



Universiteit
Leiden
The Netherlands

The effect of thermal fluctuations on elastic instabilities of biopolymers

Emanuel, M.D.

Citation

Emanuel, M. D. (2012, April 4). *The effect of thermal fluctuations on elastic instabilities of biopolymers*. *Casimir PhD Series*. Retrieved from <https://hdl.handle.net/1887/19173>

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/19173>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/19173> holds various files of this Leiden University dissertation.

Author: Emanuel, Marc David

Title: The effect of thermal fluctuations on elastic instabilities of biopolymers

Issue Date: 2012-07-04

On the organization of DNA in eukaryotes

2.1 Introducing DNA in its environment

The human genome contains approximately $6 \cdot 10^9$ basepairs, two copies of which lead to roughly 2 meters DNA per cell. On scales larger than the helical repeat length, 3.5 nm, the double helix is well described as a wormlike chain with a persistence length of about 50 nm. Disregarding volume interactions the diameter of the coil in a theta solvent would be around $\sqrt{2 \times 50\text{nm} \times 2\text{m}} \approx 450 \mu\text{m}$. That is an order of magnitude larger than the $10 \mu\text{m}$ diameter of the nucleus of a typical cell in which the DNA is always confined.

As a first level of organization the eucaryotic DNA is wrapped around protein spools, each a cylindrical wedge of diameter of 6 nm and maximal height of 6 nm. About 147 basepairs wrap along a left-handed wrapping path of 1,67 turns around the octamer. These spools are composed of 4 pairs of histone proteins, named H2A, H2B, H3 and H4. This octamer, together with the DNA wrapped around, is called the nucleosome core particle (NCP). Its structure is known in great detail from high resolution X-ray crystallography [18]. It is noteworthy to mention that the histone proteins that make up the core, although existing with some variations, are remarkably well conserved between eukaryotes. An important feature of the histones are their tails, flexible positively charged extensions. Through modifications, like acetylation and phosphorylation, it is possible to neutralize charges on the tails. These modifications form a way to regulate the organization of the DNA in the nucleus, as we will discuss in the next sections. There are also other modifications, that do not directly influence the spatial distribution of DNA but influence the binding of proteins. We will not consider them in this thesis. For a discussion of the energetics and dynamics involved in single nucleosomes we refer to other reviews [19, 20].

There is one spool for every 160 to 240 basepairs. This so called repeat length varies over species, but also over cells within one species. The stretches of DNA connecting two neighboring NCPs is called linker DNA. As a result one obtains a bead-on-a-string structure, sometimes referred to as 10 nm fiber. This structure is, however, only observed *in vitro* at

subphysiological salt concentrations. If we nevertheless assume that such a fiber exists in the nucleus with the same stiffness as for naked DNA, then we find a coil diameter of around $\sqrt{50/200} \times 450 \mu\text{m} \approx 225 \mu\text{m}$ – a value that is still much larger than the diameter of the nucleus.

In most eukaryotes a fifth histone is thought to bind the two outgoing double strands at each NCP resulting in a denser structure. From EM measurements it is concluded that the linker histone assembles the two outgoing strands in a stem of 3 nm length [21]. This linker histone has a couple of variants (named H1 and H5), but is also very well conserved. That all the histone proteins are so well conserved through evolution indicates that their functional properties are rather intricate. In the presence of linker histones and for physiological salt concentrations one observes dense fibers, usually referred to as 30 nm or chromatin fibers, that are known through *in vitro* experiments for over 30 years. They typically feature around 30 nm diameter, independent of whether they are extracted from cells or whether they are reconstituted. These fibers have a contour length that is only about 1/50th of the contour length of the DNA that it contains but seem to be much stiffer than naked DNA. Assuming a fiber persistence length of 240 nm [22] leads to a coil size of $\sqrt{2 \times 240 \text{nm} \times 2 \text{m}/50} \approx 140 \mu\text{m}$. If these stiff chromatin fibers really exist *in vivo* this would call for another level of organization/condensation before the genetic material fits into the nucleus. This larger scale organization will be the main issue we will address in this chapter.

The rest of this chapter is organized as follows. In Section 2.2 we discuss the local fiber structures of *in vitro* 30 nm fibers and the relation with the fibers *in vivo*. In Section (2.3) we discuss some of the polymer models that were introduced to describe the organization of the chromatin fiber on large scales. We argue that the simplest way to describe the fiber is by separating the condensation itself from the organization and show that the resulting model is good enough to explain the existing data. Finally, in Section (2.4) we discuss possible mechanisms for the condensation of the fiber.

2.2 The 30 nm chromatin fiber

Most of what is known about the 30 nm fiber is known through *in vitro* experiments. A large number of theoretical models for the 30 nm fiber has been put forward, all of them being more or less compatible with the experimental data. Only very recently, the assembly of very regular reconstituted fibers brings such high demands on the models that many of them can be invalidated. In this section we will comment on the various models that have been proposed. Furthermore we stress the important role of the energetics of linker DNA bending and of nucleosomal interaction in determining the fiber structure. We discuss how the tails play an important part here and how a cell via their modifications can regulate the degree of fiber condensation.

2.2.1 Old models

Since 30 nm fibers are rather dense, EM-images are open to a plethora of interpretations (cf. Ref. [23] for a recent review on some of the proposed models). Early EM-images, obtained by Finch and Klug in 1976 [24], showed repetitive bands with a spacing of around 110 Å that were almost orthogonal to the fiber axis, whereas no substructures seemed to be present along these bands. This observation suggested a solenoidal arrangement of the nucleosome and with this *solenoidal* model the first chromatin fiber model was born. In order to have successive nucleosomes along the helical path in contact, the linker DNA has to be strongly bent, especially for “short” linker lengths. Structural stability requires thus strong NCP-NCP interactions with the result that the fiber diameter is expected to be independent of the linker length. Although there is no specific argument known why the solenoidal arrangement has a specific diameter, experiments indicating a diameter that is independent of the linker length is seen as supporting the solenoid model. Recent single molecule experiments by Kruihof et al [25] showing a linear force-extension relation of the chromatin fiber over a rather large range of extensions also point in the direction of a solenoid arrangement.

