equations (6) and (7) one can describe the complexation as a function of chain stiffness. For large l_P, $B^- > 0$ and there is no wrapping; the free energy is minimized for $l = 0$. There is, however, still the possibility of more open complexes with many point contacts between the sphere and the chain which we shall discuss in section 2.3. As l_P is reduced, B^- changes sign which marks the onset of wrapping. For $B^- < 0$ and $C > 0$ the minimum of $F_1(l)$ lies inbetween l_{min} and l_{iso}. According to equation (6) the position of the free energy minimum l^* is given by

$$l^* = l_{iso} - CR_0/\xi \tag{8}$$

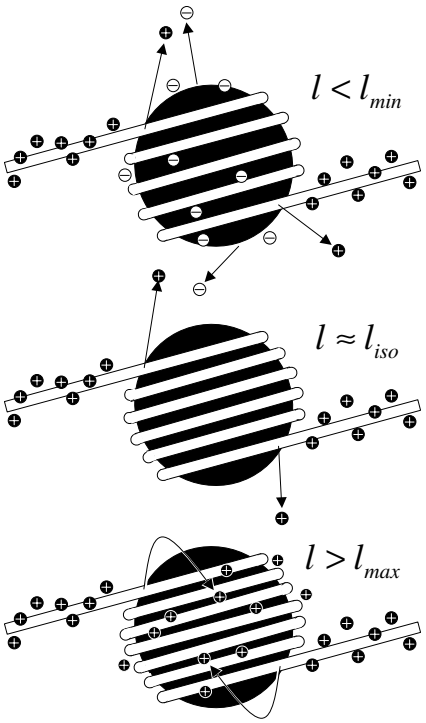

Figure 3. Schematic view of the complex between a highly charged chain and an oppositely charged sphere. Depicted are three scenarios. From top to bottom: for short wrapping lengths complexation is driven by release of counterions from sphere and chain; for intermediate values (around the isoelectric wrapping length) all counterions of the sphere have been released, further complexation still leads, however, to release of counterions from the chain; for even larger wrapping lengths there is no further counterion release.

a result first given by Park *et al* [43]. Further reduction of l_P leads to increasing l^* until the complex reaches the isoelectric point at $C = 0$. For smaller l_P, $C < 0$ and according to equation (8) $l^* > l_{iso}$ so that the complex is overcharged. Consequently, for a fully flexible chain with $l_P = 0$, the complex is *always* overcharged. The critical persistence length below which complexes are overcharged is $l_P = 2(\Omega + \ln(\kappa R_0))R_0^2/b$.

It can be clearly seen from this line of argument that it is the release of counterions from the chain that drives the overcharging. What opposes this effect is the charging energy of the complex, the repulsion between the chain and the overcharged complex and, most importantly, the bending stiffness of the chain.

2.2.2. Multi-sphere complex (highly charged case). In the previous section it was discussed how the release of counterions from the chain upon adsorption causes overcharging. Even though the focus of this chapter is on single-nucleosome properties it is instructive to consider next the case of a highly charged chain placed in a solution of oppositely, highly charged spherical macro-ions. This case has been investigated by myself, Bruinsma and Gelbart [51]. The solution is represented as a reservoir with concentration c_m of uncomplexed spheres. The chemical potential is the sum of the usual ideal solution term and the electrostatic free energy

of a spherical macro-ion with charge $Z \gg Z_{max}$ (cf equation (3)):

$$\frac{\mu_{sphere}}{k_B T} = \ln(c_m R_0^3) + Z\tilde{\Omega} \simeq Z\tilde{\Omega}. \tag{9}$$

The number of spheres that complex with the chain are determined by requiring this chemical potential to equal that of the complexed spheres. In [51] we assumed a beads-on-a-string configuration, with a mean spacing D between spheres. The Euclidean distance S between the beginning and the end of this configuration is related to the number of complexed spheres via $N = S/D$. The wrapping length l per sphere follows from S and L to be $L \simeq Nl + S - 2NR_0$. The Gibbs free energy of the bead-on-a-string configuration is given by

$$G(N, l) = NF_1(l) + F_{int}(N, l) - \mu_{sphere}N \tag{10}$$

where F_1 is the above given single-sphere complex free energy, equation (4), and F_{int} is the interaction between the complexed spheres. For sphere–sphere spacings $D(N, l) = S(N, l)/N$ with $2R_0 < D < \kappa^{-1}$, the complexed spheres feel a mutual electrostatic repulsion approximately given by (for $|Z(l)| < Z_{max}$)

$$\frac{F_{int}(N, l)}{k_B T} \simeq \Lambda N \frac{l_B Z^2(l)}{D(N, l)} \tag{11}$$

with $\Lambda \simeq 2\ln(\kappa^{-1}N/L)$. For $D \approx 2R_0$ adjacent spheres interact via a strong excluded volume interaction; for $D \gg \kappa^{-1}$ the electrostatic interaction is screened.

$G(N, l)$ has to be minimized with respect to both N and l. We showed in [51] that due to the large chemical potential of the spheres (last term in equation (10)) it is energetically favourable to keep adding spheres to the chain up to the point when $D \approx 2R_0$. At this point the hard-core repulsion terminates complexation and the chain is completely 'decorated' with spheres. It follows then that the number N of spheres and the wrapping length l per sphere are related via $N \simeq L/l$, i.e., essentially the whole chain is in the wrapped state. This argument holds for any $l_{min} < l < l_{max}$. Because of this relation (namely $N = L/l$) the Gibbs free energy depends only on the number N of complexed spheres:

$$\frac{G(N)}{k_B T} \simeq N\left\{(\Lambda + 1)\frac{l_B}{2R_0 b^2}(l_{iso} - L/N)^2 - \frac{\mu_{sphere}}{k_B T}\right\} + \text{constant}. \tag{12}$$

Clearly, the first term of equation (12) favours the isoelectric configuration $L/N = l_{iso}$. However, because of the second term, we can lower the free energy further by increasing N beyond L/l_{iso}. This is not a small effect since $\mu_{sphere}/k_B T$ is of order $Z \gg Z_{max}$ while the first term of equation (12), the capacitive energy, is of order $(l_B/R_0)Z_{max}^2 \approx Z_{max}$ (since $Z_{max} \approx R_0/l_B$). The spheres in the many-sphere complex are thus *undercharged*. The optimal wrapping length is as follows

$$l \simeq l_{iso}\left(1 - \frac{\tilde{\Omega}}{\Lambda + 1}\frac{R_0/l_B}{Z}\right). \tag{13}$$

Physically, this effect can be illustrated by first setting $L/N = l_{iso}$. In this case the complex is isoelectric. Now add one more sphere. By equally redistributing the chain length between the $N+1$ spheres, one has an individual wrapping length $l = L/(N+1)$ close to the isoelectric one. Therefore the previously condensed counterions of the added sphere are released and increase their entropy. By adding more and more spheres, while reducing $l = L/N$, more and more counterions are liberated.

In both cases, the single-sphere case of the previous section and the multi-sphere case discussed here, it is the counterion release that is the driving force which brings oppositely highly charged macro-ions together. In the first case the release of the cations of the chain is

responsible for bringing more monomers to the complex than necessary for its neutralization; in the second case it is the release of the anions of the spheres that attracts more spheres to the chain than 'optimal' and the spheres are undercharged. Comparing the similarities between the two cases it might be more appropriate to say that in the latter case the spheres overcharge the chain.

Finally, there is also the possibility of having a solution of chains and spheres in a certain stoichiometric ratio such that there are $N < L/l^*$ (with l^* being the single-sphere wrapping length, equation (8)). Then essentially all spheres will complex with the chains (due to the large contribution from the counterions to the chemical potential, equation (9)). Since in this case there is enough chain available, each sphere will be overcharged by the chain. Indeed, using equation (10) with $\mu_{sphere} = 0$ we found in [51]

$$l^* \simeq l_{iso} - C\frac{R_0}{\xi}\left(1 - \frac{2\Lambda R_0 N}{L}\right) \tag{14}$$

as the optimal wrapping length. This is the single chain wrapping length, equation (8), with a slightly reduced deviation from the isoelectric point due to the electrostatic repulsion between the complexed spheres, equation (11).

2.2.3. Weakly charged case. The first theoretical models on sphere–chain complexation were presented in 1999, each of which used quite a different approach to this problem [43–46]. Park *et al* [43] considered a semiflexible and highly charged chain and showed that counterion release leads to overcharging, as discussed in section 2.2.1, cf equation (8). The other studies [44–46] considered weakly charged chains and arrived at the conclusion that in this case overcharging should also be a common phenomenon.

Gurovitch and Sens [45] studied a point-like central charge (the 'sphere') and a connected chain of charges (the flexible chain). On the basis of a variational approach (self-consistent field theory using an analogy to quantum theory [78]) they came to the conclusion that the chain collapses on the central charge even if the total charge of the resulting complex becomes 'overcharged'. The critical polymer charge up to which this collapse occurs is 15/6 times the central charge. In a subsequent discussion [114] it became clear that this number has to be accepted with caution and that other effects, like the formation of tails, loops, etc were not included in the class of trial function which were used in that study.

Mateescu *et al* [44] used a purely geometrical approach in order to calculate the zero-temperature state of a complex of a sphere and a perfectly flexible chain in the absence of any small ions (no salt, no counterions). They divided the chain into two regions, one straight tail (or two tails on opposite sites of the complex) and a spherical shell around the macro-ion. The only approximation in that study was to uniformly smear out the monomer charges within the spherical shell. Starting from a point-like sphere and then gradually increasing its radius R_0, they found the following typical scenario (cf figure 1 in that paper). For very small R_0 the complex is slightly undercharged and shows two tails. With increasing R_0 more and more chain wraps around the sphere leading to an overcharging of the complex. For sufficiently large sphere radius the whole chain is adsorbed. Before this point is reached there are, for sufficiently long chains, two jump-like transitions: one from the two-tail to the one-tail configuration and then one from the one-tail case to the completely wrapped state.

Finally, Netz and Joanny [46] considered the complexation between a semiflexible chain and a sphere. For simplicity, they considered a two-dimensional geometry and calculated, using a perturbative approach, the length of the wrapped section and the shape of the two tails for different salt concentrations. That study focuses on the wrapping transition and its discussion (together with that of subsequent studies, [47, 54]) will be relegated to section 2.3.

The four studies mentioned above agreed in the respect that overcharging should be a robust phenomenon occurring in these systems, but there was still a transparent argument missing that would clarify the nature of the underlying mechanism that leads to overcharging. Nguyen and Shklovskii [48] bridged this gap by showing that correlations between the charged monomers induced by the repulsion between the turns of the wrapped chain can be considered as the basis for this effect. They studied again a fully flexible chain and neglected the entropy of the chain configurations. The chain is assumed to be in the one-tail configuration with the tail radially extending from the sphere. Then, as it is the case in [44], the energy of the chain-sphere complex is completely given by the electrostatic interactions between the different parts:

$$\frac{E}{k_BT} \simeq \frac{l_B(l-l_{iso})^2}{2R_0b^2} + \frac{l_Bl}{b^2}\ln\left(\frac{\Delta}{r}\right) + \frac{l_B(L-l)}{b^2}\ln\left(\frac{L-l}{r}\right) + \frac{l_B(l-l_{iso})}{b^2}\ln\left(\frac{L-l+R_0}{R_0}\right). \tag{15}$$

I use here the same symbols as in the previous sections (cf beginning of section 2.2.1; $l_{iso} = bZ$ denotes again the isoelectric wrapping length). The first term in equation (15) is the charging energy of the complex (the sphere plus the wrapped chain of length l), the second term is the self energy of the wrapped chain portion and will be discussed in detail below. The third term is the self-energy of the tail of length $L - l$, and the fourth term accounts for the interaction between the complex and the tail.

I discuss now the second term in equation (15) following the arguments given in [48]. The length Δ denotes the typical distance between neighbouring turns of the wrapped chain, i.e. $\Delta \approx R_0^2/l$. Consider an isoelectric complex, $l = l_{iso}$, and assume that $\Delta \ll R_0$ (multiple turns). Pick an arbitrary charged monomer on the wrapped chain. It 'feels' the presence of other neighbouring charged monomers up to a typical distance Δ beyond which the chain charges are screened by the oppositely charged background of the sphere. Hence the wrapped portion of the chain can be 'divided' into fractions of length Δ that behave essentially like rods of that length and of radius r having a self-energy $\sim l_B(\Delta/b^2)\ln(\Delta/r)$. One has l/Δ such portions leading indeed to $l_B(l/b^2)\ln(\Delta/r)$. Another interpretation of this term can be given as follows (again following [48]): one can consider the formation of the complex as a two-step process. First one brings in sections of length R_0 and places them on the sphere in random positions and orientations. This leads to a self-energy $\sim l_B(R_0/b^2)\ln(R_0/r)$ whereas the interaction of each segment with the random background charge can be neglected. Then, as the second step, one reorients and shifts these pieces on the ball in order to minimize their mutual electrostatic repulsion, i.e., one forms something like an equidistant coil with distance Δ between the turns. Now there is an additional contribution stemming from the attraction between each chain piece of length R_0 with a stripe on the sphere of length R_0 and width Δ leading to a gain in the electrostatic energy, scaling as $-l_B(R_0/b^2)\ln(R_0/\Delta)$. This contribution from all these R_0-sections (l/R_0 pieces leading to $-l_B(l/b^2)\ln(R_0/\Delta)$) constitutes the *correlation energy* of the wrapped chain. Together with the self-energy of these pieces (l/R_0 times $l_B(R_0/b^2)\ln(R_0/r)$) the correlation leads to $l_B(l/b^2)\ln(\Delta/r)$ which is indeed the second term of equation (15).

What is now the prediction of equation (15)? Minimization of E with respect to l leads to the following condition [48]

$$(l-l_{iso})\left(\frac{1}{R_0} - \frac{1}{L-l+R_0}\right) = \ln\left(\frac{l}{R_0}\frac{L-l}{L-l+R_0}\right) + 2 \simeq \ln\left(\frac{l_{iso}}{R_0}\right). \tag{16}$$

On the right-hand side the argument of the logarithm was simplified assuming $L - l \gg R_0$ (long tail), $l \gg R_0$ (many turns of the wrapped chain) and $l \approx l_{iso}$. Equation (16) can be interpreted as follows: the left-hand side describes the cost (if $l > l_{iso}$) of bringing in a chain segment from the tip of the tail to the surface, the simplified term on the right-hand side is the

gain in correlation energy (cf equation (15)). This leads to the following optimal wrapping length:

$$l^* \simeq l_{iso} + R_0 \ln\left(\frac{l_{iso}}{R_0}\right) \tag{17}$$

which demonstrates that the correlations induce indeed an overcharging of the complex.

Note that equation (17) gives the asymptotic value of l^* for long chains, $L \gg l^*$, the case where most of the monomers are located in the tail. As discussed in detail in [48] one encounters a discontinuous transition from the one-tail configuration to the completely collapsed state when one decreases L to such a value that the length of the tail is just of order R_0 (up to a logarithmic factor). This collapse is similar to the collapse discussed in [45] but occurs for much shorter chains (namely for chains of the length $l_{iso} + R_0 \ln(\cdots)$). The authors also considered the two-tail configuration and showed that it is formed, again in a discontinuous fashion, for very large chains of length $L > l_{iso}^2/R_0$. This might be, contrary to the claim in [48], in qualitative agreement with the prediction of [44] who found, for sufficiently large spheres, with increasing L two discontinuous jumps, from the collapsed to the one-tail configuration and from the one-tail to the two-tail state (cf figure 1 in that paper).

It is also worth mentioning that this theory can be easily extended to semiflexible chains. One has to add equation (1) to the free energy (15) and finds (following the steps that led to equation (17)) for the optimal wrapping length

$$l^* \simeq l_{iso} + R_0 \ln\left(\frac{l_{iso}}{R_0}\right) - \frac{l_P b^2}{2R_0 l_B} \tag{18}$$

i.e., the mechanical resistance of the chain against bending decreases the wrapped amount.

In [49] and [50] Nguyen and Shklovskii also considered the many-sphere case for weakly charged components. Similar to the case considered in the previous section they found in the case of an abundance of spheres undercharged complexes (a phenomenon which they call 'polyelectrolyte charge inversion' as opposed to 'sphere charge inversion'). It was shown that in this case the imbalance is mainly caused by the reduction of the self-energy of each complexed sphere (cf [49] for details).

Nguyen and Shklovskii argue that the correlation effect is the basis of all the phenomena discussed in section 2.2. The approximation given in [44], for instance, is to homogeneously smear out the charges of the wrapped chain and in this respect it overestimates the gain in electrostatic energy upon complexation, i.e., the second term in equation (15) is neglected. They call this approximation the 'metallization approach' [52], an approximation that obviously holds for sufficiently tight wrapping only. Also in that reference they argue that the overcharging via counterion release as discussed in [43] (cf section 2.2.1) is ultimately based on the correlation effect. Their argument is as follows: consider an isoelectric single-chain complex and assume that the chain is wrapped in a random fashion around the sphere. Then the electrical field close to the complex is essentially vanishing. If more chain was wrapped around the complex no counterions would be released. It is only the fact that the chain will be adsorbed in an orderly fashion due to its self-repulsion that each section is surrounded by a correlation hole that leads to counterion release even beyond the isoelectric point.

Sphere–chain complexes have also been considered in several computer simulations [59–71]. Wallin and Linse [59] studied the effect of chain flexibility on the geometry of a complex of a single sphere with a polyelectrolyte; we will come back to this problem in section 2.3. The same authors also varied the line charge density of the polyelectrolyte [60] and the radius of the sphere [61], they then considered the case when there are many chains present [62]. Chodanowski and Stoll [65] considered the complexation of a flexible chain on a sphere (assuming Debye–Hückel interaction) and found good agreement with [48] concerning

overcharging and the discontinuous transition to the one-tail configuration for longer chains. The case of multisphere adsorption was studied by Jonsson and Linse, having flexible [64] or semiflexible [69] chains and taking explicitly into account the counterions of the spheres and the chain. Their findings show the same qualitative features as discussed in this section and in section 2.2.2. A recent study by Akinchina and Linse [70] focused systematically on the role of chain flexibility on the structure of the complex; the results will be discussed below in section 2.3. Messina *et al* [67, 68] demonstrated that in the case of strong electrostatic coupling (large values of l_B) it is even possible that a polyelectrolyte chain forms a complex with a sphere that carries a charge of the *same* sign, a process which is made possible by correlation effects making use of the neutralizing counterions. Most recently Dzubiella *et al* [71] studied the polarizibility of overall neutral chain-sphere complexes in electrical fields as well as the interaction between two such complexes.

2.2.4. Physiological conditions. Up to here I have discussed only sphere–chain models for the case of weak screening, $\kappa R_0 < 1$. In physiological conditions, however, the screening length is roughly 10 nm and hence ten times smaller than the overall diameter of the nucleosome. This section is devoted to this case ($\kappa R_0 \gg 1$).

I will mainly focus here on weakly charged chains and spheres where the linear Debye–Hückel theory can be applied. For strong screening, $\kappa R_0 \gg 1$ the potential ϕ_{sphere} close to the ball looks essentially like that of a charged plane with charge density $Z/(4\pi R_0^2)$: $e\phi_{sphere}(h)/k_BT \simeq (l_BZ/2\kappa R_0^2)\mathrm{e}^{-\kappa h}$ (h: height above surface). Neighbouring turns of the adsorbed chain have locally the geometry of (weakly) charged rods for which it has been predicted that they form a lamellar phase [115, 116]. The lamellar spacing Δ follows from the competition between the chain–sphere attraction and the chain–chain repulsion. The chain–sphere attraction leads to the following adsorption energy per area:

$$\frac{f_{chain-sphere}}{k_BT} \simeq -\frac{l_BZ}{2\kappa R_0^2 b\Delta} \tag{19}$$

assuming that the chain is so thin that its adsorbed charged monomers feel an unscreened attraction to the surface, $\kappa r \ll 1$. To calculate the rod–rod repulsion one starts from the potential around a single rod: $e\phi_{chain}(R)/k_BT = -2l_Bb^{-1}K_0(\kappa R)$ (R: radial distance from rod axis); K_0 denotes the modified Bessel function that has the asymptotics $K_0(x) \simeq -\ln x$ for $x \ll 1$ and $K_0(x) \simeq (\pi/2x)^{1/2}\exp(-x)$ for $x \gg 1$. This leads to the following free energy density of the chain–chain repulsion:

$$\frac{f_{chain}}{k_BT} = \frac{2l_B}{b^2\Delta}\sum_{k=1}^{\infty} K_0(k\kappa\Delta). \tag{20}$$

To proceed further one might consider two limiting cases. If the lamellar spacing is much smaller than the screening length, $\kappa\Delta \ll 1$, the sum in equation (20) can be replaced by an integral [115, 116]:

$$\frac{f_{chain}}{k_BT} \approx \frac{2l_B}{b^2\Delta}\int_0^{\infty} K_0(k\kappa\Delta)\mathrm{d}k = \frac{\pi l_B}{b^2\kappa\Delta^2}. \tag{21}$$

The free energy density $f = f_{chain-sphere} + f_{chain}$ is then minimized for the *isoelectric* lamellar spacing: $\Delta = 4\pi R_0^2/(bZ)$ leading to the wrapping length $l \simeq 4\pi R_0^2/\Delta = bZ = l_{iso}$. Note, however, that going to the continuous limit means smearing out the charges, neglecting the correlation energy discussed after equation (15). This is similar to the approximation used by Mateescu *et al* [44] who studied the unscreened case (discussed above in section 2.2.3). There they considered, however, the self-energy of the tails, a contribution driving more

chain monomers to the sphere and hence leading to overcharging. This overcharging was overestimated since the correlation effects (included in the theory of Nguyen and Shklovskii [48], cf also equation (15)) were washed out. Here, on the other hand, for strong screening the self-energy of a chain section remains the same, whether it is adsorbed or not (on a length scale κ^{-1} it always looks straight). Hence it is here appropriate not to include the tail contribution.

This might lead one to expect that there is no overcharging for the case of strong screening. However, as mentioned above, there is an approximation involved when going from equation (20) to (21). This approximation is only good for $\Delta \simeq r$. A more careful calculation also leads here to the prediction of overcharging. To see this one has to realize that $\int K_0(k\kappa\Delta)\mathrm{d}k - \sum_k K_0(k\kappa\Delta) \approx \int_0^1 K_0(\kappa\Delta)\mathrm{d}k \simeq -\ln(\kappa\Delta)$. Taking this into account one can replace equation (21) by

$$\frac{f_{chain}}{k_B T} \simeq \frac{\pi l_B}{b^2\kappa\Delta^2}\left(1 + \frac{2}{\pi}\kappa\Delta\ln(\kappa\Delta)\right). \tag{22}$$

Minimizing the free energy density with this additional contribution (coming from correlation effects) leads to a slightly smaller lamellar spacing

$$\Delta \approx \frac{b^{-1}}{\frac{Z}{4\pi R_0^2} + \frac{\kappa}{\pi b}\ln\left(\frac{Zb}{4\pi R_0^2\kappa}\right)} \tag{23}$$

and to a wrapping length that is larger than the isoelectric one (overcharging):

$$l^* = \frac{4\pi R_0^2}{\Delta} \approx l_{iso} + 4R_0^2\kappa\ln\left(\frac{l_{iso}}{4\pi R_0^2\kappa}\right). \tag{24}$$

Lowering the ionic strength leads to smaller κ-values and hence to a reduction of the degree of overcharging. When $\kappa^{-1} \approx R_0$ one recovers equation (17), the result presented by Nguyen and Shklovskii [48] for the case of weak screening.

Netz and Joanny [46] also considered the case of spheres with an even smaller charge density $Z/4\pi R_0^2$ where $\Delta > \kappa^{-1}$. In that case only the interactions with the two next neighbouring turns count. From equation (20) follows

$$\frac{f_{chain}}{k_B T} = \frac{\sqrt{2\pi}l_B}{b^2\kappa^{1/2}\Delta^{3/2}}\mathrm{e}^{-\kappa\Delta}. \tag{25}$$

Minimizing $f = f_{chain-sphere} + f_{chain}$ leads then approximately to

$$\Delta \simeq \kappa^{-1}\left(\ln\left(\frac{R_0^2\kappa}{bZ}\right) + 1\right) \tag{26}$$

i.e., the spacing is of the order of the screening length (up to logarithmic corrections). The overcharging can then become very large. Clearly the term 'overcharging' becomes quite questionable when there is such a strong screening that charges in the complex interact only very locally over length scales of order $\kappa^{-1} \ll R_0$.

So far the bending energy was not accounted for, i.e., the chain was assumed to be perfectly flexible. Bending leads to an additional energy (per area): $f_{bend} \simeq l_P/(2R_0^2\Delta)$. This contribution scales with the lamellar spacing as $1/\Delta$ as does the chain–sphere attraction, equation (19). One can therefore interpret the bending to renormalize the sphere charge to a smaller value $\tilde{Z} = Z - l_P b\kappa/l_B$. In fact, l_{iso} in equation (24) has to be replaced by $l_{iso} - l_P b^2\kappa/l_B$.

When going to highly charged systems one encounters non-linear screening (counterion condensation). This case has been extensively discussed by Nguyen *et al* [117]. They showed

that the case $\Delta > \kappa^{-1}$ corresponds essentially to the above described case of strongly screened lamellas, equation (26), but with an effective rod line charge density $1/l_B$ (instead of the bare value $1/b$). There is nearly no counterion release in this case. The other limit $\Delta \ll \kappa^{-1}$ is highly non-linear and quite complicated. For details I refer the reader to [117] (section VI in that paper).

Concluding, what can be learned from the sphere–chain systems with regard to the nucleosome structure? It is clear that many of the assumptions entering the models do not agree with the details of the nucleosome conformation. Most of the studies discussed in sections 2.2.1–2.2.3 assume a weak screening which does not correspond to physiological conditions. Many studies neglect the bending energy which is a major energetic penalty for DNA wrapping around the octamer. Hydrogen bonds and solvent effects are not included. Most importantly, these studies do not account for the fact that the wrapping path is clearly prescribed by more or less specific binding patches on the octamer surface (cf section 2.1). And the octamer is, of course, not a sphere. Even when going to the strong screening case, discussed above, the assumed lamellar arrangement of the wrapped chain has only a vague resemblance to the one and 3/4 turns of wrapped nucleosomal DNA. Here one might at least say that the ∼28 Å pitch of the superhelical ramp is in rough agreement with the prediction $\Delta \approx \kappa^{-1}$ given in equation (26).

Nevertheless, the improved understanding of the chain–sphere complexes that has been achieved in recent years is in my opinion very helpful. Many of the investigated model systems resemble closely complexes between macro-ions (colloids [118, 119], dendrimers [120], charged micelles [121] etc) and synthetic polymers which is of technological relevance as a means of modifying macro-ion solution behaviour. The sphere–chain systems can also help us to understand the nucleosome behaviour under changing ionic conditions. Especially when the ionic strength is lowered the electrostatics becomes long range. In this case many local details become overruled by the electrostatics. The behaviour of the nucleosome at low and high ionic strength will be the subject of the next section.

2.3. Unwrapping transition

The nucleosomal complex is only stable at intermediate salt concentrations. In section 2.3.1 I will give an overview of experimental results that shed some light on the instabilities of the nucleosome core particle at low and high ionic strengths. I also report on how these instabilities can be understood in the framework of a sphere–chain model (for *short* chain length). Then I provide a thorough discussion of the instabilities of sphere–chain complexes at high ionic strength in section 2.3.2 and at low ionic strength in section 2.3.3—both for chains of *arbitrary* chain length.

2.3.1. Instabilities of the nucleosome core particle at low and at high ionic strength. Yager *et al* [72] characterized the stability of the nucleosome core particle as a function of the salt concentration (NaCl) and the concentration of core particles (measured via the 260 nm absorbance). Using a variety of experimental methods (velocity sedimentation, gel exclusion chromatography and gel electrophoresis) they arrived at the following main conclusions (cf also the schematic phase diagram, figure 4). For not too low concentrations core particles are stable for ionic strengths ranging from 2 to 750 mM (called region 1 in [72]; this includes physiological relevant salt concentrations ∼100 mM). For slightly higher salt concentrations (region 3 [72]) or low concentration of core particles (region 2) the DNA is partially dissociated; an equilibrium between histones, free DNA and core particles is observed. In particular, in region 3 there is also the occurrence of (H2A–H2B)-depleted particles. At salt concentrations

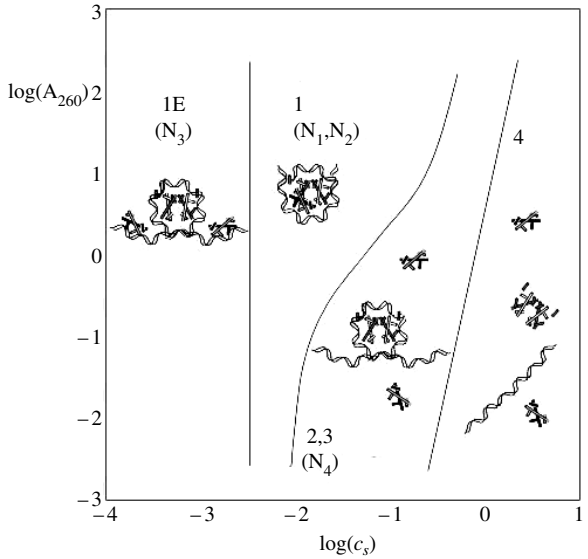

Figure 4. The effects of salt on the conformation and stability of nucleosome core particles as a function of the salt concentration c_s and the DNA concentration measured by 260 nm adsorbance (adapted from [72]). The notations of [72] for the different states are given as well as the ones of [8] (in brackets).

beyond 1.5 M the core particle is completely dissociated into histone oligomers (the (H2A–H2B) dimer and the $(H3–H4)_2$ tetramer) and free DNA (region 4). On the other hand, for very low salt concentration <1 mM (called region 1E in [72]) one finds an 'expanded' state of the nucleosome.

