Nucleosome Curtains and their application to the study of DNA condensation by condensin

Corentin J. Moevus

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019
ABSTRACT

Nucleosome Curtains and their application to the study of DNA condensation by condensin

Corentin J. Moevus

Single-molecule techniques have the potential to resolve exquisite details inaccessible by common ensemble techniques, such as the dynamic and static heterogeneities of a population. In particular, “DNA Curtains”, a technique based on total internal reflection fluorescence microscopy, allows for the high-throughput studies of DNA-based processes in real time, at the single-molecule level. The nucleosome, the basic unit of chromatin, is composed of 147 base pairs of DNA wrapped around an octamer of histone proteins. Nucleosomes have been extensively studied due to their fundamental inhibitory effect on all eukaryotic DNA-based processes. To study nucleosomes at the single-molecule level using DNA Curtains, a technique based on fluorophore-labeled recombinant nucleosomes was developed. This technique, called Nucleosome Curtains, allows for the study of hundreds of nucleosomes and their interactions with DNA-based processes at the single-molecule resolution, in real-time. Nucleosome Curtains were used to study the effect of nucleosomes on DNA condensation by condensin. Titration of nucleosomes on DNA does not result in slower DNA compaction by condensin, indicating that nucleosomes do not inhibit this process. Direct observation of the interactions between nucleosomes and ongoing DNA compaction by condensin reveals that nucleosomes are readily compacted 80% of the time without any slow-down from condensin. Hence, Nucleosome Curtains is a stable, powerful technique to study nucleosomes at the single-molecule level, and demonstrated that nucleosomes do not affect DNA condensation by condensin.
# Contents

List of Figures ........................................... v

List of Tables ............................................. x

List of Abbreviations and Acronyms ............................... xi

Introduction .................................................. 1

1  Nucleosomes ................................................. 1

1.1  Brief history review ......................................... 1

1.2  The nucleosome core particle ................................. 4

1.3  Structure of the histones in the nucleosome core particle ...... 5

1.4  Structure of nucleosomal DNA ................................ 8

1.5  Histone-DNA interactions .................................. 11

1.6  Evolutionary conservation of histones ......................... 13

1.7  Histone H1 .................................................. 15

1.8  Chromatin and higher order structures ....................... 17

1.9  Nucleosome positioning .................................... 22

1.10 Nucleosome dynamics ..................................... 27
1.11 Chromatin remodelers ........................................ 38
1.12 Histone post-translational modifications ................. 41
1.13 Histone variants ................................................. 50
1.14 Nucleosome (dis)assembly .................................... 59

2 Chromatin condensation ........................................... 64
2.1 Condensation of chromosomes ................................ 64
2.2 Structural maintenance of chromosomes (SMC) proteins .. 66
2.3 Mechanisms of chromosome morphogenesis ............... 69

3 Single molecule experimentation .................................. 76
3.1 The limitations of ensemble techniques ...................... 76
3.2 Overview of single-molecule techniques ...................... 80

4 DNA Curtains .......................................................... 86

1 Establishment of a protocol for the study of nucleosomes using DNA Curtains 92
1.1 Single-molecule studies of nucleosomes ...................... 92
1.2 Study of nucleosomes using DNA Curtains ................. 95
1.3 Selection of cysteine mutation sites for labeling .......... 100
1.4 Purification of recombinant histone octamers .............. 103
1.5 Purification of \( \lambda \) DNA ........................................ 105
1.6 Reconstitution of nucleosomes .................................. 106
1.7 Selection of organic dyes for labeling nucleosomes ....... 108
1.8 Microccocal nuclease (MNase) treatment of nucleosomal DNA gives expected pro-
tection ................................................................. 109
2 Single-molecule nucleosome curtains reveal nucleosomes are not a barrier to loop extrusion by condensin 151

2.1 Introduction 151

2.1.1 Condensin is necessary for the proper formation of chromosomes 151

2.1.2 Condensin is a large, ring-shaped ATPase protein 152

2.1.3 Condensin binds to DNA 153

2.1.4 Simulations show that DNA looping is sufficient for chromosome formation 154

2.1.5 Condensin extrudes loops of DNA 155

2.1.6 Histone chaperones are necessary for the formation of chromosomes in vitro 156

2.2 Results 157

2.2.1 Single molecule observation of condensin reveals its DNA condensation and looping activity 157

2.2.2 Condensin associates with DNA loops 159
List of Figures

1  Electron micrographs of chromatin ........................................ 2
2  Crystal structure of the Xenopus nucleosome at 2.8Å ................... 4
3  Structure of each individual histone ....................................... 6
4  Assembly of the histone octamer ............................................. 7
5  Modification of the DNA helix in nucleosomes .......................... 9
6  Structure of the contact sites between histones and DNA .......... 12
7  Rates of histone evolution and different packing structure between \textit{S. cerevisiae} and \textit{X. laevis} .......................... 13
8  Cryo-EM pictures of the effect of H1 on trinucleosome structures ... 15
9  Steps and models for higher-order chromatin structures ............ 19
10 Sequence determinants of nucleosome positioning .................... 22
11 A model for nucleosome transient site exposure ....................... 28
12 Different modes of intrinsic nucleosome dynamics .................... 36
13 Different roles of chromatin remodellers ................................. 38
14 Post-translational modifications on the globular core of the histone octamer ........ 43
15 Post-translational modifications on histone tails ....................... 47
16 Assembly of nucleosomes in vivo and in vitro .......................... 60
1.13 Competition of non-nucleosomal histones on an electrophoretic mobility shift assay (EMSA) ................................................................. 138
1.14 Titration of histone octamers on 147 bp 601 DNA .................................................. 139
1.15 Screen for the optimal reconstitution ratio for the assembly of nucleosome core particle (NCP)s. ......................................................... 140
1.16 EMSA of NCPs reconstituted with dye-labeled nucleosomes ............................... 141
1.17 Examples of Nucleosome Curtains with all three cystein-mutant histone octamers 141
1.18 Example of surface binding of histones as a function of salt .................................. 142
1.19 Effect of salt on non-nucleosomal histones on DNA ............................................. 143
1.20 Nucleosome stability at high concentration ......................................................... 144
1.21 Position distribution of nucleosomes on λ DNA .................................................. 145
1.22 Effect of nucleosome reconstitution ratio on photobleaching steps .................... 146
1.23 Number of nucleosomes as a function of reconstitution ratio ............................. 147
1.24 MNase digestion of 19x601 plasmids .................................................................. 148
1.25 MNase assay on λ 601 constructs. ................................................................. 149
1.26 Nucleosome Curtains of dye-labeled histones on a 7x601 λ DNA construct ......... 150

2.1 Presentation of the dCas9 system to study DNA condensation ............................ 176
2.2 Schematic drawing of the compaction experiment .............................................. 177
2.3 Kymograms of DNA compaction by condensin as a function of ATP hydrolysis . 178
2.4 Histogram of compaction velocity of dcas9-QDot labeled DNA by condensin 179
2.5 Representative kymograms of the directionality of DNA condensation by condensin 180
2.6 Scatterplot of condensin compaction directionality ............................................. 181
2.7 Barrier view of the compaction of YoYoI-labeled DNA by condensin 181
2.8 Kymogram of a single DNA molecule compacted by condensin 182
2.9 Quantification of the fraction of DNA molecules compacted in the presence or absence of condensin 182
2.10 Quantification of the number of puncta per compaction events 183
2.11 Kymogram of formation and compaction of DNA loops by condensin 184
2.12 Kymogram of YoYoI-labeled DNA molecule with QDot705-labeled condensin 185
2.13 Quantification of the colocalization between condensin and DNA puncta 186
2.14 SDS PAGE of WT and H2A S47C histone octamers 187
2.15 MNase assay of reconstituted nucleosomes 187
2.16 DNA Curtains experiment with H2AS47CDY530 nucleosomes 188
2.17 Kymogram of a flow on/flow off on a nucleosome-reconstituted DNA molecule 189
2.18 Number of nucleosomes per DNA molecule as a function of the reconstitution ratio 190
2.19 Nucleosomes signal per DNA molecule as a function of the reconstitution ratio 191
2.20 Nucleosome binding as function of AT-richness of the underlying sequence 192
2.21 Survival probability of nucleosomes in CAB buffer 193
2.22 Photobleaching steps of H2A S47C DY530 labeled nucleosomes 194
2.23 Kymogram of the condensation of nucleosomal DNA 1:80 WT histone octamer by 5nM condensin 195
2.24 Condensation velocity as a function of nucleosome concentration 196
2.25 Possible outcomes of the interaction between condensin and nucleosomes during loop extrusion 197
2.26 Nucleosomes remain bound through complete and reversible compaction of nucleosomal DNA by condensin. ................................. 198
2.27 Compaction of nucleosomal DNA is dependent on condensin ATPase activity. . . 199
2.28 Quantification of H2A S47C DY530 signal intensity before compaction and after induced decompaction. .................................................. 200
2.29 Kymogram of the spontaneous release of H2A S47C DY530 nucleosomal DNA by condensin. .............................................................. 201
2.30 Kymogram of continued compaction over a nucleosome by a loop of DNA being extruded by condensin ....................................................... 202
2.31 Quantification of the outcomes of collisions between a DNA loop extruded by condensin and a nucleosome ...................................................... 203
2.32 Effect of nucleosome on compaction velocity ........................................ 204
2.33 Effect of nucleosome on compaction processivity ...................................... 205
2.34 Survival distribution of DNA loops integrating a nucleosome ...................... 206
2.35 Potential mechanisms of nucleosome bypass during loop extrusion by condensin 207
List of Tables

1  Number of known histone post-translational modifications for each core histone  .  .  41
List of Abbreviations and Acronyms

a.a.  amino acids.

ABC  ATP-binding cassette.

AFM  atomic force microscopy.

ARP  actin-related protein.

bp   base pairs.

BSA  bovine serum albumin.

CHD  Chromodomain Helicase DNA-binding.

cryoEM  cryogenic electron microscopy.

EMSA  electrophoretic mobility shift assay.

FACT  Facilitates access to chromatin.