Recurring patterns where nucleosomes seem to stack along two rows, with linker DNA tracing out a *zig-zag* pattern, both in intact [26] and nuclease-digested isolated chromatin [26, 27] led Worcel et al [27] and Woodcock et al [26] to propose so-called *twisted ribbon* models. Cyro-electron microscopy indicated that linker DNAs are essentially straight [21], at least at low salt concentrations. This information gave support to the so-called *crossed-linker* models [28] that assume non-sequential folding of the NCPs connected via straight linkers that criss-cross the interior of the fiber. Such a fiber structure is set by the linkers and not by the NCPs, so that the fiber diameter depends continuously on the linker length (for details cf. the discussion of the two-angle model in Ref. [19]). Slight variations in linker length produce irregular fibers similar to the ones observed for native chromatin [29]. By cross-linking the nucleosomes before digesting the linker DNA with nuclease Dorigo et al [30] came to the conclusion that the the fiber must be of the two start type.

2.2.2 Ribbon model

The models we discussed so far do not predict the diameter of the fiber (solenoid model) or they predict a diameter that varies with the linker length (crossed-linker model). The state-of-the-art experiment [31] on the dependence of the fiber diameter on the linker length makes use of the fact that the affinity for nucleosomes depends on the specific sequence of basepairs that wrap around the octamer [32]. This fact allows one to construct DNA templates on which equally spaced nucleosomes are formed [33]. Based on this method Robinson et al [31] produced regular reconstituted chromatin fibers with varying repeat lengths of 177 to 237bp, increasing in steps of 10bp. Their EM measurements revealed a constant diameter of 33nm, and a nucleosome line density (NLD) varying between 0.9 and 1.2 nucleosome per nm, for repeat lengths from 177 up to 207 bp. For the longer repeat lengths the diameter was

Table 2.1: Relation between number of ribbons, fiber diameter and Nucleosome line density.

Ribbons	Diameter (nm)	NLD (nucleosomes/nm)
3	23	0.54
4	28	0.74
5	33	1.0
6	38	1.2
7	44	1.5

44nm, the NLD increasing to a range of $1.3 - 1.5\text{nm}^{-1}$. These results suggest that it is not the length of the linker DNA that sets the fiber diameter, contrary to the prediction of the crossed linker model. Instead these findings support the view that it is the arrangement of the nucleosomes that somehow favors a discrete set of fiber diameters (namely 33 and 44 nm). The linker DNA length enters only as a second order effect that determines which one of the two fiber diameters is chosen.

This leads to the important question of why there is a discrete set of preferred nucleosomal arrangements. To answer this question Depken and Schiessel [34] studied all possible fibers with densely packed nucleosomes. By mapping the chromatin cylinder to a strip with the long sides identified, it is straightforward to show that dense packings are achieved by placing the nucleosomes in ribbons. Different possible dense packings can then be characterized by the number of ribbons. This leads to a discrete set of ribbon models, all of them, however, still having an infinite, continuous range of possible diameters. What changes locally on the nucleosome level, when the diameter is changed is the effective wedge angle between neighboring nucleosomes in a ribbon. It is known from experiments that NCPs under certain conditions form arcs made out of a stack of NCPs where each NCP acts as a wedge with a wedge angle of 8 degrees [35]. Assuming that this is also the ideal wedge angle inside a chromatin fiber, one can predict a discrete set of diameter/NLD combinations depending exclusively on the number of ribbons, see Table 2.1. Note that two of the predicted fibers, the 5 ribbon and the 7 ribbon structures, have diameters that coincide precisely with the experimental values 33 and 44 nm. The 5 ribbon nucleosomal shell can be seen in Figure 2.1.

In addition, one has also to specify the connectivity of the nucleosomes. If one assumes a regular connection which is identical for each pair of connected nucleosomes and where the fiber is not built out of disjoint filaments, the connectivity can be characterized by two numbers, N and m . An (N, m) -fiber is then a fiber with N ribbons where the linker DNA connects ribbons that are m steps apart. One can then show that for fibers with 1, 2, 3, 4 and 6 ribbons the only possible connectivity is the one where neighboring ribbons are

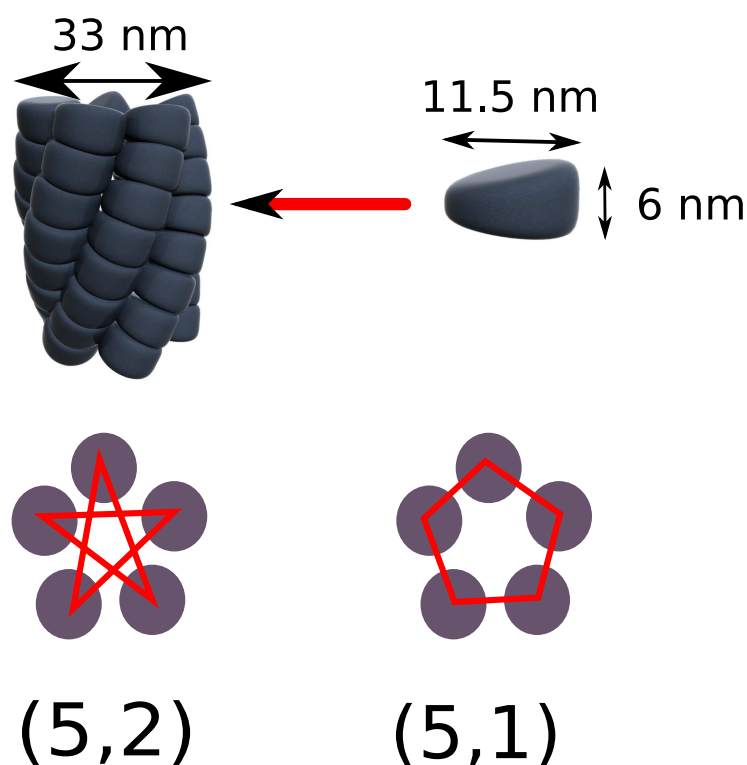


Figure 2.1: A 5-ribbon fiber with two connection schemes, (5,2) and (5,1)

connected, $m = 1$. In this notation the solenoid model is a (1, 1)-fiber and a twisted ribbon is a (2, 1)-fiber. The 5 ribbon structure has besides its nearest neighbor connectivity, (5, 1), also the next nearest neighbor, (5, 2), available. The 7 ribbon model allows for (7, 1), (7, 2) and (7, 3).