Using a different experimental approach (measurement of the fluorescence of the aa tyromisine in the histone proteins) Khrapunov *et al* [8] came to similar conclusions (cf figure 4). For ionic strengths between 5 and 600 mM the core particle is intact but one finds different degrees of contact between the core histones (the resulting two forms are called N_1 and N_2 in [8]). At larger ionic strength ($\approx$1.2 M) the terminal regions of the DNA unwrap and the (H2A–H2B) dimers are dissociated (N_4) and at an even larger value ($\approx$1.5 M) the $(H3–H4)_2$ tetramer leaves the DNA. Finally, at low salt concentrations one encounters an open state (called the N_3-form): the dimers break their contact with the tetramer and the DNA termini unwrap (with the dimers still attached to them).

The key features of the behaviour of core particle DNA (neglecting the substructure of the octamer) are indeed recovered in the framework of the sphere–chain models. Most clearly this has been demonstrated in a study by Kunze and Netz [47, 54] (cf also [46]). They considered the complexation of a charged, semiflexible chain with an oppositely charged sphere. All charges in their system interact via a standard Debye–Hückel potential. The optimal DNA configuration $\boldsymbol{r}(s)$ $(0 \leqslant s \leqslant L)$ follows then from the minimization of the free energy functional

$$\frac{\mathcal{F}\{\boldsymbol{r}(s)\}}{k_B T} = \frac{l_P}{2}\int_0^L \mathrm{d}s\left(\frac{1}{R(s)}\right)^2 - \frac{l_B Z}{b(1+\kappa R_0)}\int_0^L \frac{\mathrm{e}^{-\kappa(|\boldsymbol{r}(s)|-R_0)}}{|\boldsymbol{r}(s)|} + \frac{l_B}{b^2}\int_0^L \mathrm{d}s \int_s^L \mathrm{d}s' \frac{\mathrm{e}^{-\kappa(|\boldsymbol{r}(s)|-|\boldsymbol{r}(s')|)}}{|\boldsymbol{r}(s)-\boldsymbol{r}(s')|}. \tag{27}$$

The first term on the right-hand side is the bending energy of the chain where $R(s)$ denotes its curvature at point s along the contour. The second and third term account for

the electrostatic attraction between monomers and the sphere and the monomer–monomer repulsion, respectively. The symbols are the same as used above (cf section 2.2.1). Kunze and Netz chose the parameters (l_P, L, b and R_0) such as to mimic the values of the core particle and varied Z and κ^{-1} as 'free' parameters. The optimal shape $\boldsymbol{r}(s)$ was found by numerical minimization of equation (27) and characterized by two order parameters, a rotational and a torsional one. They found the following overall picture: for reasonable values of the sphere charge Z one finds for vanishing ionic strength ($\kappa^{-1} \to \infty$) an open, planar configuration where only a small fraction of the chain is wrapped whereas the two tails (of equal length) are extended into roughly opposite directions. This is reminiscent of the open structures reported in the experimental studies (region 1E in [72], N_3-form in [8]). Upon addition of salt the structure stays first extended but loses at some point its rotational geometry (in the form of a transition from a two- to a one-tail configuration) and then at even higher ionic strength the chain goes from a planar to a non-planar configuration. It begins then to wrap more and more and, at some point, regains its rotational symmetry (the wrapping path resembles then some kind of 'tennis-ball seam pattern'). At that point (which is well below physiological ionic strengths) the chain is already almost completely wrapped. It stays in this wrapped state up to very high salt concentrations. Only then does the chain unwrap in a discontinuous fashion because the chain–sphere attraction is sufficiently screened. Again these features of the complex (the wrapped compact state in a wide range around physiological conditions and the unwrapping at high salt content) reflect findings in [8] and [72]. The behaviour of such complexes for chains of *arbitrary* length is the subject of the next two sections.

2.3.2. The rosette state at high ionic strength. For physiological conditions (or for even higher salt concentrations) the electrostatic interaction between the DNA chain and the octamer can be considered as short-range. It is then usually sufficient to assume some attractive short-range interaction between the chain and the octamer with a range of interaction $\kappa^{-1} \ll R_0$. In this spirit Marky and Manning [73] considered the wrapping of a semiflexible chain around a cylinder. They came up with the picture of a simple 'all or nothing' unwrapping transition. Denote with λ the adsorption energy per length of the chain on the cylinder (in units of k_BT). Then for each additional wrapped length Δl, one gains $\Delta E_{ads}/k_BT = -\lambda \Delta l$. On the other hand, in order to wrap the chain it has to be bent with a curvature R_0^{-1} with R_0 being the radius of the octamer; this leads to an energetic cost $\Delta E_{elastic}/k_BT = l_P \Delta l/2R_0^2$, cf equation (1). From this follows that if the adhesion energy λ is larger than

$$\lambda_c = \frac{l_P}{2R_0^2} \tag{28}$$

then more and more chain will wrap around the histone spool (up to a point when there are no longer adsorption sites available). On the other hand, for $\lambda < \lambda_c$, the chain unwraps completely. This unwrapping transition can also be induced by increasing the persistence length of the chain, cf equation (28).

This, however, is not the complete picture. The numerical study by Wallin and Linse [59] has already indicated that with increasing chain stiffness one encounters a *gradual* change of the conformations of the complexed chain towards more extended structures. In [74] myself, Rudnick, Bruinsma and Gelbart showed in a systematic analytical study that there is indeed a wide range of parameters in which more open, multi-leafed ('rosette') states occur in this system. The results of this model study, together with some additional material, will be presented in this section. Following that reference I will first discuss the ground-state configurations of the sphere–chain complex and then account for thermal fluctuations. Then I will present the general phase diagram that includes wrapped ($\lambda > \lambda_c$) and open ($\lambda < \lambda_c$) structures.

In [74] we started from the popular worm-like chain model (WLC) which provides a good description of the mechanical properties of DNA (for reviews cf [122, 123]). The chain molecule is represented by a semiflexible tube of radius r characterized by two elastic moduli, the bending and torsional stiffnesses. The elastic energy of a WLC of length L can be expressed as

$$E_{elastic} = \frac{1}{2}\int_0^L \mathrm{d}s\left[A\left(\frac{1}{R(s)}\right)^2 + C\left(\frac{\mathrm{d}\Theta}{\mathrm{d}s}\right)^2\right]. \tag{29}$$

Here A is the bending stiffness and $1/R(s)$ the curvature of the chain at point s along its contour. The stiffness is usually expressed as $A = k_B T l_P$ where l_P is the orientational persistence length of the chain—as given above in section 2.1, cf equation (1). The torsional angle of the chain is Θ and the torsional stiffness is C, for DNA $C \simeq k_B T \times 750$ Å [99]. In addition to this bending contribution there is a short-range attraction between the chain and the sphere (or cylinder) with a range of interaction $\delta \ll r$.

As mentioned above, for strong attraction $\lambda > \lambda_c$ the chain is wrapped around the sphere. On the other hand, if $\lambda < \lambda_c$, the WLC can only make point contacts to the sphere. The energy of a contact point (in units of $k_B T$)

$$\mu \simeq \lambda\sqrt{R_0\delta} \tag{30}$$

follows from the length $\sqrt{R_0\delta}$ of chain portion around that point contact that is located within the distance δ from the sphere. In [55] I gave the quantities λ and δ in terms of strongly screened electrostatics. In that case $\delta = \kappa^{-1}$. The adsorption energy per length can be estimated from the Debye–Hückel electrostatic potential close to the surface (cf beginning of section 2.2.4). From equation (19) it follows that

$$\lambda = \frac{l_B Z}{2\kappa R_0^2 b} \tag{31}$$

and hence the unwrapping into the rosette takes place when λ reaches the critical value given by equation (28), i.e., when the persistence length

$$l_P = \frac{l_B Z}{\kappa b} \tag{32}$$

is reached. That this is the upper bound for l_P for having a stable wrapped complex has been predicted by Netz and Joanny [46] (cf equation (35) in that paper).

For $\lambda > \lambda_c$ the optimal number M^* of point contacts as well as the preferred configuration of the chain is obtained by a minimization of the energy

$$E = E_{elastic} - k_B T \mu M \tag{33}$$

with $E_{elastic}$ given by equation (29). In [55] we searched first for the minima of the elastic energy (zero temperature conformations), neglecting thermal fluctuations that were included in a second step (see below).

The search can be performed systematically by applying the *Kirchhoff analogy* [124, 125] which relates stationary points of the WLC energy, equation (29), to the well-studied classical mechanics problem of the trajectory of a supported, symmetric spinning top in a gravity field. It can be easily shown that the action of the spinning top has precisely the same form as equation (29) with time playing the role of the arc length, the orientation of the top corresponding to the tangent vector of the WLC, the gravitational force being a tension acting on the rod[2], etc. This analogy has been repeatedly applied to DNA related problems

[2] The tension comes here from the 'sticky' sphere that induces the rosette structure discussed below. Later, in section 2.4.2, I will give an example where it becomes more obvious how the rod tension formally enters the Hamiltonian in form of a Lagrange multiplier T, leading to a term that resembles the potential energy of the spinning top.

over the last 20 years (e.g. see [126–135]). For a nice visual review of the spinning top–elastic rod analogy the reader is referred to [125].

In [74] we presented solutions of the corresponding Euler–Lagrange equation that combine the following features:

(i) the WLC closes on itself[3],
(ii) it is possible to inscribe a sphere of radius R_0 inside the WLC that touches the WLC at M points,
(iii) there is no self-intersection of the WLC chain with itself if it is surrounded by a tube of radius r, and
(iv) the solution is stable against small perturbations.

The resulting rosette-type configurations can be characterized by the number of loops M. Figure 5 shows such a rosette (computed numerically); as indicated in the figure it is indeed possible to inscribe a sphere (or a cylinder) in the central hole of the rosette. For each M, we adjusted the linking number of the loop to minimize the elastic energy. By varying the degree of spatial distortion (characterized by the so-called 'writhe') a family of solutions was obtained, for given M, with different hole diameters. Solutions with the maximum amount of writhe have the smallest central hole diameter as well as the lowest elastic energy. The inset of figure 5 shows the elastic energy of a loop of length L, in dimensionless units, for the rosette state, as a function of the degree of writhe.

These solutions are actually saddle points of the WLC energy, i.e., there is a finite subset of infinitesimal distortions that lower the elastic energy of the WLC, the rest raises the energy or leaves it unchanged. The role of the sphere is to stabilize this saddlepoint and turn it into a real maximum (for a detailed discussion cf [133]). The energy of a minimum-hole rosette depends on the overall chain length L and the number of point contacts M (= number of leaves) as

$$\frac{E_{min}(M)}{k_B T} \simeq \frac{2\chi l_P M^2}{L} - \mu M \tag{34}$$

with $\chi = 7.02$ (see inset of figure 5). This result can be understood making use of results of the earlier work by Yamakawa and Stockmayer (YS) [136] who showed that a loop of length l, formed by imposing common endpoints on a WLC strand, assumes the form of a lemniscate-shaped leaf with an 81° apex angle. The elastic bending energy of a leaf is $e(l) = 2\chi A/l$ and $E_{min}(M)$ given above just equals $Me(L/M)$ plus the adhesion energy. We verified numerically that the leaves of the rosette indeed have apex angles close to 81°. The energy $E_{min}(M)$ exhibits a minimum as a function of the number M of rosette leaves for

$$M^* = \frac{\mu L}{2\chi l_P}. \tag{35}$$

We thereby obtained an open, multileafed structure controlled by the adhesion energy and the persistence length that competes with the wrapped state. Packing considerations imply an upper limit for the number of loops of order $M_{max} \approx (R_0/r)^{3/2}$; this is the maximal number of contacts, each excluding an area $\approx r\sqrt{R_0 r}$, that can be closely packed on the surface of the sphere.

Note that the above results do not depend critically on the assumption that the chain forms a closed loop. In fact, for a chain with open ends we expect the same number of rosette leaves, each again having approximately the shape of a YS-loop with an 81° apex angle. This means

[3] This assumption is only for technical reasons. As discussed below, the behaviour for an open chain is essentially identical. The solutions for closed WLC can be characterized in terms of the topologically conserved linking number. Loops with a non-zero linking number show a configuration that combines twist and spatial distortion (known as 'writhing').

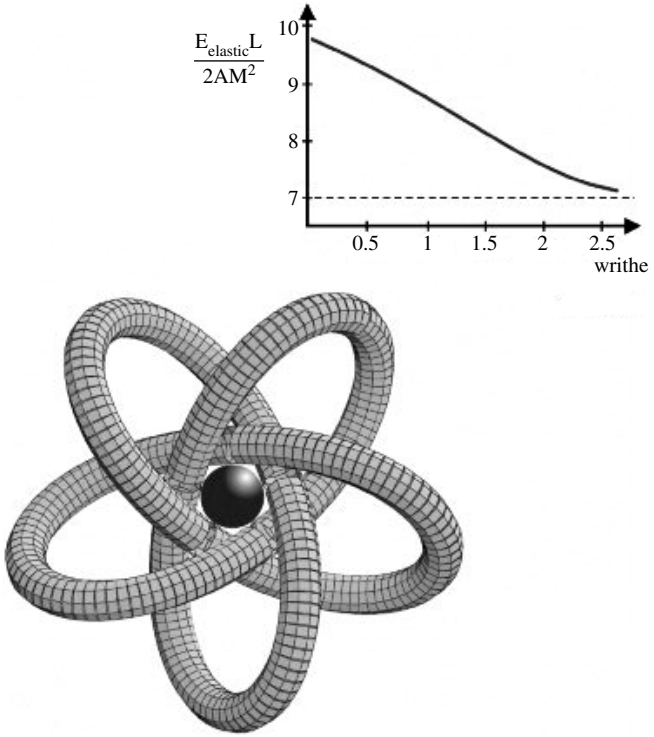

Figure 5. Five-leafed rosette. This configuration corresponds to the minimal energy solution of the Euler–Lagrange equation with a maximum amount of writhe and the smallest central hole (cf text for details).

that here neighbouring leaves will have a relative orientation of $\sim 180°–81° = 99°$. In addition, the leaves will be slightly twisted with respect to each other (like propeller blades) to account for the mutually excluded volume. I give an overview over the results in a 'phase diagram' (ground state configurations) in figure 6 (adapted from [55]; an alternative presentation with a different choice of axes is presented in [74]).

We next studied in [74] the stability of the rosette against *thermal fluctuations*. We started from a single, large loop of length L and constructed the rosette step by step, by attaching to the sphere lemniscate-shaped leaves of variable length of the kind examined by YS [136]. The finite-temperature free energy cost $F(l_{leaf}, \phi)$ of introducing a single leaf of length l_{leaf} and apex angle ϕ into a large strand was computed by YS. Using path-integral methods they found [136]

$$\frac{f(l_{leaf}, \phi)}{k_B T} \simeq \begin{cases} 2\chi \dfrac{l_P}{l_{leaf}} + \ln \dfrac{l_{leaf}}{l_P} + W(\phi) + \cdots & \text{for } l_{leaf} \ll l_P \\ \dfrac{3}{2} \ln \dfrac{l_{leaf}}{l_P} + \cdots & \text{for } l_{leaf} \gg l_P. \end{cases} \tag{36}$$

The function $W(\phi)$ has a minimum when the apex angle of the leaf is approximately 81°. The logarithmic contribution to $f(l_{leaf}, \phi)$, associated with the configurational entropy of the loop neglecting excluded volume interaction (θ solvent), imposes a free energy penalty for *large* leaves, $l_{leaf} \gg l_P$ (corresponding to the entropy of a closed random walk in three dimensions). The enthalpic $1/l_{leaf}$ contribution imposes an energy penalty for *small* leaves. Hence, for

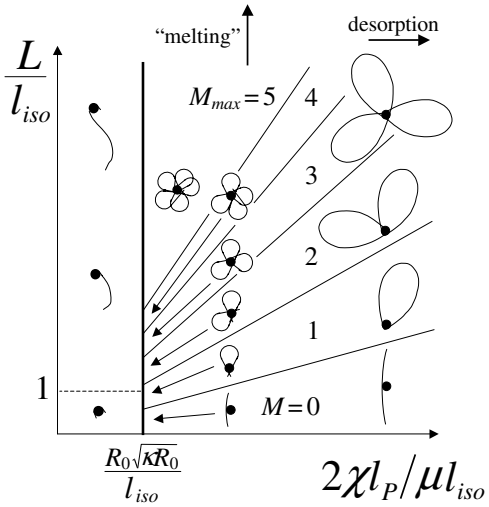

Figure 6. The sphere–chain complex in the case of short-range attraction (for instance, at high ionic strength). Depicted is the diagram of states as a function of the total length L of the chain and its persistence length l_P divided by the point contact energy μ (both axes in units of l_{iso}). The thick vertical line indicates the sharp unwrapping transition from the wrapped to the rosette-type complexes.

given ϕ, $f(l_{leaf}, \phi)$ as a function of l_{leaf} has a shallow minimum, its value being close to that of the persistence length l_P. The total free energy cost $F_M(\{l_i\})$ of introducing into a large loop an M-leafed rosette for a fixed distribution $\{l_i\}$ of leaf lengths is then given by

$$\frac{F_M(\{l_i\})}{k_BT} \simeq \sum_{i=1}^{M} \frac{f(l_i)}{k_BT} - \mu M. \tag{37}$$

Here

$$\frac{f(l)}{k_BT} = \frac{2\chi l_P}{l} + \frac{3}{2}\ln\left(\frac{l}{l_P}\right) \tag{38}$$

constitutes an interpolation formula between the large and small l limits of $f(l, 81°)$ as given by equation (36). The partition function Z_M for an M-leafed rosette follows by integration over all possible leaf distributions

$$Z_M = \int_0^\infty \prod_{i=1}^{M} \frac{dl_i}{r} \exp\left[-\frac{1}{k_BT}\left(\sum_{i=1}^{M} f(l_i) + P\sum_{i=1}^{M} l_i\right)\right] = (Z_1)^M. \tag{39}$$

Assuming $\mu \gg 1$ (strong sticking points) one can assume the number of leaves always to be maximal, $M = M_{max}$ (the index is dropped here and in the following for simplicity of notation). In equation (39) a Lagrangian multiplier P is introduced in order to satisfy the constraint $\sum_{i=1}^{M} l_i = L$. Physically, P is the overall tension of the loops induced by their adhesion to the sphere. The free energy is then

$$G(P) = -k_BT \ln Z_M = Mk_BT\left\{2\sqrt{\frac{2\chi l_P P}{k_BT}} + \ln\left(\frac{r/l_P}{\sqrt{\pi/2\chi}}\right)\right\}. \tag{40}$$

It is interesting to note that $G(P)$ is mathematically identical to the free energy of a one-dimensional many-body system of M particles under a 'pressure' P confined to a circular track

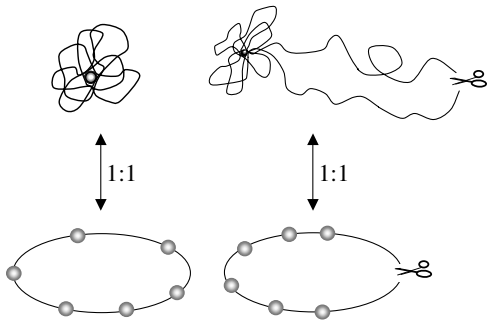

Figure 7. Molten rosettes of closed (left) and open chains (right). These structures are extremely fragile; cutting one loop leads to two long tails at the expense of the other loops. Also shown is the corresponding problem of particles on a one-dimensional closed (or open) track that interact via a nearest-neighbour pair potential.

of length L. The particles are interacting via a 'nearest-neighbour pair potential' $f(l)$ while $k_BT\mu$ is the 'chemical potential' of the particles. According to equation (38), the effective pair potential $f(l)$ is concave (i.e., $\mathrm{d}^2 f/\mathrm{d}l^2 < 0$) for interparticle spacings exceeding a spinodal threshold spacing of order l_P. Experience with mean-field theory of many-body systems suggests that one should expect phase decomposition if the average spacing between particles exceeds a spinodal threshold.

Using $L = \mathrm{d}G/\mathrm{d}P$ one finds the non-linear 'tension–extension' curve $P(L) = 2\chi l_P k_B T (M/L)^2$. Using equations (39) and (40), it is straightforward to compute the first and (reduced) second moments of the leaf size distribution:

$$\langle l_{leaf}\rangle = -k_BT\frac{\mathrm{d}}{\mathrm{d}P}\ln Z_1 = \frac{L}{M} \tag{41}$$

and

$$\frac{\sqrt{\langle(l_{leaf} - \langle l_{leaf}\rangle)^2\rangle}}{\langle l_{leaf}\rangle} = \frac{1}{2}\sqrt{\frac{L}{\chi M l_P}}. \tag{42}$$

Thus one encounters no phase-coexistence: the leaf size grows with chain length in the same manner as the '$T = 0$' solution. This does not mean, however, that the rosette structure is not altered when the mean leaf size exceeds l_P, because the reduced second moment starts to exceed one at that point (actually, at $\langle l_{leaf}\rangle$ exceeding $4\chi l_P$). One can therefore identify $L/M \approx l_P$ as the onset point of heterogeneity of the leaf size distribution; the orderly, symmetric rosette is starting to 'melt'. It must be emphasized though that $G(P)$ is analytic and that there is no true thermodynamic singularity.

This heterogeneous rosette state is very fragile. For instance, the loop size distribution will change drastically if one considers an *open* chain, i.e., if one allows for two free ends. Consider a chain of length L with M leaves of length l_i ($i = 1, \ldots, M$) and the two chain ends of length l_0 and l_{M+1}. One requires $\sum_{i=0}^{M+1} l_i = L$. The calculation of the partition function goes along similar lines as above. One has just to multiply Z_M in equation (39) with $\int_0^\infty \mathrm{d}l_0\,\mathrm{d}l_{M+1}\exp(-P(l_0 + l_{M+1})/k_BT)/r^2$ which accounts for the free ends. This leads to $G(P)$, equation (40), with an additional additive term $2k_BT\ln(rP/k_BT)$. The 'pressure' has to be chosen such that $L = \mathrm{d}G/\mathrm{d}P = M\sqrt{2\chi l_P k_B T/P} + 2k_BT/P$, i.e.

$$\sqrt{P} = \sqrt{\frac{2k_BT}{L} + \frac{\chi k_BT l_P M^2}{2L^2}} + \frac{M}{L}\sqrt{\frac{\chi k_BT l_P}{2}}. \tag{43}$$

The average loop size is now given by

$$\langle l_{leaf}\rangle = -k_BT\frac{\mathrm{d}}{\mathrm{d}P}\ln Z_1 = \sqrt{\frac{2\chi k_BTl_P}{P}} \simeq \begin{cases} \dfrac{L}{M} & \text{for } \frac{L/M}{l_P} \ll M \\ \sqrt{\chi l_P L} & \text{for } \frac{L/M}{l_P} \gg M. \end{cases} \tag{44}$$

The second moment of the loop size distribution obeys

$$\frac{\sqrt{\langle(l_{leaf}-\langle l_{leaf}\rangle)^2\rangle}}{\langle l_{leaf}\rangle} = \left(\frac{k_BT}{8\chi l_P P}\right)^{1/4} \simeq \begin{cases} \dfrac{1}{2}\sqrt{\dfrac{L/M}{\chi l_P}} & \text{for } \frac{L/M}{l_P} \ll M \\ \dfrac{1}{2}\left(\dfrac{L}{\chi l_P}\right)^{1/4} & \text{for } \frac{L/M}{l_P} \gg M. \end{cases} \tag{45}$$

Hence one recovers the result for the closed-loop case but only for $L/M \ll Ml_P$. In the opposite case, $L/M \gg Ml_P$, one finds a different scaling $\langle l_{leaf}\rangle \propto \sqrt{l_PL}$ (equation (44)) which shows that the mean leaf size is now small compared to L/M. Most of the chain is part of the two free ends that emerge from the rosette; the average length of a free end is $\langle l_0\rangle = k_BT/P$ which leads indeed to $L/2 - M\sqrt{\chi l_PL}/2$. A schematic view of molten rosettes formed from closed and open chains and the corresponding analogy of a 1D gas of particles is shown in figure 7.

In appendix A I present some new results that extend the above calculations to the case of rosette formation in d-dimensional space. This will shed some light on the nature of the 'phase coexistence' within molten rosette structures. In the next section 2.3.3 I shall show that rosette structures do also occur at low ionic strength. In that section I will also contrast the unwrapping transitions into the rosette at low and high ionic strength. Furthermore, I will speculate if rosette structures could occur in DNA–histone complexes.

2.3.3. The rosette state at low ionic strength. The rosette configuration discussed in the last section is a way of bringing at least a small fraction of the chain in close contact with the ball (in the form of point contacts). The majority of the monomers resides in the loops that do not 'feel' the presence of the sphere but are needed to connect the point contacts via small curvature sections. At first sight one might thus expect the rosettes to be a special feature of chain–sphere complexes with short-range attraction.

This is, however, not true. Rosettes are quite robust and also occur in systems with a much larger range of interaction. Recently this became clear in a Monte Carlo study by Akinchina and Linse [70]. They considered the complexation of a semiflexible charged chain with an oppositely charged ball that carries the same absolute charge as the chain (isoelectric complex). No small ions were present so that the charged monomers were attracted to the sphere via long-range $1/r$-interaction. The authors simulated systems with chains of different persistence lengths and linear charge densities as well as spheres with different radii. Depending on the choice of parameters they encountered a multitude of structures, ranging from collapsed structures with a 'tennis-ball seam pattern' or a solenoid arrangement of the wrapped chain [47, 54] to open multi-leafed structures very much resembling the ones discussed in the previous section. The rosette structures occur for stiffer chains on smaller spheres. The example configurations in [70] clearly show some rosette structures with one, two and three leaves (cf figure 1, system II in that paper). That these are representative example configurations can most clearly be seen in the adsorption probability of monomers as a function of the monomer index (figure 3 in [70]).

To better understand why rosettes also occur in the long-range case I developed a scaling theory for the system in [55] that I will outline in the following. Consider first sufficiently

short chains $L = bN \leqslant bZ = l_{iso}$ where the chain charge is smaller than (or equals) the sphere charge. The energy of the rosette with M leaves is then approximately given by

$$\frac{E_{rosette}}{k_BT} \simeq \frac{l_P}{L}M^2 - \frac{l_BZ}{b}M. \tag{46}$$

The first term is the bending energy of M leaves of length L/M and typical curvature $\sim M/L$. This has, of course, the same form as the elastic contribution to the energy of the rosette at strong screening, the first term of equation (34) (up to a numerical constant that I do not consider here). The second term is the attraction between the ball charge Z and the chain charge L/b over the typical distance L/M. Remarkably this term shows the same scaling with M as the second term of equation (34) that describes the energy of the point contacts! One has just to identify the point contact energy μ (for the strong screening case) with leaf-sphere attraction

$$\mu = \frac{l_BZ}{b} \tag{47}$$

for the unscreened case. The optimal leaf number is thus again (cf equation (35))

$$M^* \simeq \frac{\mu L}{l_P} \tag{48}$$

and the leaf size is

$$l_{leaf} \simeq \frac{L}{M^*} \simeq \frac{l_P}{\mu}. \tag{49}$$

The rosette state competes with the wrapped state that was already discussed in section 2.2. The rosette state is expected to transform *continuously* into the wrapped structure when $L/M^* \simeq R_0$; then the leaves become so small that they touch with their contour the surface of the sphere. Indeed, by setting $M = L/R_0$ in equation (46) one finds

$$\frac{E_{wrap}}{k_BT} \simeq \frac{l_PL}{R_0^2} - \frac{l_BZL}{bR_0} = \left(\frac{l_P}{R_0^2} - \frac{l_BZ}{bR_0}\right)L \tag{50}$$

which can be considered as the free energy of the wrapped state: the first term is the bending energy, equation (1), and the second accounts for the electrostatic chain–sphere attraction. All other electrostatic contributions (as written down in equation (15)) are much smaller and do not occur on this level of approximation.

On the right-hand side of equation (50) I arranged the terms in such way that one can deduce directly an unwrapping transition at $l_P/R_0^2 = l_BZ/bR_0$, i.e., at

$$l_P = \mu R_0. \tag{51}$$

At first sight one might expect that at that point the chain unwraps in a strongly discontinuous fashion similar to the cases discussed above (the highly charged case, equation (7), and the short-range case discussed at the beginning of section 2.3.2). However, this 'unwrapping point' corresponds just to the point $L/M^* \simeq R_0$ when loops form on the sphere. This leads to a smooth transition as pointed out before equation (50). That the unwrapping transition occurs rather smoothly at low ionic strength and sharply at high ionic strength has been predicted by Netz and Joanny [46] even though the authors did not allow in their study for rosette structures.