FCS  fluorescence correlation spectroscopy.

FRET  Forster energy resonance transfer.

HAT  histone acetyl transferase.

HDAC  histone deacetylase.

HEAT  “huntingtin, elongation factor 3, protein phosphatase 2A and TOR1 kinase”.
ISWI  Imitation Switch.

MNase  microccocal nuclease.

NCP  nucleosome core particle.

NuRD  Nucleosome remodeling and deacetylase.

NURF  Nucleosome remodeling factor.

PARP1  Poly(ADP-ribose) polymerase 1.

PEG  poly(ethylene glycol).

PHD  Plant homeodomain.

PTM  post-translational modification.

RSC  Remodels the structure of chromatin.

SAGA  Spt-Ada-Gcn5 acetyltransferase.

SAXS  small angle X-ray scattering.

SHL  superhelix axis location.

SMC  structural maintenance of chromosomes.

SWI/SNF  Switch/Sucrose Non-Fermentable.

TAD  topologically associating domain.

TCEP  tris(2-carboxyethyl)phosphine.

TIR  total internal reflection.

TIRF  total internal reflection fluorescence.

WHD  winged-helix domain.

WT  wild-type.
Acknowledgements

I am thankful to Sy Redding, for introducing me to biophysics and mentoring me in my first steps, to Daniel Duzdevich for his constant support, discussions and insights, and to Fabian Erdel for reintroducing me to a proper work-life balance and for his constant support, help and wisdom. All three participated greatly to the unraveling of my thesis and I appreciated every laugh and argument we had together. I am thankful to Eric Greene for giving me the space and time to explore crazy projects and ideas, and for putting me back in the right path when they would go too far, or nowhere. He has created many opportunities for me to learn and grow that have transformed me as a scientist and as a person. I thank Gregg Gundersen and Yinghui Mao for giving me a chance to learn with them, as well as my thesis committee, Hynek Wichterle, Steve Goff, Ruben Gonzalez, and Shan Zha, for their help, input, and support. I am also grateful to Shan Zha, Gregg Gundersen, Yinghui Mao, and Livio Pellizzoni for interviewing me and giving me a chance at Columbia University, as well as to Ron Liem. I thank Zaia Sivo for her constant support, for navigating the administration system for me, and for bearing with my complaints. I also thank the Boeringher Ingelheim Fonds team, Claudia, Sandra and Anja, for their financial support and mentorship. BIF is an amazing fellowship to receive and a great family, thanks to which I have met amazing people and learned tremendously. Finally, I would like to thank my friends, family, and especially Itziar, for their constant attention and support, especially in the last few months.
Introduction

My thesis revolves around the study of nucleosomes at the single-molecule model, with the bulk of my work consisting in the development of a protocol to study nucleosomes using DNA Curtains, and in the implementation of this protocol to study the effect of nucleosomes on DNA condensation by condensin. I will first present nucleosomes, DNA condensation, and single-molecule biology. I will present more thoroughly the single-molecule technique “DNA Curtains”, which I used throughout my thesis.

The two chapters will explain my work on nucleosomes and condensin, and the significance of my results. The first chapter presents the development of the single-molecule technique Nucleosome Curtains, an extension of DNA Curtains to study nucleosomes. The second chapter will present my work on the study of DNA condensation by condensin using Nucleosome Curtains, and the breakthroughs that came out of this work.

1 Nucleosomes

1.1 Brief history review

Almost a century and a half ago, Walther Flemming named chromatin the thread-like structure that would heavily stain with his basophilic dyes. He said of it “the word chromatin may stand until
its chemical nature is known, and meanwhile stands for that substance in the cell nucleus which is readily stained.” (Flemming, 1882; D. E. Olins & Olins, 2003). It took about a century to define chromatin as a polymer of nucleosomes, but the name remains.

Hewish and Burgoyne (1973) observed that, when digesting chromatin with endonucleases, it would form a “ladder” on a gel: a repeating lattice of DNA bands all spaced by the same amount. This was the first indication of chromatin as a polymer. Subsequently, A. L. Olins and Olins (1974) published beautiful micrographs of the “beads-on-a-string” model of chromatin, showing clear “bead” units linked by a main scaffold (fig. 1). They also report that Woodcock et al. captured the same images at the same time, independently. Woodcock, however, faced harsher criticism from a Nature reviewer: “A eukaryotic chromosome made out of self-assembling 70 Å units, which could perhaps be made to crystallize, would necessitate rewriting our textbooks on cytology and genetics! I have never read such a naive paper purporting to be of such fundamental significance. Definitely it should not be published anywhere!” (D. E. Olins
While this questions the authority and trust one should put in reviewers, Woodcock was able to publish the micrographs, a few years later (C. Woodcock, Safer, & Stanchfield, 1976). This burst of discoveries was crowned by Roger Kornberg’s proposal for chromatin as a repeating structure of histone proteins with about 200 bp of DNA (Kornberg, 1974). A year later, Oudet, Gross-Bellard, and Chambon (1975) combined electron microscopy and biochemistry to demonstrate the model and coined the term “nucleosome” as a complex of DNA and histone proteins. A more thorough, interestingly illustrated, and pleasantly narrated account of the discovery of the nucleosome was published by D. E. Olins and Olins (2003).

Histones, first discovered in 1884 (Kossel, 1884), form a family of small proteins (11-22 kDa), generally heavily positively charged, that are extremely conserved among eukaryotes. There are five main histones, in increasing order of mass: H4, H2B, H2A, H3, and H1 (Panyim & Chalkley, 1969; Phillips & Johns, 1965). They form the majority of the protein content of chromatin (Van Holde, 1989). Four histones, H2A, H2B, H3, and H4, each present in two units, form the protein part of the nucleosome, named the histone octamer due to its stoichiometry (J O Thomas & Kornberg, 1975; Jean O. Thomas & Butler, 1977). The histone octamer was first shown to bind 145-148 bp of DNA (Grellet, Penon, & Cooke, 1980; Lohr & Van Holde, 1979; Mirzabekov, Shick, Belyavsky, & Bavykin, 1978; Shindo, McGhee, & Cohen, 1980; Simpson & Künzler, 1979; Kensal E. van Holde, 1989b), but crystal structures indicate more precisely that it binds 147 bp (see section 1.4 and Davey, Sargent, Luger, Maeder, and Richmond, 2002). A nice historical perspective of the discovery of histones is given in Kensal E. van Holde (1989a).
Figure 2: Crystal structure of the *X. laevis* nucleosome at 2.8Å. Ribbon representation. H2A, H2B, H3 and H4 are depicted in yellow, red, blue and green, respectively. DNA is colored in turquoise and orange. Adapted from Luger, Mäder, Richmond, Sargent, and Richmond (1997a).

1.2 The nucleosome core particle

The nucleosome core particle, often abbreviated nucleosome core particle (NCP), consists of 147 base pairs (bp) of DNA with 2 copies of each histone H2A, H2B, H3, and H4 (the core histones). It is called so because it is the structure that is isolated after nuclease digestion of chromatin (Hewish & Burgoyne, 1973; Noll, 1974), indicating tighter binding of DNA to the structure. The nucleosome core particle is what is generally referred to as the “nucleosome”, although this is a misnomer. A nucleosome also contains the linker DNA between two NCPs, making its actual DNA length longer than 147 bp. If histone H1 is added to the nucleosome core particle with the 10-20 bp of DNA it protects (see section 1.7 and Nikitina, Wang, Gomberg, Grigoryev, and Zhurkin, 2013)—a sequence called the linker DNA, because it links two nucleosomes together—it is called a chromatosome. Chromatin is actually a polymer of chromatosomes, that is, of nucleosome core particles separated
by H1 bound to linker DNA of varying length (Hörz & Zachau, 1980).

The definitive structure and composition of the nucleosome core particle was obtained by X-ray crystallography by Luger, Mäder, Richmond, Sargent, and Richmond (1997a). The nucleosome is composed of about 146 bp of DNA (instead of 147 bp, due to a slight asymmetry in the structure, see section 1.4) forming a left-handed superhelix of 1.67 turns around a protein spool of core histones (fig. 2). The nucleosome has a diameter of about 11 nm and a thickness of about 5.5 nm. The structure revealed key information about the histones and the DNA that form the NCP.

1.3 Structure of the histones in the nucleosome core particle

The ternary structure of the histone octamer was first determined in the absence of DNA by Arents, Burlingame, Wang, Love, and Moudrianakis (1991). A key contribution of this structure was the discovery of the histone fold motif. The histone fold is conserved in all four core histones and found in many more DNA-associated proteins (Andreas D. Baxevanis, Arents, Moudrianakis, & Landsman, 1995). It occupies most of the mass of the core histones and consists of an α-helix of 11 residues (α1), followed by a short loop and a β-strand (L1), a 27 residues helix (α2), one more short loop and β-strand (L2), and finally another helix of 11 residues (α3) (Arents & Moudrianakis, 1995). Each histone is depicted in fig. 3. One can see the histone fold as a similarly elongated, tripartite structure in each of the core histones (on each corner) versus the globular domain of H1 (at the center). The histone fold motif organizes heterodimerization of histones in a characteristic handshake interaction, where the chains literally “clasp” each other and form a large interaction surface in an antiparallel manner. The interactions are mostly stabilized by hydrophobic residues (Arents et al., 1991; Luger et al., 1997a; Ramakrishnan, 1997).
In solution, histones form dimers of H2A-H2B and tetromers of H3-H4 that assemble into an octamer in the presence of high salt (> 1.5M) or of DNA (Ruiz-Carrillo & Jorcano, 1979). While the dimer interactions are mediated by the histone fold “handshakes”, the tetramerization of the H3-H4 dimers is mediated by a 4-helix bundle on the C-terminal tail of H3. The association of the H2A-H2B dimers with the H3-H4 tetramer is mediated by another 4-helix bundle between H2B and H4 (fig. 2) (Luger et al., 1997a). These interactions are largely driven by hydrophobic interactions, and the presence of a hydrophobic pocket at the core of the histone octamer is probably involved in its stabilization at high salt concentration, as the highly charged environment would disfavor the exposition of the hydrophobic residues (Ramakrishnan, 1997).