Why do the experiments of Robinson et al [31] indicate fiber diameters compatible to a dense 5 ribbon fiber for shorter linker lengths and of a 7 ribbon fiber at longer lengths? In Ref. [34] it is argued that the best of all possible structures is (5, 2) since this, at least for the shortest repeat length, allows for relatively straight linkers. However, when a repeat length of 217 bp is reached the estimated volume of the linker DNA becomes larger than the volume available inside a 5 ribbon shell. At that value of the linker length the linker is just long enough to form a (7, 3) structure with straight linkers whereas the only available 6 ribbon fiber would require strong linker bending. Finally, let us mention that in a recent experiment also a 167 bp repeat length was probed, resulting in a 21.3 nm wide fiber with a NLD of 0.56 nucleosome/nm [36]. This might point to a 3 ribbon structure. The small discrepancy can be attributed to the approximations involved in the model that start to matter at such small fiber diameters.

2.2.3 Energetics: elasticity and electrostatics

One important property we need to address is the energetics involved in the models. It is usually inferred from experiments that the linker histones are crucial [37, 36]. For example Routh et al [36] deduced from EM measurements and sedimentation rates, that reconstituted chromatin fibers with 197 bp repeat length *without* linker histones do not condense into a 33 nm fiber, but into a less dense structure. Pulling experiments in the picoNewton range [25] suggest nonetheless that at low forces such fibers have the same force extension behavior with or without linker histones. Fibers without linker histone show, however, a transition to a more open structure at a lower force value than fibers with linker histones. This suggests a stabilizing role of the linker histones. The conflicting results of Routh et al might reflect the preparation for the EM measurements and the floppiness of the structure in case of the sedimentation measurements.

In both cases, solenoid and crossed-linker model, the energy needed to bend the linker DNA can be estimated, assuming that DNA behaves like a wormlike chain for these contour lengths. The linker lengths range from 3.3 nm to 13 nm for the 30 nm fibers with linker histones and are, except for the shortest value, clearly longer than the helical repeat of the double helix. For the solenoid model the elastic energy can be as high as 35 kT per nucleosome. This high value actually suggests that the solenoidal structure, especially its diameter, would change with linker length to reduce the elastic penalty. At first sight a 5 ribbon structure would encounter similar problems, but here the ribbons can slide with respect to each other without changing the fiber diameter or NLD, lowering the elastic energy to a few kT's per nucleosome, the exact value depending on parameters like stem length and the exit angle by which the linker DNA leaves the stem.

It is not only the elastic energy that needs to be accounted for, but there is also a strong electrostatic repulsion between the negatively charged linkers. The above mentioned picoNewton pulling experiments by Kruithof et al indicate that the presence of magnesium ions is indispensable for the formation of a 30 nm fiber. This can be most likely attributed to the strong electrostatic repulsion that affects the linkers in a dense fiber. What the high density of nucleosomes suggest is that there is an effective attraction between NCPs that dictates the fiber condensation. This is the subject of the next section.

2.2.4 Energetics: Nucleosome attraction

In the experiments by Dubochet and Noll in 1978 [35] it was found that isolated NCPs in controlled ionic conditions show a strong tendency to self-assemble into arcs and cylinders, and this in the absence of magnesium. The NCPs are stacked "face-to-face", indicating the importance of nucleosome interactions in the formation of higher-order structures. As a possible mechanism behind the nucleosomal attraction was later put forward the formation of tail bridges where positively charged tails of one nucleosome interact with negative charges on the core of a NCP closeby.

The tail conformations result from a competition between electrostatics and entropy. For low salt concentrations tails are condensed onto their nucleosome, whereas for higher salt concentrations, the entropic contribution of the tails to the free energy becomes more important. The tails gradually unfold with increasing ionic strength and the effective diameter of the NCP saturates around physiological salt concentrations [38]. Osmometric measurements on dilute solutions of NCPs show a minimum in the 2nd virial coefficient around the same salt concentration [39]. This led Mangenot et al [39] to suggest that the tails give the dominant contribution to the interaction between nucleosomes. More recently Bertin et al [40] found that tail-intact NCPs show attraction in the absence of magnesium, with a 2nd virial coefficient in qualitative agreement with Mangenot et al, whereas it approached the hard-sphere repulsive interaction for NCPs where the tails were removed with trypsin. Moreover, for trypsinized oligonucleosomes it has been observed that no higher order folding occurred for an increase in monovalent salt [41] or magnesium concentration [42].

To understand better how the tails induce an effective attraction Mühlbacher et al [43] modelled the NCPs as freely rotating spheres with a homogeneous surface charge distribution representing the histone-DNA core of the NCP. The eight tails were modelled as identical flexible chains grafted onto the sphere at the vertices of a cube inscribed in the sphere. The screened electrostatics was approximated by the Debye-Hückel interaction. For an appropriately chosen effective sphere charge the pair potential showed an effective attraction of a few $k_B T$ around physiological salt concentrations. Importantly this attraction disappeared when a small fraction of the tail charges was removed – hinting at a possible biochemical mechanism through which the cell can control the nucleosomal interaction.

2.2.5 The 30 nm fiber *in vivo*

Until now we have discussed an idealized 30 nm fiber, mostly based on very regular reconstituted fibers in clean static conditions. It is an open question how relevant these structures are for the properties of chromatin within the nucleus. Bystricky et al [44] measured how the spatial distance (SD) depends on the chemical distance (CD) along the DNA double helix in budding yeast, using fluorescence in situ hybridization (FISH) for CDs from 14 kb (kilo basepairs) up to 100 kb. We will adhere here to the convention of expressing the CD in the number of basepairs, each basepair contributing a "length" of 0.35 nm.