To complete the picture one has also to study chains that are longer than the isoelectric length, $L > l_{iso}$. Then at least three terms are needed to capture the essential physics of the rosette state:

$$\frac{E_{rosette}}{k_BT} \simeq \frac{l_P}{l}M^2 - \frac{l_BZ}{b}M + \frac{l_Bl}{b^2}M. \tag{52}$$

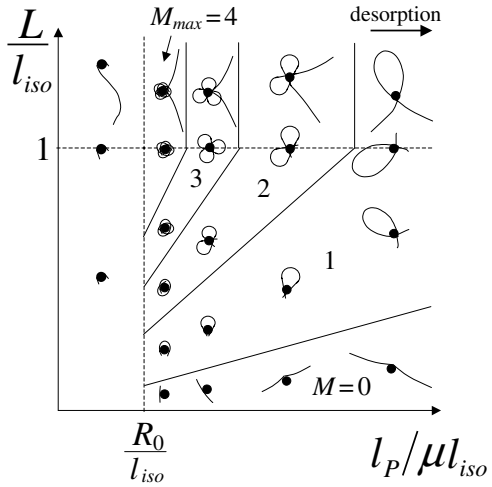

Figure 8. The sphere–chain complex at low ionic strength. The axes are chosen as in figure 6 with μ now being the leaf-sphere attraction, equation (47). The unwrapping transition is smooth in this case.

Here the monomers are 'allowed' to distribute between the rosette of length l and a tail of length $L - l$. The first two terms are the same as above, equation (46), the last term describes the self-repulsion of the monomers that constitute the rosette (additional logarithmic terms accounting for the self-energy of the tail and its interaction with the rosette, cf equation (15), are smaller and neglected here). Minimization with respect to l leads to the optimal rosette length

$$l^* \simeq b\sqrt{\frac{l_P M}{l_B}}. \tag{53}$$

The free energy (52) with the optimal wrapping length l^* is minimized for the following number of leaves:

$$M^* = \frac{\mu l_{iso}}{l_P}. \tag{54}$$

Hence, on this level of approximation, $l^* \simeq b\sqrt{l_P M^*/l_B} \simeq l_{iso}$, i.e., the rosette monomers just compensate the central ball charge; the remainder of the monomers extends away from the rosette in a tail of length $L - l_{iso}$. Each leaf is of size

$$l_{leaf} \simeq l_P/\mu. \tag{55}$$

The rosette disappears at $l_{iso}/M^* = R_0$, i.e., when equation (51) is fulfilled. It is then replaced by a wrapped chain of length l_{iso} (plus additional correction terms such as the one given in equation (18)) and a tail of length $L - l_{iso}$.

In figure 8 I depict the complete diagram of the sphere–stiff chain complexes to be expected in the long-range case. I again plot L versus $x = l_P/\mu$ (in units of l_{iso}), which is for rosettes of just the leaf size, equation (49). When one starts in this diagram at a large x-value and goes towards smaller values (with some arbitrarily fixed value $L < l_{iso}$) then all leaves shrink and more and more leaves can form. At $x = R_0$ the maximal number of leaves (for that given value of L) is reached and at the same time the leaves disappear simultaneously in a continuous fashion. For $x < R_0$ the chain wraps around the sphere. For $L > l_{iso}$ the excess charges are

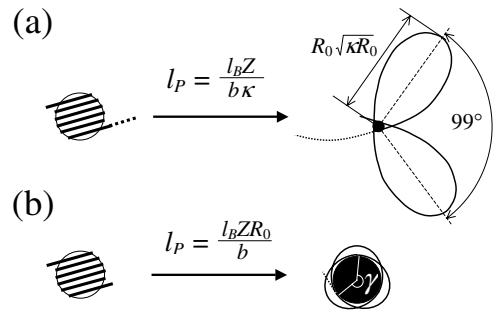

Figure 9. The unwrapping transition for (a) short-range and (b) long-range interaction. In the former case the chain unwraps discontinuously into large-leafed YS loops, in the latter case the transition is continuous.

accommodated in tails and all rosettes have the same length l_{iso}. The borderlines between different rosette ground states are then independent of the total length of the chain and thus appear as vertical lines. Desorption occurs when the free energy, equation (46) or (52), equals the thermal energy $k_B T$. This point is reached when

$$\frac{l_P}{\mu} \simeq \begin{cases} \sqrt{l_P L} & \text{for } L \leqslant l_{iso} \\ \sqrt{l_P l_{iso}} & \text{for } L > l_{iso}. \end{cases} \tag{56}$$

An arrow in figure 8 indicates the direction in which desorption takes place.

Before comparing these results with the Monte Carlo simulations by Akinchina and Linse [70] and with the properties of the nucleosomal complex, it is instructive to take a closer look at the *unwrapping transition* and to contrast the short- and the long-range case. The former case, $\kappa R_0 \gg 1$, is depicted in figure 9(a). As discussed in section 2.3.2 the unwrapping transition is expected to occur at $\lambda_c \simeq l_P/2R_0^2$ (equation (28)) which leads to equation (32). At this point the structure jumps in a strongly discontinuous fashion into a *large*-leafed rosette with leaves of size $l_{leaf} = L/M^*$. Using equations (30) and (35) one indeed finds

$$l_{leaf} \simeq R_0\sqrt{\kappa R_0} \gg R_0. \tag{57}$$

As discussed in the paragraph after equation (35), neighbouring leaves have a relative orientation of $\sim 99°$ which I also indicated in figure 9(a).

The unwrapping at low ionic strength is depicted in figure 9(b) and goes as follows. When the chain becomes so stiff that $l_P/R_0^2 > l_B Z/(bR_0)$ the wrapped state is no longer stable (cf equation (50)). At that time many small leaves ($M^* = L/R_0$ ones) form simultaneously in a continuous fashion. Their size scales as $l_{leaf} \simeq R_0$, the precise prefactor being not accessible to our scaling argument. The typical opening angle γ of the loop at the point of its formation scales as $(L/R_0)/M^* \approx 1$, again with an unknown numerical value. A multi-leafed configuration slightly above the unwrapping point is depicted in figure 9(b).

Additional insight can be gained by generalizing the attractive force between a given chain charge and the sphere by a power law $-AZ/r^\alpha$ with an arbitrary exponent $\alpha > 0$. An integer value $\alpha = d - 2$ with $d = 3, 4, \ldots$ can be interpreted as a charged chain that adsorbs onto a d-dimensional ball in d-dimensional space. The electrostatic term for the rosette in equation (46) then takes the form $-AZN^\alpha/(bL^{\alpha-1})$ and the one for the wrapped state scales as $-AZL/bR_0^\alpha$. Unwrapping takes place at $l_P^* \simeq AZ/(bR_0^{\alpha-2})$. At this critical value the energy of the rosette $E_{rosette}(l_P = l_P^*)$ has (as a function of M) a minimum at $M^* \simeq L/R_0$ for

$\alpha < 2$ ($d < 4$), suggesting a rather smooth unwrapping transition similar to the one depicted in figure 9(b). This minimum turns into a maximum at $\alpha = 2$ ($d = 4$). For larger values of α we find $M^* = 0$, i.e., the unwrapping transition is sharp, similar to the short-range case discussed in the previous section.

Let me also quickly come back to the highly charged case reviewed in section 2.2.1. In that case the dominant contribution to the complexation energy is the release of counterions which is a rather short-range interaction and consequently the unwrapping transition should be expected to be discontinuous, even at low ionic strength. To determine the unwrapping point one has to use equation (7) (with the '$-$' sign) from which it follows that unwrapping into the rosette occurs at $l_P \simeq (\Omega + \tilde{\Omega})R_0^2/b$. The point contact energy is of order $(\Omega + \tilde{\Omega})\sqrt{R_0\lambda_{GC}}/b$ where the so-called Guoy–Chapman length $\lambda_{GC} \simeq 1/(\sigma l_B)$ is the thickness of the layer of condensed counterions around the sphere ($\lambda_{GC} \ll R_0$ for strong counterion condensation [51]). The leaf size at the unwrapping point is then given by

$$l_{leaf} \simeq \sqrt{\frac{R_0}{\lambda_{GC}}}R_0 \gg R_0 \tag{58}$$

which indeed indicates a sharp unwrapping transition for highly charged systems.

We compare now the results of this section to the Monte Carlo simulations by Akinchina and Linse [70]. As mentioned above the simulated systems were always at the isoelectric point, i.e. $L = bZ = l_{iso}$. Furthermore there was no screening. The simulation results have thus to be compared with figure 8. Four systems have been considered, each having a fixed set of parameters b, Z and R_0 but with seven different values of l_P. This means that for each case the systems were located on the dashed horizontal line at $L = l_{iso}$ in figure 8. In one system (called system II in [70]) the continuous development from a wrapped to the rosette configuration has been seen very clearly. Example configurations are shown in figure 1, system II in [70]. For $l_P = 7$ Å the chain is wrapped, at $l_P = 60$ Å there is already a slight indication of very small loops ($N = 4$ or 5, cf the small oscillations in figure 3, system II, open squares). The next system depicted already has a much stiffer chain, $l_P = 250$ Å, and shows very clearly three leaves, then two leaves at $l_P = 500$ Å and one leaf for the stiffest chain, $l_P = 1000$ Å. In figure 8 I have chosen the parameters such that $N_{max} \simeq l_{iso}/R_0$ equals 4 so that this corresponds roughly to system II in [70]. To compare with the simulations one has to follow the $L = l_{iso}$-line in figure 8: one starts with wrapped structures for $x = l_P/\mu < R_0$ and finds then the continuous evolution of rosettes when the line $x = R_0$ is crossed. The leaves grow at the expense of their number (first 4, then 3, 2 leaves), just as has been observed in the simulations. The other three systems considered have different sets of parameters R_0 and b. The observed behaviour of these systems is also in good agreement with the theoretical predictions. For details I refer the reader to the last section of [55].

Finally let me speculate up to what extend the results of the previous and the current section apply to the nucleosome system under different ionic conditions. In particular, let me ask if rosette structures could in principle occur in DNA–histone complexes. First consider the core particle (147 bp DNA). As already discussed in section 2.2.4, at physiological salt concentrations ($\sim$100 mM) one has $\kappa^{-1} \simeq 10$ Å so the short-range case of the previous section applies. However, the estimate for λ from equation (31) is not reliable since $r \simeq 10$ Å (partial screening), since the binding sites between DNA and the histones are quite specific and since the linear Debye–Hückel theory is not reliable for such highly charged components. As mentioned in section 2.1 λ can instead be derived experimentally from competitive protein binding to nucleosomal DNA [27] to be of order 6 k_BT per sticking point, i.e., $\lambda \approx (1/5)$ Å^{-1}. This is *roughly* five times larger than that one would expect from equation (31) (but note also that contact is made mainly between DNA minor grooves which are 10 bp apart!). In any case,

equation (32) predicts an unwrapping for sufficiently small values of κ^{-1} but the numbers are not reliable. On the other hand, the unscreened long-range case of the current section applies when $c_s < 1$ mM ($\kappa^{-1} > 100$ Å). As mentioned in section 2.3.1, the completely wrapped configuration is no longer stable at this point. This is not surprising since the nucleosomal DNA overcharges the protein octamer by at least 74 negative charges (if not by 160 charges since 86 charged residues are inside the octamer, cf section 2.1). Hence it is expected that a considerable part of the terminal DNA unwraps and is part of one or two tails. In figure 8 this corresponds to the wrapped chain structures with tails that are found for small values of $l_P/\mu < R$ and large values of $L > l_{iso}$.

It would be interesting to redo the experiments [8, 72] mentioned in section 2.3.1 with complexes between histone octamers and DNA segments that are longer than 147 bp. For sufficiently large salt concentrations I expect the formation of DNA rosettes (if there is no interference with the partial disintegration of the octamer which could be avoided by introducing covalent linkages between the core histones). For the other limit (low salt) it might be appropriate to use the argument for highly charged chains and spheres as given before, equation (58). The linear charge density of DNA is very high (two phosphate groups per bp, i.e., 3.4 Å^{-1}). The Manning theory [111] indeed predicts that counterion condensation reduces the linear charge density to $-e/l_B$ with $l_B = 7$ Å. Also the charge of the histone octamer is so high that counterion condensation is important. It was argued above that the unwrapping occurs around $l_P = (\Omega + \tilde{\Omega})R_0^2/b$ (Ω and $\tilde{\Omega}$ are numbers of order (but larger than) one). However, since the DNA persistence length is so small that $l_P < R_0^2/b \approx (50\ \text{Å})^2/1.7\ \text{Å} \simeq 1500$ Å I expect the wrapped state to be stable at low ionic strength and there should be no formation of rosettes in this limit.

2.4. Nucleosome repositioning

It has been shown that nucleosome repositioning occurs spontaneously via thermal fluctuations (under certain conditions). This autonomous repositioning is the subject of this section. I first review the relevant experiments (section 2.4.1) and then discuss three theoretical models proposed to account for this effect: bulge diffusion (section 2.4.2), large loop repositioning (section 2.4.3) and twist diffusion (section 2.4.4). I contrast the three cases in section 2.4.5 and speculate that similar modes might be catalysed by remodelling complexes that use the energy of ATP hydrolysis. Nucleosome repositioning induced through transcription on short DNA segments is also briefly discussed in that section.

2.4.1. Experiments. An early study of uncatalysed nucleosome repositioning was presented by Beard [137]. He constructed 'chromatin–DNA-hybrids' where segments of radioactively labelled naked DNA were covalently joined to sections of chromatin fibres (derived from a simian virus). Remarkably the experiments suggested that nucleosomes spontaneously reposition themselves and invade the naked DNA. This was shown by several experimental methods:

(i) the occurrence of a radioactive component in the ∼175-bp band of a gel electrophoresis experiment that separated the products of a nuclease digestion,
(ii) electron micrographs showed an increased spacing of the nucleosomes on the chromatin pieces, and
(iii) binding of radioactive components of the recut hybrids on nitrocellulose filters (to which naked DNA would not bind under the given conditions).

All these methods indicated that something 'happens' on the timescale of hours in a 150 mM NaCl solution at elevated temperatures (37 °C). The methods used gave, however, not more quantitative information about the nucleosome migration rate. Furthermore, the conditions were not well-defined enough to exclude ATP-driven processes. Finally, the histone tails belonging to the virial DNA were strongly acetylated compared to the 'normal' chromatin of its host cell (which, as shown much later, does not have a big effect on nucleosome mobility, see below).

In that paper Beard also discussed three different possible modes of nucleosome repositioning:

(a) jumping where the octamer dissociates *completely* from the DNA and complexes at another position,
(b) sliding or rolling in which the octamer moves along the DNA without dissociation, and
(c) displacement transfer where a naked segment of the DNA displaces a nucleosomal DNA section and takes over.

Beard argues that two of the mechanisms (jumping and displacement transfer) imply that in the presence of competing DNA nucleosomes would be transferred from one DNA chain to another, a fact that was not observed in his experiment when chromosomes and radioactively labelled, naked DNA were mixed together. Nucleosomes could only migrate onto the naked DNA when the chromosome and DNA pieces were covalently joined. Beard concluded that 'sliding and rolling modes' should be responsible for octamer repositioning.

Spadafora *et al* [138] showed via gel electrophoresis that nucleosomal rearrangement occurs in fragments of rat liver chromatin when certain conditions are fulfilled: either one needs to go to high ionic strength above 600 mM NaCl or the fibre has to be depleted of the linker histone H1; in that case rearrangement also occurred around physiological conditions. It was observed that when either of these conditions is fulfilled *and* if temperatures are elevated (again 37 °C), then nucleosomes seem to move closer to each other (in the timescale of hours), away from the natural 200 bp repeat length towards a ∼140 bp repeat length. That length corresponds to a close packing of nucleosomes, a fact that the authors assigned to internucleosomal attraction. Furthermore, since H1 is dissociated from the nucleosome at around 600 mM salt content they concluded that one role of H1 is to prevent nucleosome mobility.

Similar conclusions were also drawn by Watkins and Smerdon [139] from corresponding experiments on human chromatin. An interesting additional experiment studied the exchange of histone proteins between different DNA chains. This was shown by mixing ^{14}C, labelled chromatin, and ^{3}H, labelled naked DNA. It became clear that at *high* ionic strength (600 mM) there is a formation of (probably intact) nucleosomes on the competing DNA whereas at physiological conditions this was not unambiguously detected. Moreover, the double-radioactive component showed the footprint of a ∼146 bp repeat length. This might indicate that a combination of histone protein transfer and subsequent nucleosome 'sliding' into tight packing took place.

An important series of experiments on nucleosome repositioning under rather well-defined conditions was performed by Pennings *et al* [29–31]. Even though the original focus of this study was to understand better the *positioning* of nucleosomes on special natural sequences that have a high affinity to octamers [140], the authors came up with elegant methods to monitor the nucleosome *repositioning*. It was found [29] that on tandem repeats of 5s rDNA positioning sequences (each of length 207 bp) nucleosomes assemble in one dominant position surrounded by minor positions multiples of 10 bp apart. The most interesting observation was that there is a dynamic redistribution between these positions. This was shown by cutting

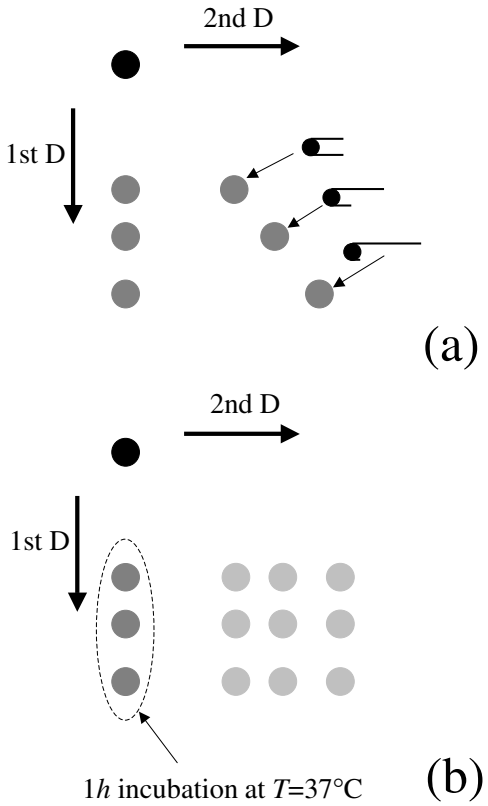

Figure 10. Schematic view of the two-dimensional gel electrophoresis experiment by Pennings *et al* [29] that allowed the demonstration of autonomous nucleosome repositioning. (a) Under conditions where no repositioning takes place the final products line up on a diagonal, each spot corresponding to a certain nucleosome position. (b) If the mononucleosomes are incubated the final products form a square of dots.

the 207_{18} chromatin into its repeating subunits and then studying the nucleosome dynamics on such 207 bp fragments. The authors took advantage of the fact that different nucleosome positions on the chain give rise to different electrophoretic mobilities[4] and that the motion of the nucleosome along the chain can be suppressed by subphysiological temperatures or ionic strengths, and by the presence of Mg^{2+}. The 207 bp mononucleosomes were first separated by an initial dimension of the electrophoresis in conditions where mobility is suppressed. In this way they obtained different bands indicating a set of preferred positions. An entire track from such a gel was then incubated for some period of time in new conditions where mobility may occur, then changed back to physiological conditions where mobility is suppressed again, and run in a *second*, equivalent, dimension of gel electrophoresis. Essentially, the first dimension of electrophoresis created a non-equilibrium distribution. Depending on the conditions this distribution relaxed during the subsequent incubation, which in turn was detected as products moving off the diagonal in the second dimension of electrophoresis; cf figure 10 for a schematic depiction of the two-dimensional separation technique.

[4] One reason is the bend induced by the octamer; it is known that a bend on a naked DNA fragment affects its mobility in the gel in very much the same way [141]. Another reason is the inhomogeneity of the charge distribution along the chain that traps such a polymer in the gel in a U-like conformation [142, 143] (cf figure 3 in [142]).

In [29] it was found that substantial redistribution took place when the sample was incubated for 1 h at 37 °C but not at 4 °C. The experiments were carried out in low ionic conditions in a tris-borate buffer (0.5 × TBE; cf [144] for a discussion of the effects of this buffer on naked DNA). There was a set of preferred positions, all multiples of 10 bp (the DNA helical pitch) apart. This means that the nucleosomes had all the same rotational positioning with respect to the DNA. Another feature that was observed is that the nucleosomes have a preference for positioning at the ends of the DNA fragments, a typical feature for nucleosomes on short DNA that was recently discussed by Sakaue *et al* [66]. The 5s rDNA positioning sequence itself, however, is located more towards the middle. When gel-separating the mononucleosomes directly after they have been excised from the tandemly repeated nucleosomes, this position indeed led to the strongest band. After incubation, however, the end positions showed the highest probability.

The authors extended their study to head-to-tail dimers of 5s rDNA (207_2) [30]. In the first dimension of a 2D gel electrophoresis the mononucleosomes were separated according to their position on the dimer. This was followed by an incubation at 4 or 37 °C and a subsequent cutting of the dimer into its monomers. The resulting product was then separated through electrophoresis in a second dimension. Again for the sample incubated at elevated temperatures a repositioning of the nucleosomes was found. Interestingly, however, the study indicated that the repositioning took place only within a cluster of positions around each positioning sequence but not between them, a fact that was shown by radioactive labelling of one half. This finding indicates that there is no 'long-range' repositioning at low ionic strength. Other systems studied in [30] were fragments of H1-depleted native chromatin and nucleosomes reconstituted on *Alu* repeats; in these cases a repositioning was also detected as a result of an elevated temperature incubation. The authors concluded that the repositioning 'may be visualized as following a corkscrew movement within the superhelical path of the DNA' [30]. The same authors studied in [31] the nucleosome mobility on the 207_2 dimer in the presence of linker histone H1 (or its avian counterpart H5) and found that the mobility of nucleosomes was dramatically reduced.

Ura *et al* [145, 146], following [30], studied nucleosome mobility on the 207_2 dimer under varying conditions, namely in the presence of various chromosomal proteins and in the case when the core histones are acetylated. In the former case mobility was suppressed (depending on the type and concentration of the chromosomal protein), in the latter case the mobility was not changed much.

Flaus *et al* [147] developed a different strategy to determine nucleosome positioning and repositioning. In their method they used a chemically modified H4 histone that induces, after the addition of some chemical, a cut on the nucleosomal DNA 2 bp away from the dyad axis. Via gel electrophoresis of the resulting product they were able to determine the nucleosome position with bp resolution. Using this method Flaus and Richmond [148] studied the nucleosome dynamics on a mouse mammary tumor sequence which revealed several features of repositioning more clearly. The longest fragment of this sequence studied was 438 bp long and had two positioning sequences where two nucleosomes assembled, each at a unique position. These positions were also found when mononucleosomes were assembled on shorter fragments that included only one of the two positioning sequences. The authors studied the degree of repositioning of the mononucleosomes on such shorter fragments (namely nucleosome A on a 242 bp fragment and nucleosome B on a 219 bp fragment) as a function of heating time (ranging from 20 to 80 min) and temperature (ranging from 0 to 50 °C). It was found that the repositioning rates, as estimated from the occurrence and intensity of new bands, increase strongly with temperature but also depend on the positioning sequence (and/or length of the fragment). The difference in repositioning for the two sequences is remarkable: at 37 °C one has to wait ∼90 min for the A242 and more than 30 h for the B219 to have half

of the material repositioned. Another feature found was again a preference for end positions (roughly 70 bp from the dyad axis, similar to the finding in [29]). For nucleosome B, which showed a slower repositioning, the set of new positions were all multiples of 10 bp apart (namely at a 20, 30, 40, 50 bp-distance from the starting position), i.e., they all had the same rotational phase. On the other hand, nucleosome A did not show such a clear preference for rotational positioning. It was argued that these differences reflect specific features of the underlying bp sequences involved. Nucleosome B is complexed with a DNA sequence that has AA/AT/TA/TT dinucleotides that show a 10 bp periodicity inducing a bend on the DNA whereas nucleosome A is positioned via homonucleotide tracts.

The authors speculate in [148] that the preference for end positions might be caused by one (or several) of the following mechanisms:

(1) direct histone interaction with a special structure at the DNA terminus,
(2) relief of the repulsion between entering and exiting strand and
(3) entropy gain by having a long unbound DNA extension.

Recently Hamiche *et al* [149] demonstrated that the (uncatalysed) nucleosome mobility along DNA depends on the presence of histone tails. In particular, in the absence of the N-tail of H2B that passes inbetween the two turns of the nucleosomal DNA [10], spontaneous repositioning of the nucleosomes was detected (here *during* a gel electrophoresis in a second direction).

How does repositioning work? The studies seem to indicate that the octamer is not transferred to competing DNA (at least hardly under physiological conditions, cf [139]). It is also clear that repositioning should preferably not involve the simultaneous dissociation of all the 14 binding sites which would be too costly to be induced by thermal fluctuations. A mechanism that requires only the breakage of a few contacts is loop diffusion. Strictly speaking one has to distinguish here between two different kinds of loops: small loops or bulges (section 2.4.2) and big loops (section 2.4.3) that show a qualitative different energetics and lead also to a different picture of the overall nucleosome repositioning dynamics. Another possibility could be the above mentioned corkscrew motion as suggested in [30] (and even earlier by van Holde and Yager [150]). This could be facilitated through the twist diffusion of small defects that only require the breaking of one or two contacts at a time, a mechanism discussed in section 2.4.4.

2.4.2. Bulge diffusion. In [75] myself, Widom, Bruinsma and Gelbart argued that the repositioning of nucleosomes without dissociation from the DNA chain that wraps them might be possible through the diffusional motion of small intranucleosomal loops. This biological process is analogous to the familiar physical situation of reptation of 'stored length' in polymer chains. Thirty years ago de Gennes [78] discussed the motion of a flexible chain trapped in a gel, modelled by a matrix of fixed point-like obstacles that cannot be crossed by the polymer. Figure 11 depicts schematically the mechanism whereby diffusion of these 'defects' of stored length ΔL gives rise to overall translation of the chain. Specifically, when the loop moves through the monomer at B, this monomer is displaced by a distance ΔL. de Gennes wrote down a conservation equation for this motion of defects along the trapped chain and calculated its overall mobility, and thereby, in particular, the molecular-weight dependence of the overall translation diffusion coefficient. In our present situation the reptation dynamics do not arise from obstacles due to a host matrix (as in a gel) or to other chains (as in a melt), but rather to loops associated with unsaturated adsorption of the DNA on the protein complex. Similar physics arise in the lateral displacements of a linear polymer adsorbed on a bulk solid surface. Sukhishvili *et al* [151, 152], for example, have measured the translational motion of adsorbed

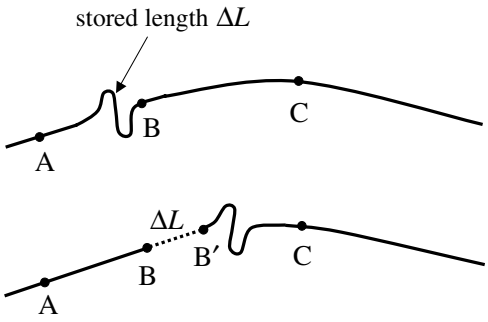

Figure 11. According to de Gennes and Edwards the translational diffusion of trapped chains can be envisaged by a reptation mechanism. When a defect of stored length ΔL passes through a monomer at B it is moved by that amount to a new position B'.

polyethylene glycol (PEG) on functionalized (hydrophobic) silica, specifically the dependence of its centre-of-mass diffusion constant on molecular weight. They find an unusual scaling behaviour, but one that can be accounted for by 'slack between sticking points', so that lateral motion of the polymer proceeds via a caterpillar-like diffusion of chain loops. In the case of intranucleosomal loops considered in this section, one is essentially in the limit of infinite molecular weight, because of the chain length being large compared to the bead (solid substrate) diameter. Furthermore, one is dealing with a lower-dimensional problem, since the DNA chain is wrapped (absorbed) on a 1D path rather than a 2D surface. But the basic features of loop formation and diffusion, and subsequent motion of the overall chain, in particular the exclusive role of equilibrium fluctuations in driving these processes, are the same in both cases.

As shown in earlier studies of competitive protein binding to nucleosomal DNA [27, 28, 98], thermal fluctuations lead to lengths of the chain becoming unwrapped at the ends of its adsorbed portion. If some length of linker is pulled in before the chain re-adsorbs, then an intranucleosomal loop is formed, see figure 12(b). In [75] we calculated first the equilibrium shape and length distribution of these loops in terms of the chain bending stiffness A, adsorption energy per unit length λ, and protein aggregate size R_0 (to be more precise, the radius of curvature of the DNA centreline). We then considered the diffusion of these loops from one end of the nucleosome to the other. Finally, treating this motion as the elementary step in the diffusion of the nucleosome itself along the wrapping chain, we were able to make estimates of the nucleosome repositioning rates as a function of A, λ, R_0 and solvent viscosity η.

I present here (in more detail than in our letter [75]) the calculation of the equilibrium statistical mechanical probability associated with the formation of a *small* intranucleosomal loop (large loops are considered in the next section). When no loops are present, a nucleosome consists of a length l of DNA chain wrapped continuously around the octamer, see figure 12(a). In reality, as discussed in section 2.1, the configuration of the adsorbed chain is a left-handed superhelix (of contour length l) spanning the full height of a cylinder. One can proceed, however, without making any explicit assumptions about the shapes of either the histone octamer or the wrapped DNA.