The octamer assembly reaction is depicted in fig. 4. The associated equilibrium constants are estimated to be $9 \times 10^{-8}$, $3 \times 10^{-9} \text{M}$, $9 \times 10^{-6} \text{M}$, $5.2 \times 10^{-23} \text{M}$, $2 \times 10^{-6} \text{M}$, and $2 \times 10^{-6} \text{M}$, for $K_{d,1}$, $K_{d,1}$, $K_{d,2}$, $K_{d,3}$, $K_{d,4}$, and $K_{d,5}$, respectively, the three first ones being at 0.1 M salt, and the two last ones at 2 M salt (Rippe, Mazurkiewicz, & Kepper, 2008). The tetramerization of H3-H4 is extremely favorable, making it unlikely to find isolated H3-H4 dimers in solution, whereas the dimerization of H2A and H2B, as well as its association with the H3-H4 tetramer, are comparatively weak. Note that at very low ionic strength, histones are unstructured, and that at physiological ionic strength, they aggregate, making these measurements particularly difficult (Karantza, Freire,
& Moudrianakis, 1996; Van Holde, 1989). “The histones are commonly regarded as unpleasant proteins for rigorous studies.” (Luck et al., 1956)

Histone N-terminal tails account for about 25% of the protein mass. They are unstructured in the crystal structure and contain a large amount of arginine and lysines. They protrude outside of the octameric core. While all the other core histones have a short C-terminal tail (3-8 residues in X. laevis) after the histone fold, H2A also has an unstructured C-terminal tail (28 residues in

![Diagram of histone octamer assembly](image)

Figure 4: Assembly of the histone octamer. Adapted from Rippe, Mazurkiewicz, and Kepper (2008).
X. laevis) presenting a similarly high amount of lysine and arginine (Luger et al., 1997a). While histone tails are mostly involved in contacts with DNA (see section 1.5), the H4 tail has been observed to interact with an acidic patch on the surface of the H2A-H2B dimer of an adjacent nucleosome, both in vitro (Dorigo, Schalch, Bystricky, & Richmond, 2003) and in vivo (Wilkins et al., 2014), as well as in the crystal structures (Luger et al., 1997a; White, Suto, & Luger, 2001). In the crystal structure of Saccharomyces cerevisiae nucleosomes, the H2B C-terminal tail is also seen to interact with that of a neighboring nucleosome. This interaction is prevalent enough that it causes the S. cerevisiae crystal to pack differently from the X. laevis crystal (see section 1.6 and White et al., 2001).

Overall, the histone octamer forms a spool-shaped structure assembled as a tetramer of H3-H4 with two dimers of H2A-H2B. The octamer is mostly held together by interactions between the H3-H4 tetramer and H2A-H2B dimer subunits, while the heterodimers are held together by interactions, mostly hydrophobic, between the histone fold domains of the monomers in a characteristic “handshake” manner. The histone N-terminal tails, as well as the C-terminal tail of H2A, are unstructured and extend outside of the main protein spool.

### 1.4 Structure of nucleosomal DNA

The nucleosome incorporates 147 bp of DNA making 1.67 turns around the histone octamer complex, forming a left-handed superhelix with a radius of 41.9 Å and a pitch of 25.9Å (Davey et al., 2002; Richmond & Davey, 2003). This extremely tight curvature—considering that the persistent length of DNA is around 150 bp—leads to a compaction factor of 4.4-times compared to linear DNA. The DNA is primarily in the B-form conformation. Interestingly, the curvature between
Figure 5: Modification of the DNA helix in nucleosomes. (a) Representation of the concepts of roll, tilt, shift, slide and twist between DNA base pairs. Each “domino” is one base pair, with each side of the domino being one base. (b) Values for roll, tilt, shift, slide and twist in the structure of the DNA of a 147 bp nucleosome core particle. The base-pair-step values plotted are the means for the two halves of the symmetrical sequence (shown above, with the dyad position labelled ‘0’). The iCAT and iSAT curves (red) show the roll and tilt contributions to the ideal superhelix curvature. The minor-groove blocks show base-pair-step shift alternation (pink in shift and tilt) or kinks (orange in roll, slide and twist). The primary bound-phosphate groups are indicated above the base sequence by pointers (numbered as the 5′ phosphate of the adjacent base) showing the strand direction (dark, 3′→5′; light, 5′→3′) and the interacting histone motif (L1, loop 1; L2, loop 2; A1, α1). Adapted from Richmond and Davey (2003).
the base pairs in the DNA superhelix is twice that necessary to form 1.67 turns compared with an idealized superhelix. This is explained by the significant amount of twist, roll, shift, and twist in between the base pairs (fig. 5), with the actual roll curvature being 222% that of the idealized superhelix and happening mostly inside the major and minor grooves (Richmond & Davey, 2003). Twist and slide contribute to the superhelix pitch, while roll and shift, to the superhelix radius. A thorough description is given by Richmond and Davey (2003).

The central position of the DNA on the nucleosome is called the dyad. It sits directly on a base pair in the 147 bp structure, making it a symmetric structure with 73 bp on each side of the central base pair (Davey et al., 2002). The dyad is often labeled as position 0, or superhelix axis location (SHL)0. Every DNA helix turn is labeled relative to the dyad as SHL1, SHL2, etc. on one side and SHL-1, SHL-2, etc. on the other side. These locations happen to be every time the minor groove faces away from the histone octamer (Luger et al., 1997a).

Interestingly, the first crystal structure was obtained with 146 bp of DNA. This caused a twist-defect in the DNA structure that resulted in poor DNA resolution, This defect was corrected in the 147 bp structure, leading to much higher DNA resolution (Davey et al., 2002). This twist-defect lead to an average 10.2 bp per turn of the DNA helix, versus 10.5 for B-form DNA and 10.43 for the 147 bp DNA structure (Richmond & Davey, 2003). In vivo, nucleosome structures in the chicken erythrocyte have been observed around 10.17 bp per turn (Travers & Klug, 1987). This indicates that nucleosomes commonly cover less than 147 bp, and that while 147 bp is the coverage of the “ideal” nucleosome, genomic nucleosome core particles may protect the 145-148 bp range initially discovered (Grellet et al., 1980; Lohr & Van Holde, 1979; Mirzabekov et al., 1978; Shindo et al., 1980; Simpson & Künzler, 1979; Kensal E. van Holde, 1989b) by stretching or compressing the helix. The in vivo discrepancy, however, could also be explained by the sequence preference of the
microccocal nuclease (MNase) (Nikitina et al., 2013).

In brief, the DNA is wrapped 1.67-times around the histone core, forming a left-handed super-helix. While the DNA is mostly in B-form conformation, significant base pair step transformations (roll, shift, slide, tilt) are observed and more than account for the sharp curvature. The DNA is separated evenly around the dyad, leading to a close-to-B-form 10.43 bp/turn, while the double helix of DNA sequences of even base pair number is compressed to account for the asymmetry, for example to about 10.2 bp/turn in the case of a 146 bp sequence.

1.5 Histone-DNA interactions

While the flexibility of the sequence affects the preference of the histone octamer for DNA (see section 1.9), the histones are the major player in shaping the superhelix. For example, while it is thought that poly(dAT·dT) sequences next to promoters inhibit nucleosome deposition, only slight variations are observed in vitro (Jeffrey J. Hayes, Bashkin, Tullius, & Wolffe, 1991) and in the crystal structure of a nucleosome with a 16 bp poly(dAT·dT) sequence (Bao, White, & Luger, 2006). Hence, the DNA sequence only has a minor contribution in defining the structure of the nucleosome, which is dictated by the histone octamer.

Importantly, there is no base pair-specific contact within the nucleosome. All of the 14 contact points between the octamer and the DNA happen with a minor groove facing the octamer. An arginine is inserted into the minor groove every time it faces the histone core. (Luger et al., 1997a). Each histone monomer contributes to at least one arginine, with two from H2A and three from H3. Ten of these arginines come from the histone fold domains, while four come from the N-terminal tail of H2B and H3. These arginines interact with the phosphate backbone as well as with
the deoxyribose via hydrogen bonds and non-polar interactions. This causes a bend in the minor
groove, which narrows to a width of about 3 Å, versus a usual 12 Å (Richmond & Davey, 2003).
The histones fold domains organize 129 of the 147 bp protected in the nucleosome. Each histone
fold heterodimer organizes about 27-28 bp of DNA, with a 4 bp linker in between each (Richmond
& Davey, 2003). The contacts are orchestrated by the L1 and L2 loops, as well as the α1 helices
(see fig. 3 for nomenclature). Each histone monomer contacts the DNA at two locations: its α1
helix and its L1 or L2 loop. The other 18 bp of DNA is organized by the N-terminal tails of H2B
and H3.

The N-terminal tails of the core histones, as
well as the C-terminal tail of H2B, are heavily
positively charged due to their high lysine and
arginine content. The N-terminal tails of H3
and H2B extrude outside of the histone spool
via a channel in between two minor grooves of
the DNA superhelix. The N-terminal tail of H3
extrudes at SHL6.5 and -6.5, at the entry and
exit of the DNA onto the histone spool. This ar-
gues for an important role of the N-terminal tail
of H3 in controlling the wrapping and unwrapping
of DNA (see sections 1.10 and 1.12). The
N-terminal tails of H2A and H4 also extend out-
side of the superhelix and contacts a DNA mi-
nor groove, but not from in between the gyres

Figure 6: Structure of the contact sites between hi-
stones and DNA. (a) Side chains that make hydro-
gen bonds or hydrophobic interactions with the DNA
backbones, and arginines inserted in the minor groove
are shown. Main-chain-to-DNA-phosphate hydrogen
bonds are shown in magenta. Leucine 65 is in con-
tact with a thymidine methyl group (yellow bond).
Two of three arginine–threonine pairs are hydrogen
bonded (cyan bond). (b), H2A–H2B histone-fold pair.
Adapted from (Luger et al., 1997a)
Figure 7: (A) Rate of evolution of histones and various proteins. Data from 1978. Adapted from Van Holde (1989). (B and C) Different packing of the *X. laevis* versus *S. cerevisiae* nucleosome crystals. (B) *X. laevis* crystal packing of the NCPs viewed approximately down the superhelical axis (left) and rotated 90° around the b-axis (right). Only the DNA is shown for clarity; the same colors denote nucleosomes that lie within the same plane. (C) *S. cerevisiae* crystal packing of the NCPs in the same views as (A). Adapted from White, Suto, and Luger (2001).

of the superhelix (Luger et al., 1997a). While these tails are unstructured, and hence unseen in the crystal structure, they bind to DNA and account for 103 of the 203 salt bridges between histones and DNA at physiological salt concentration (Walker, 1984). They also participate in the thermal stability of the structure (Ausio, Dong, & van Holde, 1989).