Before we discuss the FISH data we need to address the reliability of 3d-FISH measurements. The main steps of FISH consists of a fixation step (using buffered formaldehyde), a denaturalization step (by heating up the sample to 75 °C for 2 minutes) followed by a hybridization step where a fluorescent probe is matched to a specific sequence of the (denatured) DNA. Obviously these procedures are rather harsh. Recent experiments [45, 46] set up to evaluate the reliability of FISH experiments came to the conclusion that large scale structures, down to light microscope resolution, are reasonably well preserved for this 3d-FISH. The older "2d-FISH", based on methanol/acetic acid fixation seems to be significantly less reliable. On a smaller scale both procedures are destructive and reliability suffers.

Keeping these restrictions in mind, Bystricky et al [44] could fit their data with the curve of a wormlike chain with a persistence length of 170 – 200 nm and a NLD of 0.64 – 0.91 nucleosomes/nm. Apparently a rather stiff condensed fiber is formed but how it relates to the original *in vivo* structure is an open question. Nonetheless the persistence length inferred from the experiment is in relatively good agreement with coarse grained models of the 30 nm fiber [22, 47]. Especially from the simulation in Ref. [22] one can draw the not so surprising conclusion that the worm like chain model is only appropriate for small bends. Kinks can be formed with relative low cost. This last point has to be kept in mind when discussing large scale structures to which we will turn in the next session.

2.2.6 Conclusion: the 30 nm fiber

The existing models of the 30nm fiber should be regarded as a description of a groundstate of the chromatin fiber in the nucleus. The reconstituted fibers form an interesting playground to get a grip on the possible energies that play a role in the formation. It is important to realize that in the nucleus the chromatin fiber gets synthesized in a way different from that of reconstituted fibers. The *in vivo* fiber is also not as regular and static as the reconstituted fibers. Although the fiber is often thought of as a wormlike chain, it should be kept in mind that a notion of persistence length is only valid as some kind of rough average. The flexibility has most likely large variations and the fiber is highly extensible. We will not rely on any specific persistence length for the next section. When there is a need to compare values with the value of the Kuhn length of the chromatin fiber we will use a value of 300 nm, realizing that one could argue as well for lengths ranging from 60 [48] to 400 [44] nm.

2.3 Large scale structures

By analyzing local radiation damage of chromosomes [49] Cremer et al concluded – contrary to the general picture at that time – that during interphase chromosomes are segregated within their own domain. Most of the evidence for the existence of these separate chromosome domains comes from FISH data. As we have discussed before, the resolution of these data is restricted to the optical one, not only because of the measurement apparatus but also due to the preparation procedure. Although there is some evidence that intermingling of chromatin domains does occur [50], recent experiments point in the direction of domains that at least do not intertwine [51]. This section discusses the structure of chromosomes in their own domain.

2.3.1 Experiments and models

The use of FISH to examine the distribution of chromatin in interphase nuclei dates back to van den Engh et al [52]. They measured via 2d-FISH the SD versus the CD for many pairs of probes. In a first analysis of their data they noticed that they could fit it to a Gaussian chain up

to 2 Mbp of CD. For longer CDs a flattening of the curve was found indicating the existence of some constraint. This analysis was worked out in greater detail by Hahnfeldt et al [53] where the data were fitted to a Gaussian chain within a spherical confinement. Extension of these 2d-FISH measurements on the same chromosomes up to 190 Mbp [54] showed for larger CD's again the footprint of a random walk but with a considerable smaller slope of CD versus the squared SD. The data were fitted to a fixed Mbp giant loop model with the loop nodes forming a long distance random walk. It was argued that this is the simplest model to explain the new data.

These same data led Münkkel and Langowski [55] to propose their multi-loop subcompartment model as a Gaussian chain with a non-hardcore volume interaction together with harmonic bonding into loops. A compartment of loops is then attached to the next compartment by a chromatin link. Their idea was first of all to emulate the formation of clearly separated subcompartments within a chromosome territory. Several characterizations of these subcompartments exist. We will use the terminology euchromatin for the less dense, gene active regions and heterochromatin for the denser, inactive regions. Other not fully compatible terms are R-band/G-band (based on Giemsa staining) [56], ridge (regions of increased gene expression)/antiridge [57], depending on ones taste, measurement technology or functionality. A special feature of this model is that, according to the authors, it replicates the scaling of the SD as a function of the CD over large distances, *not* as a polymer at its θ -point (a random walk with exponent $1/2$), *nor* as a polymer in a good solvent (a self avoiding random walk with exponent $3/5$), but with an exponent of $1/3$, from which they inferred that it behaves as a globule.

More recent 3d-FISH measurements [58, 59] differentiated between distances within euchromatin and heterochromatin regions of chromosomes in human fibroblast cells. The authors noted that they could fit their data to a globular state with exponent $1/3$ but observed also a levelling off to a constant value for larger CDs. The size of this levelling off was significantly different for different regions, strengthening the idea that non active regions are considerably denser than active ones. The above mentioned levelling off led Bohn et al [60] to propose yet another loop based model: the random loop model. A Gaussian chain without volume interactions but with harmonic attraction of the same strength as the chain links is introduced between non-neighboring beads whose separation is chosen with a fixed probability. The authors claimed that such a random loop configuration is needed to explain the levelling off.

2.3.2 The loop and the globule

In this subsection we attempt to critically analyze the models that we briefly discussed in the previous subsection. Our starting point is the fact that the chromatin fibers within the nucleus and also within their compartments are highly confined. We now separate this notion of confinement from the spatial distribution of the chromatin, somewhat in the spirit of the spherical confinement model. The possible mechanism behind this confinement will be

discussed in the next section. The logic behind this approach is that it is not known what causes the confinement, let alone the details behind it. As will become clear we also do not need this knowledge to explain the data presented so far. This also means that we will not assume any loop formation, leaving the causality between confinement and loop formation open.