Consider a fluctuation in which some length of the chain becomes unwrapped (this can *only* happen at the end of the adsorbed portion of chain) and simultaneously some length, say ΔL, of linker (i.e., previously unadsorbed chain) is 'pulled in' before the chain re-adsorbs. The fluctuation has then produced a loop of contour length

$$L = L^* + \Delta L \tag{59}$$

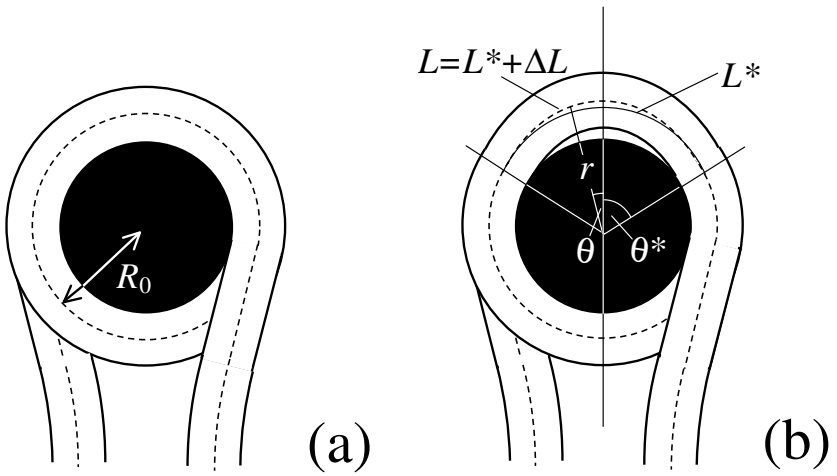

Figure 12. (a) Top view of the defect-free nucleosome looking down the superhelical axis. (b) Geometry of an intranucleosomal bulge.

where L^* is the 'exposed' length of nucleosome associated with the loop, see figure 12(b). Note that ΔL shows a strong preference for values that are multiples of 10 bp lengths, since this is the periodicity of the DNA helical pitch (binding sites are located where minor grooves face inwards to the octamer, cf section 2.1); other values require a twisting of the loop DNA which is energetically costly. For the energy associated with forming a loop of this kind, one can write [75]:

$$\frac{\Delta U}{k_B T} = \frac{E_{elastic}}{k_B T} - \left(\frac{l_P}{2R_0^2} - \lambda\right) L^* = \frac{l_P}{2} \int_{loop} \frac{\mathrm{d}s}{R^2(s)} - \left(\frac{l_P}{2R_0^2} - \lambda\right) L^*. \quad (60)$$

The first term is the bending energy of the loop with $1/R(s)$ being the local curvature at distance s along its contour, cf also equation (29). The second term accounts for the length L^* which has been adsorbed and bent with curvature $1/R_0$ *before* loop formation. The mathematical details of the functional minimization of ΔU, equation (60), are presented in appendix B. There it is shown that the formation energy of an optimal small loop for given extra length ΔL is approximately given by[5]

$$\frac{\Delta U}{k_B T} \simeq \frac{6}{5}(20\pi^4 l_P \lambda^5)^{1/6}(R_0 \Delta L)^{1/3} - \frac{l_P}{2R_0^2}\Delta L. \quad (61)$$

Using equation (61) one finds that the energy required to form a loop of minimal size of 34 Å (one helical pitch) is roughly $27\,k_B T$ (assuming $R_0 = 43$ Å, $\lambda = 0.176$ Å^{-1}, $l_P = 500$ Å); larger bulges are even more expensive. Here a conceptional problem of the small loop mechanism seems to arise: the formation energy of a small loop is of the same order as the complexation energy of the nucleosome itself (about $30\,k_B T$, cf section 2.1)! However, note two points:

(i) the WLC model is not very reliable for such strong curvatures and might overestimate the actual bending energy, and

[5] In [75] we made one further approximation, namely when inserting θ^* into equation (B.12), we neglected the second and third term that nearly cancel out when one inserts the typical parameters of the nucleosome. This led us to equation (2b) in that paper. Here I present the full expression, equation (61).

(ii) we assumed in the calculation leading to equation (61) that $TR_0^2/A \simeq \pi^2/\theta^{*2} \gg 1$ (cf appendix B).

From equation (B.13) it follows, however, that $\theta^* \simeq 1.75$ and hence $\pi^2/\theta^{*2} \approx 3$. So one is actually in the crossover region between small and large 'tensions' T.

To account for this fact, the calculation can be redone using a refined estimation for $\lambda_\pm$, namely $\lambda_+ = 1/\lambda_- = \sqrt{2} + \sqrt{TR_0^2/A}$ which is a much simpler expression than equation (B.6) but is still a rather good approximation; in particular it shows the right asymptotic behaviour, equation (B.7). This leads to $\theta^* \approx 1.3$ and $\Delta U \approx 20\, k_B T$ for $\Delta L = 10$ bp.

Now the probability distribution for the formation of loops of size ΔL is then simply given by the corresponding Boltzmann factor, normalized such that the maximal number of loops (expected for the unphysical case $\Delta U = 0$) is the geometrically possible one, i.e. l/L^*:

$$n_{eq}(\Delta L) \simeq \frac{l}{L^*} \mathrm{e}^{-\Delta U/k_B T}. \tag{62}$$

Even the value $20\, k_B T$ calculated above shows that the spontaneous formation of a small loop is a *very* rare event. I will check in the following if it occurs sufficiently often that it could account for the experimentally observed autonomous repositioning rates discussed in the previous section. In order to proceed here the dynamics of the loops needs to be considered and that of the resulting nucleosome repositioning.

The key idea [75] here is that diffusion of the histone octamer along the DNA is achieved by formation and annihilation of loops. Let D denote the diffusion constant relevant to this motion of the ball along the chain, and let w be the rate at which loops are formed (by incorporation of linker length) $D = w\Delta L^2$. These loops "disappear' due to their diffusion 'off' the ball, at a rate that is proportional to the instantaneous number of loops, i.e., at a rate $C_A n$. Accordingly, the overall rate of formation of loops is given by $w - C_A n$, which must vanish at equilibrium, implying $n_{eq} = w/C_A$. Since this number is much smaller than unity, we are justified in assuming that only one loop at a time needs to be considered in treating the diffusion of intranucleosomal loops. It follows from the Boltzmann expression for n_{eq} (see (62)) that w, the rate of loop formation, is given by

$$w \simeq C_A (l/L^*) \exp(-\Delta U/k_B T) \tag{63}$$

and D by $w\Delta L^2$.

It remains only to evaluate C_A, characterizing the rate of diffusion of loops 'off' the ball. Let D^+ denote the diffusion constant associated with this motion (D^+ characterizes the diffusion of loops *through* a wrapped ball, as opposed to the coefficient D that describes diffusion of the *ball* along the chain). Since the distance which the loop must move to leave the ball is l, the wrapping length, one can write $C_A^{-1} \simeq l^2/D^+$. From the Stokes–Einstein relation one has furthermore that $D^+ = k_B T/\zeta$ where $\zeta \simeq \eta L^*$ is the friction coefficient of the loop, with η the effective solution viscosity. L^*, as before, is the exposed length of the octamer associated with the loop, and hence provides the loop size relevant to its diffusion along the (1D!) nucleosome path of the chain. This hydrodynamic description is justified by the fact that loop diffusion requires unbinding of only a single sticking site, whose binding energy is of order of a few $k_B T$. Combining all of the results from this and the preceding paragraph then gives

$$D \approx \frac{k_B T}{\eta l} \left(\frac{\Delta L}{L^*} \right)^2 \exp(-\Delta U/kT) \tag{64}$$

with L^* given by equation (B.13) and ΔU by equation (61).

Recalling $\theta^* = 1.3$ and $\Delta U = 20\, k_B T$ for $\Delta L \simeq 34$ Å and taking reasonable estimates for η (a centipoise), R_0 (43 Å) and l (500 Å) we find that D is of order 10^{-16} cm^2 s^{-1}. Hence

typical repositioning times are of the order of an hour; furthermore there is a strong dependence on the temperature. A closer comparison of these theoretical estimates with the experiments (discussed in section 2.4.1) will be given in the discussion in section 2.4.5, after I have also presented the theories for large loops and twist diffusion.

2.4.3. Large loop repositioning. The perturbation calculation presented in [75] and reviewed in the previous section allows us to only study small loops that store an amount of excess length of 10 or 20 bp. To also describe large loops a different approach is necessary as was presented by Kulić and myself in [76]. In that paper we made use of the Euler–Kirchhoff theory for the static equilibrium of rods which allowed us to describe loops of *any* given excess lengths. The outcome of that paper changed our view of how repositioning via loop formation should work; besides the local repositioning based on bulge diffusion there should also occur a long-range hopping via large loops, at least for the case of very low nucleosome line densities as is often encountered in *in vitro* experiments.

In [76] we started again from the Hamiltonian given by equation (60). The section of the DNA constituting the loop has a contour length L and is parameterized by its arc length s ranging from $-L/2$ to $L/2$. L is the sum of the exposed length $L^* = 2\theta^* R_0$ ($2\theta^*$: opening angle) and the excess length ΔL, see equation (59). In order to compute the ground state for a trapped intranucleosomal loop the total energy (60) has to be minimized under two constraints:

(i) the excess length ΔL is prescribed so that the following relation between the opening angle θ^* and the total loop length L has to be fulfilled

$$\Delta L = L - 2\theta^* R_0 = \text{constant}, \tag{65}$$

(ii) at the two ends $s = \pm L/2$ the rod has to be tangential on an inscribed circle of given radius (representing the nucleosome):

$$R_0 = \left| \frac{y(L/2)}{-x'(L/2)} \right| = \text{constant}. \tag{66}$$

Here $x(s)$ and $y(s)$ are the Cartesian coordinates of the rod axis as a function of the arc-length parameter s (cf figure 13). The absolute value in the second constraint needs to be introduced formally for dealing with crossed rod solutions (which are considered later on) and can be omitted for simple uncrossed loops.

For the analytical description of the loop geometry it is convenient to introduce the angle $\theta = \theta(s)$ between the DNA tangent and the Y-axis (cf figure 13) that describes the DNA centreline (note that this is the same angle θ as the one introduced in the previous section, cf figure 12). Integrating the sine (cosine) of θ over the arc length parameter s yields the X (Y) Cartesian coordinate of any point along the rod, and the derivative θ' gives the rod curvature R^{-1}. Furthermore the nucleosome opening angle θ^* is simply related to θ at the boundary, namely $\theta^* = \theta(L/2)$ for simple loops and $\theta^* = \pi - \theta(L/2)$ for crossed loops (see below).

The two constraints equations (65) and (66) can be rewritten in terms of θ and then be introduced into the minimization by two Lagrange multipliers T_1 and T_2. We then arrive at the following functional

$$\mathcal{F}\{\theta(s)\} = A \int_0^{L/2} (\theta')^2 \mathrm{d}s - \left(\frac{A}{2R_0^2} - k_B T \lambda \right) L^* + T_1[L - (\Delta L + L^*)] + T_2 \left[\int_0^{L/2} \cos\theta \mathrm{d}s - R_0 \sin\theta^* \right]. \tag{67}$$

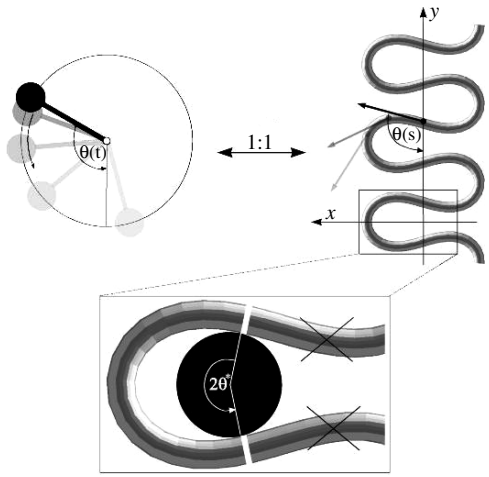

Figure 13. The Kirchhoff analogue for the case of a planar pendulum and a planar semiflexible rod under tension. The inset displays how to construct a intranucleomal loop by inscribing a circular disc representing the octamer.

Here the first term is the bending energy, the second accounts for the exposed length $L^* \equiv 2\theta^* R_0$ and the third and fourth term are the imposed length and tangency constraints. Equation (67) can be rearranged in the more familiar form

$$\int_0^{L/2} (A(\theta')^2 + T_2 \cos\theta) \mathrm{d}s + \text{b.t.} \tag{68}$$

where b.t. denotes boundary terms (depending on $\theta(L/2)$ only) that obviously do not contribute to the first variation inside the relevant s interval. The integral in equation (68) is analogous to the action integral of the plane pendulum with $A(\theta')^2$ corresponding to the kinetic and $-T_2 \cos\theta$ to the potential energy. The latter analogy is nothing else but Kirchhoff's kinetic mapping between deformed rods and the spinning top that contains the present problem as a simple special case (cf the paragraph after equation (33) for a brief discussion of the Kirchhoff analogy).

Kirchhoff's analogy provides one directly with explicit expressions for DNA shapes subjected to twisting, bending and various geometric/topological constraints. Here, for the case of planar untwisted rods, also called the *Euler elastica*, where the corresponding 'spinning top' reduces to the plane pendulum, the rod conformations are most generally given by

$$\cos\theta(s) = 1 - 2m\,\mathrm{sn}^2\left(\frac{s}{\Lambda}\middle| m\right). \tag{69}$$

This can be integrated to obtain the general planar rod shapes in Cartesian coordinates:

$$x(s) = 2\sqrt{m}\Lambda\,\mathrm{cn}\left(\frac{s}{\Lambda}\middle| m\right) \tag{70}$$

$$y(s) = 2\Lambda E\left(\frac{s}{\Lambda}\middle| m\right) - s \tag{71}$$

with sn, cn$(\cdot|m)$ (and later below dn) denoting the Jacobi elliptic functions with the parameter m and $E(u|m)$ being the incomplete elliptic integral of the second kind in its 'practical' form [153]. The two parameters $m > 0$ and $\Lambda > 0$ in equations (70) and (71) characterize the shape and the scale of the solution, respectively. These solutions are up to trivial plane

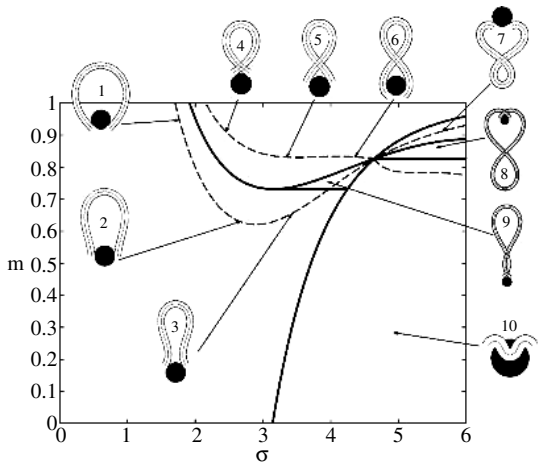

Figure 14. Diagram of the solutions of the Euler elastica providing an overview of the possible loop shapes as a function of the shape parameter m and the contact parameter σ. Loops of constant excess length $\Delta L = 10$ bp are located on the dashed curves. The solid curves separate regions with different geometrical characteristics: simple loops ('1' to '3'), crossed loops ('4' to '6') and more exotic shapes ('7' to '10').

rotations, translations, reflections and shifting of the contour parameter $s \to s + s_0$, the most general solutions to the Euler elastica. For different parameters m one obtains different rod shapes corresponding to different solutions of the plane pendulum motion [125]. The case $m = 0$ describes a pendulum at rest which corresponds to a straight rod. For $0 < m < 1$ one has strictly oscillating pendulums corresponding to point symmetric rod shapes where the turning points of the pendulum have their counterparts in points of inflection of the rod. For $m < 0.72$ the rod is free of self-intersections like the one depicted in figure 13. For m larger than 0.72 the rods show varying complexity with a multitude of self-intersections and for $m = 1$ one has the homoclinic pendulum orbit corresponding to a rod solution with only one self-intersection that becomes asymptotically straight for $s \to \pm\infty$. For even higher values of m, i.e., for $m \geqslant 1$ one has revolving pendulum orbits corresponding to rods with self-intersections lacking point symmetry. Finally, the limiting case $m \to \infty$ corresponds to the circular rod shape.

In order to describe a trapped loop one needs to use equations (70) and (71) imposing the constraints (65) and (66). For details of this calculation I refer the reader to [76]. There we present explicit solutions for the scaling parameter Λ, the opening angle θ^* and the excess length ΔL as functions of the 'contact parameter' $\sigma = L/2\Lambda$ and the shape parameter m, i.e., $\Lambda = \Lambda(\sigma, m)$, $\theta^* = \theta^*(\sigma, m)$ and $\Delta L = \Delta L(\sigma, m)$. Inserting $\Lambda(\sigma, m)$ and $\theta^*(\sigma, m)$ into equation (60) leads to the final expression for the loop formation energy

$$\Delta U(\sigma, m) = \frac{4A}{R_0}\left|\frac{(E(\sigma|m) + (m-1)\sigma)(2E(\sigma|m) - \sigma)}{\mathrm{sn}(\sigma|m)\mathrm{dn}(\sigma|m)}\right| - 2R_0\left(\frac{A}{2R_0^2} - k_B T\lambda\right)\arccos[\pm(2\mathrm{dn}^2(\sigma|m) - 1)] \tag{72}$$

with $\pm = \mathrm{sign}(2E(\sigma|m) - \sigma)$.

Now the problem of finding the ground state loop for given excess length ΔL reduces to a two variable (σ, m) minimization of equation (72) under the constraint $\Delta L(\sigma, m) \stackrel{!}{=} \Delta L$. This final step was performed numerically [76]. An overview over the different solutions can

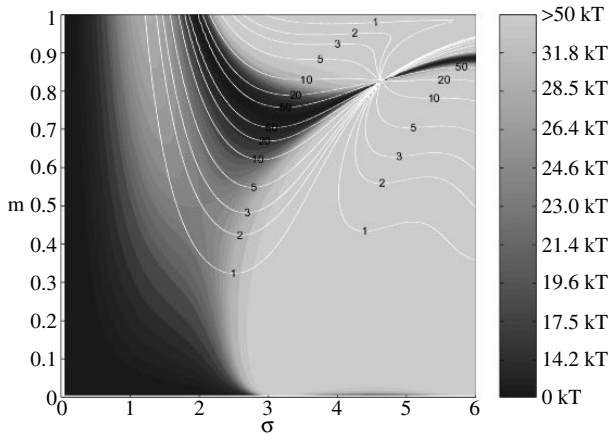

Figure 15. Density plot of the total loop energy, equation (72), as a function of m and σ (same parameter range as in figure 14). The white curves denote lines of constant excess length (in multiples of 10 bp).

be obtained by inspecting some loop geometries in the resulting (σ, m) parameter plane. Both parameter values σ and m vary between 0 and ∞, though the loops of practical importance are all found within the range $0 < m < 1$ and $0 < \sigma < 5$. In figure 14 this relevant section of the parameter space is depicted together with a few example loops. The dashed curves indicate parameter values which lead to constant excess length $\Delta L = 10 \times 3.4$ nm (corresponding to 100 bp). On these curves are located the loop shapes '1' to '7' that are examples of such 100 bp-loops. The whole parameter plane is subdivided into regions of structurally different solutions that are separated by solid lines. The large region starting at $\sigma = 0$ contains exclusively simple loops (like '1', '2' and '3') without self-intersections and nucleosome penetration. Above that simple-loop region there is a region that contains loops with a single self-intersection; it includes the branch of 100 bp-loops with the example configurations '4', '5' and '6'. To the right there are non-physical cases where the loops penetrate the nucleosome, like example '10'. There are also three other regions with single and double crossing points ('7', '8', '9') where the loop can be founds on the 'wrong' side of the nucleosome like in '7' and '8'.

In [76] we determined the energy minimizing loop for each value of excess length ΔL. Figure 15 shows a contour plot of the loop energies, equation (72), in the (σ, m) plane for the same range of parameters as in figure 14. The nucleosome parameters chosen in figure 15 are the same as in [76], namely $R_0 = 40$ Å, $\lambda = 0.23$ Å^{-1}, $l_P = 500$ Å; these values are close (but not identical) to the ones used in the previous section. Shown in this figure are also the corresponding lines of constant ΔL (with $\Delta L = 1, 2, \ldots, 50 \times 3.4$ nm). As already observed in figure 14 there are, for any given ΔL, different branches of (σ, m) values corresponding to uncrossed, simply crossed and other, more exotic, structures. For short excess lengths one finds that the loops with the smallest formation energy ΔU, equation (72), belong to the simple, uncrossed kind. For example, the optimal loop geometry for a 100 bp-loop is structure '2' in figure 14 which is located at the point where the lines of the 100 bp loops in figure 15 encounter the total energy minimum. Interestingly, the optimal loop shape switches from the simple to the crossed type when an excess length of ∼500 Å is reached.

The ground state energy as a function of the excess length ΔL is given in figure 16. Consider first the *simple loops* that are energetically preferable for $\Delta L < 50$ nm. Inspecting

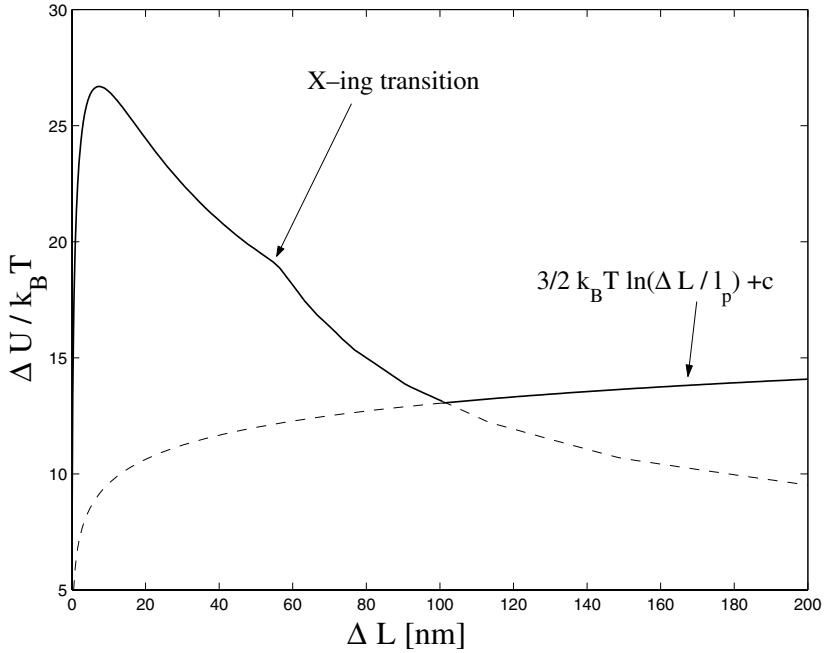

Figure 16. The energy of the optimal loops for given excess length ΔL. The kink in the curve reflects the switch from simple to crossed loops. The dashed curve gives the free energy for entropic loops that are much longer than their persistence length.

figure 16 one finds the remarkable fact that the loop formation energy is non-monotonous in that range. First it increases sharply (namely as $\Delta U \propto \Delta L^{1/3}$, cf [76], in accordance with the small loop behaviour equation (61)). Then at some critical excess length $\Delta L = \Delta L_{crit}$ (which is approximately $\Delta L_{crit} \approx 2.2 \times 3.4$ nm for the above given parameters) the loop energy reaches a maximum $\Delta U(\Delta L_{crit}) \approx 26\, k_B T$. Beyond that the energy decreases with increasing ΔL.

In order to explain this behaviour one might naively argue as follows: for excess lengths shorter than the DNA persistence length it is energetically unfavourable to store additional length into the loop because it requires increasing deformation of the loop DNA. On the other hand, for loops longer than l_P the bending energy contribution becomes very small; to add more length should even decrease this energy since the loop can lower its curvature. However, the occurrence of the maximum of ΔU at ∼22 bp excess DNA lengths, a value that is *considerably* smaller than the persistence length, is surprising at first sight.

The explanation for this small value is given in [76]. There it is shown that the condition for the critical excess length ΔL_{crit} is given by a simple geometric distinction between two loop shapes: the subcritical loop (figure 17(a)) that has none of its tangents parallel to the X-axis (i.e. $\theta(s) \neq \pi/2$ for all s) and the supercritical loop (figure 17(b)) that has two or more tangents parallel to the X-axis where $\theta(s) = \pi/2$. Now it can be easily envisaged that adding some extra length $\mathrm{d}L$ to a subcritical loop increases its energy ΔU (cf [76]) whereas for supercritical loops additional length decreases its energy. In the latter case one might just cut the loop at the two points P_L and P_R in figure 17(b) and introduce there the additional length (this operation does not change the energy) and *then* relax the shape by letting it evolve to the new equilibrium while keeping θ^* constant. It was demonstrated in [76] that this condition of the parallel tangents indeed leads to the above given small value of ΔL_{crit}.

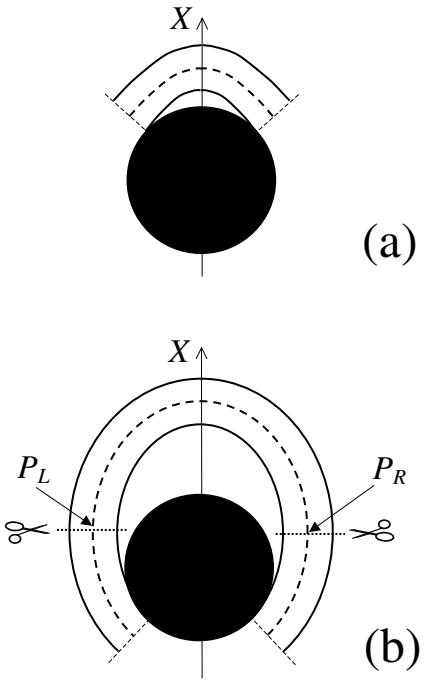

Figure 17. Subcritical (a) and supercritical (b) loops show a different behaviour when additional length is added. In the former case the energy goes up, for the latter it goes down. This allows us to understand why the maximum of ΔU is reached already for a very short excess length, cf figure 16.

The ground state of loops switches from simple uncrossed loops to *crossed loops* when one reaches an excess length of ∼500 Å. Here, however, arises an additional complication. As can be seen by inspecting loops '4', '5' and '6' in figure 14 these structures contain a self-intersection at the crossing point. Therefore in principle a planar theory cannot capture the geometry of crossed loops. One might thus leave the plane and describe the self-contacts of the rod with corresponding point forces in 3d as done by Coleman *et al* [134] in a general theory of rod self-contacts. However, since the self-avoiding crossed loops stay close to a plane in all cases of practical interest (namely loops for which the self-contact point is not too close to the nucleosome), it is here sufficient to treat the self-interaction as a perturbation (cf [76] for details). One finds an additional small contribution for crossed loops due to the out-of-plane bending caused by the self-contact. In figure 16 this contribution (a few k_BT) has already been added; as can be seen from this figure crossed loops are still favoured for sufficiently long excess lengths $\Delta L > \Delta L_{cross}$ (here ∼600 Å). This can be rationalized by the fact that for long enough loops the adsorption energy (proportional to θ^*) starts to dominate over the bending energy so that loops with smaller θ^*, namely crossed ones, become favourable.

Increasing the length even further one leaves the energy-dominated regime in which entropic effects can be neglected due to short loop length $L < l_P$. For larger lengths entropic effects become important and one enters the *entropic loop* regime (cf the discussion of large-leafed rosettes, equation (36), and [136]). In the large loop limit where the loop is longer than several l_P the chain loses its 'orientational memory' exponentially and behaves as a random

walk which starts from and returns to the same point. The entropic cost for gluing the ends of this random walk together is approximately given by

$$\Delta U \simeq 3/2 k_B T \ln(\Delta L / l_P) + E_0. \tag{73}$$

Here $E_0 \approx 6.5\ k_B T$ denotes the bending plus adsorption energy contributions of the overcrossing DNA segments that enter and leave the nucleosome in the large-loop limit $\Delta L \to \infty$. As already discussed after equation (36), the free energy minimum occurs at the crossover between the elastic ($\Delta L < l_P$) and entropic ($\Delta L \gg l_P$) region where the decreasing elastic energy is overtaken by the increasing entropic contribution, as can also be seen in figure 16.

The free energy, equation (73), leads to an algebraically decaying probability $w(\Delta L)$ for the jump lengths scaling as $w \propto \exp(-\Delta U / k_B T) \propto (\Delta L)^{-3/2}$. In general, power law distributions of the form $w \propto (\Delta L)^{-\gamma}$ with $\gamma > 1$ lead to superdiffusive behaviour of the random walker (here the nucleosome displacement *along* the DNA). According to Levy's limit theorem the probability distribution of the random walker (more precisely, the distribution of the sums of independent random variable drawn out from the same probability distribution $w \propto (\Delta L)^{-\gamma}$) converges to a stable Levy distribution of index $\gamma - 1$ [154–156]. This so-called *Levy-flight* [157] differs in many respects from the usual diffusion process, as for short time intervals big jumps are still available with significant probability. Moreover, all moments (besides possibly the first few ones) diverge. For the present case $\gamma = 3/2$ even the first moment does not exist. Note that the value $3/2$ is based on the assumption of an ideal chain (no excluded volume); in general the excluded volume leads to self-avoiding-walk statistics with a slightly larger value of γ around 2.2 [156] (cf also [158]). In that case one has a finite value of the first moment, i.e., a finite average jump length.