### 1.6 Evolutionary conservation of histones

Histone sequences are heavily conserved among eukaryotes, and are among the slowest evolving proteins, as pictured in fig. 7 (Van Holde, 1989). The highest conserved histone is H4, with around 95% sequence identity among species, while H1 is the least conserved, at about 31% (A D Baxeva-
nis & Landsman, 1998). In general, yeast histones present the most difference, which may be related to the fact that the yeast genome is mostly expressed, as it is a single-cell organism, versus a larger amount of the genome of multi-cellular organisms being silenced. Those mild sequence difference barely show in the structure, with very few differences between yeast and metazoan nucleosome core particles. Indeed, superimposing the *S. cerevisiae* and *X. laevis* crystal structures gives a root mean square deviation of 1.57 Å. The main structural differences lie in the tails of the histones and in the packing of the crystal (White et al., 2001).

*X. laevis* and *S. cerevisiae* histones H2A, H2B, H3 and H4 share 72%, 67%, 84% and 92% sequence homology, respectively, with most of the difference in the histone tails. The N-terminal tail of H4 in *X. laevis* forms contacts with an acidic path on the surface of a nearby nucleosome. In *S. cerevisiae*, however, these contacts are not observed. Instead, the H4 N-terminal tail interacts with the DNA of a neighboring nucleosome. A similar behavior is observed for the C-terminal tail of H2A, with its α helix forming contacts with DNA of a neighboring particle. In *X. laevis*, only a few residues of the α-helix are involved in this contact, however, in *S. cerevisiae*, residues along the whole length are involved, participating in a tighter interaction. In addition to the different behavior of the tails, in the *S. cerevisiae* nucleosome crystal, the first α-helix of H3 interacts with the L1 loop of H2B in the neighboring NCP via a divalent ion (Mn\(^{2+}\)) interface. Overall, this causes a drastic change in the packing of the *S. cerevisiae* nucleosome crystal (see fig. 7) versus that of *X. laevis*. While this difference may not be relevant in vivo, it is indicative of potential ways a putative 30nm-fiber could stack (White et al., 2001 and section 1.8).
1.7 Histone H1

The study of H1 is complicated by its high variability between organisms and variants: *Homo sapiens* and *Mus musculus* have 11 H1 variants, *Caenorhabditis elegans* has 8 and *X. laevis* has 5. These variants can be tissue-specific or constitutively expressed, replication-dependent or -independent (Harshman, Young, Parthun, & Freitas, 2013). Moreover, it has been suggested that H1 would bind chromatin through different modes, adding to the complexity (Brown, Izard, & Misteli, 2006). Still, H1 and its variants were shown to fulfill several functions in the cell. First H1 stabilizes the linker DNA at the entry and exit of the nucleosome, promoting the assembly of higher order chromatin structures (James Allan, Mitchell, Harborne, Bohm, & Crane-Robinson, 1986; Syed et al., 2010; Thoma, Koller, & Klug, 1979). H1 has also been involved in nucleosome spacing (Y. Fan et al., 2003), replication origin firing (Thiriet & Hayes, 2009), gene expression (Shen & Gorovsky, 1996) and silencing (Xingwu Lu et al., 2013).

H1 is also known as the “linker histone” because it binds to the linker DNA in between the
nucleosome core particles in the chromatin fiber. It has been shown to organize and protect the linker DNA in 10 bp intervals, potentially up to 10 bp on each side of the nucleosome, forming what looks like a stem at the site of entry/exit of the DNA from the nucleosome (see fig. 8 and Nikitina et al., 2013; Syed et al., 2010). In addition, H1 was shown to protect the DNA at the dyad, probably because it sits in between the entry and exit sites of the nucleosome (refer to the nucleosome structure in fig. 2 and Syed et al., 2010). H1 binds in a 1:1 stoichiometry with the nucleosome, meaning that unlike the core histone, there is only 1 copy of H1 in the chromatosome (J J Hayes & Wolffe, 1993). In vivo, the range goes from 0.74 to 1.35 H1 molecules per nucleosome depending on organisms and tissue types, liver being on the lower side (Bates & Thomas, 1981).

Histone H1 is the least conserved of the histone family (A D Baxevanis and Landsman, 1998, and fig. 7), and as a consequence, its structure varies greatly between organisms. Metazoans H1 exhibit a tripartite structure, with a central globular domain flanked by two probably unstructured tails, the C-terminal one being highly basic (see fig. 3, the central histone, and Hartman, Chapman, Moss, and Bradbury, 1977). Avian erythrocyte cells also express a variant of H1 called H5 that follows the same structure, the globular domain exhibiting a winged-helix structure with 3 α-helices and a β-hairpin (Aviles, Chapman, Kneale, Crane-Robins, & Bradbury, 1978; Ramakrishnan et al., 1993). The S. cerevisiae homolog of H1, Hho1p, displays a different structure, with either only one helix-wing-helix domain and a possibly unstructured tail (K. Ono et al., 2003) or two globular domains (Landsman, 1996). Interestingly, and in contrast with mammals (Y. Fan et al., 2003; Y. Fan et al., 2005), Hho1p was shown no be non-essential, with no observed defect in viability, silencing or meiosis in cells depleted of Hho1p (Patterton, Landel, Landsman, Peterson, & Simpson, 1998). Newer reports indicate that the higher-order organization of chromatin is affected in Hho1p mutant yeasts (Georgieva, Roguev, Balashev, Zlatanova, & Miloshev, 2012), but the fact that yeast cells
presents no perceivable defect indicates that higher-order chromatin structures may not be essential in *S. cerevisiae*. In addition, it was shown that Hho1p is involved in the suppression of homologous recombination (Jessica A Downs, Kosmidou, Morgan, & Jackson, 2003).

Histone H1 binds to DNA more weakly than the core histones (Van Holde, 1989). It has also been shown to be more mobile than core histones in vivo, with a residence time in the order of minutes, indicating that H1 comes on and off chromatin relatively fast (Misteli, Gunjan, Hock, Bustin, & Brown, 2000). Several residues are involved in DNA binding, with the most important of them in the C-terminal tail (Hendzel, Lever, Crawford, & Th’ng, 2004), while the globular domain uses at least 8 arginines and lysines to bind DNA via electrostatic interactions (Brown et al., 2006). Mutation of either the C-terminal tail or the essential arginines and lysines of the globular domains leads to a loss of residence time on chromatin, down from minutes to several seconds to milliseconds (identical to non-interacting proteins) (Brown et al., 2006; Hendzel et al., 2004). In vivo, the stability of H1 binding to chromatin is modulated by local phosphorylation and dephosphorylation (Raghuram et al., 2013). The exact structure and binding sites of H1 on linker DNA and nucleosomes remains to be determined.

### 1.8 Chromatin and higher order structures

The polymer of nucleosomes is called chromatin. In its linear form, chromatin exists as the so-called “10 nm fiber”, named after the width of the fiber as observed by electron microscopy. It is the “beads-on-a-string” fiber of histone octamers on a molecule of DNA (see fig. 9A). The 10 nm fiber is actually 11 nm wide, since this is the width of a nucleosome (Luger et al., 1997a).

Histones can interact with each other to give a “secondary structure” to chromatin. One of
the most important interaction for chromatin folding into higher-order structures is that of the N-terminal tail of histone H4 with an acidic patch on the surface of the H2A-H2B dimer of a neighboring nucleosome (Dorigo et al., 2003; Pepenella, Murphy, & Hayes, 2014), as observed in the crystal structure of the *X. laevis* nucleosome (Luger et al., 1997a), but not in the *S. cerevisiae* nucleosome structure (White et al., 2001 and fig. 7). Outside of the crystal, this acidic patch was demonstrated to control chromatin folding and transcription repression (Zhou, Fan, Rangasamy, & Tremethick, 2007). It was also shown that lysine 16, on the N-terminal tail of H4, but not the other lysines, induces folding of chromatin. When H4 K16 is acetylated, the interaction is lost and chromatin decompacts (Allahverdi et al., 2011). It was also shown that all N-terminal tails of the histones contribute to chromatin folding (Fletcher & Hansen, 1995; Gordon, Luger, & Hansen, 2005). In the absence of nucleosome tails, no compaction is observed (Allan, Harborne, Rau, & Gould, 1982). Notably, the N-terminal tail of H3 was shown to transfer from intra-molecular to inter-molecular interactions as chromatin compacts (Zheng, Lu, Hansen, & Hayes, 2005). Another key interaction lies in the C-terminal of H2B. While it is unclear whether these peptides contribute to folding, their ubiquitination leads to the disruption of chromatin in a manner independent from H4 acetylation, indicating another folding (or unfolding only) pathway (Fierz et al., 2011). Overall, it appears that the N-terminal tails of histones are responsible for the higher-order folding of chromatin, although histone H1 also plays an essential role (James Allan et al., 1986; Georgieva et al., 2012; Thoma et al., 1979).