To set the stage let us estimate the density of the chromatin fiber in a human cell. Human lymphocytes have a nuclear volume between 380 and $525 \mu\text{m}^3$ [61]. Let us assume that the chromatin is spread throughout the nucleus, neglecting the space taken up by nuclear organelles like the nucleoli or other non chromatin domains. We expect this approximation to be sufficient in view of the precision of the experiments we aim to describe here. With a NLD of 0.7 nm^{-1} , repeat length of 200 bp and a fiber diameter of 33 nm , a $380 \mu\text{m}^3$ nucleus corresponds to a chromatin volume fraction of almost 0.1 . The first fact that needs to be understood is why a pure random walk can explain the data this well (up to lengths before confinement/loop forming sets in). After all, the chromatin fiber has a high cross section. It could be a coincidence that attractive forces just balance the repulsive forces and the Kuhn length could be much shorter than what we would conclude from a persistence length of around 150 nm . In that case it makes one wonder how this could be achieved with varying density. It is appropriate to introduce some (old) polymer physics at this point.

Let us suppose that chromatin is highly confined but in a reasonably good solvent. When we start to follow the chromatin fiber from a given point, for length scales up to its persistence length, the relation between SD and CD will be linear. For longer contour lengths there will be a crossover first to a random walk followed by a crossover to a self avoiding random walk. For length scales larger than that correlation length, density-fluctuations disappear and there is no preferred direction. The correlation length is just set by the distance where collisions with segments closeby in CD and segments far away, i.e. for large CD's, are as likely. This length scale can be derived following Ref. [62] that we use in the following.

The expectation value for the segment density, C_0 , is constant within a compartment. This is in fact only true neglecting boundary effects, but with the relatively high densities in the nucleus this is a reasonable assumption. Starting from one segment at the origin, in its proximity the segment distribution can be divided into a contribution coming from segments close by along the backbone, c_c , and segments far away, c_f , both strongly fluctuating on small scales:

$$C_0 = \langle c_c(r) + c_f(r) \rangle = \langle c_c(r) \rangle + \langle c_f(r) \rangle. \quad (2.1)$$

Suppose now that the solution is dilute enough such that excluded volume effects become important before the semi-diluteness of the solution becomes relevant. Close to the origin the density is dominated by the first term and the presence of far away segments can be neglected.

N segments of the chain with volume interactions have then a radius of gyration that scales as

$$R \approx a \left(\frac{v}{a^3} \right)^{1/5} N^{3/5}, \quad (2.2)$$

where a is the Kuhn length and v the excluded volume. From this expression follows as segment density:

$$\langle c_c(r) \rangle \approx \left(\frac{a^3}{v} \right)^{1/3} a^{-3} \left(\frac{r}{a} \right)^{-4/3} \quad (2.3)$$

The far away contribution kicks in when the first term becomes of the order of the overall density. This defines the correlation length ξ :

$$\frac{\xi}{a} \approx \left(\frac{a^3}{v} \right)^{1/4} (C_0 a^3)^{-3/4}. \quad (2.4)$$

On length scales larger than ξ , the density fluctuations are minimal and the chromatin is in the globular state. Although it is true that the size of a globule scales as $(\#\text{monomers})^{1/3}$ it is a misconception that locally the distance between segments scales with $(\#\text{ of connecting segments})^{1/3}$. It is an old argument by Flory [3] that points to an exponent of $1/2$ in this case: namely because of the homogeneity of the globule is the pressure from all directions the same and so there are no effective volume effects. The chain forms thus a random walk of "blobs" with the blob size given by the correlation length. Although there are some deviations at short length scales to this "Flory ideal chain hypothesis", it was recently shown by Lua et al [63] – using an ingenious counting algorithm on a lattice – to hold more than well enough. The only drawback of that simulation when applied to our problem is that it presupposes that the system is ergodic and that the system has time to thermalize. That last point might be important, as we will discuss at the end of this section.

Let us consider in this context the more recent 3d-FISH data for chromosome 1 in human fibroblast [58]. The only data available are the average SD as function of the CD *and* the standard deviation, both depicted in Figure 2.2. Knowing other moments might reveal more details. As mentioned before the SD levels off at larger CDs. We expect that the height of the plateau corresponds to the average distance between two points in the compartment:

$$\langle SD \rangle = \frac{1}{V^2} \int_V d^3 \mathbf{r}_1 \int_V d^3 \mathbf{r}_2 |\mathbf{r}_1 - \mathbf{r}_2| \quad (2.5)$$

For example for a spherical compartment of radius R we find an average of $(36/35)R$ and a standard deviation of $\sigma = (\sqrt{174}/35)R \approx 0.38R$, while for a flat disk with radius R the average distance is $0.9R$ with a standard deviation of $0.42R$. Human fibroblast has an almost 2d nucleus as can be inferred also from the FISH measurements over the whole chromosome 1 [58]. For a length of 30 Mb there is a levelling off around $1.8 \mu\text{m}$. Using a volume fraction