We presented in [76] some numerical estimates of the dynamics of the nucleosome repositioning on DNA fragments of different lengths. The basic idea is that the transition rate $w(\Delta L)$ for a jump of length ΔL is proportional to $C_A \exp(-\Delta U(\Delta L))$ with ΔU being the loop formation energy. The Arrhenius constant C_A^{-1} involved in the loop formation has in principle to be determined experimentally. A theoretical estimate [75] was reported in the previous section, cf equation (63), where it was shown that C_A corresponds roughly to the inverse lifetime of the loop. Hence $C_A^{-1} \approx l^2 \eta R / k_B T \approx 10^{-5}$–$10^{-6}$ s. In [76] we considered two DNA lengths: (147 + 90) bp (short segment) and (147 + 300) bp (intermediate length). For the short piece the octamer repositioning occurs on the timescale of hours (in accordance with the previous section); on the intermediate segment the repositioning time is of the order of seconds to minutes. Important in both cases is where the nucleosome initially starts. If the start position is at an end of the DNA fragment then the nucleosome jumps preferentially to the other end since large jumps are energetically favoured, cf figure 16. This leads to a fast relaxation of the initial position. Smaller jumps also take place but less frequently; these jumps, however, lead on long timescales to an equal distribution (in accordance to Boltzmann's law) of the octamer along the DNA fragment. On the other hand, if the initial position is chosen in the middle of the DNA piece then the relaxation process is slower since a smaller loop is initially required; this first jump is then preferential to an end position.

As mentioned in section 2.4.1 the repositioning is often followed by gel electrophoresis (cf figure 10). It is therefore helpful to ask how the resulting band structure evolves with time [76]. Let us start again with an end positioned nucleosome. In this case, as just mentioned, the octamer initially exhibits mainly jumps back and forth between the two ends. Since the mobility is symmetric with respect to the middle position, this leads to the interesting conclusion that in standard gel electrophoresis these jumps would not be detected at all! Only slowly shorter jumps will allow the nucleosome to inhabit positions away from the ends. But

this slower process cannot be distinguished easily from short-range diffusion (away from the initial end), a process like the one discussed in the previous section. On the other hand, when starting from the middle position the preference to jump to the end positions will lead to an initial population gap in the band structure between the fast band (end positions) and the slow one (middle position). This gap would not occur for short-range repositioning. For more details on the expected band structures, the reader is referred to [76].

2.4.4. Twist diffusion. Let me now discuss twist diffusion, which might be another possible mechanism for nucleosome repositioning. I will give here some first theoretical quantitative estimates for this mechanism (without any free parameters) based on recent calculations by Kulić and myself; the full presentation will be given elsewhere [77]. If a 1 bp twist defect (one missing or one extra bp) forms through thermal activation at one end and manages to get through to the other end, this results in a 1 bp step of the nucleosome along the DNA and at the same time in a rotational motion by $\sim 36^\circ$, i.e., the nucleosome performs a short fraction of a corkscrew motion.

The possibility of twist defects was demonstrated as soon as the high resolution crystal structure of the core particle was resolved [10]. In that study the core particles were prepared from a palindromic 146 bp DNA and core histones assuming that the resulting complex would show a perfect two-fold symmetry. However, it turned out that one bp is localized directly on the dyad axis so that one half of the nucleosomal DNA is of length 73 bp whereas the other is only 72 bp long. The missing bp of the shorter half is, however, not localized at its terminus but instead at a 10 bp stretch close to the dyad axis (cf figure 4(d) in [10] that shows a superimposition of the two DNA halves). The reason is presumably the attraction between the DNA termini of adjacent particles in the crystal (cf figure 4(c) in that paper) that try to come close to mimicking a bp step at the cost of forming a twist defect far inside the wrapped chain portion. In fact, crystals of core particles with 147 bp DNA do not show this defect [11].

To proceed further we describe the DNA chain within a Frenkel–Kontorova model, i.e., we view it as a chain of particles connected by harmonic springs in a spatially periodic potential. The original Frenkel–Kontorova model was introduced more than 60 years ago in order to describe the motion of a dislocation in a crystal [159]. In the meantime variants of this model were applied to many different problems including charge density waves [160], sliding friction [161, 162], ionic conductors [163, 164], chains of coupled Josephson junctions [165] and adsorbed atomic monolayers [166, 167]. Here, in the context of DNA adsorbed on the protein octamer, the beads represent the bp. The springs connecting them have an equilibrium distance of $\bar{b} = 0.6$ nm (which is here taken to be the distance along the DNA sugar–phosphate backbone, *not* the distance $b = 0.34$ nm along the fibre axis), and the spring constant is chosen such that it reflect the DNA elasticity. Specifically

$$E_{elastic} = \sum_n \frac{K\bar{b}^2}{2}\left(\frac{x_{n+1} - x_n}{\bar{b}} - 1\right)^2 \tag{74}$$

with x_n being the position of the nth bp measured along the helical backbone and $K\bar{b}^2/2 \simeq 70$–$100\ k_BT$ accounts for the coupled twist-stretch elasticity [168–170]. Finally, the external potential comes from the contact points to the octamer. The distance between neighbouring contact points is 10 bp which corresponds to 60 nm along the arc length of the minor groove. A contact point at position x_0 is here modelled by the following function

$$E_{ads} = -U_0\left(\left(\frac{x - x_0}{a}\right)^2 - 1\right)^2 \theta(a - |x - x_0|) \tag{75}$$

with $\theta(x) = 1$ for $x \geqslant 0$ and zero otherwise. The two parameters, the depth U_0 of the potential and its width a, can be estimated as follows. U_0 represents the pure adsorption energy per point contact which can be estimated from competitive protein binding [27, 28] to be of order $6\,k_BT$ (cf section 2.1). The other parameter, a, can then be estimated from the fluctuations of the DNA in the crystal (measured by the B-factor, cf figure 1(b) in [10]). The fluctuations of the DNA at the binding sites are much smaller than in the middle. Using a quadratic expansion of equation (75) one finds from a straightforward normal mode analysis [77] that $a \approx \bar{b}/2$, i.e. the adsorption regions lead to a strong localization of the DNA.

Now having all the numbers at hand we can answer the question whether the twist defects are localized between two contact points. The deformation energy of the defect localized along a 10 bp stretch is of order $7\,k_BT$. On the other hand, by distributing the defect homogeneously over 20 bp the elastic energy goes down by $\sim 7/2\,k_BT$ at the cost of releasing the adsorption point in the middle (roughly $6\,k_BT$). The smearing out of the defect costs, therefore, $\sim 3\,k_BT$. This already shows that the kink is not so strongly localized which points towards a high mobility of the twist defect.

Let us assume now that a kink with one missing bp is located between two binding sites. When this kink jumps to the neighbouring 10 bp (say to the right) it has to cross a barrier of height $\sim 3\,k_BT$. Using equations (74) and (75) we obtain an explicit form for the barrier and, together with the single bead friction $\zeta \approx 10^{-9}\,k_BT\,\mathrm{s\,nm^{-2}}$ [171, 172], are able to calculate the Kramer's escape rate for the kink from the given localized 10 bp stretch to a neighbouring one. This leads us to a typical time $t_{step} \approx 10$ ns for going from one stretch to the next (cf [77] for details).

To determine the rate at which twist defects are formed at the entry–exit points of the DNA one can now use an argument similar to the one presented in [75] (cf also section 2.4.2): the ratio of the lifetime t_{life} of a kink to the time interval t_{inj} between two kink injection events at the end of the wrapped DNA portion equals the probability of finding a defect on the nucleosome, i.e. $t_{life}/t_{inj} \simeq N_{site}\mathrm{e}^{-\Delta U/k_BT} \approx 10^{-2}$. Here $N_{site} = 13$ denotes the number of possible positions of the defect between the 14 binding sites.

How is the average lifetime t_{life} of a defect related to t_{step}, the typical time needed for one step? It is possible to calculate the mean first passage time for a defect that starts at one end (say the left one) and leaves the nucleosome at the same end, τ_{left}, or at the other end, τ_{right}. From [173] one finds $\tau_{left} = (25/6)t_{step}$ and $\tau_{right} = 28t_{step}$. Furthermore, the probability to leave at the left end is $p_{left} = 12/13$ and at the right end $p_{right} = 1/13$ [173] which gives the lifetime as the weighted average $t_{life} = 6t_{step}$. Only a fraction p_{right} of the defects reaches the other end and will lead to a repositioning step, i.e., the time of one diffusion step of the nucleosome along the DNA is of order $T = t_{inj}/p_{right} \approx 10^{-4}$ s where use was made of the above presented relations between the timescales. From this follows directly the diffusion constant D of the nucleosome along the DNA: $D = b^2/(2T) \approx 7 \times 10^{-12}\ \mathrm{cm^2\,s^{-1}}$ (with $b = 0.34$ nm).

Therefore we find a diffusion constant that is *much* larger than the one expected for the repositioning via bulges ($D \approx 10^{-16}\ \mathrm{cm^2\,s^{-1}}$, cf equation (64)). Most importantly, it is also orders of magnitude larger than the diffusion constant observed in the experiments. How can this apparent inconsistency be resolved?

Most likely, the diffusion is considerably slowed down due to the quenched disorder stored in the bp sequence of the DNA. In fact, the bulk of the repositioning experiments has been made on DNA with rather strong positioning sequences leading to a strong rotational positioning of the nucleosome. Starting from a preferred position the nucleosome would arrive after five steps to the left (or to the right) on the 'wrong' site of the nucleosome, forcing the DNA to be bent in an unfavourable direction. This means that the nucleosome needs to cross a barrier

in order to reach a position 10 bp apart. For instance, in the case of the 5S rDNA sequence theoretical estimates indicate a barrier height of the order 10 k_BT [174, 175]. This leads to a strong reduction of the effective diffusion constant, namely $D_{eff} \approx D\mathrm{e}^{-10} \approx 10^{-16}$ cm^2 s^{-1}, a value comparable to the one found for bulge diffusion. In [77] we formally include these effects into the Frenkel–Kontorova framework by introducing an octamer-fixed bending field and by attributing 'bending charges' to the beads. This allows us to give a rough quantitative treatment of the nucleosome mobility as a function of the underlying bp sequence.

2.4.5. Discussion: bulge versus twist diffusion. Comparing the different repositioning mechanisms presented in the previous sections one has to conclude that they lead to very different 'sliding' scenarios. On *short* DNA fragments repositioning could in principle work via bulge (section 2.4.2) or twist diffusion (section 2.4.4); large loops cannot occur because there is not enough free DNA length available. Bulge diffusion is rather slow (timescale of hours) since the formation of a small loop is costly, mainly because the opening angle of a bulge is rather large, leading to several open binding sites. The repositioning rates should show a strong temperature dependence as well as a strong dependence on the adsorption strength (i.e., a strong dependence on the ionic conditions). The preferred repositioning steps are multiples of 10 bp.

On the other hand repositioning via twist defects should be much faster (timescale of seconds). The nucleosome should slide in a corkscrew motion along the DNA and should forget its initial position rather quickly. However, if the underlying DNA sequence induces a strong rotational positioning signal the timescale becomes comparable to that of small loop repositioning. Futhermore, due to the underlying bp sequence one should expect a 10 bp spacing between the dominant positions that is, however, here not the result of 10 bp jumps but just reflects the relative Boltzmann weights of favourable and unfavourable positions. The estimates of the diffusion constants of these two mechanisms are too unreliable (activation energies appear in the exponent!) to allow one to predict which of the mechanisms should be favoured. If on the other hand a rather homogeneous DNA sequence is used, our prediction is that corkscrew motion is the much faster and therefore predominant mechanism. The experiment by Flaus and Richmond [148] already goes in this direction; comparing their experimental results (cf section 2.4.1) with the theoretical pictures seems to point towards sliding motion for one of the positioning sequences (the one that has a homonucleotide tract), whereas it is not clear whether the nucleosome escapes from the rotational positioning trap via bulge or via twist diffusion.

On *long* DNA fragments single nucleosomes could also be repositioned via large loops (section 2.4.3). Our theoretical model suggests that large loop repositioning would be much faster than bulge diffusion. Also it should be expected that a similar mechanism allows the nucleosome to be transferred to competing naked DNA chains. As discussed in the experimental section above there is, however, not much evidence for such processes. Only Watkins and Smerdon [139] report such a nucleosome transfer to free DNA at higher ionic strength. This again allows us to speculate that each of the different mechanisms might play a dominant role in a certain parameter range. To come to more definite conclusions more systematic experiments have to be made on short as well as long DNA molecules with and without positioning sequences under varying ionic conditions.

Repositioning *in vivo* might be actively facilitated by chromatin remodelling complexes whose action is currently studied *in vitro* (reviewed in [7, 32–35]). There are two major families: SWI/SNF and ISWI. They both burn ATP to enhance nucleosome dynamics but their underlying modes of action seem to be fundamentally different. The SWI/SNF class disrupts many of the DNA–nucleosome contacts making the nucleosomal DNA vulnerable to DNA

digestion. It has been even observed that some of these complexes are capable of transfering the octamer to another DNA chain [176]. This might indicate that their mode of action is the creation of large loops (similar to the one discussed in section 2.4.3) that could lead to large repositioning steps. In fact, Bazett-Jones *et al* [177] observed that the SWI/SNF complex creates loops on naked DNA as well as on bead-on-a-string nucleosome fibres (cf the electron spectroscopic images, figures 1 and 3, in that paper). On the other hand, the mode of action of the ISWI family seems not to interrupt the nucleosome–DNA contact on an appreciable level. Since these complexes induce nucleosome repositioning it has been speculated that they might work via twist or bulge diffusion. In the meantime the latter mechanism seems to be more likely since ISWI induced nucleosome sliding appears even if the DNA is nicked and hence a torsion cannot be transmitted between the complex and the nucleosome to be shifted [178].

Another interesting and very prominent system known to mediate nucleosome repositioning is unexpectedly the RNA polymerase. It is found to be able to transcribe DNA through nucleosomes without disrupting their structure, yet moving them *upstream* the DNA template, i.e., in the opposite direction of transcription [36–40]. To rationalize this seemingly paradoxical finding, Felsenfeld *et al* [40] introduced a model which assumes that the polymerase crosses the nucleosome in a loop. This would indeed explain the backwards directionality of repositioning. Note that such a loop would have a different shape than the ones discussed above since polymerases induce a kink at the DNA with a preferential angle of $\sim 100^\circ$ [179–181]. This means, however, that an RNA polymerase sitting in an intranucleosomal loop would soon get stuck since it transcribes the DNA in a corkscrew fashion; this would complicate this mechanism [40]. It might well be that this effect only occurs on short DNA fragments as used in the experiments. If so, it would be an artefact that would not work *in vivo*. In that case another mechanism, namely induced corkscrew motion of the bound DNA towards the polymerase and subsequently the recapturing of the nucleosome at its exposed binding sites by the other end of the DNA fragment might also be a possible scenario [182].

3. 30 nm fibre

3.1. Solenoid versus crossed-linker model

Whereas the structure of the core particle has been resolved up to atomic resolution [10], there is still considerable controversy about the nature of the higher-order structures to which they give rise. When stretched, the string of DNA–histone complexes has the appearance of 'beads-on-a-string'. This basic structure can be seen clearly when chromatin is exposed to very low salt concentrations, and is sometimes referred to as the '10 nm fibre' [12]. When the ionic strength is increased towards physiological values (100 mM), the fibre appears to thicken, attaining a diameter of 30 nm [13]. Linker histones (H1 or H5) play an important role in this compaction mechanism: in their absence fibres form more open structures [12]. These strongly cationic proteins act close to the entry–exit point of the DNA. They carry an overall positive charge and seem to bind the two strands together, leading to a stem structure [23]; in fact, this stem is missing in the absence of linker histones.

There is a longstanding controversial discussion concerning the structure of the 30 nm fibre [14–17]. There are mainly two competing classes of model: the solenoid models [12, 18, 19] and the zig-zag or crossed-linker models [20–24, 183]. In the solenoid model (depicted in figure 18(a)) it is assumed that the chain of nucleosomes forms a helical structure with the axis of the core particles being perpendicular to the solenoid axis. The DNA entry–exit side faces inward towards the axis of the solenoid. The linker DNA

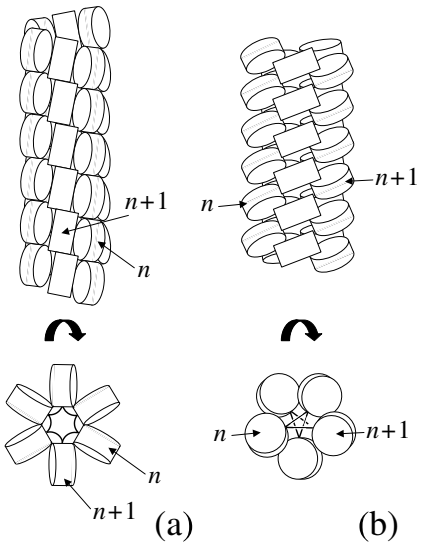

Figure 18. The two competing models for the 30 nm fibre: (a) the solenoid model and (b) the crossed-linker model. Both are shown from the side and from the top; the latter view allows us to distinguish the different linker geometries. Note that these are idealized models; real fibres are believed to be less regular [16].

(shown as a thick lines at the bottom of figure 18(a)) is required to bend in order to connect neighbouring nucleosomes in the solenoid which in turn requires strong nucleosome–nucleosome interactions to hold this structure together. The other class of model posits straight linkers that connect nucleosomes located on opposite sides of the fibre. This results in a three-dimensional zig-zag-like pattern of the linker (cf figure 18(b)).

Images obtained by electron cryomicroscopy [23] should in principle be able to distinguish between the structural features predicted by the two different models. The micrographs show a zig-zag motif at lower salt concentrations and they indicate that the chromatin fibre becomes more and more compact when the ionic strength is raised towards the physiological value. A similar picture also emerges from atomic force microscopy [22, 184]. However, neither method allows us to identify the linker geometry at physiological ionic conditions, so that one still cannot exclude the possibility that the fibre folds close to physiological conditions into a solenoid-like structure by a bending of its linkers. This is in fact the structure that is depicted in most of the standard textbooks on cell biology (e.g. [1]). X-ray diffraction data that constituted the basis for many models also lead to controversial interpretations, cf [16] for a critical discussion.

In view of this fact it is an important recent experimental achievement that single chromatin fibres can be stretched via micromanipulation techniques [81–83]. The force–extension curves allow us in principle to discern between the different structures. So far, computer simulations [86] as well as analytical approaches [24, 84] to chromatin fibre stretching seem, when comparing their predictions to the experimental data, to support the crossed linker models.

Another intriguing way that might allow us to discriminate experimentally between these two types of structure is to measure the fibre orientation in strong magnetic fields, as has already been done long ago [185]. Such a method has been used successfully to determine the persistence length of naked DNA [186, 187]. One makes use of the anisotropic magnetic susceptibility of the bp that causes the DNA double helix to orient its axis perpendicular to

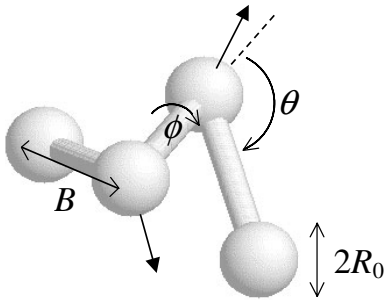

Figure 19. Fraction of a two-angle fibre containing four nucleosomes. The two angles, the deflection angle θ and the dihedral angle ϕ are depicted together with the nucleosome diameter $2R_0$ and the 'linker length' B. The arrows denote the nucleosomal superhelix axis, cf figure 2.

the field. As a consequence, single core particles orient their DNA superhelix axis parallel to the field [185]. Since the nucleosome axes in the two fibre models are oriented in different directions with respect to the fibre axis (cf figure 18), an external field would induce orientations of the two fibre models in different directions.

In the following I will focus on analytical models and computer modelling of the chromatin fibre which all belong to the class of crossed-linker models. In the next section the possible geometrical structures that follow from regular two-angle fibres (a 'generalization' of crossed-linker models) are presented. Section 3.3 gives speculation about the 'optimal' fibre design from a biological point of view. Then in section 3.4 I will give a detailed account of the mechanical properties of the fibre comparing analytical results and computer models to recent stretching experiments. Finally, in section 3.5 I report on a recent model that relates the degree of fibre swelling to the ionic strength.

3.2. Structure diagram of the two-angle fibre

To address the folding problem of DNA at the level of the 30 nm fibre myself, Gelbart and Bruinsma [24] introduced a mathematical description for the different possible folding pathways which was based on Woodcock's crossed-linker model [20] (cf also a related study on closed minichromosomes [188]). At the simplest level, we assumed that the geometric structure of the 30 nm fibre can be obtained from the intrinsic, single-nucleosome structure. The specific roles of linker elastic energy, nucleosome–nucleosome interaction, preferred binding sites, H1 involvement, etc were then treated afterwards as 'corrections' to this basic model [24, 91]. To see how single-nucleosome properties can control the fibre geometry, consider the fact that DNA is wrapped a non-integral number of turns around the nucleosome, e.g., 1- and-3/4 times (147 bp) in the case of no H1. This implies that the incoming and outgoing linker chains make an angle θ with respect to each other; the entry–exit angle π–θ is non-zero. In the presence of the histone H1 (or H5) the in- and outcoming DNA are glued together along a short section resulting in a stem-like structure [23]. While the precise value of the resulting exit-angle depends on salt concentration, presence or absence of linker histones, degree of acetylation of the histones, etc (discussed in section 3.4) one may nevertheless assume θ to be a quantity *that is determined purely at the single-nucleosome level*.

Next, there is a rotational (dihedral) angle ϕ between the axis of neighbouring histone octamers along the necklace (see figure 19). Because nucleosomes are rotationally positioned along the DNA, i.e., adsorption of DNA always begins with the minor groove turned in towards

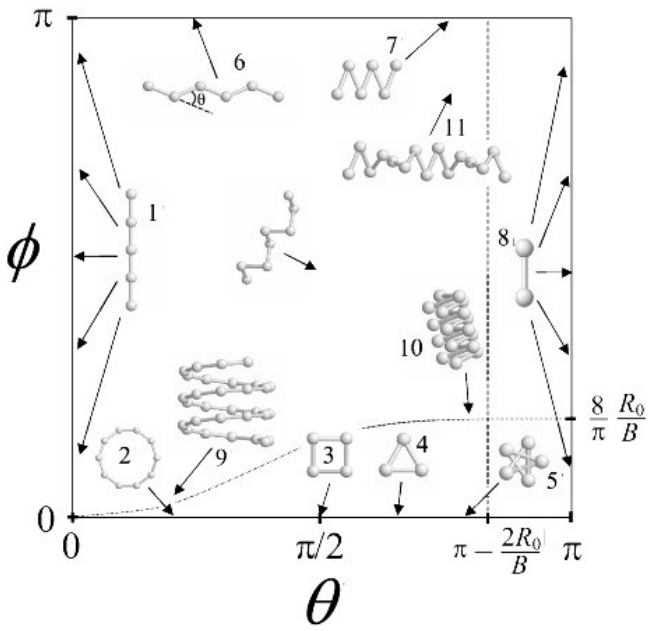

Figure 20. Diagram of geometries of two-angle fibres in the (θ, ϕ)-plane. Shown are some example configurations with the arrows denoting their position in the plane. The lines give the boundaries to the forbidden structures due to short-range excluded volume (large θ values) and long-range excluded volume (small ϕ-values).

the first histone binding site, the angle ϕ is a *periodic* function of the linker length B, with the 10 bp repeat length of the helical twist of DNA as the period. There is experimental evidence that the linker length shows a preferential quantization involving a set of values that are related by integral multiples of this helical twist [189], i.e., there is a preferred value of ϕ.

Treating the pair of angles (θ, ϕ), together with the linker length B, as *given* physical properties (even though *in vivo* they are likely under biochemical control), the geometrical structure of the necklace is determined entirely by θ, ϕ and B. The model only describes linker geometry and does not account for excluded volume effects and other forms of nucleosome–nucleosome interaction; it assumes that the core particles are point-like ($R_0 = 0$) and that they are located at the joints of the linkers. The model also assumes that the linkers are *straight*. The (θ, ϕ) model is similar to the freely rotating chain model encountered in polymer physics literature (see, for instance, [80]). The main difference is that in the present case there is no free rotation around the linker and so torsion is transmitted (see also [190]).

Before giving a detailed discussion of two-angle geometries let me provide a short overview of the possible structures in the (θ, ϕ)-space that is shown in figure 20. Both angles θ and ϕ can each vary over the range 0–π. At the edges of the diagram where one of the angles assumes an extremal value, the configurations are always planar. On the line $\phi = 0$ are located circles (see structure '2' in figure 20) and star-type polygons (that are closed for specific values of θ like '5'). The planar zig-zag structures are found on the line $\phi = \pi$ (cf '6' and '7'); for $\theta = 0$ one has straight configurations ('1') and for $\theta = \pi$ 'dimer' structures ('8'). If one moves from the line $\phi = 0$ towards larger values of ϕ the circles and star-like polygons stretch out into the direction perpendicular to their plane, forming solenoids ('9') and fibres with crossed linkers ('10'), respectively. On the other hand, if one starts at the top

of the diagram ($\phi = \pi$) and decreases the value of ϕ the planar zig-zag structure extends into the third dimension by becoming twisted ('11'). Various examples of two-angle fibres were displayed by Woodcock *et al* [20] in their figure 2, namely fibres with $\theta = 150°$ and many different values of ϕ, corresponding to a vertical trajectory on the right-hand side of figure 20. Three different configurations with a fixed value of ϕ and different values of θ are displayed in figure 3(c) in another paper by these authors [23].

An analytical description of the structures can be achieved as follows (cf [24]): it is possible to construct a spiral of radius R and pitch angle ψ such that the nucleosomes, but not necessarily the linker chain, are located on this spiral. The nucleosomes are placed along the spiral in such a way that successive nucleosomes have a fixed (Euclidean) distance B from one another. From straightforward geometrical considerations we derived in [24] analytical expressions that relate pitch angle ψ and radius R of the solenoid as well as s_0 (defined as the vertical distance between successive 'nucleosomes' along the helical axis) to the pair of angles θ, ϕ and linker length B. The corresponding relations $B = B(\psi, R, s_0)$, $\theta = \theta(\psi, R, s_0)$ and $\phi = \phi(\psi, R, s_0)$ are equations (32)–(34) in [24]. I present here the reverse relations that have the advantage that they allow the direct calculation of the overall fibre geometry from the local geometry. Specifically, the radius R of the master solenoid is given by

$$R = \frac{B \sin(\theta/2)}{2 - 2\cos^2(\theta/2)\cos^2(\phi/2)} \tag{76}$$

and its pitch angle ψ by

$$\cot\psi = \frac{\tan(\theta/2)\arccos(2\cos^2(\theta/2)\cos^2(\phi/2) - 1)}{2\sin(\phi/2)\sqrt{1 - \cos^2(\theta/2)\cos^2(\phi/2)}}. \tag{77}$$

Finally, the distance s_0 of neighbouring nucleosomes along the fibre axis is obtained from

$$s_0 = \frac{B \sin(\phi/2)}{\sqrt{\sec^2(\theta/2) - \cos^2(\phi/2)}}. \tag{78}$$

Using these relations, it is straightforward to construct a catalogue of structures.

If either one of the angles θ or ϕ assumes the value 0 or π, then the resulting structure is *planar*. Consider first the line $\phi = 0$. If one also has $\theta = 0$ the fibre forms a straight line ('1' in figure 20). For small non-vanishing θ the structure forms a *circle* of radius $R \simeq B/\theta$ (as follows directly from equation (76)). For the special case $\theta = 2\pi/n$, with n an integer, the ring contains n monomers before it repeats itself and one obtains a *regular polygon* ('2'). The special case $\theta = \pi/2$ corresponds to the square ('3'). With increasing θ the radius of the circle shrinks and approaches asymptotically the value $B/2$. For $\theta = \pi(n-1)/n$ with n being an odd integer one encounters a series of closed *star-like polygons* with n tips. In particular, $n = 3$ corresponds to the equilateral triangle ('4') and $n = 5$ to the regular pentagram ('5').

Next consider the case $\phi = \pi$ and θ is arbitrary. This case corresponds to 2D *zig-zag-like structures*, as shown by '6' and '7' at the top of figure 20. The length of a fibre consisting of N monomers is given by $L = s_0 N = B\cos(\theta/2)\ N$ (cf equation (78)) and the diameter is given by $2R = B\sin(\theta/2)$ (cf equation (76)). Note that the length of the fibre increases with decreasing θ. Finally, there are two remaining cases of planar structures: $\theta = 0$ with an arbitrary value of ϕ leads to the straight line mentioned earlier ('1'); $\theta = \pi$ and arbitrary ϕ corresponds to linkers that go back and forth between two positions ('8').

Structures with $\theta \neq 0$ *and* $\phi \neq 0$ form *three-dimensional fibres*. For small angles, $\theta \ll 1$ and $\phi \ll 1$, structures resemble *solenoids* (see '9') where the linkers themselves follow closely a helical path corresponding to that of the master solenoid. For these structures one finds from

equations (76) and (78) the following limiting behaviour of the fibre radius and length (for N monomers):

$$R \simeq \frac{B\theta}{\phi^2+\theta^2}, \qquad L \simeq \frac{BN\phi}{\sqrt{\phi^2+\theta^2}}. \tag{79}$$

Furthermore, the pitch angle ψ is given by

$$\cot\psi \simeq \frac{\theta}{\phi}. \tag{80}$$

This suggests a classification of solenoids into dense helices with small pitch angle $\psi \simeq \phi/\theta$ for $\phi \ll \theta$ and open helices with large pitch angle $\psi \simeq \pi/2 - \theta/\phi$ for $\phi \gg \theta$. Other geometrical information can be obtained easily. For instance, the vertical distance d between two turns follows from $d = 2\pi R/\cot\psi$ to be

$$d \simeq \frac{2\pi\phi B}{\phi^2+\theta^2}. \tag{81}$$

Dense helices, $\phi \ll \theta$, are characterized by $d \ll R$ and open ones by $d \gg R$.