The secondary structure of the 10 nm fiber has been a subject of contention for more than 40 years. It has been proposed that the 10 nm fiber folds into a 30 nm fiber, but the structure of that fiber and its in vivo relevance have been debated. Finch and Klug (1976) first observed a 30 nm wide chromatin structure stabilized by histone H1. Its existence was later confirmed in vivo by
Figure 9: (A) Proposed steps of chromatin folding. DNA becomes 10 nm fiber once assembled into nucleosomes, that maybe become 30 nm fiber through the interaction between nucleosomes, this fiber itself folding until the mitotic chromosome structure is reached. Bottom right is a human nucleus. (B) Models of 30 nm structures, with nucleosomes represented as spheres and linker DNA as rods. Left: “one-start” model. The nucleosomes form a solenoid shape, interacting with the next nucleosome in the chain. The width of the fiber depends on the length of the linker. Right: “two-start” model. The nucleosomes rise in pairs, forming a helix similar to the DNA helix. Nucleosomes interact with the second next nucleosome. Nucleosomes are numbered as “N_x”. (C) Model for the discrepancy between in vivo and in vitro observations of the 30 nm fiber. At low concentrations, nucleosomes are more likely to interact within the same fiber. However, as concentration rises, such as in the nucleus and in mitotic chromosomes, the trans interactions between chromatin strands become more likely, leading the nucleosomes to behave like a viscous liquid or a “polymer melt”. Adapted from Maeshima, Hihara, and Eltsov (2010).

Small angle X-ray scattering (SAXS) (Langmore & Paulson, 1983). Several models were proposed for the formation of the fiber, among which 2 gathered the most evidence: a two-start model, with the helix forming in a fashion similar to that of DNA, and a one-start model following a solenoidal shape (fig. 9B). The two-start model gained the most traction. Evidence emerged from electron microscopy (C. L. Woodcock, Frado, & Rattner, 1984), X-ray crystallography (Schalch, Duda, Sargent, & Richmond, 2005), single-molecule spectroscopy (Kruithof et al., 2009), and cryogenic electron microscopy (cryoEM) (Song et al., 2014). Contenders of the one-start model demonstrated
it using electron microscopy (Finch & Klug, 1976; P. J. J. Robinson, Fairall, Huynh, & Rhodes, 2006) and SAXS (J. Widom & Klug, 1985). It appeared that main difference between the two models was the nucleosome repeat length (P. J. Robinson & Rhodes, 2006; Routh, Sandin, & Rhodes, 2008). Computer simulations confirmed that the two structures could be obtained at different repeat lengths, and that some, but not all repeat lengths could give rise to both structures (Grigoryev, Arya, Correll, Woodcock, & Schlick, 2009; Stehr, Kepper, Rippe, & Wedemann, 2008; Wong, Victor, & Mozziconacci, 2007). Hence, it appears that depending on the length of the linker DNA between histones, chromatin could either fold into a one-start or two-start 30-nm fiber, resolving the controversy by agreeing with both hypotheses.

As the controversy about the structure of the 30 nm fiber seemed to resolve, its relevance in vivo was questioned (Eltsov, MacLellan, Maeshima, Frangakis, & Dubochet, 2008). Most studies of the 30nm fiber have been done using chromatin reconstituted in vitro (K. van Holde & Zlatanova, 1995). In addition, most in vivo studies were flawed. Joti et al. (2012) demonstrated that the 30nm peak observed by SAXS by Langmore and Paulson (1983) was coming from ribosomes, and not from a putative 30 nm fiber. Dubochet and Sartori Blanc (2001) were also questioning the usage of electron microscopy for the study of cellular structures. Indeed, fixation techniques tend to cause the subject of study to aggregate, forming artifactual structures. Strukov and Belmont (2009) showed that the folding of chromosomes is not reproducible, arguing against a neatly-packed structure. Finally, building of the work from McDowall, Smith, and Dubochet (1986) using vitrified samples (cryoEM) and correcting for artifacts, Eltsov et al. (2008) denied the observation of the 30 nm fiber in chromosomes. Using the same tools used for the demonstration of the existence of the 30 nm fiber, electron microscopy and SAXS, Nishino et al. (2012) also concluded that the 30 nm fiber is not observable in vivo.
As the importance of the 30 nm fiber was tuned down, another line of evidence emerged to explain the structure of chromatin in vivo, presenting it as a disorganized soup of 10 nm fiber (fig. 9C). As Maeshima, Hihara, and Eltsov (2010) explain it:

However, at the high nucleosome concentrations that occur in vivo, inter-fibre nucleosome interactions become increasingly dominant...Nucleosome fibres are forced to interdigitate with one another. This interferes with the formation and maintenance of the 30-nm chromatin fibre, leading to the ‘polymer melt’ state...This means that the nucleosome does not ‘know’ to which fibre it belongs.

The first evidence of the “polymer melt” structure came from vitrified electron microscopy, proposing that “…the chromosome is formed by the compact association of 11 nm filaments, or portions thereof, interacting in a manner akin to the molecules of a liquid.” (McDowall et al., 1986) It was also shown that chromatin organizes into random megabase pair loops (Yokota, van den Engh, Hearst, Sachs, & Trask, 1995), and that chromatin can be found in open and closed states depending on its gene content (Gilbert et al., 2004). Computer simulations observed the polymer melt model (Stehr et al., 2008), in vivo diffusions of probes argued for the fractal structure of chromatin (Bancaud et al., 2009), and modern Hi-C techniques merged it all and further demonstrated the genome to be organized as a “molten globule” fractal, a loose globule of chromatin composed of loose globules of chromatin (Lieberman-Aiden et al., 2009). Observation of the chromatin of developing embryos did not lead to the observation of 30 nm fiber, even in condensed DNA regions, where 10 nm fiber was predominant (Ahmed et al., 2010).

Overall, while the question of the 30 nm fiber structure obsessed most of the higher-order chromatin field for the last few decades, it is now becoming clearer that the 30 nm fiber does not exist,
at least in significant amount, in vivo. It also makes sense. As the picture of chromatin is becoming clearer, with its swarms of post-translational modifications, histone variants and other protein complexes related to DNA’s functions and maintenance, it seems less and less likely that chromatin would fold into a neatly pitched helix and more and more likely it would resemble a molten globule. As a consequence, it is apparent that any structure above the 10 nm fiber will not be mediated by nucleosomes, but by other factors. While nucleosomes may participate in the stabilization of such structure, they would not induce it (Maeshima, Imai, Tamura, & Nozaki, 2014).

1.9 Nucleosome positioning

DNA has a persistence length of about 150 bp (P. J. Hagerman, 1988), meaning that significant energy is required to bend it and twist it so that 147 bp make 1.67 turns around the histone octamer (see section 1.4). It has been estimated that the free-energy cost of DNA bending and twisting in the octamer is about 76 kcal/mol and 6.6 kcal/mol, respectively (J. Widom, 2001). In comparison, the hydrolysis of ATP provides 7.6 kcal/mol. As a
result, any sequence that can alleviate some of that energetic cost, and hence lower the activation energy of the assembly reaction, would be preferred by the histone octamer. Bendable, bended, twistable and twisted DNA sequences, and in particular, sequences that can accommodate significant compression of the minor groove where the arginines of the octamer insert (see section 1.5), as well as the large amount of roll in the DNA molecule (refer to section 1.4) would provide an easier “landing pad” for histone octamers, compared to stiffer DNA sequences.

For a molecule of DNA \( L \) base pairs long, there is \( L - 146 \) potential nucleosome binding sites. If nucleosome assembly happens at equilibrium, the octamer will select the most favorable of the \( L - 146 \) sites for forming a nucleosome. The selection of a more favorable binding site for the nucleosome is called “nucleosome positioning”. Note that the equilibrium reached in vitro does not necessarily match equilibrium reached in vivo. The equilibrium is defined by the lowest energy state of the system, not of the nucleosome. A sequence is considered a stronger positioning sequence if, at equilibrium, it is more likely occupied by nucleosomes than another sequence.

Early studies have showed that repeated \((A/T)_3\)-NN-(G/C)_3-NN motifs are preferentially bound by histone octamers. The authors observed increased protection of the sequence in the A/T part of the decamers (Shrader & Crothers, 1989). The 10 bp periodicity, we now know, corresponds to where the octamer binds the DNA, every time the minor groove faces the octamer. To complement these studies, P. Lowary and Widom (1998) established a system to selectively enrich for stronger binding sequences. They started with a random pool of \( 5 \times 10^{12} \) 220 bp sequences and reconstituted nucleosomes with it, selecting only the top 10% by using a ratio of octamer to DNA of 1:10. Because reconstitution of nucleosomes in vitro reaches equilibrium (Shrader & Crothers, 1989), the 10% DNA sequences selected by the octamer should be the most favorable ones. They purified the nucleosomes on a sucrose gradient, extracted the DNA, amplified it by PCR, and then
repeated the process. This method is known as “SELEX”: systematic evolution of ligands by exponential enrichment (Irvine, Tuerk, & Gold, 1991). After 15 rounds of enrichment, they compared the relative binding free energy of the 30–50 remaining sequences versus that of the 5S sequence, a naturally occurring sequence that shows strong nucleosome positioning (Simpson & Stafford, 1983). Strikingly, the octamers showed a preferential binding to the SELEX sequences, with the strongest binding sequence, clone 601, exhibiting a $\Delta \Delta G^0$ of $-2.9 \pm 0.33$ kcal/mol versus 5S. Calculating the equilibrium preference:

$$
\Delta G^0 = -RT \ln K_{eq} \\
\Delta \Delta G^0 = -RT (\ln K_{eq,601} - \ln K_{eq,5S}) \\
= -RT (\ln \frac{K_{eq,601}}{K_{eq,5S}}) \\
\frac{\Delta \Delta G^0}{-RT} = \ln \frac{K_{eq,601}}{K_{eq,5S}} \\
\exp\left(\frac{\Delta \Delta G^0}{-RT}\right) = \frac{K_{eq,601}}{K_{eq,5S}}
$$

$$R = 0.001987 \text{kcal/mol} \quad T = 298K \quad \Delta \Delta G^0 = -2.9 \text{kcal/mol}$$

$$\frac{K_{eq,601}}{K_{eq,5S}} \approx 134$$

The 601 sequence is 134 times more likely to be assembled into a nucleosome than the 5S sequence, at room temperature. Note that here, $K_{eq} = \frac{[AB]}{[A][B]}$. The equilibrium constant is in the format used in chemistry, unlike the usual $K_{eq} = \frac{[A][B]}{[AB]}$ used in biochemistry. The 5S sequence itself is about 2.5 times more likely to be selected over random sequences, and the lowest known nucleosome binding sequence is 7.5 times less likely to be selected than the 5S (Thåström et al., 1999). Hence, overall, the 601 sequence is 1000 times more likely to be assembled into a nucleosome than the
Weakest nucleosome binding sequence.