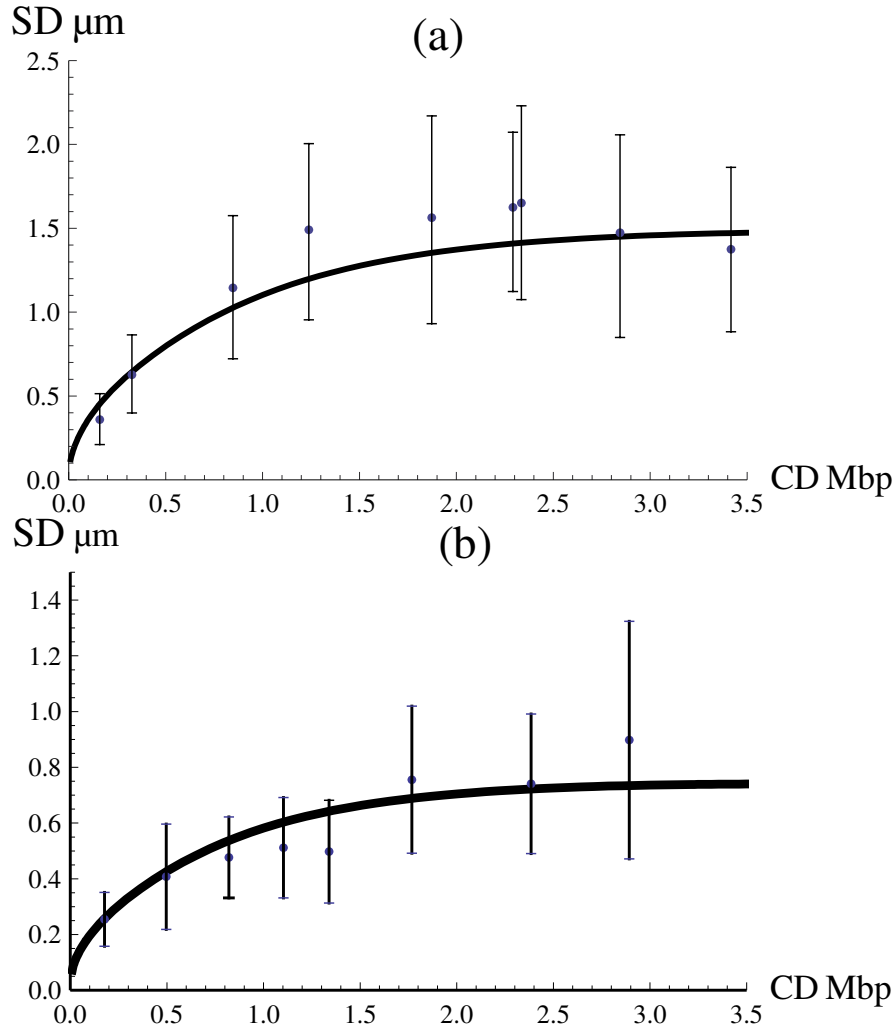


Figure 2.2: 3d-FISH data from [58] for the (a) active and (b) non active regions of chromosome 1 in fibroblasts. The error bars are the standard deviations. The curves use (2.7) with the parameters given in the main text.

of 0.1, a repeat length of $r_l = 200$ bp, and an NLD of 0.7 nm^{-1} one predicts a value of the thickness for a disk-shaped compartment of 150 nm.

It is important to realize that although the globular state has local Gaussian behavior, it is not to be confused with a Gaussian chain confined to the region of a globule, the structure assumed in [53]. The latter has a strong density peak in the center of the confinement [3], while here we have a constant density profile. The reason for this difference is that the confined Gaussian is Gaussian because of the *lack* of volume interactions while in the case of a globule it is Gaussian *because* of volume interactions. This distinction is not always appreciated. Would we repeat the calculation for the average distance and standard deviation using the

confined Gaussian probability, $\rho(r) = \sin(\frac{\pi r}{R})/(\pi R r^2)$, we would find for the height of the plateau and its standard deviation:

$$\langle SD(CD \rightarrow \infty) \rangle_{\text{Gauss}} \approx 0.7R \quad \sigma_{\text{Gauss}} \approx 0.14R \quad (2.6)$$

We would clearly get a higher estimate of the size of the compartment. It is interesting to note that the *relative* standard deviation is 0.2, considerably smaller than the 0.38 for the globule. When we compare this with the plateau of chromosome 1, for both eu- and heterochromatin compartments, we see a relative standard deviation of 0.37 indicating that the picture of a globule describes this in a satisfactory way.

In the following we model the chromatin within its compartment as a Gaussian chain within a rectangular box. To ensure a flat density profile we have to choose reflecting boundary conditions. The mean squared SD is then given by:

$$\langle (SD(CD))^2 \rangle = \frac{1}{V} \left(\frac{3}{2\pi L(CD)b} \right)^{3/2} \int_V d^3 \mathbf{r}_1 \int_V d^3 \mathbf{r}_2 (\mathbf{r}_1 - \mathbf{r}_2)^2 \quad (2.7)$$

$$\sum_{n_x, n_y, n_z = -\infty}^{\infty} \left[\exp \left(-\frac{3(\mathbf{r}_1 - f_{\mathbf{n}}(\mathbf{r}_2))^2}{2L(CD)b} \right) \right], \quad (2.8)$$

where

$$f_{\mathbf{n}}(\mathbf{r})_i := (-1)^{n_i} (r_i - n_i L_i) \quad (2.9)$$

takes care of the reflections. The number of reflections is counted by \mathbf{n} in each direction. The quantity b is the step length of the random walk and $L(CD)$ is its length. A natural choice for b is the correlation length. In that case $L(CD)$ would be the contour length of the random walk of correlation blobs. The problem here is that we do not know the functional relationship $L = L(CD)$. Let us use the contour length of the chromatin fiber, using the same values as above. The size of the box we choose to be $3.1 \times 3.1 \times 0.15 \mu\text{m}$. Fitting the curve of $\sqrt{\langle (SD(CD))^2 \rangle}$ to the 3d-FISH data [58] we extract a step size of $b = 80 \text{ nm}$. We can also estimate the correlation length using equation (2.4). We assume $a = 300 \text{ nm}$ as Kuhn length and a rigid rod excluded volume a la Onsager [16] of $v = a^2 d$ where we take $d = 30 \text{ nm}$ as diameter. This results in an effective segment volume of $\pi d^2 a / 4 \text{ nm}^3$. If we assume again a volume fraction of 0.1 we recover a correlation length of 80 nm. This seems to be good to be true, especially after realizing that the correlation length is *shorter* than the Kuhn length on which we based the calculation. In fact we find for the thermal blob size, the size up to which a non-confined chain can be considered ideal, using the same fiber parameters as above, a value larger than $1 \mu\text{m}$. We have apparently overestimated the segment density in deriving equation (2.4). Using instead ideal chain scaling up to the correlation length results in a correlation length as follows

$$\frac{\xi_{\text{ideal}}}{a} \approx \frac{1}{C_0 a^3} \quad (2.10)$$

With the same values as before we now find a correlation length of 24 nm, a value much smaller than before. This point illustrates nicely that the data available at the moment do not justify detailed models.