Structures where ϕ is still small but where the entry–exit angle θ is large, i.e. $\pi - \theta \ll \pi$, form *fibres with crossed linkers*. As discussed above for $\phi = 0$ one encounters star-shape polygons that are closed for $\theta = \pi(n-1)/n$ with n odd. For *non*-vanishing $\phi \ll 1$ the star-shaped polygons open up in an accordion-like manner into a three-dimensional fibre with the following radius and length (for N monomers):

$$R \simeq \frac{B}{2\sin(\theta/2)}, \qquad L \simeq \frac{BN\phi}{2}\cot(\theta/2). \tag{82}$$

Assume now that $\theta_n = \pi(n-1)/n$ so that the projection of the fibre is a closed polygon (this is only strictly true for $\phi = 0$ but it is still a good approximation for $\phi \ll 1$). For this set of angles monomers i and $i+n$ come very close in space; their distance d follows from the master solenoid that has $n-1$ turns inbetween these two monomers:

$$d \simeq \frac{2\pi(n-1)R}{\cot\psi} \simeq \frac{\pi\phi B}{4}. \tag{83}$$

Finally structures with a rotational angle ϕ close to π, say $\phi = \pi - \delta$ with $\delta \ll 1$, lead to *twisted zig-zag structures*, see '11'. In this case monomer $i+1$ is located nearly opposite to the ith monomer, but slightly twisted by an angle δ. Monomer $i+2$ is then on the same side as monomer i but slightly twisted by an angle 2δ and so on. The geometrical properties of the resulting fibre are the following

$$R \simeq \frac{B}{2}\sin(\theta/2), \qquad L \simeq BN\cos(\theta/2) \tag{84}$$

and show only a higher order dependence on ϕ that we gave explicitly in [24]. For $\phi = \pi$ one recovers the planar zig-zag structure for which equation (84) becomes exact.

If one takes into account the excluded volume of the core particles, then certain areas in that phase diagram are forbidden; reminiscent of the familiar Ramachandran plots used in the study of protein folding [191]. For simplicity we assume in the following that the core particles are spherical with a radius R_0 and that their centres are located at the joints of two linkers, cf figure 19. There are two different types of interactions. One is between monomers at position i and $i \pm 2$ (short-range interaction), and leads to the requirement that the entry angle must be sufficiently small:

$$\theta < 2\arccos(R_0/B). \tag{85}$$

This condition excludes a vertical strip on the right-hand side of the diagram, as indicated in figure 20 by a dashed line.

There is *also* a long-range excluded volume interaction that comes into play for small values of ϕ. This is apparent for the case $\phi = 0$ where one finds planar structures that run into themselves. Starting with a circular structure one has to increase ϕ above some critical value so that the pitch angle of the resulting solenoid is large enough so that neighbouring turns do not interact. This leads to the requirement $d > 2R_0$ with d given by equation (81) (using $\phi \ll \theta$), i.e.,

$$\phi > \frac{1}{\pi}\frac{R_0\theta^2}{B}. \tag{86}$$

For the large θ-case (fibres with crossed linkers) one finds from equation (83) the condition

$$\phi > \frac{8}{\pi}\frac{R_0}{B}. \tag{87}$$

The two conditions, equations (86) and (87), shown schematically as a dotted curve in figure 20, lead to a forbidden strip in the structure diagram for small values of ϕ.

Figure 20 does not show the interesting 'fine structure' of the boundary of that forbidden strip that is due to commensurate–incommensurate effects. I already noted that there are special θ values for which the projection of the linkers forms a regular polygonal star ($\theta_n = \pi(n-1)/n$) or a regular polygon ($\theta'_n = 2\pi/n$) (for small values of ϕ). In these cases the nucleosomes i and $i + n$ 'sit' on top of each other. On the other hand, for other values of θ, monomers of neighbouring loops will be displaced with respect to each other. In this case monomers of one loop might be able to fill in gaps of neighbouring loops so that the minimum allowed value of ϕ is smaller than estimated above. We are currently exploring the interesting mathematical problem of the exact boundary line that is also sensitive to the exact nucleosome shape [192]. The dotted line in figure 20 only represents the upper envelope of the actual curve.

The above given discussion of the two-angle model was based on the assumption of a perfectly homogeneous fibre where B, θ and ϕ are constant throughout the fibre. For a discussion of the effect of randomness in these values on the fibre geometry I refer the reader to [24].

3.3. Chromation fibre: optimization of design?

If one assumes that the chromatin fibre has a relatively regular structure and that the linker DNA is straight, then the two-angle model might be a good description of the fibre geometry. In that case the question arises where in the structure diagram, figure 20, the 30 nm fibre is actually located. The diagram on its own, does not favour any structure over another. However, the diagram plus the formulae given above allow the study in [24] to invoke the following two criteria to optimize the structure of the 30 nm fibre and to check *a posteriori* their usefulness. The two suggested criteria are: (i) maximum compaction, and (ii) maximum accessibility. The first criterion is obvious: inactive chromatin should be packed as dense as possible because of the very large ratio of DNA length to nucleus size (cf also [193] to see how severe this packing problem actually is). By the second criterion we meant that a *local accessibility* mechanism is required for gene transcription and that this mechanism should somehow be optimized (see below).

In order to attain maximum compaction one needs structures that lead to high bulk densities $\rho = 1/(2\sqrt{3}R^2 s_0)$ (assuming that the 30 nm fibres are packed in parallel forming a hexagonal lattice). A comparison of the 3d densities of all possible structures shows that fibres with

internal linkers have highest densities ρ, namely (cf equation (82))

$$\rho \simeq \frac{16}{2\sqrt{3}\phi(\pi-\theta)B^3}. \tag{88}$$

In particular, the highest density is achieved for the largest possible value of θ and the smallest possible value of ϕ that is still in accordance with the excluded volume condition. This set of angles is located at the point where the dotted curve and the dashed line in figure 20 cross each other. Apparently this *also* represents the only region in the phase diagram where excluded volume effects are operative on a short-range and a long-range scale at the *same* time, i.e., nucleosome i is in close contact with nucleosome $i-2$ and $i+2$ *as well as* with nucleosomes father apart along the contour length of the necklace. This unique set of angles is given by $\theta_{max} = 2\arccos(R_0/B)$, cf equation (85), and $\phi_{min} \simeq (8/\pi)(R_0/B)$, cf equation (87).

In order to achieve maximum accessibility we looked in [24] for structures that, for a given entry–exit angle $\pi-\theta$ of a highly compacted structure, achieve the maximum reduction in nucleosome line density $\rho_L = s_0^{-1}$ for a given small change $\Delta\theta$ of the angle θ. In other words, we looked for a maximum of $\mathrm{d}\rho_L/\mathrm{d}\theta$ which we called the 'accessibility'. *Interestingly, the accessibility is maximized at the same unique pair of angles* $(\theta_{max}, \phi_{min})$. This can be seen from its angle dependence for fibres with crossed linkers

$$\frac{\mathrm{d}\rho_L}{\mathrm{d}\theta} \simeq \frac{4}{\phi(\pi-\theta)^2 B}. \tag{89}$$

Note that this change in ρ_L with θ is achieved by changing the number of monomers per vertical repeat length d. The length d itself is only weakly dependent on n according to equation (83).

The above given formulae are now compared with experimental results. For chicken erythrocyte chromatin one has $B \approx 20$ nm (centre-to-centre distance of nucleosomes, [17]). Together with $R_0 \approx 5$ nm this leads to $\theta_{max} \simeq 151°$, $\phi_{min} \simeq 36°$ and $\rho_L \simeq 6.9$ nucleosomes per 11 nm (using equations (78), (85) and (87); the approximate formula, equation (82) gives $\rho_L \simeq 6.8$). The theoretically derived values can now be compared with the experimental ones reported by Bednar *et al* [23] for chicken erythrocyte chromatin fibres. From their table 1 one finds that for an ionic strength of 80 mM (which is close to the physiological value) $\theta \approx 145°$ and $\rho_L = 6.0$ nucleosomes per 11 nm. Furthermore, electron cryotomography constructed stereo pair images of an oligonucleosome (cf figure 3(b) in [23]) indicate that the chromatin fibre might indeed have the structure of a fibre with crossed linkers, with $n \approx 5$; this would correspond to $\theta = \pi(n-1)/n \approx 144°$.

Information concerning the preferred value for ϕ can be obtained from the measured statistical distribution of the nucleosome repeat lengths. This distribution shows statistically preferred linker lengths equal to $10k + 1$ bp with k a positive integer [189], which, in turn, indicates that the rotation angle ϕ corresponds to a change in helical pitch associated with 1 bp, i.e. $360°/10 = 36°$. This value coincides with ϕ_{min}, the value that we estimated for maximum compaction. However, the statistical uncertainty around the expectation values for the nucleosome repeat length is sufficiently large to make this estimate for ϕ less reliable.

The second feature, the local accessibility, can be monitored *in vitro* by changing the salt concentration. Bednar *et al* report, for example, that θ decreases with decreasing ionic strength, namely $\theta \approx 145°$ at 80 mM, $\theta \approx 135°$ at 15 mM and $\theta \approx 95°$ at 5 mM [23]. In the biochemical context the change of θ is accomplished by other mechanisms, especially by the depletion of linker histones and the acetylation of core histone tails (cf my discussion in section 3.5), both of which are operative in transcriptionally active regions of chromatin. These mechanisms lead effectively to a decrease of θ.

As pointed out below equation (89), the decrease of θ is accompanied by a decrease of the line-density $\rho_L = n/d$ of nucleosomes at an essentially fixed value of d. In other words, the

number of vertices of the projected polygon decreases significantly with decreasing θ because $\theta_n = \pi(1-1/n)$. In that respect the effect of reducing θ below the optimal packing value might be best viewed as an 'untwisting' of the 30 nm fibre. Using the experimentally determined values of θ one finds that the line density (the number of nucleosomes per 11 nm) is given by $\rho_L \approx 5.7$ for $\theta \approx 145°$, $\rho_L \approx 4.3$ for $\theta \approx 135°$ and $\rho_L \approx 2.0$ for $\theta \approx 95°$, which is in reasonable agreement with the experimental values $\rho_L \approx 6.0$, 3.2 and 1.5 [23]. Furthermore, the number of polygonal vertices $n = \pi/(\pi - \theta)$ decreases as follows: $n \approx 5.1$ for $\theta \approx 145°$, $n \approx 4.0$ for $\theta \approx 135°$ and $n \approx 2.1$ for $\theta \approx 95°$, consistent with the stereo pair images by Bednar *et al*, suggesting $n \approx 5$ at an ionic strength of 80 mM and $n \approx 3$ at 5 mM (cf figures 3(a) and (b) in [23]).

Let me close this section with a cautionary remark [24]. The 3d density and the line density of the fibre cannot only be changed by changing θ or ϕ but also by changing the linker length (in multiples of 10 bp). A variation in B changes the location of the point $(\theta_{max}, \phi_{min})$ in the diagram of geometrical states, and thus the values of the maximum 3d and line densities that can be achieved, namely

$$\rho^{(max)} \simeq \frac{16}{2\sqrt{3}\phi_{min}(\pi - \theta_{max})B^3} \simeq \frac{\pi}{2\sqrt{3}R_0^2 B} \tag{90}$$

and

$$\rho_L^{(max)} \simeq \frac{4}{B\phi_{min}(\pi - \theta_{max})} \simeq \frac{\pi}{4}\frac{B}{R_0^2}. \tag{91}$$

This shows that fibres with smaller values of B can achieve higher 3d densities but have a smaller maximal line density (and accessibility $\mathrm{d}\rho_L/\mathrm{d}\theta \propto B^2$). From this one might infer that active cells should have larger nucleosome repeat lengths in order to maximize the accessibility to their genetic material. An overview of nucleosome repeat lengths in different organisms and tissues is given in table 7-1 of van Holde's book [14]. The data shown there do not follow this rule, unfortunately. In fact, very active cells like yeast cells and neuronal cells have in general short nucleosome repeat lengths while inactive ones like sperm cells have large ones. This shows that the optimization principle of high density and accessibility has to be used with caution.

3.4. Mechanical properties of the two-angle model

The two-angle model, as discussed in the previous sections, is purely geometrical. Could it also be useful for predicting *physical properties* of the 30 nm fibre? The response of the 30 nm fibre to *elastic stress* was indeed one of the major issues in our paper on the two-angle model [24]. In an independent study on the two-angle model by Ben-Haïm *et al* [84] this question has been the major focus. By combining in this section results from both papers we will for the first time be able to give an analytical expression for the elastic properties of the two-angle model as a function of the underlying pair of angles θ and ϕ.

Before doing so let me remark that the elastic stress can either be of external or of internal origin. External stresses are exerted on the chromatin during the cell cycle when the mitotic spindle separates chromosome pairs [194]. The 30 nm fibre should be both highly flexible and extensible to survive these stresses. The *in vitro* experiments by Cui and Bustamante demonstrated that the 30 nm fibre is indeed very 'soft' [81]. The 30 nm fibre is also exposed to *internal stresses*. Attractive or repulsive forces between the nucleosomes will deform the linkers connecting the nucleosomes. For instance, electrostatic interactions, either repulsive (due to the net charge of the nucleosome core particles) or attractive (bridging via the lysine-rich core histone tails [10]) could lead to considerable structural adjustments of the model.

Before considering the elastic properties of the two-angle model, it is helpful to briefly recall some results concerning the large-scale elasticity of the DNA itself [168, 195]. The measured force–extension curve of naked DNA breaks up into two highly distinct regimes: the 'entropic' and 'enthalpic' elastic regimes. For very low tension F ($\leqslant$pN), the restoring force is provided by 'entropic elasticity' [79]. In the absence of any force applied to its ends, the DNA's rms end-to-end distance (chain length, L) is small compared to its contour length (L_0) and the chain enjoys a large degree of conformational disorder. Stretching DNA reduces its entropy and increases the free energy. The corresponding force f increases linearly with the extension L:

$$F \simeq \frac{3k_BT}{l_P}\frac{L}{L_0}, \qquad L \ll L_0 \tag{92}$$

with $l_P \approx 500$ Å being the thermal persistence length of DNA [99].

For higher forces ($F > 10$ pN), the end-to-end distance L is close to L_0 and the elastic restoring force is due to distortion of the internal structure of DNA. In this regime, the force extension curve can be approximated by

$$F \simeq k_BT\gamma\frac{L - L_0}{L_0}, \qquad L > L_0. \tag{93}$$

The stretching modulus $\gamma = (\partial f/\partial L)L_0/k_BT$ of DNA is about 300 nm^{-1} [195, 196], i.e., almost four orders of magnitude larger than the corresponding value $3/l_P$ obtained from equation (92).

In the following I shall first discuss how the mechanical properties of the linker backbone (modelled as a two-angle fibre) can be derived analytically from its geometry. Then, in section 3.4.2 the influence of nucleosome–nucleosome interaction is considered before I compare in section 3.4.3 the theoretical results with that of stretching experiments on chromatin fibres [81–83].

3.4.1. The elasticity of the linker backbone. That the chromatin fibre is highly flexible due to the large amount of twistable and bendable linker DNA has been pointed out by myself, Gelbart and Bruinsma [24]. For a few special cases we were also able to calculate the stretching modulus of the two-linker model. A complete analysis of the elastic properties of the two-angle model has been given by Ben-Haïm *et al* [84]. In that paper the authors managed to relate the macroscopic mechanical properties of the fibre to the geometrical properties of the master solenoid (i.e. to quantities like R, s_0 and ψ). Their underlying microscopic geometrical model was more complicated since it was assumed that the linker DNA leaves the octamer as a straight line so that entering and exiting strands are displaced with respect to each other. A similar arrangement has also been assumed in the original study by Woodcock *et al* [20]. From cryo-EM pictures it is known, however, that in the presence of linker histones the entering and exiting strands are glued together in a stem [23] and this is also the situation encountered in the mechanical stretching experiments by Cui and Bustamante [81]. Therefore it might be more appropriate to model the influence of the nucleosome on the linker DNA just as inducing a kink on the DNA, as modelled in the above discussed variant of the two-angle model. Since for this case we have the exact relations between the geometrical parameters of the master solenoid (R, s_0 and ψ) and the underlying two-angle geometry, equations (76)–(78), the problem of calculating the mechanical properties of the two-angle fibre is now completely analytically solved.

Let me sketch in the following the elegant line of arguments used by Ben-Haïm *et al* [84] to determine the mechanical parameters. The basic idea is that the two-angle fibre can be

described as an extensible WLC [168–170] as already suggested in [81, 86]. In the linear response regime the relation is

$$\begin{pmatrix} F \\ M_t \\ M_b \end{pmatrix} = \begin{pmatrix} k_B T\tilde{\gamma} & k_B T\tilde{g} & 0 \\ k_B T\tilde{g} & \tilde{C} & 0 \\ 0 & 0 & \tilde{A} \end{pmatrix} \begin{pmatrix} u \\ \Omega \\ R^{-1} \end{pmatrix}.$$

Here F, M_t and M_b denote the external force and torque components: F is the force along the fibre axis, M_t the torsional torque (the torque component of $\boldsymbol{M}$ parallel to the fibre axis) and M_b denotes the flexural torque which is the torque component perpendicular to the fibre axis. These stresses are linearly related to the strains: u is the relative extension, Ω the twist rate and R^{-1} is the curvature of the fibre. The components of the stress–strain tensor give the mechanical properties of the fibre: the stretching modulus $\tilde{\gamma}$, the bending stiffness $\tilde{A}$, the torsional stiffness $\tilde{C}$ and the twist–stretch coupling constant $\tilde{g}$. These quantities follow from the underlying properties of the linker DNA that is modelled as a *non*-extensible WLC (like in equation (29)). In [84] the authors wrote down the energy density as a function of the stresses (and not of the strains as usual). Then they compared the resulting energy per linker with the energy that follows from a microscopic calculation of the fibre elastic energy (again as a function of the applied stresses). The microscopic calculation was based on the equilibrium condition for Kirchhoff rods (WLCs) [125], applied to the linker DNA: the force $\boldsymbol{f}(s)$ on the linker at any given point $\boldsymbol{r}(s)$ (s: arc length) equals the external tension

$$\boldsymbol{f}(s) = \boldsymbol{F} \tag{94}$$

and the local torque obeys

$$\boldsymbol{m}(s) = \boldsymbol{M} - \boldsymbol{v} \wedge \boldsymbol{F} \tag{95}$$

where $\boldsymbol{v}$ is the vector pointing from the fibre axis to the point $\boldsymbol{r}(s)$. With these assumptions it was possible to obtain analytical expressions for the mechanical fibre properties [84]. Specifically:

$$\tilde{\gamma} = \frac{s_0}{k_B T B} \frac{C + \Delta S \cos^2 z}{R^2 \cos^2(\eta/2)} f(\eta, z) \tag{96}$$

$$\tilde{A} = \frac{A s_0}{B} \frac{2C}{A + C - \Delta S \cos^2 z} \tag{97}$$

$$\tilde{C} = \frac{s_0}{B}\left(\frac{C}{3}\tan^2(\eta/2) + A - \Delta S \cos^2 z\right) f(\eta, z) \tag{98}$$

and

$$\tilde{g} = -\frac{s_0}{k_B T B} \frac{\Delta S \cos z \sin z}{R \cos(\eta/2)} f(\eta, z) \tag{99}$$

where[6]

$$f(\eta, z) = \frac{3A}{3A + \tan^2(\eta/2)(C + \Delta S \cos^2 z)}. \tag{100}$$

Note that all the parameters occurring in equations (96)–(100) can be deduced analytically from the two-angle geometry. Specifically R is the fibre radius, equation (76), and z denotes the angle between the fibre axis and the linker, $z = \arccos(s_0/B)$ with s_0 given by equation (78). Furthermore $\eta = \cot(\psi) s_0/R$ is the angle between neighbouring nucleosomes as seen when viewed down the fibre axis, i.e. η/s_0 is the twist rate of the unperturbed fibre. From equations (76)–(78) it follows:

$$\eta = \arccos(2\cos^2(\theta/2)\cos^2(\phi/2) - 1). \tag{101}$$

[6] In the original work [84] this function is called $K(\eta, 3)$. Note that there is a printing error in the denominator (last line of equation (16) in that paper). The factor $1/3A$ has to be removed.

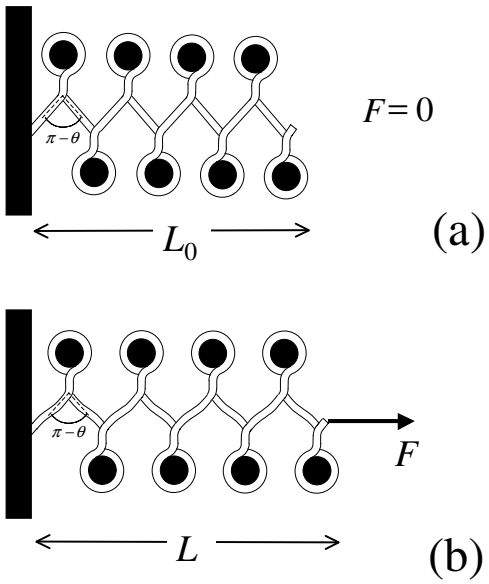

Figure 21. Stretching of a zig-zag chain. (a) The unperturbed chain, $F = 0$, with a total length L_0 has straight linker DNA. (b) The same fibre under tension $F > 0$ stretches to an end-to-end distance $L > L_0$ via the bending of its linkers.

The other parameters describe the mechanical properties of the DNA: the bending stiffness A and the torsional stiffness C as well as their difference $\Delta S = A - C$. Therefore we know now the macroscopic mechanical properties of the two-angle fibre as explicit functions of the microscopic parameters.

These functions, $\tilde{\gamma} = \tilde{\gamma}(\theta, \phi)$, $\tilde{A} = \tilde{A}(\theta, \phi)$, $\tilde{C} = \tilde{C}(\theta, \phi)$ and $\tilde{g} = \tilde{g}(\theta, \phi)$ are, however, rather unwieldy. To get an idea of the overall behaviour one might resort to numerical calculations as done in [84], where the mechanical moduli of the fibre were calculated as a function of ϕ for two values of θ, cf figure 11 in that paper. It was found that the moduli vary strongly with ϕ (and thus with the linker length) and it was argued that this strong dependence might be used in the biological context as a regulatory factor.

Having the analytical relations at hand, another approach has now become available, namely to look at limiting cases (solenoids, fibres with crossed linkers and zig-zag structures) which show simple dependences on the underlying geometry (i.e., on the angles θ and ϕ), as discussed in section 3.2. Not surprisingly $\tilde{\gamma}$, $\tilde{A}$, $\tilde{C}$ and $\tilde{g}$ are also simple functions of these underlying angles in all the limiting cases. We will give a complete overview in a forthcoming publication [192]. Here I will restrict myself to two limits only.

Let me start with the *planar zig-zag fibre*. Such a chain can be stretched via the bending of its linkers, maintaining the deflection angle θ at each kink, cf figure 21. It is also clear that the linker will not be twisted in this planar geometry. In [24] we calculated the stretching modulus $\tilde{\gamma}$ for this special arrangement. In order to do so we wrote down the elastic energy of the linker (similar to equation (29)) and determined the deformed shape (for small perturbations) from the corresponding Euler–Lagrange equation taking the boundary conditions into account. We found

$$\tilde{\gamma} = \frac{12A\cos(\theta/2)}{k_B T B^2 \sin^2(\theta/2)}. \tag{102}$$

This result can now also be obtained directly from the general formula, equation (96), by simply setting $\phi = \pi$. It is evident from equation (102) that the stretching occurs via linker bending since $\tilde{\gamma}$ depends on A only whereas the DNA torsional stiffness C does not enter the expression. I also note that the planar zig-zag fibre shows other interesting features, especially two different persistence lengths for bending in plane and bending out of plane as discussed in [24, 84, 197]. General features of such polymers with highly anisotropic bending rigidities have been considered by Nyrkova *et al* [198].

Now I consider the case that might be of importance for 30 nm fibres: the chains with crossed linkers ($\phi \ll 1$, $\pi - \theta \ll \pi$). Starting from the general expressions it is straightforward to show that the linker geometry leads in this case to the following overall mechanical properties:

$$\tilde{\gamma} \simeq \frac{3A}{k_B T B^2}\phi(\pi - \theta) \tag{103}$$

$$\tilde{A} \simeq \frac{AC}{A + C}\frac{\phi(\pi - \theta)}{2} \tag{104}$$

$$\tilde{C} \simeq \frac{A\phi(\pi - \theta)}{4} \tag{105}$$

and

$$\tilde{g} \simeq -\frac{3A\Delta S}{16k_B T C B}\phi^2(\pi - \theta)^3. \tag{106}$$

It can be seen from equation (103) that stretching occurs via linker bending (as in the case of zig-zag fibres) and from equation (105) that also the *twisting* of the overall fibre is achieved via the *bending* of the linkers. The dependence of $\tilde{A}$ on the DNA parameters, equation (104), shows that fibre bending involves both bending and twisting of the linkers, a fact that is due to the different orientations of individual linkers with respect to the bending direction. Finally, the twist-stretch coupling is very small (cf the angle dependence in equation (106)).

The elasticity of the linker backbone is predicted to be very soft. For instance, the stretching modulus $\tilde{\gamma}$ scales for fibres with crossed linkers and zig-zag chains as $A/(k_B T B^2)$. This is of the order one (per nanometre) for an effective linker length of 20 bp as compared to a γ of $\sim$300 nm^{-1} for free DNA (see above). Of course, depending on the values of θ and ϕ, this value varies over a wide range. The other mechanical parameters of the two-angle fibre also indicate an extremely soft structure. Because of this it is evident that the presence of the nucleosomes play a crucial role in determining the mechanical properties of the 30 nm fibre. The excluded volume will not allow a strong bending that would lead to overlapping nucleosomes and the nucleosome–nucleosome attraction counteracts the stretching of the fibre under external tension. Therefore before I compare in section 3.4.3 the theoretical expressions and the results from fibre stretching experiments, it is indispensable to discuss first how the nucleosomes modify the mechanical properties of the two-angle fibre.

3.4.2. Role of the nucleosome interaction. The effect of attractive interaction between nucleosomes is to cause a *compression* of the 30 nm fibre. Phase behaviour studies of linker-free nucleosome solutions, i.e., solutions of disconnected nucleosomes [199] (cf also [200]) indicate that nucleosome core particles spontaneously form fibre-like *columnar structures*, presumably due to attractive nucleosome–nucleosome interaction. Attractive nucleosome interaction could be mediated for instance by the lysine-rich core histone tails [10], as mentioned above.

Let me first discuss the role of this internucleosomal attraction on the stretching elasticity of a fibre. Following [24] the special case of a planar zig-zag structure with elastic linkers is considered where a short-range interaction between nucleosomes is also assumed. This

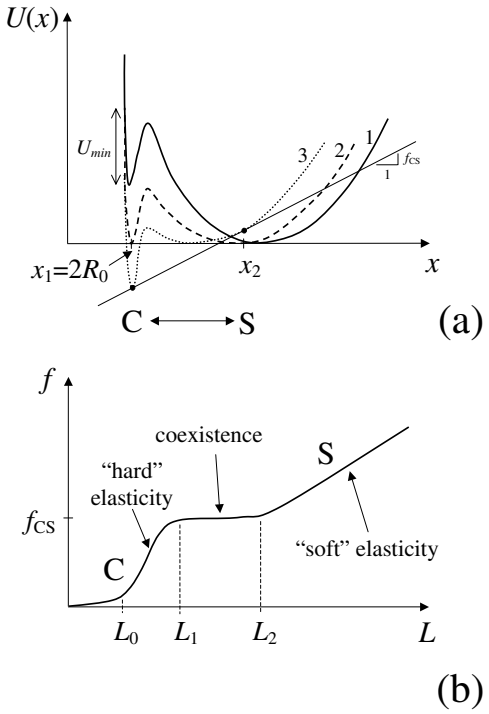

Figure 22. (a) Internucleosomal interaction potential U between nucleosomes i and $i+2$ as a function of distance x. In addition to the elastic contribution there is a short-range attraction for nucleosomes within close contact, $x = 2R_0$. The different curves correspond to different values of the angle θ. Curve '1' has the global minimum at large x (swollen state 'S') whereas curve '3' has the minimum for nucleosomes in close contact (condensed state 'C'). Curve '2' corresponds to the transition point. Also depicted is the common tangent for curve '3'. Its slope corresponds to the critical stretching force f_{CS} at which nucleosomes are transferred from the state C to state S. (b) Force-extension curve of a condensed fibre. For extensions L with $L_1 < L < L_2$ one finds a coexistence plateau with the restoring force f_{CS}.

interaction, denoted by U_{inter}, is taken to be a short-range attraction, of strength $-U_{min}$, that acts only when the nucleosomes are in close contact, i.e., at a distance $x \approx 2R_0$ of the order of the hard-core diameter. For a given nucleosome, say the ith, the closest nucleosomes in space are numbers $i+2$ and $i-2$ (cf figure 21). The interaction between other pairs is disregarded. The elastic interaction $U_{elastic}$ follows directly from equation (102) applied to a trinucleosome ($N = 2$):

$$U_{elastic}(x) = \frac{3}{\sin^2(\theta/2)}\frac{A}{B^3}(x - x_0)^2 = \frac{\tilde{K}}{2}(x - x_0)^2 \tag{107}$$

where $x_0 = 2B\cos(\theta/2)$ denotes the distance between nucleosomes i and $i+2$ for straight linkers (cf equation (84)). The total internucleosomal $U(x)$ equals $U_{inter}(x) + U_{elastic}(x)$.