All strong nucleosome positioning sequences share key features of bendability and twistability (Kevin Struhl and Segal, 2013; J. Widom, 2001 and fig. 10). First, a periodic positioning of bendable dinucleotides occurs every 10 bp, when the minor groove interacts with the octamer, and the octamer insert an arginine in the groove (see section 1.5). These dinucleotides are often AT or TA. Second, GC dinucleotides are often found out of phase with the AT/TA dinucleotides, also with a 10 bp periodicity. They are found where the major groove faces away from the histone octamer (like at the dyad). (Kevin Struhl & Segal, 2013; J. Widom, 2001). The combination of optical tweezers and single-molecule Forster energy resonance transfer (FRET) confirmed that more bendable sequences provide a tighter binding. The authors used the 601 sequence, which is asymmetric, and pulled on it from both sides. They observed that the stiffer side would always unwrap first, indicating less tight binding (C. J. Moevus & Greene, 2015; T. T. Ngo, Zhang, Zhou, Yodh, & Ha, 2015). In the crystal structure of the 147 bp nucleosome, kinking always happens at a CA/TG dinucleotide, when the minor groove faces the octamer. The DNA for the crystal structure was from human α-satellite repeats (Davey et al., 2002). Simulations indicate that CA/TG and TA dinucleotides are the most bendable (Tolstorukov, Colasanti, McCandlish, Olson, & Zhurkin, 2007). Studies of the nucleosome positioning sequences in the mouse genome have shown that these sequences are enriched in the CA dinucleotide too, arguing for its role in helping bend nucleosomal DNA (Widlund et al., 1997). It was also shown that G/C content is the most predictive factor for nucleosome positioning in vivo (Chua, Vasudevan, Davey, Wu, & Davey, 2012). A comprehensive review of the effect of DNA sequence on nucleosome positioning was written by J. Widom (2001).

While in vitro positioning depends on sequence affinity, what defines the most favorable binding site can be different in vivo and in vitro, and as such, sequence affinity does not equate nucleo-
some positioning. For example, albeit its extremely strong positioning capacity, the 601 sequence
does not pause transcription or position nucleosomes strongly in vivo (Perales, Zhang, & Bentley,
2011). Moreover, natural sequences found in genomes are at best 6 times weaker than artificial nu-
cleosome positioning sequences, indicating that little evolutionary pressure was put on this function
(Thåström et al., 1999). Hence, some forces other than sequence affinity are involved in defining
the position of nucleosomes in vivo.

Strong nucleosome positions in vivo are mostly related to the function of the genome. For
example, the +1 and -1 nucleosomes around a promoter (named so because they are the nucleo-
somes directly before and after the promoter) are positioned with remarkable consistency (Yuan
et al., 2005). It was proposed that these tightly positioned nucleosomes induce the positioning of
all surrounding nucleosomes, as the position of nucleosomes becomes fuzzier as a function of the
distance from the +1 and -1 nucleosomes (Mavrich et al., 2008). Nucleosomes are often thought
to be aligned by the presence of a poly(dAT·dT) sequence next to the promoter. Poly(dAT·dT)
are strongly inhibitory to nucleosome deposition due to their inherent stiffness (H. C. M. Nelson,
Finch, Luisi, & Klug, 1987; Struhl, 1985; Suter, Schnappauf, & Thoma, 2000). However, in vitro,
nucleosomes can form on poly(dAT·dT) sequences (Bao et al., 2006; Jeffrey J. Hayes et al., 1991).
One way to explain this discrepancy would be that in vivo, due to the wealth of sequence space,
the poly(dAT·dT) sequences are avoided, resulting in a consistent nucleosome-less space. Alterna-
tively, the avoidance of the poly(dAT·dT) sequences may be catalyzed by nucleosome remodelers
or by the preferential binding of other proteins on poly(dAT·dT) tracks. Some studies favor the
first explanation (J. D. Anderson & Widom, 2001; Segal & Widom, 2009) while some others favor
the second (A. L. Hughes, Jin, Rando, & Struhl, 2012; Krietenstein et al., 2016; Yong Zhang et al.,
2009), indicating that it probably is a contribution of both. Terminator sequences are also depleted
of nucleosomes, however, in that case, the sequence seems to play very little role (X. Fan et al., 2010).

The exact contribution of sequence to nucleosome positioning in vivo remains debated. While it was shown that nucleosomes preferentially bind to the *S. cerevisiae* genome versus the *Escherichia coli* one, reports vary on how much of the in vivo positioning can be recapitulated simply by sequence prediction (Segal et al., 2006; Yong Zhang et al., 2009). Interestingly, in vitro studies at the single molecule level of different nucleosome positioning prediction algorithms, as well as correlation between in vivo and in vitro sequence positions, indicate that sequence does play some role (Visnapuu & Greene, 2009). The exact contribution of sequence and function in nucleosome positioning remains to be determined, but it is clear that such studies will have to consider the energetic minima of the system, rather than that of the nucleosome, if they want to determine strong nucleosome positioning sequences in vivo.

1.10 Nucleosome dynamics

Nucleosomes are often seen as static entities that bar access to chromatin by steric hindrance. This vision may have been influenced by the impact the crystal structure had on the field (Andrews & Luger, 2011; Luger et al., 1997a). Indeed, crystal structures tend not to move. However, several lines of evidence indicate that nucleosomes are dynamic entities, with both intrinsic and extrinsic processes involved. First, nucleosomes are inherently dynamic, as they display breathing, sliding, and gaping. Second, nucleosomes are actively handled, remodeled and modified by other proteins.
Seminal work by the laboratory of Jonathan Widom introduced the concept of spontaneous site exposure (Polach & Widom, 1995). Spontaneous site exposure, also known as nucleosome breathing, dynamic unwrapping or transient site exposure, is the process through which a nucleosome randomly releases one/several of its 14 histone-DNA interactions (see section 1.5). If the release happens around the entry/exit of the nucleosome, the DNA is fully accessible—it becomes an extension of the linker DNA—and the protected sequences are accessible for other factors to bind (see fig. 11). Such a mechanism would allow nucleosome to prevent weakly-binding factors to access DNA. Indeed, weak binders with a low $k_{a}$ would be less likely to bind the small amount of time the DNA is free, and those with a low $k_{d}$ would release fast and have to wait for another chance that the nucleosome releases their binding site. To test this model, Polach and Widom (1995) inserted restriction enzyme cut sites at various locations within the nucleosomal DNA, moving from the entry/exit site of the nucleosome all the way to the dyad. They then compared the enzymatic activity of the endonucleases on naked DNA versus the same DNA in a nucleosome. Strikingly, the rates of
digestion were 3 to 6 orders of magnitude lower in the case of nucleosomes, but not fully inhibited. This confirmed the idea that nucleosomes limit access to DNA, but that transient site exposure can give other factors access to nucleosomal DNA. In addition, the rates of digestion decreased three orders of magnitude between the entry/exit of the nucleosome and the dyad, supporting the model in which the DNA gradually peels off, making peeling off all the way to the dyad less likely than at the edges.

The nucleosome breathing rate is expected to be dependent on the DNA sequence. Indeed, a sequence that is bound more tightly by the octamer should have a lower probability of being released and becoming accessible (Eslami-Mossallam, Schiessel, & van Noort, 2016). For example, it was observed that sequences that are more flexible bind more tightly to the octamers and require more force to release (T. T. Ngo et al., 2015 and section 1.9). To test this possibility, J. Anderson and Widom (2000) probed the site exposure dynamics of a variant of the 601 sequence, termed 601.2 and the 5S sequence. The 601.2 sequence has an affinity for octamers around 5 times higher (hence binding more tightly) than the 5S sequence. As hypothesized, the 601.2 sequence offered a protection up to 100 times greater than the 5S sequence. Interestingly, this difference was reduced as the cuts happened closer to the dyad. At the edges of the nucleosome, the 601.2 sequence would release the DNA 100 times less than the 5S sequence. On the other hand, at the dyad, the difference between the two sequences is not significant. This could be due to the probability being so low that it cannot be discerned through this set of experiments, or to the 601.2 sequence binding less tightly at the dyad, or the 5S sequence binding more tightly at the dyad.

Although the evidence for dynamic unwrapping of the nucleosome is strong, it still is possible that this transient site exposure actually comes from the nucleosome translocating rather than the DNA transiently releasing. As the nucleosome moves on the DNA, it would expose the cut site and
give a chance to the endonucleases to cut. The low site exposure rate could thus be a low translocation rate instead. To test this hypothesis, J. D. Anderson, Thåström, and Widom (2002) established three tests. First, they looked whether adding extra DNA around a nucleosome core particle makes digestion of DNA more likely. However, they observed no difference between the 174 bp, 225 bp or 275 bp sequences, indicating that extra sequence space, which would increase site exposure by translocation—the longer the sequence the more likely the nucleosome not to be on the cut site—does not lead to increased loss of protection. Second, they showed that nucleosomes to not translocate in their assays. Simply, they separated the nucleosomes on a non-denaturing polyacrylimide gel and only obtained one band in the case of their positioning sequence (601 derivatives), but several bands in the case of 5S, as expected. Indeed, it is wildly believed that nucleosome species at different positions on a DNA molecule migrate at different speeds on a gel, as was observed on the 5S sequence (Meersseman, Pennings, & Bradbury, 1992). They furthered their assay by isolating specific bands of 5S nucleosomes, incubating them in the same conditions as their digestion assay, and then having them migrate again on a native gel. Once again, they observed no movement from the nucleosomes, indicating that their experimental setup does not induce translocation. Overall, these results indicate that the effect of nucleosomes on DNA protection does not come from nucleosome translocation, and argues for transient site exposure.