We can repeat the same calculations for a heterochromatin region. The resulting fit for a correlation length of 20 nm is depicted in Figure 2.2 (b). The compaction of the heterochromatin is, depending on the excluded volume parameter, $4^{(3/4)} \cong 2.8$, equation (2.4), to 4, equation (2.10), times higher in this naive view.

Although the concept of a globule seems to fit the data well, this is somewhat deceptive. First of all FISH data are not very reliable. Higher moments might reveal more structure than a globular state. One should also not forget that the nucleus is never in equilibrium. Active processes do influence the distribution of chromatin and the cell is continuously evolving through its cell cycle. An interesting example is a recent simulation by Rosa and Everaers [64] where the interphase structure is a consequence of the fast decondensation of structured mitotic chromosomes within the confinement of the nuclear envelope, whereas the timescale of equilibration is much longer than the duration of interphase. The consequence is that the chromatin fiber is not thermally relaxed. Although they compare their simulations with rather unreliable 2D-FISH data, the authors demonstrate how the nonequilibrium conditions in the nucleus can influence the spatial distribution of chromatin.

2.3.3 Conclusion: large scale structure

The techniques available, in the time the paper on which this chapter is based was written, to reveal the large scale organization of chromatin do in our opinion not justify any detailed polymer model. Most features one can capture with the statement that chromatin exists in a condensed state. A new development is the 5C (chromosome conformation capture carbon copy) technique, developed by Job Dekker [65], where the cell is fixated for some time allowing parts of the chromatin that are in close proximity to attach. Statistical analysis on the contact points gives an indication of what parts of the chromosome are structurally in close proximity. Based on this analysis it is argued, that the chromatin fiber is ordered as a fractal. The benefit would be the lack of entanglement. This possible fractal structure was hinted at by Grosberg [66] already in 1988. His ideas were based on theoretical arguments from deGennes [67] that a collapsed linear polymer does not collapse immediately to a globule in the Flory sense, but collapses first, in a cascade of smaller globules that have no time to thermally mix, into a fractal globule. The weak point of this line of thought is that in reality the chromatin never collapses from a good solvent, but is continuously in a more or less condensed globular state. A weaker statement is that the chromatin happens to be in a non-entangled conformation as a result of its history, and that the time scale of entanglement is much longer than the cell cycle referring again to the simulations [64] mentioned in the last subsection

2.4 Mechanism behind compactification

As we have seen the chromosomes within the nucleus are more or less confined, each in its own domain. These domains are much smaller than the typical size of a chromatin fiber in a good solvent. The amount of compactification in interphase is furthermore not constant but varies considerably between regions containing active genes and less active regions.

Within the cell cycle a dramatic increase of compactification is apparent when the cell enters metaphase. This change from dispersed interphase chromosomes towards compactified mitotic chromosomes during prophase is fast and remarkably synchronous.

The question one has to ask is what are the forces involved in compactifying the chromatin inside the nucleus, depending on its activity during interphase and into mitotic chromosome during mitosis, while keeping different chromosomes separated. At the same time these forces might be also somehow relevant for sister chromatid segregation after replication.

One factor we will not consider here are active processes that could take part in condensing or decondensing chromatin. We like to think that these factors are working on top of a more general mechanism.

2.4.1 Enclosure

At first thought one might think that the chromatin is confined by the nuclear envelope, the double membrane encapsulating the nucleus. This was assumed in a recent simulation where it was also pointed out that the time scale of the cell cycle is too short for the expanded chromatins to thermalize [64] and especially to mix. This is an attractive scenario, especially reminding us that the organization of chromatin is probably never in equilibrium.

Purely by enclosing the nucleus it is, however, not possible to regulate the densities depending on gene activity. Especially, some regions in the nucleus seem to be chromatin-free without any membrane separating them from the chromatin containing regions [68, 51]. Finally there is still the need for a mechanism to condense the chromatin for mitosis.

2.4.2 Specific binding sites

Two somewhat overlapping models start from a more or less rigid backbone of proteins or protein filaments.

Nuclear Matrix

The nuclear matrix (for a review see [69]) is conceived to be a network of filaments that span the nucleus reminiscent of the cytoskeleton in the cytoplasm. The idea is that this network is responsible for the organization in chromatin compartments. The chromatin is supposed to be attached to this network through specific sequences along the genome appropriately named MAR's or Matrix Attachment Regions, some not precisely known, presumably AT

rich regions. An important problem with this model is that the network itself has never been indisputably detected and it is often defined as “the stuff that is left behind within the nuclear envelope when all other material has been extracted.” The electron microscope images suggesting such a structure could suffer from artifacts caused by sample preparation. A second objection is the variable position of each chromosome from cell to cell [49] that makes the existence of such a well defined nuclear matrix less likely.

Finally a large number of specific MAR's slightly contradicts the notion of robustness under mutation and of the constrained diffusion of chromatin observed during interphase [70]. It is nonetheless very well possible that some parts of the chromosome are localized with respect to the nuclear envelope. Conceptually this seems likely at least for the centromere, being the only part of the chromatin that needs to be localized during mitosis.

Scaffold

The idea that a rigid protein scaffold organizes the chromosomes originates from EM pictures of mitotic chromosomes depleted from histones using high salt concentrations [71]. A core of non-histone proteins in the shape of the original mitotic chromosome remains with a halo of bare DNA loops attached to it. The main constituents of this scaffold were later found to be topoisomerase II [72] and condensin [73], the former being a protein complex responsible for disentangling DNA, the latter being a complex closely related to the cohesin complex that keeps the two sister chromatids bound together up to telophase [74]. The picture that evolved was that of loops of chromatin attached to the protein scaffold through specific AT rich regions called SAR's (scaffold attachment regions). The notion of a scaffold has some overlap with the previously mentioned nuclear matrix and not too surprisingly the weak points of the matrix concept carry over to the scaffold.