Figure 22(a) shows $U(x)$ for different values of θ. Let me assume for simplicity that the interaction energy U_{inter} remains unchanged. Curve '1' in figure 22(a) shows $U(x)$ for a small value of θ where the global minimum of $U(x)$ is located at $x = x_0$ denoted by 'S' (swollen state). Curve '2' corresponds to an intermediate value of θ at which the minima at 'S' and 'C' have the same value. For this value of θ, $\theta = \theta_c$, the energy minimum shifts from 'S' to a new minimum, representing the condensed state 'C'. The change in θ produced a *structural*

transition from a swollen state to a condensed state. Finally, curve '3' depicts $U(x)$ for a deflection angle $\theta > \theta_c$ with the minimum at 'C'. The critical angle for the 'S' to 'C' transition can be determined by comparing the bending energy at close contact, $U_{elastic}(2R_0)$, and the strength U_{min} of the short-range attraction. Equating both leads to the following condition for θ_c:

$$\cos(\theta_c/2) - \sqrt{\frac{BU_{min}}{12A}}\sin(\theta_c/2) = \frac{R_0}{B}. \tag{108}$$

In the swollen state the elastic properties should be in principle the ones that were discussed in the previous section. In the condensed state, the elastic properties are determined by the detailed form of the nucleosome interaction potential.

If the condensed state has a lower free energy, i.e. if $\theta > \theta_c$, then an external stretching force f can induce a transition from the condensed to the swollen state. The transition point f_{CS} follows from a 'common-tangent' construction. The conditions are $U'(x_1) = U'(x_2) = f_{CS}$ and $(U(x_2) - U(x_1))/(x_2 - x_1) = f_{CS}$ (cf figure 22(a)) leading to [24]

$$f_{CS} = \sqrt{2KU_{min}} - \tilde{K}(x_0 - 2R_0). \tag{109}$$

The corresponding force–extension curve has a 'coexistence plateau', cf figure 22(b). If the imposed end-to-end distance is smaller than L_0 (the contour length of the condensed fibre) then the restoring force is entropic. For $L_0 < L < L_1$ the force rises sharply with increasing L. This 'hard elasticity' is governed by the nucleosomal interaction potential U_{inter}. Then at $L = L_1$ the coexistence plateau is reached. Between $L = L_1$ and $L = L_2$ parts of the fibre are in the 'S' state and parts are in the 'C' state. For larger extensions, $L > L_2$, the fibre shows soft elasticity due to the bending and twisting of the linkers as discussed in the previous section.

Katritch *et al* [86] presented a Monte Carlo simulation of the chromatin fibre that was based on a model very similar to the two-angle model. The nucleosomes were modelled as spheres and attached to the kinks in the linker backbone via a short stem. The only difference from the above discussed two-angle model was that the rotational angle ϕ between each pair of nucleosomes was chosen randomly from the interval $-\pi$–π. These fibres were then stretched as in a micromanipulation experiment [81] and their force–extension relationships were measured. The values for $\tilde{\gamma}$ were in good agreement with what is expected on theoretical grounds (a detailed discussion is given in appendix D of [24]). What is of special interest here is that they also studied the effect of a short-range nucleosome–nucleosome attraction. Using a value U_{min} of order 2 k_BT (or larger) they observed very clearly the occurrence of a pseudoplateau in the force–extension curve similar to figure 22(b).

The nucleosomes also have a big effect on the persistence length $\tilde{l}_P$ of the fibre. This has been demonstrated most clearly in a computer simulation by Wedemann and Langowski [87] (cf also an earlier preliminary study of this group [201]). Their model is again very closely related to the two-angle model discussed above. Differences are that the entering and exiting DNA at the nucleosome are slightly displaced in the direction of the nucleosome axis and that the screened electrostatic interaction between linkers was taken explicitly into account. Nevertheless, equation (97) should be expected to give a good estimation of the contribution of the linker DNA to the fibre persistence length. Using the values of that simulation ($\theta \simeq 143°$, $\phi \simeq 80°$, $B = 10$ bp) gives $\tilde{l}_P \simeq 13$ nm. However, the persistence length observed in the simulation is 265 nm, i.e. 20 times larger! This is clearly an effect of the nucleosomes. The role of the linkers is to bring the nucleosomes into contact. The nucleosomes (modelled here as ellipsoids) experience then in addition an attractive force where U_{min} has been chosen to be of order k_BT. This leads to a very dense structure with the nucleosomes in contact so that

there is hardly any space for fibre bending. Most clearly this is seen in figure 9 of [87] that shows a contraction of a fibre that has been stretched out first. As long as the nucleosomes are not in contact the fibre shows sharp bends and strong shape fluctuations. The fibres stiffen very strongly as soon as the dense state is reached.

This all shows that the nucleosome interaction is a crucial element determining the mechanical fibre properties. Therefore a more microscopic model taking details of the nucleosome structure into account might be important for a theoretical prediction of the properties of chromatin fibres. A first step in this direction has been made by Beard and Schlick [202]. They performed a molecular dynamics simulation where the nucleosomes were represented by disks made of several hundred charges that were chosen to match the crystal structure [10]. Di- and trinucleosomes as well as whole fibres have been studied. The authors demonstrated also that a fibre with a crossed linker geometry unfolds into an open zig-zag fibre as a result of changing ionic conditions. However, in their nucleosome model they neglected most of the histone tails that constitute very likely the crucial ingredient for the nucleosome–nucleosome attraction.

3.4.3. Stretching chromatin. It is now possible to measure the mechanical behaviour of single chromatin fibres via micromanipulation techniques as has been demonstrated in three studies [81–83]. Each of the studies focused on a different variant of the 30 nm fibre. Cui and Bustamante [81] stretched native chicken erythrocyte chromatin fibres containing linker histones and contrasted the cases of low and high ionic strength. For low ionic strength (5 mM NaCl) it was found that fibres are very soft. By fitting their data to that of an extensible WLC they found a stretching modulus of $k_BT\tilde{\gamma} \approx 5$ pN and a persistence length of $\tilde{l}_P \approx 30$ nm. The theoretical values are $k_BT\tilde{\gamma} \simeq 6.3$ pN (from equation (96)) and $\tilde{l}_P \simeq 16$ nm (from equation (97)) for $\phi = 36°$ and $\theta = 95°$ (cf section 3.3) and for a linker length of 20 bp (chosen on the basis of the 210 bp repeat length of chicken erythrocyte chromatin [14] minus roughly 190 bp associated with the core and linker histones). The theoretical and experimental values are close which indicates that the mechanical properties of a swollen fibre at low salt concentrations are mainly determined by the elasticity of its linker backbone. The nucleosomes are less important since they are not close enough in such a swollen fibre. For high stretching forces around 20 pN there is an irreversible change in the overall length of the fibre due to histone 'evaporation' which has been seen more clearly in the other two stretching experiments (see below).

Stress–strain curves for fibres at higher ionic strength (40 mM NaCl) are also reported in [81]. In this case the fibre is much denser and nucleosomes approach each other closely. Attractive short-range forces and the increase of θ associated with higher ionic strength should favour the condensed phase. Indeed a plateau appears at 5 pN in the force–extension curve (cf figure 4 of [81]). From the extent of the plateau, 0.6 μm, its height, 5 pN, and the number of nucleosomes in the stretched fibre, ≈280, it was estimated that there is an attractive interaction energy of $\sim 3\ k_BT$ per nucleosome [81]. In [24] we used equation (109) to independently estimate the strength of the internucleosomal attraction from the value of the critical force alone. Neglecting the second term in that equation one finds $U_{min} \approx f_{CS}^2/(2\tilde{K}) \approx 3\ k_BT$ assuming $\theta = 140°$ and again $B = 20$ bp. Note that one finds with these values that the stretching modulus of the linker backbone, equation (96), is of order $k_BT\tilde{\gamma} \simeq 4.7$ pN, i.e., even lower than the value 6.3 pN predicted above for low ionic strength. The fact that these fibres appear much stiffer with respect to (small) deformations indicates that its mechanical properties are mainly determined by the nucleosome–nucleosome attraction and *not* by the backbone elasticity. This is also in accordance with the estimated large value of the persistence length of condensed chromatin fibres (∼260 nm, cf [203]).

Bennink *et al* [82] assembled chromatin fibres by exposing λ-DNA to *Xenopus* egg extract before stretching them[7]. This extract contains core histones but no linker histones. There is, however, an abundance of non-histone proteins, some of which act close to the DNA entry–exit point similar to linker histones. For small forces a stretch modulus ∼150 pN was extracted from the data. This indicates again that chromatin fibres are quite stiff at high salt concentrations (here 150 mM NaCl) compared to the pure linker backbone elasticity but still quite soft compared to naked DNA ($k_BT\gamma \simeq 1200$ pN). For stretching forces of order 20 pN irregular sawtooth-like fluctuations were observed, each being a result of a sudden fibre lengthening by multiples of ∼65 nm. This was attributed to the unravelling of single or multiple nucleosomes.

Finally, Brower-Toland *et al* [83] used well characterized fibres that were prepared from tandem repeats of the 5s rDNA positioning sequence and core histones (no linker histones). As shown in [29] (cf my discussion of that paper in section 2.4.1) most of the nucleosomes are localized at the preferred positions. Around 20 pN the force–extension curve showed a regular sawtooth pattern reminiscent of the one observed during the unravelling of tandem repeat domains in the protein titin [206, 207]. The spacing of the peaks, ∼27 nm, is indicative of the unravelling of only one turn of the nucleosomal DNA. The outer sections of the DNA were detected to be released at much smaller forces. Brower-Toland *et al* [83] gave also a theoretical explanation of this observation: they speculate that the first 76 bp are unwrapped much more easily due to weaker binding between DNA and the octamer whereas strong binding sites occur as soon as these first 76 bp are unwound. Only when a sufficiently large force is applied are these binding sites broken on the timescale of the experiment.

However, this explanation is questionable since it does not take into account the actual unwinding geometry. Cui and Bustamante [81] have already pointed out that it requires a twisting of the core particle to unwrap the inner part of the nucleosomal DNA. The free DNA has to be bent strongly close the nucleosome which leads to a considerable barrier that has to be crossed during unwrapping. We are currently calculating this barrier analytically in the WLC framework [208]. This high barrier might explain why the nucleosomes are dissociated only at surprisingly high tensions, even in the absence of linker histones. In fact, as Marko and Siggia pointed out in [90], one would find nucleosome release (and a corresponding plateau in the force–distance curve) around 2 pN, if such a barrier could be neglected. This value follows from the comparison of adsorption energy (∼30 k_BT, cf section 2.1) to wrapping length (∼50 nm): $f \approx 30\,k_BT/50\text{ nm} \approx 2$ pN.

3.5. Fibre swelling

In this section I discuss how the entry–exit angle $\alpha = \pi - \theta$ of the DNA at the nucleosomes is controlled via electrostatics. A more detailed account on this subject is provided in [91]. As mentioned above it can be seen in cryo-EM studies [23] that the fibres open up and therefore become more accessible when the ionic strength is reduced and that this opening is directly linked to an increase in α. It was suggested that via other mechanisms (for instance, the acetylation of the lysine-rich histone tails [17], as explained in more detail below) the angle α and therefore the degree of swelling can be changed for a given section of the fibre and that this constitutes a biochemical means of controlling the transcriptional activity of genes.

Whereas the x-ray studies of the core particle [10] permit detailed knowledge of the wrapped part of the DNA it does not give insight into the conformational properties of the entering and exiting strands. One has therefore to refer to electron cryomicrographs. In these

[7] A detailed discussion of the kinetics of the chromatin assembly in this kind of experiment is given in [204]; cf also [205].

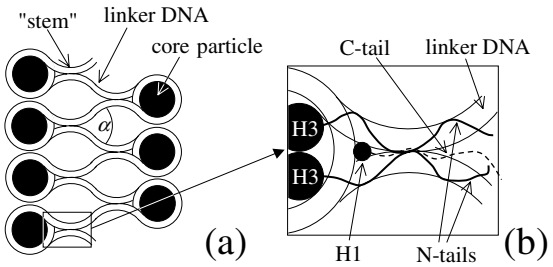

Figure 23. (a) Schematic view of a section of the 30 nm fibre (for simplicity shown here as a two-dimensional zig-zag). (b) Enlarged view of the stem region showing a speculative model of the role of the H1 histone and some N-tails from the core histones, cf also figure 7 of [184].

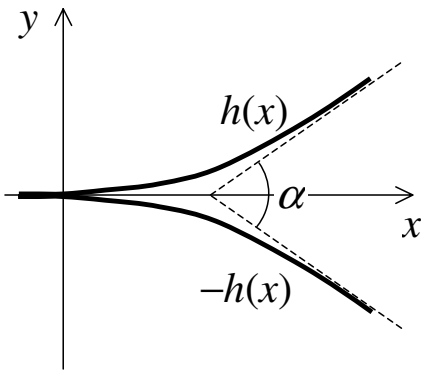

Figure 24. Idealized model for the entry–exit region of the DNA at the nucleosome. The thick curved lines represent the entering and exiting DNA that enclose a well-defined angle α that in turn determines the overall geometry of the 30 nm fibre.

micrographs it can be seen clearly that 10 nm stretches of the entering and exiting DNA strands are glued together forming a unique 'stem motif' [23] (cf also figure 23(a)). The gluing of the two equally charged chains is accomplished, amongst other things, via the linker histone H1 as shown schematically in figure 23(b).

At physiological concentrations the electrostatics is essentially short-range ($\kappa^{-1} \simeq$ 10 Å for 100 mM salt). It seems therefore reasonable to assume that α is set within the small region where the two linker DNA are in close contact, i.e., within the stem region. This value of α in turn controls the large-scale secondary structure of chromatin, the 30 nm fibre, as discussed in section 3.2. To mimic this situation I assumed in [91] a geometrical arrangement where two parallel DNA strands are held together tightly at $y = 0$ for $x \leqslant 0$ and are free for $x > 0$, cf figure 24. Because of their mutual electrostatic repulsion the two strands bend away from each other. When the two strands are far enough from each other their interaction is screened so that they asymptotically approach straight lines defining the opening angle α as indicated in figure 24.

The conformation of the upper DNA chain can be described by the height function $h(x)$. By symmetry the position of the lower strand is then given by $-h(x)$. To mimic the stem two boundary conditions are imposed at $x = 0$, namely $h(0) = h'(0) = 0$. The entry–exit angle α is related to the slope of $h(x)$ at infinity via $\tan(\alpha/2) = h'(\infty)$. The two DNA chains are modelled as semiflexible polymers with persistence length l_P and line-charge density $-e/b$

that interact via a screened electrostatic potential. The free energy of the system is then given by

$$\frac{\mathcal{F}\{h(x)\}}{k_B T} \simeq \int_0^\infty \mathrm{d}x \left[l_P \left(\frac{\mathrm{d}^2 h}{\mathrm{d}x^2} \right)^2 + \frac{2 l_B}{b^2} K_0(2\kappa h(x)) \right]. \tag{110}$$

The first term in the integral accounts for the bending of the *two* DNA strands and the second term describes the interaction between the two chains ($K_0(x)$ being the $0th$ order modified Bessel function). Here the interaction of a given charge on one chain with all the charges on the other chain is approximated by the interaction of this charge with a *straight* chain at the distance $2h$[8]. The conformation of the upper chain, $h(x)$, is then the solution of the corresponding Euler–Lagrange equation

$$l_P \frac{\mathrm{d}^4 h}{\mathrm{d}x^4} - \frac{2 l_B \kappa}{b^2} K_1(2\kappa h) = 0 \tag{111}$$

together with four boundary conditions: two are given at the origin (see above) and two follow from the condition of straight 'linkers' at infinity: $h''(\infty) = h'''(\infty) = 0$. Defining $\tilde{h} = 2\kappa h$ and introducing the dimensionless quantity $\tilde{x} = (4 l_B \kappa^2 / b^2 l_P)^{1/4} x$, equation (111) can be rewritten as $\mathrm{d}^4\tilde{h}/\mathrm{d}\tilde{x}^4 = K_1(\tilde{h})$ with a solution showing the dimensionless asymptotic slope $c_0 = \mathrm{d}\tilde{h}/\mathrm{d}\tilde{x}|_{\tilde{x}=\infty}$. It follows immediately that $\tan(\alpha/2)$ is given by

$$\tan(\alpha/2) = h'(\infty) = \frac{c_0}{\sqrt{2}} \left(\frac{l_B}{l_P} \right)^{1/4} \left(\frac{1}{\kappa b} \right)^{1/2} \tag{112}$$

where c_0 is of order one (cf [91] for details).

This result might help to understand better how the local electrostatics controls the geometry of the chromatin fibre *in vitro* and, on a more tentative level, *in vivo*. The *in vitro*-experiments show that chromatin fibres 'open up' with decreasing salt concentrations. As already mentioned in section 3.3 it was estimated from electron cryomicrographs that $\alpha_{exp} \approx 85°$ for $c_s = 5$ mM and $\alpha_{exp} \approx 45°$ for $c_s = 15$ mM and from electron cryotomography that $\alpha_{exp} \approx 35°$ for $c_s = 80$ mM [23] (cf [17] and [210] for other approaches to determine α). One expects from equation (112) that $\alpha \simeq 2 \arctan(C c_s^{-1/4})$ with C being a constant. Let me take the angle at the highest salt concentration, $c_s = 80$ mM, as the reference value. From this follows $C = 0.94$. With this value of C the prediction is $\alpha \approx 51°$ for $c_s = 15$ mM and $\alpha \approx 64°$ for $c_s = 5$ mM. Whereas the theoretical value $\alpha \approx 51°$ at intermediate ionic strength is close to $\alpha_{exp} \approx 45°$, the value $\alpha \approx 64°$ for low salt concentrations is noticeably too low ($\alpha_{exp} \approx 85°$). However, as mentioned above, for such a large value of α the chain–chain repulsion is underestimated by ∼20%.

How can the degree of swelling of the chromatin fibre be controlled *in vivo*? Under the assumption that the above mentioned geometry is valid the only parameter that might be under biochemical control is the linear charge density b^{-1}. It is known that the formation of a dense chromatin fibre is dependent on the presence of several components, especially of the cationic linker histones and of some of the lysine-rich (i.e., cationic) N-tails of the core histones that appear to be long, flexible polyelectrolyte chains [10]. In figure 23(b) I give a tentative picture of the conformation of two N-tails that protrude from the histone core. It is known that if either of these components is missing the fibre does not fold properly (cf [17] and references therein). As indicated in the figure the tails might form a complex with the entering and exiting

[8] This can be shown [91] to be a good approximation as long as $l_P \gg l_{OSF} = l_B/4b^2\kappa^2$ (the Odijk–Skolnick–Fixman length that describes the electrostatic stiffening of the chain, cf [209]). In fact, it is also precisely that limit at which the *intra*molecular interaction can be neglected (as done here). One can reformulate the above condition in the simple form $\alpha \ll 1$ (cf equation (112) below); for instance it is found that for $\alpha = 45°$ the approximation is still excellent and for $\alpha = 90°$ the chain–chain repulsion is overestimated by ∼20%.

linker DNA in such a way that they effectively reduce its linear charge density b^{-1}. It is known that transcriptionally active regions in chromatin show an acetylation of the core histone tails (i.e., the cationic groups of the lysines are neutralized). In that tentative picture this acetylation mechanism would increase b^{-1} and according to equation (112) this would lead to an opening of the entry–exit angle α. The acetylation might therefore be the first step in the decondensation of a stretch of the chromatin fibre that needs to be accessed for transcription.

Let me note that processes that are involved in the acetylation and de-acetylation are quite specific and involved as, for instance, discussed in [211]. The histone tail modifications might serve specific functions via the modification of their secondary structure that in turn modifies their interaction with certain proteins [212]. Recently there have even been attempts to decipher a specific 'language' of covalent histone modifications [213]. It might be that such specific processes act in concert with the more basic charge neutralization principle discussed here.

4. Conclusions and outlook

Chromatin is of fundamental importance for a host of biological processes ranging from gene expression to cell division. Consequently there is huge research activity among biologists in this area. For physicists chromatin has also become of interest since there are more and more experiments available that work under quite well-defined conditions. Such experiments typically involve only a few components (DNA, histone proteins, etc) but no active protein 'machines'. These experiments either focus on elucidating properties of single nucleosomes or of beads-on-a-string complexes ('chromatin fibres'). They study the behaviour of these systems under changing ionic conditions and/or under an externally applied tension. Also the dynamics of these systems that is solely driven by thermal fluctuations is investigated. Theoretical treatments and computer simulations that capture the essential features of the chromatin system are now possible and thus allow us to estimate the energy and timescales occurring in chromatin. Other approaches look at simplified model systems and try to identify general physical principles that govern complexes of charged chains and macro-ions. With the better understanding of the mechanical and dynamical properties of nucleosomes and chromatin fibres one hopefully also gains a deeper insight into more complicated questions like the working of chromatin remodelling complexes, the interaction between RNA polymerase and nucleosomes, etc.

To proceed in this direction it is crucial to obtain reliable numbers from experiments. One energy scale that dominates many processes is the adsorption energy of DNA on the octamer that has now been measured quite directly through stretching experiments. Another important feature, especially in 30 nm fibres, is the nucleosome–nucleosome interaction energy. Again detailed experimental studies have been performed and await a detailed theoretical treatment.

Chromatin is a very active and exciting field in biology where tremendous progress has been made in recent years. I hope that at least a few of the ideas gained from the physical models will be of help to biologists to develop a clearer picture of the working of chromatin and to design appropriate experimental setups.

Acknowledgments

The author thanks R Bruinsma, W M Gelbart, I Kulić and J Widom for many valuable discussions. Useful conversations with R Everaers, K Kremer, J Langowski, F Livolant, S Mangenot, G Maret and B Mergell are also acknowledged.

Appendix A. Rosette in d dimensions

The heterogeneity found for open loops (cf equation (44)) is reminiscent of a phase coexistence. To clarify why the loop sizes are so sensitive to small details, especially why there is no 'phase separation' for closed chains (cf equation (41); even though $\mathrm{d}^2 f/\mathrm{d}l^2 < 0$ at larger separations), I will present here some unpublished results in which the free energy is formally recomputed for arbitrary space dimensions d. In order to do so one has to replace the entropy term in equation (38) by $(d/2)\ln(l/l_P)$. Also in that case one obtains an analytical expression for the partition function, namely (for an open chain)

$$Z_M = \left[\frac{2l_P^{\frac{d+2}{4}}}{r}\left(\frac{2\chi k_B T}{P}\right)^{\frac{2-d}{4}} K_{\frac{d}{2}-1}\left(\sqrt{\frac{8\chi l_P P}{k_B T}}\right)\right]^M \left(\frac{k_B T}{rP}\right)^2 \tag{A.1}$$

with $K_\nu(x)$ denoting the modified Bessel function of νth order. In the case of large 'pressure' P, $l_P P/k_B T \gg 1$, it follows from the asymptotic form of $K_{d/2-1}(x)$ for large x, $K_{d/2-1}(x) \simeq \sqrt{\pi/2x}\mathrm{e}^{-x}$, that the leading term of the resulting free energy $G(P)$ is independent of d. Therefore one recovers the 3D case, i.e., equations (43)–(45) for $L/M \ll Ml_P$. For the low 'pressure' regime, $l_P P/k_B T \ll 1$, I use the asymptotics $K_{d/2-1}(x) \simeq 2^{-1}\Gamma(|d/2-1|)(2/x)^{|d/2-1|}$ for $x \ll 1$ and $d \neq 2$. This leads to the following asymptotic behaviour of the partition function

$$Z_M \simeq \begin{cases} \left[\Gamma\left(\dfrac{2-d}{2}\right)\dfrac{l_P^{d/2}}{r}\left(\dfrac{k_B T}{P}\right)^{\frac{2-d}{2}}\right]^M \left(\dfrac{k_B T}{rP}\right)^2 & \text{for } d < 2 \\ \left[\Gamma\left(\dfrac{2-d}{2}\right)\dfrac{l_P}{r}(2\chi)^{\frac{2-d}{2}}\right]^M \left(\dfrac{k_B T}{rP}\right)^2 & \text{for } d > 2. \end{cases} \tag{A.2}$$

It follows then from $L = \partial G/\partial P$ that P is given in leading order by

$$P \simeq \begin{cases} \left(\dfrac{2-d}{2}M + 2\right)\dfrac{k_B T}{L} & \text{for } d < 2 \\ \dfrac{2k_B T}{L} & \text{for } d > 2. \end{cases} \tag{A.3}$$

The average leaf size can in principle be calculated, as before, from Z_1. Here, however, it turns out to be more convenient to calculate $\langle l_{leaf}\rangle$ directly:

$$\langle l_{leaf}\rangle = \frac{\int_0^\infty \mathrm{d}l\, l^{1-d/2}\exp\left(-\frac{2\chi l_P}{l} - \frac{Pl}{k_B T}\right)}{\int_0^\infty \mathrm{d}l\, l^{-d/2}\exp\left(-\frac{2\chi l_P}{l} - \frac{Pl}{k_B T}\right)} = \sqrt{\frac{2\chi l_P k_B T}{P}}\frac{K_{2-d/2}(\sqrt{8\chi l_P P/k_B T})}{K_{1-d/2}(\sqrt{8\chi l_P P/k_B T})}. \tag{A.4}$$

Now using the above given power law behaviour of the Bessel function together with equation (A.3) the average leaf size follows:

$$\langle l_{leaf}\rangle \simeq \begin{cases} \dfrac{L}{M + \frac{4}{2-d}} & \text{for } d < 2 \\ \dfrac{\Gamma(2-d/2)}{\Gamma(d/2-1)}\dfrac{\chi^{d/2-1}}{2^{3-d}}\left(\dfrac{l_P}{L}\right)^{\frac{d-2}{2}} L & \text{for } 2 < d < 4 \\ \dfrac{4\chi l_P}{d-4} & \text{for } d > 4. \end{cases} \tag{A.5}$$

First note that for $d < 2$ the leaf size is set by the overall length of the chain but does not depend on l_P; on the other hand, for $d > 4\langle l_{leaf}\rangle$ it is solely determined by l_P. Speaking in the

picture of interacting particles on a track of length L one can explain these two extreme cases as follows. For $d < 2$ the increase of the 'nearest neighbour pair potential' beyond a distance l_P (given by $(d/2)\ln(l/l_P)$) is too small to keep the particles together; instead they explore all available space. For higher space dimensions than 4 the prefactor of the log-term is large enough to keep neighbouring particles close to the ideal distance $\sim l_P$ given by the shallow minimum of $f(l)$, equation (38). The case $d = 3$, which I have already given in equation (44) (a result recovered in equation (A.5)), is an intermediate case where $\langle l_{leaf}\rangle$ reflects the overall chain length L as well as the position $\sim l_P$ of the shallow minimum. Note further that in the limit $L \to \infty$ the average size per leaf goes to infinity for $d < 4$ (but the 'particles' will only be spread out over the whole volume, $M\langle l_{leaf}\rangle \approx L$, for $d < 2$).

Concerning the role of dimensionality one also gains some insight by the following simple argument (similar to the famous Onsager–Manning argument for the condensation of counterions on an infinitely long charged rod [111]). Consider a pair of particles at distance l in one dimension that attract each other via $f(l) = k_BT(d/2)\ln(l/l_P)$. Now assume that the particles move further apart from the distance l_1 to the distance $l_2 > l_1$. This leads to an increase in energy by $\Delta E = k_BT(d/2)\ln(l_2/l_1)$. On the other hand the particles gain entropy since they are now less confined: $-k_BT\Delta S = k_BT\ln(l_1/l_2)$. Hence for $d < 2$ the particles will 'lose' each other since their attraction to the nearest neighbours is overruled by the gain in entropy, as derived rigorously in equation (A.5).

Finally, I mention that the same extension to arbitrary dimensions d can be performed for closed chains. One finds then phase separation for molten rosettes if $d > 4$. More specifically, loops have a preferred spacing L/N for $d < 4$ and $4\chi l_P/(d-4)$ for $d > 4$. This is different from the results on open chains, equation (A.5), in the interval $2 < d < 4$; hence in $d = 3$ molten rosettes respond strongly to a cutting of the chain.

Appendix B. Formation energy for small intranucleosomal loops

Since the configurations of small loops are essentially planar, it is convenient to describe them in terms of the function $r(\theta)$, cf figure 12(b), where r and θ are the polar coordinates of an arbitrary point on the loop (with the origin chosen on the cylinder axis and the X-axis running through the centre of the loop). In these terms the line element ds takes the form $\mathrm{d}s = \mathrm{d}\theta\sqrt{r^2 + (\mathrm{d}r/\mathrm{d}\theta)^2}$ and the loop excess length is given by

$$\Delta L = \int_{-\theta^*}^{\theta^*} \mathrm{d}\theta\sqrt{r^2 + (\mathrm{d}r/\mathrm{d}\theta)^2} - 2\theta^* R_0 \tag{B.1}$$

with $2\theta^*$ being the aperture angle of the bulge, $2\theta^* R_0 = L^*$. The local curvature $1/R = |\mathrm{d}^2\boldsymbol{r}(s)/\mathrm{d}s^2|$ of the loop takes the form $1/R = |r''r - 2r'^2 - r^2|/(r^2 + r'^2)^{3/2}$, with primes and double primes denoting the first and second derivatives, respectively, with respect to θ. Restricting ourselves here to *small* loops, one can write $r(\theta) = R_0 + u(\theta)$ with $u \ll R_0$ everywhere. Keeping only quadratic terms in u and its derivatives one obtains $\mathrm{d}s \simeq R_0\mathrm{d}\theta(1 + u/R_0 + u'^2/(2R_0^2))$ and

$$\mathrm{d}s\frac{1}{R^2(s)} \simeq \mathrm{d}\theta\left[\frac{1}{R_0^3}\left(\left(u''^2 + \frac{3}{2}u'^2 + u^2\right) + 4uu''\right) - \frac{2u''}{R_0^2} - \frac{u}{R_0^2} + \frac{1}{R_0}\right]. \tag{B.2}$$

The bending energy of the loop is then

$$E_{elastic} \simeq \frac{1}{2}A\int_{-\theta^*}^{\theta^*}\mathrm{d}\theta\left[\frac{1}{R_0^3}\left(u''^2 - \frac{5}{2}u'^2 + u^2\right) - \frac{u}{R_0^2} + \frac{1}{R_0}\right] \tag{B.3}$$

where the boundary conditions $u(\theta) = \mathrm{d}u(\theta)/\mathrm{d}\theta = 0$ at $\theta = \pm\theta^*$ have been used.