It remains possible that restriction enzymes cut into the nucleosome with a lower probability, without the need for transient site exposure. To further demonstrate their model, G. Li and Widom (2004) used FRET (a fluorescent-based technique used to measure relative distances, see section 3.2 and Roy, Hohng, and Ha, 2008) to show that the histone octamer and the DNA separate from each other when proteins bind. They generated two pieces of DNA with a donor dye at one end and a LexA binding site on either end of the DNA molecule. They put the acceptor dye at the dyad on
the histone octamer, on either histone H3-V35 or H2A-K119. Hence, in properly formed nucleosomes, the two dyes should be in close proximity and the FRET efficiency should be maximal. However, if the nucleosome transiently releases the section of DNA which contains the dye, the FRET efficiency should decrease for that amount of time. Using nucleosomes reconstituted with these DNA molecules and octamers, they titrated in the DNA-binding protein LexA and observed the energy transfer efficiency, as an ensemble measurement. If the site exposure theory is right, in the presence of LexA, there is a probability that the transiently released DNA containing the LexA binding site gets bound by LexA, and hence that the nucleosome remains trapped in the low FRET efficiency state. On the other hand, if LexA binds nucleosomes in the closed conformation, then the FRET efficiency should not change. They observed that as the concentration of LexA increase, the energy transfer decreases, indicating that less and less nucleosomes are in the “closed” conformation, and that more and more are trapped in an exposed state by LexA, up to a concentration of LexA of 100 μm, where the reaction saturates. This argues in favor of the transient site exposure theory, and against direct nucleosome binding. They then used the second construct, where the dye on the DNA is on the opposite side of the LexA binding site. With this construct, if LexA binds, the FRET signal should not change, as the binding site is on the opposite site of the FRET pair. However, if the nucleosome translocation hypothesis for site exposure is true, then the FRET efficiency should decrease, as the whole nucleosome will have moved away, not just the side with the binding site. At the same concentrations of LexA as they used with the first construct, they never observed a change of FRET with this construct. This experiment further argues against the model of site exposure by nucleosome translocation.

While these experiments agree with the “nucleosome breathing” model, and disagree with most competing models, they do not demonstrate that sites are transiently exposed. Five months after the
FRET paper, G. Li, Levitus, Bustamante, and Widom (2004) published nucleosome breathing rates of $k_{12} \simeq 4 s^{-1}$ and $k_{21} \simeq 20 - 90 s^{-1}$ (refer to fig. 11). To obtain these rates, they reused the same system as presented above, but used a stop-flow FRET experiment, as well as FRET fluorescence correlation spectroscopy (FCS). In the stopped flow FRET experiment, LexA was injected at once and the injection was recorded. They measured the change in FRET caused by LexA as a function of time, which gave them the on rate $k_{12}$, as they showed $k_{23}$ is large and hence $k_{12}$ is the rate-limiting step in LexA binding (refer to fig. 11). Using the $K_{eq}$ they measured in the previous paper (G. Li & Widom, 2004), they obtained $k_{21}$. To confirm their results, they used FRET FCS, which gave them similar values even in the absence of LexA, indicating that this wrapping and unwrapping is independent of the presence of transcription factors. Finally, using the same tools, Tims, Gurunathan, Levitus, and Widom (2011) calculated the rates of wrapping and unwrapping for LexA site positioned deeper into the nucleosome. As expected, these rates greatly decreased when away from the entry/exit of the nucleosome. Overall, this set of experiments, combining FRET and FCS, capped a decade and a half of research on nucleosome breathing, nicely demonstrating the theory.

Polach, Lowary, and Widom (2000) also tested the effect of histone tails on spontaneous site exposure. Indeed, they have been shown to be involved in half of the salt bridges between DNA and histones (section 1.5 and Walker, 1984). To test the effect of the tails, they trypsinized the nucleosomes, which is a commonly used technique to obtain “globular histones” that are devoid of tails but otherwise perfectly normal, albeit for the contribution of the tails (L. Böhm & Crane-Robinson, 1984; Vettese-Dadey, Walter, Chen, Juan, & Workman, 1994). The trypsinized nucleosomes were then compared to canonical nucleosomes for transcription factor binding, restriction enzyme digestion and exonuclease digestion. Interestingly, only a minor difference was observed,
with the tailless nucleosomes providing 1.5-14 times less protection—still 100 times higher than naked DNA. The highest difference of protection was observed on DNA sites more within the nucleosomes. However, because the authors only gathered two positions within the nucleosome, a trend cannot be extrapolated. This is a surprising results, as histone tails are thought to control access to the nucleosomes. It is possible that a more dramatic effect would have been observed in the context of chromatin, as histone tails have also been hypothesized to control the folding of chromatin (see section 1.8), or that this modest change is sufficient to induce a larger effective difference (K. van Holde & Zlatanova, 2006).

Finally, (Poirier, Bussiek, Langowski, & Widom, 2008) demonstrated that spontaneous site exposure behaves similarly in the context of chromatin. Whether there is one or several nucleosomes, the cut rate from restriction enzymes barely varies. On the other hand, the cut rate in the linker DNA decreases around 50 times in chromatin—even in the absence of H1—versus naked DNA, indicating that chromatin protects the linker DNA, potentially via the closed conformation of the fiber.

Overall, this set of experiments by the laboratory of Jonathan Widom can be considered one of the most important pillar of our understanding of nucleosomes. Indeed, it brought out the nature of nucleosomes as a barrier against non-specific DNA binding, rather than just a unit for DNA condensation or a barrier to gene expression. Moreover, it provided the scientific community with a quantitative, non-static view of nucleosomes and chromatin.

**Nucleosome sliding**

While Widom & coworkers did not observe any nucleosome movement in their assays, nucleosome sliding had been previously characterized (Meersseman et al., 1992; Pennings, Meersseman, & Bradbury, 1991). It was observed by 2D native gel electrophoresis, with an incubation in be-
tween the perpendicular runs, that a single band, indicating a single nucleosome conformation, would become multiple. This multiplication of the bands is temperature and magnesium dependent, arguing for a conformational change in the nucleosome that was estimated to be movement of the nucleosome on the DNA substrate, although this was not shown. It has been shown, however, that at least some of the different positions retain all of their histones, arguing against nucleosomes falling apart (Hamiche, Kang, Dennis, Xiao, & Wu, 2001). It is also interesting to note that consequent nucleosome sliding was observed on telomeric sequences, in vitro, although a role for this effect in vivo remains to be determined (Pisano et al., 2007).

Flaus and Richmond studied the movement of nucleosomes at the base-pair resolution using site directed hydroxyl radical double-strand breaks (Flaus, Luger, Tan, & Richmond, 1996; Andrew Flaus & Richmond, 1998). They mutagenized the H4 histone with a cysteine at position 47 (H4-S47C), and then attached an EDTA molecule to it via a thiol bond. This EDTA molecule is hence positioned at the dyad of the nucleosome. Upon addition of Fe$^{2+}$ and H$_2$O$_2$, the iron cation is captured by the EDTA molecule on histone H4, and can proceed to form hydroxyl radicals at this position via the Fenton reaction. Hydroxyl radicals cause DNA double-strand breaks at the vicinity of the dyad, since this is where they are generated, hence allowing for precise determination of the location of the dyad on the DNA (Flaus et al., 1996). Using a DNA sequence with two nucleosome positioning sequences, they studied the effect of time and temperature on nucleosome translocation. They found that nucleosome translocation is dependent on time, sequence and temperature. While nucleosomes were preferentially reconstituted on the first of the two positioning sequences, they were also more likely to translocate away from it, moving up to 60 bp away in 80 minutes, at 37°C. On the other hand, nucleosomes on the second positioning sequence were slower to move, and exhibited a 10 bp periodicity in their positions. Analysis of the underlying sequences showed a 10
bp periodicity of AA/TT and AT/TA in the second nucleosome positioning sequence, but not in the first one, indicating potentially different mechanisms of translocation (Andrew Flaus & Richmond, 1998).

Theoretical analyses of nucleosome sliding were performed to estimate a putative mechanism (Eslami-Mossallam et al., 2016). It was estimated that if nucleosomes simply “roll” on DNA, the energetic cost would be about $75k_BT$, or about 44.5 kcal/mol at room temperature, which makes it heavily unlikely without some enzyme to catalyze the reaction. It was instead rationalized that movement must happen from defect in the DNA wrapping that can spread around the nucleosome. A first solution is that a section of the nucleosomal DNA releases by spontaneous site exposure, and then either rebinds another site on the octamer, or at another sequence position on the same octamer contact. This would cause a bulge of DNA that could diffuse onto the octamer, eventually leading to the displacement of the octamer on the DNA (although it rather is the displacement of the DNA on the octamer). A 10 bp bulge moving around the octamer was estimated to have a $20k_BT$ (11.86 kcal/mol at room temperature) energetic cost. Another DNA defect could be a twist defect (see section 1.4). Instead of a whole loop being formed, the DNA would be stretched (-1 bp, like in the 146 bp nucleosome structure) or compressed (+1 bp). Just like the loop, this defect could diffuse between the contact sites, eventually potentially reaching the other end of the nucleosome, resulting in a net movement of 1 bp. Such a mechanism would cost about $10k_BT$ (5.93 kcal/mol at room temperature) for a single base pair. While twist defects are more common, and expected to lead to faster nucleosome displacement, the exact contribution of each defect will depend on the DNA sequence. On a DNA sequence with strong 10 bp AT or AA periodicity, like the second nucleosome positioning sequence in Andrew Flaus and Richmond (1998), the contribution of the bulge displacement will be stronger, leading to 10 bp steps, while on other sequences, like the
Figure 12: Nucleosomal DNA (octamer not shown) is represented in three possible modes of intrinsic nucleosome dynamics. (A) Breathing, where DNA segments are released from the nucleosome, (B) sliding, where the whole octamers moves 1 base pair or more, and (C) gaping, where the superhelix pitch varies, opening “like a clamshell”. The green and red dots represent the position of the FRET dyes used by T. T. M. Ngo and Ha (2015) to demonstrate the existence of nucleosome gaping. Adapted from T. T. M. Ngo and Ha (2015).

first positioning sequence of the same study, twist defect will be the leading mode of diffusion (Eslami-Mossallam et al., 2016).