A strong argument against the structural importance of a mitotic scaffold can be drawn from experiments by Poirier and Marko in 2002 [75], where it was found that the mitotic chromosome loses its structural integrity by gradually cutting the DNA with nuclease, showing that the chromatin fiber defines the structure of the chromosome, making the existence of a scaffold unlikely. Moreover, it was already known that topoisomerase II was not needed for the formation of the mitotic structure [76]. In 2006 it was found, surprisingly, that the same holds for condensin. The main function of condensin seems to be the stabilization of the mitotic chromosome during telophase [77].

2.4.3 Non specific compactification

Electrostatics

An alternative to these local types of mechanisms is the idea that attractive interactions between segments of the chromatin fiber dominate the hardcore like repulsive forces in such a way that the 2nd virial coefficient becomes negative, turning the effective background of the chromatin into a poor solvent. This attraction could be caused by a tail bridging effect, as

we have discussed above between nucleosomes, and/or mediated by dynamically chromatin binding proteins.

There are some suggestions in the literature [78] that the environment in the nucleus is such that most counterions of even strong polyelectrolytes are condensed, caused by the lack of free water. Measurements on several cell types do not support this view:

- The water content of the nuclei of amphibian oocytes was measured to be around 74 – 80% of mass [79] and of rat liver cells over 85% [80].
- NMR measurements [81] of frog (*Xenopus Laevis*) oocytes showed that almost 90% of the water present in the nucleus can be considered free.
- Even the water content of the highly condensed mitotic chromosome was estimated to be at least one third of the volume [82].

There is at this stage no reason to believe that the electrochemical conditions within the nucleus differ much from what one considers physiological:

- The nuclear envelope contains a large number of nuclear pores, typically around 3000, but in certain oocytes up to 50 million [83]. These pores are fully permeable for particles up to 9 nm in diameter [84].
- The 30 nm chromatin fiber as seen *in vitro* in “physiological” conditions closely resembles the fiberlike structure seen in FISH experiments [44], as explained before.
- As mentioned above NCPs seem to be tuned for minimizing their second virial coefficients at ‘physiological’ salt concentrations.
- Extracted mitotic chromosomes appear to have their *in vivo* dimensions at “physiological” conditions, reversibly condensing or decondensing with a change in salt concentration [82].
- Upon hypertonic shock the chromatin partly condenses [51], indicating a strong dependence on salt concentration.

The change in the level of compactification is correlated with some of the histone modifications. A typical example is an H3 phosphorylation during mitosis [85]. The same phosphorylation is correlated with the formation of heterochromatin [86]. It is possible that some of the modifications have a direct effect on the interaction between segments of the chromatin fiber through changes in the charge of the tails. Alternatively they could affect the binding of proteins that mediate the electrostatic interaction.

A somewhat puzzling feature is the formation of the aforementioned hypercondensation caused by an increase of salt concentration, a condition that weakens the range of the electrostatic forces through increased screening. Another point that still is hard to understand is how this subtle balance between attractive and repulsive forces can be maintained over such a large variety of cells and organisms.

Crowding

The notion of crowding in biology [87] corresponds roughly to the notion of depletion in colloid physics. For the case that the colloids are much larger than the polymers, the so-called colloid limit, the effective attraction between the colloids is reasonably well understood [8]: When two colloids are separated by a distance shorter than the size of the polymer coils, the osmotic pressure of the polymers pushes the two colloids together. Subtleties arise from the ease with which two polymers can overlap [3] and from the appearance of a repulsive barrier when incorporating higher order interaction terms, but the main picture remains.

In the protein limit where the colloid is much smaller than the polymer, the situation is not that clear. This is the limit of interest for the condensation of chromatin in the nucleus. If we think of the chromatin fiber as being dilute in a good solvent or in a semi-dilute solution, the relevant length scale characterizing the fiber, namely the radius of gyration or the correlation length, is considerably larger than the typical size of the proteins (around 2.5 nm). In that case it was shown by DeGennes [1] that purely by excluded volume effects polymers and nanoparticles with the size of an average protein are highly miscible. The conceptual reason is that then the range of the depletion is set by the size of the protein [88]. Since the cost of placing the nanoparticle within the polymer is expected to be proportional to the segment concentration, a scaling argument shows that this requires little work. In that case is also the minimum of the depletion potential between two nanoparticles not deep enough to cause phase separation. This was also confirmed in experiments, where non-DNA binding proteins from E-coli extracts alone could not condense DNA [89], even not with concentrations considerably higher than what is found in eukaryotic nuclei. Another experiment that indicates that nanoparticles of the size of the average protein easily diffuse within the densest chromatin regions, even within the mitotic chromatin fibers, was conducted by Verschure et al. [90].

Unless other effects decrease the excluded volume parameter of the chromatin fiber and/or increase the interaction between DNA and nanoparticles, depletion interactions seem to be too small to cause condensation of the chromatin fiber. A recent paper [91] discusses the possibility that charged nanoparticles, with the same charge as the polymer, could cause condensation in the spirit of Odijk's osmotic compaction [92] of the supercoiled DNA in E-coli. The negatively charged protein fraction needed seems to be too high to be realizable, without further ingredients, in the nucleus, where overall non-histone protein fractions are 0.1 – 0.15.

2.5 Conclusion

It is clear that the eukaryotic nucleus is an extremely complex system. The biophysical approaches have made progress in the description of chromatin by building the models from a detailed picture. At some point the barriers to overcome seems, however, to be too hard to continue in that way. On the other hand it seems logical that the robustness of the eukaryotic nucleus against changing conditions and mutations must have a general physical explanation.

In this chapter we have discussed the strengths and the weaknesses of the common approaches towards an understanding of the cell cycle. We have shown that no detailed model is needed to explain the large scale organization of chromatin observed so far. This notion makes it possible to separate the physics behind the organization from the detailed structure. We hope that this approach will lead in the near future to a better understanding of the cell cycle.