The variational energy of the loop $F = E_{elastic} - T\Delta L$, for a *given* aperture angle θ^* and subject to the constraint of a fixed loop contour length $L = L^* + \Delta L$, follows from equations (B.1) and (B.3) to be:

$$\mathcal{F}\{u(\theta)\} = \frac{1}{2}\frac{A}{R_0^3}\int_{-\theta^*}^{\theta^*} \mathrm{d}\theta\left[u''^2 - \frac{5}{2}u'^2 + u^2\right] - \frac{T}{2R_0}\int_{-\theta^*}^{\theta^*}\mathrm{d}\theta u'^2. \tag{B.4}$$

Here T is the Lagrange multiplier that constrains an extra length ΔL to be adsorbed as the loop is formed (T can be interpreted as a 'tension' pulling in extra length). In equation (B.4) the constant terms as well as the term linear in u have been dropped since solutions to F with and without the linear u term differ just by a constant. The optimal loop shape (with given values of ΔL and θ^*) obeys the Euler–Lagrange equation $\delta F/\delta u = 0$:

$$u'''' + \left(\frac{5}{2} + \frac{TR_0^2}{A}\right)u'' + u = 0. \tag{B.5}$$

Solutions of equation (B.5) are of the form $u \propto \mathrm{e}^{\mathrm{i}\lambda\theta}$ with four possible values of $\lambda = \pm\lambda_\pm$ where

$$\lambda_\pm^2 = \frac{1}{2}\left(\frac{5}{2} + \frac{TR_0^2}{A}\right) \pm \frac{1}{2}\sqrt{\left(\frac{5}{2} + \frac{TR_0^2}{A}\right)^2 - 4} \tag{B.6}$$

that show the following asymptotics:

$$\lambda_+ \simeq \frac{1}{\lambda_-} \simeq \begin{cases} \sqrt{2} & \text{for } T \ll A/R_0^2 \\ \sqrt{TR_0^2/A} & \text{for } T \gg A/R_0^2. \end{cases} \tag{B.7}$$

One expects symmetric solutions of the form $u(\theta) = C_1\cos(\lambda_+\theta) + C_2\cos(\lambda_-\theta)$. The boundary conditions $u = 0$ and $u' = 0$ at $\theta = \theta^*$ have then the form $C_1\cos(\lambda_+\theta^*) + C_2\cos(\lambda_-\theta^*) = 0$ and $C_1\lambda_+\sin(\lambda_+\theta^*) + C_2\lambda_-\sin(\lambda_-\theta^*) = 0$. The solubility condition leads to the transcendental equation

$$\frac{\lambda_-}{\lambda_+} = \frac{\tan(\lambda_+\theta^*)}{\tan(\lambda_-\theta^*)}. \tag{B.8}$$

For vanishing 'tension', $T = 0$, one finds from equation (B.7) that the condition (B.8) is of the form $2\tan(\sqrt{2}\theta^*) = \tan(\theta^*/\sqrt{2})$ that has no non-trivial solution. For large T, $T \gg A/R_0^2$, one obtains from equations (B.7) and (B.8):

$$\frac{1}{(TR_0^2/A)^{3/2}} \simeq \frac{1}{\theta^*}\tan\left(\sqrt{TR_0^2/A}\theta^*\right). \tag{B.9}$$

The left-hand side of equation (B.9) is small and hence solutions are approximately given by $\sqrt{TR_0^2/A}\theta^* \simeq \lambda_+\theta^* \simeq k\pi$ with $k = 1, 2, 3, \ldots$. In the following I consider only the $k = 1$ solution which is the solution that leads to the smallest elastic energy.

Using partial integration and equation (B.5) the loop formation energy, equation (60), can be cast into the form

$$\Delta U = 2k_BT\lambda R_0\theta^* + \frac{T}{R_0}\int_0^{\theta^*}\mathrm{d}\theta u'^2 - A\int_0^{\theta^*}\mathrm{d}\theta\frac{u}{R_0^2}. \tag{B.10}$$

To proceed further one makes use of the explicit solution given below equation (B.7). Assume the large tension case $T \gg A/R_0^2$ (the quality of this approximation will be checked *a posteriori*); then $\lambda_- \ll 1$. The condition $u(\theta^*) = 0$ takes then the form $C_2 \simeq -C_1\cos(\lambda_+\theta^*)$

which leads, together with $\lambda_+\theta^* \simeq \pi$, to $C_2 = C_1$. The loop shape is thus approximately given by $u(\theta) \simeq C_1(1 + \cos(\pi\theta/\theta^*))$ from which follows its formation energy

$$\Delta U = 2k_BT\lambda R_0\theta^* + \frac{\pi^2 T C_1^2}{2R_0\theta^*} - \frac{AC_1\theta^*}{R_0^2} \tag{B.11}$$

and excess length, equation (B.1), $\Delta L \simeq 2C_1\theta^* + \pi^2 C_1^2/(2R_0\theta^*) \simeq 2C_1\theta^*$ where I used the fact that small loops have small amplitudes: $C_1 \ll R_0(\theta^*)^2$. For fixed ΔL and θ^* follows $C_1 \simeq \Delta L/(2\theta^*)$. Inserting this into equation (B.11) and using $T \simeq \pi^2 A/(R_0\theta^*)^2$ leads to

$$\frac{\Delta U}{k_BT} \simeq 2\lambda R_0\theta^* + \frac{\pi^4}{8}\frac{l_P(\Delta L)^2}{R_0^3(\theta^*)^5} - \frac{l_P\Delta L}{2R_0^2}. \tag{B.12}$$

Now minimizing ΔU with respect to θ^* (for ΔL fixed) gives the optimal aperture angle

$$\theta^* \simeq \left(\frac{5\pi^4}{16}\frac{l_P}{\lambda}\right)^{1/6}\frac{\Delta L^{1/3}}{R_0^{2/3}}. \tag{B.13}$$

Combining equations (B.12) and (B.13) one arrives at the final expression for the formation energy of a (small) loop of excess length ΔL, equation (61).

References

[1] Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson J D 1994 *Molecular Biology of the Cell* 3rd edn (New York: Garland)
[2] Bloomfield V A 1996 *Curr. Opin. Struct. Biol.* **6** 334
[3] Olson N H, Gingery M, Eiserling F A and Baker T S 2001 *Virology* **279** 385
[4] Cerritelli M E, Cheng N, Rosenberg A H, McPherson C E, Booy F P and Steven A C 1997 *Cell* **91** 271
[5] Kindt J, Tzlil S, Ben-Shaul A and Gelbart W M 2001 *Proc. Natl Acad. Sci. USA* **98** 13671
[6] Kornberg R D 1974 *Science* **184** 868
[7] Kornberg R D and Lorch Y 1999 *Cell* **98** 285
[8] Khrapunov S N, Dragan A I, Sivolob A V and Zagariya A M 1997 *Biochim. Biophys. Acta* **1351** 213
[9] Raspaud E I, Chaparon I, Leforestier A and Livolant F 1999 *Biophys. J.* **77** 1547
[10] Luger K, Mäder A W, Richmond R K, Sargent D F and Richmond T J 1997 *Nature* **389** 251
[11] Davey C A, Sargent D F, Luger K, Maeder A W and Richmond T J 2002 *J. Mol. Biol.* **319** 1097
[12] Thoma F, Koller Th and Klug A 1979 *J. Cell. Biol.* **83** 403
[13] Widom J 1986 *J. Mol. Biol.* **190** 411
[14] van Holde K E 1989 *Chromatin* (New York: Springer)
[15] Widom J 1989 *Annu. Rev. Biophys. Chem.* **18** 365
[16] van Holde K and Zlatanova J 1995 *J. Biol. Chem.* **93** 8373
[17] van Holde K and Zlatanova J 1996 *Proc. Natl Acad. Sci. USA* **93** 10548
[18] Finch J T and Klug A 1976 *Proc. Natl Acad. Sci. USA* **73** 1897
[19] Widom J and Klug A 1985 *Cell* **43** 207
[20] Woodcock C L, Grigoryev S A, Horowitz R A and Whitaker N 1993 *Proc. Natl Acad. Sci. USA* **90** 9021
[21] Horowitz R A, Agard D A, Sedat J W and Woodcock C L 1994 *J. Cell Biol.* **131** 1365
[22] Leuba S H, Yang G, Robert C, Samori B, van Holde K, Zlatanova J and Bustamante C 1994 *Proc. Natl Acad. Sci. USA* **91** 11621
[23] Bednar J, Horowitz R A, Grigoryev S A, Carruthers L M, Hansen J C, Koster A J and Woodcock C L 1998 *Proc. Natl Acad. Sci. USA* **95** 14173
[24] Schiessel H, Gelbart W M and Bruinsma R 2001 *Biophys. J.* **80** 1940
[25] Workman J L and Kingston R E 1998 *Annu. Rev. Biochem.* **67** 545
[26] Widom J 1998 *Annu. Rev. Biophys. Struct.* **27** 285
[27] Polach K J and Widom J 1995 *J. Mol. Biol.* **254** 130
[28] Polach K J and Widom J 1996 *J. Mol. Biol.* **258** 800
[29] Pennings S, Meersseman G and Bradbury E M 1991 *J. Mol. Biol.* **220** 101
[30] Meersseman G, Pennings S and Bradbury E M 1992 *EMBO J.* **11** 2951
[31] Pennings S, Meersseman G and Bradbury E M 1994 *Proc. Natl Acad. Sci.* **91** 10275
[32] Varga-Weisz P D and Becker P B 1998 *Curr. Opin. Cell Biol.* **10** 346

[33] Guschin D and Wolffe A P 1999 *Curr. Biol.* **9** R742
[34] Peterson C L 2000 *FEBS Lett.* **476** 68
[35] Flaus A and Owen-Hughes T 2001 *Curr. Opin. Genet. Dev.* **11** 148
[36] Studitsky V M, Clark D J and Felsenfeld G 1994 *Cell* **76** 371
[37] Studitsky V M, Clark D J and Felsenfeld G 1995 *Cell* **83** 19
[38] Felsenfeld G, Boyes J, Chung J, Clark D and Studitsky V 1996 *Proc. Natl Acad. Sci. USA* **93** 9384
[39] Studitsky V M, Kassavetis G A, Geiduschek E P and Felsenfeld G 1997 *Science* **278** 1960
[40] Felsenfeld G, Clark D and Studitsky V 2000 *Biophys. Chem.* **86** 231
[41] Luger K and Richmond T J 1998 *Curr. Opin. Genet. Dev.* **8** 140
[42] Grosberg A Yu, Nguyen T T and Shklovskii B I 2002 *Rev. Mod. Phys.* **74** 329
[43] Park S Y, Bruinsma R F and Gelbart W M 1999 *Europhys. Lett.* **46** 454
[44] Mateescu E M, Jeppesen C and Pincus P 1999 *Europhys. Lett.* **46** 493
[45] Gurovitch E and Sens P 1999 *Phys. Rev. Lett.* **82** 339
[46] Netz R R and Joanny J-F 1999 *Macromolecules* **32** 9026
[47] Kunze K-K and Netz R R 2000 *Phys. Rev. Lett.* **85** 4389
[48] Nguyen T T and Shklovskii B I 2001 *Physica* A **293** 324
[49] Nguyen T T and Shklovskii B I 2001 *J. Chem. Phys.* **114** 5905
[50] Nguyen T T and Shklovskii B I 2001 *J. Chem. Phys.* **115** 7298
[51] Schiessel H, Bruinsma R F and Gelbart W M 2001 *J. Chem. Phys.* **115** 7245
[52] Nguyen T T, Grosberg A Yu and Shklovskii B I 2001 *Electrostatic Effects in Soft Matter and Biophysics* ed C Holm, P Kékicheff and R Podgornik (Dordrecht: Kluwer) p 469
[53] Gelbart W M 2001 *Electrostatic Effects in Soft Matter and Biophysics* ed C Holm, P Kékicheff and R Podgornik (Dordrecht: Kluwer) p 53
[54] Kunze K-K and Netz R R 2002 *Phys. Rev.* E **66** 011918
[55] Schiessel H 2003 *Macromolecules* at press
[56] Odijk T 1980 *Macromolecules* **13** 1542
[57] Pincus P A, Sandroff C J and Witten T A 1984 *J. Physique* **45** 725
[58] von Goeler F and Muthukumar M 1994 *J. Chem. Phys.* **100** 7796
[59] Wallin T and Linse P 1996 *Langmuir* **12** 305
[60] Wallin T and Linse P 1996 *J. Phys. Chem.* **100** 17873
[61] Wallin T and Linse P 1997 *J. Phys. Chem.* B **101** 5506
[62] Wallin T and Linse P 1998 *J. Chem. Phys.* **109** 5089
[63] Haronska P, Vilgis T A, Grottenmüller R and Schmidt M 1998 *Macromol. Theory Simul.* **7** 241
[64] Jonsson M and Linse P 2001 *J. Chem. Phys.* **115** 3406
[65] Chodanowski P and Stoll S 2001 *Macromolecules* **34** 2320
[66] Sakaue T, Yoshikawa K, Yoshimura S H and Takeyasu K 2001 *Phys. Rev. Lett.* **87** 078105
[67] Messina R, Holm C and Kremer K 2002 *Phys. Rev.* E **65** 041805
[68] Messina R, Holm C and Kremer K 2002 *J. Chem. Phys.* **117** 2947
[69] Jonsson M and Linse P 2001 *J. Chem. Phys.* **115** 10975
[70] Akinchina A and Linse P 2002 *Macromolecules* **35** 5183
[71] Dzubiella J, Moreira A G and Pincus P A 2003 *Macromolecules* **36** 1741
[72] Yager T D, McMurray C T and van Holde K E 1989 *Biochemistry* **28** 2271
[73] Marky N L and Manning G S 1991 *Biopolymers* **31** 1543
[74] Schiessel H, Rudnick J, Bruinsma R and Gelbart W M 2000 *Europhys. Lett.* **51** 237
[75] Schiessel H, Widom J, Bruinsma R F and Gelbart W M 2001 *Phys. Rev. Lett.* **86** 4414
Schiessel H, Widom J, Bruinsma R F and Gelbart W M 2002 *Phys. Rev. Lett.* **88** 129902
[76] Kulić I and Schiessel H 2003 *Biophys. J.* at press
[77] Kulić I and Schiessel H 2003 *Phys. Rev. Lett.* submitted
[78] de Gennes P G 1971 *J. Chem. Phys.* **55** 572
[79] de Gennes P-G 1979 *Scaling Concepts in Polymer Physics* (Ithaca, NY: Cornell University Press)
[80] Doi M and Edwards S F 1986 *The Theory of Polymer Dynamcis* (Oxford: Clarendon)
[81] Cui Y and Bustamante C 2000 *Proc. Natl Acad. Sci. USA* **97** 127
[82] Bennink M L, Leuba S H, Leno G H, Zlatanova J, de Grooth B G and Greve J 2001 *Nat. Struct. Biol.* **8** 606
[83] Brower-Toland B D, Smith C L, Yeh R C, Lis J T, Peterson C L and Wang M D 2002 *Proc. Natl Acad. Sci. USA* **99** 1960
[84] Ben-Haïm E, Lesne A and Victor J-M 2001 *Phys. Rev.* E **64** 0519219
[85] Victor J-M, Ben-Haïm E and Lesne A 2002 *Phys. Rev.* E **66** 060901(R)
[86] Katritch V, Bustamante C and Olson W K 2000 *J. Mol. Biol.* **295** 29

[87] Wedemann G and Langowski J 2002 *Biophys. J.* **82** 2847
[88] Mangenot S, Leforestier A, Vachette P, Durand D and Livolant F 2002 *Biophys. J.* **82** 345
[89] Mangenot S, Raspaud E, Tribet C, Belloni L and Livolant F 2002 *Eur. Phys. J.* E **7** 221
[90] Marko J F and Siggia E D 1997 *Biophys. J.* **73** 2
[91] Schiessel H 2002 *Europhys. Lett.* **58** 140
[92] Houchmandzadeh B, Marko J F, Chatenay D and Libchaber A 1997 *J. Cell. Biol.* **139** 1
[93] Poirier M, Eroglu S, Chatenay D and Marko J F 2000 *Mol. Biol. Cell* **11** 269
[94] Poirier M G, Nemani A, Gupta P, Eroglu S and Marko J F 2001 *Phys. Rev. Lett.* **86** 360
[95] Poirier M G, Monhait T and Marko J F 2002 *J. Cell. Biochem.* **85** 422
[96] Marko J F and Siggia E D 1997 *Mol. Biol. Cell* **8** 2217
[97] Arents G, Burlingame R W, Wang B-C, Love W E and Moudrianakis E N 1991 *Proc. Natl Acad. Sci. USA* **88** 10148
[98] Anderson J D and Widom J 2000 *J. Mol. Biol.* **296** 979
[99] Hagerman P J 1988 *Annu. Rev. Biophys. Biophys. Chem.* **17** 265
[100] Harries R A and Hearst J E 1966 *J. Chem. Phys.* **44** 2595
[101] Strauss J K and Maher L J 1994 *Science* **266** 1829
[102] Braunlin W H, Strick T J and Record M T 1982 *Biopolymers* **21** 1301
[103] Mascotti D P and Lohmann T M 1990 *Proc. Natl Acad. Sci. USA* **87** 3142
[104] Rädler J O, Koltover I, Salditt T and Safinya C R 1999 *Science* **275** 810
[105] Koltover I, Salditt T and Safinya C R 1999 *Biophys. J.* **77** 915
[106] Record M T, Anderson C F and Lohmann T M 1978 *Q. Rev. Biophys.* **11** 103
[107] Bruinsma R 1998 *Eur. Phys. J.* B **4** 75
[108] Harries D, May S, Gelbart W M and Ben-Shaul A 1998 *Biophys. J.* **75** 159
[109] Sens P and Joanny J-F 2000 *Phys. Rev. Lett.* **84** 4862
[110] Oosawa F 1971 *Polyelectrolytes* (New York: Dekker)
[111] Manning G S 1978 *Q. Rev. Biophys.* **11** 179
[112] Rouzina I and Bloomfield V A 1996 *J. Phys. Chem.* **100** 4292
[113] Alexander S, Chaikin P M, Grant P, Morales G J, Pincus P and Hone D 1984 *J. Chem. Phys.* **80** 5776
[114] Golestanian R 1999 *Phys. Rev. Lett.* **83** 2473
Sens P 1999 *Phys. Rev. Lett.* **83** 2474
[115] Netz R R and Joanny J-F 1999 *Macromolecules* **32** 9013
[116] Schiessel H and Aranda-Espinoza H 2001 *Eur. Phys. J.* E **5** 499
[117] Nguyen T T, Grosberg A Yu and Shklovskii B I 2000 *J. Chem. Phys.* **113** 1110
[118] Ganachaud F, Elaïssari A, Pichot F, Laayoun A and Cros P 1997 *Langmuir* **13** 701
[119] Sukhorukov G B, Donath E, Davis S, Lichtenfeld H, Caruso F, Popov Vi I and Möhwald H 1998 *Polym. Adv. Technol.* **9** 759
[120] Kabanov A V, Sergeyev V G, Pyshkina O A, Zinchenko A A, Zezin A B, Joosten J G H, Brackman J and Yoshikawa K 2000 *Macromolecules* **33** 9587
[121] McQuigg D W, Kaplan J I and Dubin P L 1992 *J. Phys. Chem.* **96** 1973 and references therein
[122] Frank-Kamentskii M D 1997 *Phys. Rep.* **288** 13
[123] Schlick T 1995 *Curr. Opin. Struct. Biol.* **5** 245
[124] Love A E 1944 *A Treatise of the Mathematical Theory of Elasticity* 4th edn (New York: Dover) ch 18 and 19
[125] Nizette M and Goriely A 1999 *J. Math. Phys.* **40** 2830
[126] Benham C J 1977 *Proc. Natl Acad. Sci. USA* **74** 2397
[127] Benham C J 1979 *Biopolymers* **18** 609
[128] Le Bret M 1979 *Biopolymers* **18** 1709
[129] Le Bret M 1984 *Biopolymers* **23** 1835
[130] Shi Y M and Hearst J E 1994 *J. Chem. Phys.* **101** 5186
[131] Fain B and Rudnick J 1997 *Phys. Rev.* E **55** 7364
[132] Swigdon D, Coleman B D and Tobias I 1998 *Biophys. J.* **74** 2515
[133] Fain B and Rudnick J 1999 *Phys. Rev.* E **60** 7239
[134] Coleman B D, Swigdon D and Tobias I 2000 *Phys. Rev.* E **61** 747
[135] Zandi R and Rudnick J 2001 *Phys. Rev.* E **64** 051918
[136] Yamakawa H and Stockmayer W 1972 *J. Chem. Phys.* **57** 2843
[137] Beard P 1978 *Cell* **15** 955
[138] Spadafora C, Oudet P and Chambon P 1979 *Eur. J. Biochem.* **100** 225
[139] Watkins J F and Smerdon M J 1985 *Biochemistry* **24** 7279
[140] Simpson R T and Stafford D W 1983 *Proc. Natl Acad. Sci. USA* **50** 51

[141] Wu H-M and Crothers D M 1984 *Nature* **308** 509
[142] Loomans D, Schiessel H and Blumen A 1997 *J. Chem. Phys.* **107** 2636
[143] Schiessel H, Sokolov I M and Blumen A 1997 *Phys. Rev.* E **56** R2390
[144] Schwinefus J J and Bloomfield V A 2000 *Biopolymers* **54** 572
[145] Ura K, Nightingale K and Wolffe A P 1996 *EMBO J.* **15** 4959
[146] Ura K, Kurumizaka H, Dimitrov S, Almouzni G and Wolffe A P 1997 *EMBO J.* **16** 2096
[147] Flaus A, Luger K, Tan S and Richmond T J 1996 *Proc. Natl Acad. Sci. USA* **93** 1370
[148] Flaus A and Richmond T J 1998 *J. Mol. Biol.* **275** 427
[149] Hamiche A, Kang J-G, Dennis C, Xiao H and Wu C 2001 *Proc. Natl Acad. Sci. USA* **98** 14316
[150] van Holde K E and Yager T D 1985 *Structure and Function of the Genetic Apparatus (NATO ASI series A)* vol 98, ed C Nicolini and P O P Ts'o p 35
[151] Sukhishvili S A, Chen Y, Müller J D, Gratton E, Schweizer K S and Granick S 2000 *Nature* **406** 146
[152] Sukhishvili S A, Chen Y, Müller J D, Gratton E, Schweizer K S and Granick S 2002 *Macromolecules* **35** 1776
[153] Abramowitz M and Stegun I 1972 *Handbook of Mathematical Functions* (New York: Dover)
[154] Bouchaud J P and Georges A 1990 *Phys. Rep.* **195** 127
[155] Klafter J, Shlesinger M F and Zumofen G 1993 *Chem. Phys.* **177** 821
[156] Sokolov I M, Mai J and Blumen A 1997 *Phys. Rev. Lett.* **79** 857
[157] Mandelbrot B B 1982 *The Fractal Geometry of Nature* (San Francisco, CA: Freeman)
[158] de Cloizeaux J and Jannick G 1990 *Polymers in Solution* (Oxford: Clarendon)
[159] Frenkel J and Kontorova T 1938 *Zh. Eksp. Teor. Fiz.* **8** 1340
Frenkel J and Kontorova T 1939 *J. Physique (Moscow)* **1** 137
[160] Floria L M and Mazo J J 1996 *Adv. Phys.* **45** 505
[161] Braun O M, Dauxois T, Paliy M V and Peyrard M 1997 *Phys. Rev. Lett.* **78** 1295
[162] Strunz T and Elmer F-J 1998 *Phys. Rev.* E **58** 1601
[163] Pietronero L, Schneider W R and Strässler S 1981 *Phys. Rev.* B **24** 2187
[164] Aubry S 1983 *J. Phys.* **44** 147
[165] Watanabe S, van der Zant H S J, Strogatz S H and Orlando T P 1996 *Physica* D **97** 429
[166] Uhler W and Schilling R 1988 *Phys. Rev.* B **37** 5787
[167] Schiessel H, Oshanin G, Cazabat A M and Moreau M 2002 *Phys. Rev.* E **66** 056130
[168] Marko J F 1998 *Phys. Rev.* E **57** 2134
[169] Marko J F 1997 *Europhys. Lett.* **38** 183
[170] Kamien R D, Lubensky T V, Nelson P and O'Hern C S 1997 *Europhys. Lett.* **38** 237
[171] Levinthal C and Crane H 1956 *Proc. Natl Acad. Sci. USA* **42** 436
[172] Nelson P *Proc. Natl Acad. Sci. USA* **96** 14342
[173] van Kampen N G 1992 *Stochastic Processes in Physics and Chemistry* (Amsterdam: North-Holland)
[174] Mattei S, Sampaolese B, De Santis P and Savino M 2002 *Biophys. Chem.* **97** 173
[175] Anselmi C, Bocchinfuso G, De Santis P, Savino M and Scipioni A 2000 *Biophys. J.* **79** 601
[176] Lorch Y, Zhang M and Kornberg R D 1999 *Cell* **96** 389
[177] Bazett-Jones D P, Côté J, Landel C C, Peterson C L and Workman J L 1999 *Mol. Cell. Biol.* **19** 1470
[178] Längst G and Becker P B 2001 *Mol. Cell* **8** 1085
[179] Rees A R, Keller R W, Vesenka J P, Yang G and Bustamante C 1993 *Science* **260** 1646
[180] Schulz A, Mucke N, Langowski J and Rippe K 1998 *J. Mol. Biol.* **283** 821
[181] Rivetti C, Guthold M and Bustamante C 1999 *EMBO J.* **18** 4464
[182] Kulić I 2002 private communication
[183] Woodcock C L and Dimitrov S 2001 *Curr. Opin. Genet. Dev.* **11** 130
[184] Zlatanova J, Leuba S H and van Holde K 1998 *Biophys. J.* **74** 2554
[185] Maret G 1976 *PhD thesis* Konstanz University
[186] Maret G, v. Schickfus M, Mayer A and Dransfeld K 1975 *Phys. Rev. Lett.* **35** 397
[187] Maret G and Weill G 1983 *Biopolymers* **22** 2727
[188] Martino J A, Katritch V and Olson W K 1999 *Struct. Fold. Des.* **7** 1009
[189] Widom J 1992 *Proc. Natl Acad. Sci. USA* **89** 1095
[190] Plewa J S and Witten T A 2000 *J. Chem. Phys.* **112** 10042
[191] Stryer L 1995 *Biochemistry* 4th edn (New York: Freeman) p 420
[192] Mergell B, Schiessel H and Everaers R in preparation
[193] Daban J-R 2000 *Biochemistry* **39** 3861
[194] Janninck G, Duplantier B and Sikorav J L 1996 *Biophys. J.* **71** 451
[195] Cluzel P, Lebrun A, Heller C, Lavery R, Viovy J L, Chatenay D and Caron F 1996 *Science* **271** 792
[196] Smith S B, Cui Y and Bustamante C 1996 *Science* **271** 795

[197] Mergell B, Ejtehadi M J and Everaers R 2002 *Phys. Rev.* E **66** 011903
[198] Nyrkova I A, Semenov A N, Joanny J-F and Khokhlov A R 1996 *J. Physique* II **6** 1411
[199] Livolant F and Leforestier A 2000 *Biophys. J.* **78** 2716
[200] Fraden S and Kamien R D 2000 *Biophys. J.* **78** 2189
[201] Ehrlich L, Münkel C, Chirico G and Langowski J 1997 *CABIOS* **13** 271
[202] Beard D A and Schlick T 2001 *Structure* **9** 105
[203] Münkel C and Langowski J 1998 *Phys. Rev.* E **57** 5888
[204] Ladoux B, Quivy J-P, Doyle P, du Roure O, Almouzni G and Viovy J-L 2000 *Proc. Natl Acad. Sci. USA* **97** 14251
[205] Bennink M L, Pope L H, Leuba S H, de Grooth B G and Greve J 2001 *Single Mol.* **2** 91
[206] Rief M, Gautel M, Oesterhelt F, Fernandez J and Gaub H 1997 *Science* **276** 1109
[207] Oberhauser A, Hansma P, Carrion-Vazquez M and Fernandez J 2001 *Proc. Natl Acad. Sci. USA* **98** 468
[208] Kulić I and Schiessel H *Preprint*
[209] Barrat J-L and Joanny J-F 1996 *Adv. Chem. Phys.* **94** 1
[210] Tóth K, Brun N and Langowski J 2001 *Biochemistry* **40** 6921
[211] Grunstein M 1997 *Nature* **389** 349
[212] Hansen J C, Tse C and Wolffe A P 1998 *Biochemistry* **37** 17637
[213] Strahl B D and Allis D 2000 *Nature* **403** 41