**Nucleosome gaping**

A third intrinsic mode of nucleosome dynamics is called “gaping” and was demonstrated by T. T. M. Ngo and Ha (2015). Nucleosome gaping was first proposed theoretically to explain the compaction of the 30nm fiber (Mozziconacci and Victor, 2003 and section 1.8 for more on the 30nm fiber). By using single-molecule FRET, they demonstrated that DNA can move on the axis of the superhelix, meaning that the pitch of the superhelix can vary (fig. 12).

They placed the donor and acceptor fluorophore one superhelix pitch away from each other, and observed three FRET states: high, middle and low, with dwell times of about 100 seconds, 418 seconds (about 7 minutes), and 43 seconds, respectively. Their relative occupancy were 12%, 85%, and 3%, respectively, meaning that nucleosomes spend most of their time in the middle FRET state. They hypothesized four possible mechanisms explaining the various distances in the nucle-
osomal DNA: breathing, sliding, tightening/loosening (which would be the mechanism through which DNA sliding happens, a twist defect), and gaping. They ruled out breathing, because as salt concentration increases, breathing increases also, leading to a lower FRET (higher distance between the FRET pair), but in their experiments, the FRET efficiency increases (the distance decreases) with salt concentration. Next, they positioned the donor dye either up or down on the DNA sequence (see fig. 12, the three green dyes). They reasoned that if sliding happens, the FRET signal should increase either for the lower one or for the upper one, and that they should be anti-correlated. If sliding happens by a twist defect, when the entry/exit DNA compresses, it should get closer to the upper dye and further away from the lower, and if it stretches, it should get further away from the upper dye and closer to the lower dye. However, when increasing the salt concentration (note that it is unsure that the salt concentration would cause sliding, but their goal was to explain the increased FRET efficiency in higher salt concentration), they observed an increase in FRET signal in both cases, indicating that the DNA was not compressing or stretching. Hence, only nucleosome gaping could explain their result.

Overall, intrinsic nucleosome dynamics can be resumed as such: any interaction has an off-rate, and that off-rate can be measured. While nucleosome breathing was demonstrated as a potential way for DNA binding factors to access the otherwise unaccessible nucleosomal DNA, and hence exposed nucleosomes a barrier against non-specific binding (which could have dramatic effects for a multi-cellular organism), rather than just general inhibitors, the role nucleosome sliding and gaping remains to be determined. Indeed, in vivo, nucleosome dynamics have to be much faster than what was recorded, for example, in nucleosome sliding. Nucleosomes inhibit transcription (Bondarenko et al., 2006; Izban & Luse, 1991; Kulaeva, Hsieh, Chang, Luse, & Studitsky, 2013) and many other genomic processes like replication or repair. Their proper positioning is required
In order to ensure proper regulation of the nucleosomes, many extrinsic, for the most part ATP-dependent factors are involved. Nucleosome dynamics can be modulated actively by chromatin remodelers (section 1.11), by the post-translational modification of histones (section 1.12), by the incorporation of histone variants in the nucleosome (section 1.13), or by the regulated assembly and disassembly of nucleosomes (section 1.14).

1.11 Chromatin remodelers

There are four main subfamilies of ATP-dependent chromatin remodelers: (i) Switch/Sucrose Non-Fermentable (SWI/SNF), (ii) Chromodomain Helicase DNA-binding (CHD), (iii) INO80 and (iv) Imitation Switch (ISWI). Each family is involved in one of the four key functions of remodelers: regular spacing of nucleosomes (ISWI, CHD), repositioning of nucleosomes (SWI/SNF), nucleo-
some (dis)assembly (SWI/SNF, ISWI, CHD) or handling of histone variants (INO80) (see fig. 13). Through these functions, remodelers can catalyze or inhibit access to DNA.

How chromatin remodelers achieve their function has been, and still is, an active area of research. Indeed, they are fighting against relatively high energy barriers (see section 1.10). Chromatin remodelers contain two RecA-like lobes, similar to DNA translocases (Clapier, Iwasa, Cairns, & Peterson, 2017; Singleton, Dillingham, & Wigley, 2007). In DNA translocases, each lobe binds to the same strand of the DNA molecule and “walks” on the DNA, the front lobe first making a 1-2 bp step, followed by the second lobe. Each cycle consumes one ATP molecule. While this may work for a DNA translocase in the absence of obstacles, an histone octamer makes 14 contacts with the DNA, each contact estimated to be 1 kcal/mol, for a total of 14 kcal/mol (the bending energy of the DNA is stored in the superhelix). This means that a remodeler would need about 2 ATP molecules simply to break the octamer contacts, and another ATP molecule to move the octamer by 1-2 bp. Current evidence indicate that rather than simply pushing the nucleosome, chromatin remodelers function via a “wave-ratchet-wave” model (Saha, Wittmeyer, & Cairns, 2005). The remodeler attaches to the histone octamer and bind to the nucleosomal DNA around 20 bp after the dyad—about 97 bp away from the DNA exit site and 50 bp away from the DNA entry site of the nucleosome. From there, the remodeler uses its RecA-like domains to pull from the short side of the nucleosome. This pulling action can induces a twist-defect of a few base-pairs inside of the nucleosome, with the entry side being stretched and the exit side being compressed, or bulged. It is unclear whether DNA contacts are broken at this point, but it appears that the remodeler simply lets the nucleosome correct the twist defect by forcing the direction of the bulge towards the exit site, in a manner similar to a ratchet. It seems that the bulge would simply diffusive towards the exit, growing by 1-2 bp every ATP hydrolysis cycle, forming a “wave” (although the nucleosome
could potentially have several bulges in it). Once the wave reaches the exit site, it leads to a nucleosome movement ranging from 1 bp to 50 bp (Deindl et al., 2013; Harada et al., 2016; Zofall, Persinger, Kassabov, & Bartholomew, 2006). This way, the remodeler does not have to pay the energetic cost of breaking the 14 contact sites within the nucleosome, and potentially only has to pay the cost of inducing a twist-defect (estimated to be 5.93 kcal/mol, less than one ATP molecule, see Eslami-Mossallam et al., 2016). Hence, it appears that chromatin remodelers simply catalyze nucleosome sliding.

Other than nucleosome sliding, chromatin remodelers can also catalyze nucleosome or histone ejection (Clapier et al., 2017). Two mechanisms were proposed. The first one suggests that the decision between nucleosome ejection or nucleosome translocation is defined by the speed at which the remodeler functions. Several factors are involved in controlling the ATPase activity of chromatin remodelers, as well as their coupling between the ATPase activity and the translocation activity. In the case of Remodels the structure of chromatin (RSC), a SWI/SNF-family remodeler, it was shown that actin-related protein (ARP)s facilitate nucleosome ejection by increasing the coupling between ATP hydrolysis and DNA translocation. In addition, decreasing the ATPase rate by inserting mutations in its regulatory subunits lead to slow nucleosome translocation instead of ejection (Clapier et al., 2016). The second mechanism stipulates that the nucleosome adjacent to the one being remodeled is the one being evicted. Using a dinucleosome template, Dechassa et al. (2010) tethered SWI/SNF to one nucleosome and observed that as the first nucleosome is translocated, one H2A-H2B dimer is rapidly lost on the second nucleosome, and then, at a lower rate, the whole nucleosome is evicted. They propose that as SWI/SNF reel linker DNA in the first nucleosome, the second nucleosome is pushed into the remodeler and ultimately kicked out of the template, potentially because it reached too high of a barrier to stay on it (in this case, the DNA end,
but in vivo, it could be other nucleosomes, or transcription factors). Since these two mechanisms are not mutually exclusive, it is possible that they both contribute to nucleosome eviction. Properly regulated to stop at the ejection of the H2A-H2B dimer, these same mechanisms could be involved in the insertion of histone variants (see section 1.13) by the INO80 remodeler family (Clapier et al., 2017).

Chromatin remodelers occupy a vast and complex functional and regulatory landscape that extends beyond the reach of this presentation of nucleosome dynamics. An extensive presentation of their mechanisms and functions was recently published by Clapier et al. (2017).

### 1.12 Histone post-translational modifications

The effect of histone post-translational modifications on transcription was first described by Allfrey, Faulkner, and Mirsky (1964), ten years before histones were known to form nucleosomes (Kornberg, 1974; Oudet et al., 1975). It was observed that acetylation of histones would reduce histones’ inhibition of transcription. Today, more than 170 histone modifications have been identified on the 4 core histones, with many more on H1 and the histone variants. Of these 170 post-translational modification (PTM)s, 68 have been characterized, rather than just identified by mass spectrometry (see table 1 and Zhao and Garcia, 2015).

<table>
<thead>
<tr>
<th>Histone</th>
<th>Characterized</th>
<th>Identified only</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2A</td>
<td>16</td>
<td>27</td>
<td>33</td>
</tr>
<tr>
<td>H2B</td>
<td>9</td>
<td>33</td>
<td>42</td>
</tr>
<tr>
<td>H3</td>
<td>30</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>H4</td>
<td>13</td>
<td>22</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>68</td>
<td>112</td>
<td>170</td>
</tr>
</tbody>
</table>

Table 1: Number of known histone post-translational modifications for each core histone. Data gathered from Zhao and Garcia (2015).
The more commonly studied histone modifications are acetylation, methylation, and phosphorylation. Acetylation occurs on lysines, thereby reducing its charge, for a final net charge of 0. Acetyl groups are deposited by histone acetyl transferase (HAT)s and removed by histone deacetylase (HDAC)s (X.-J. Yang & Seto, 2007). Phosphorylation adds a negative charge on the non-charged residues serine, threonine, or tyrosine. The reaction is catalyzed by kinases and reversed by phosphatases (Rossetto, Avvakumov, & Côté, 2012). Methylation is found on both lysines and arginines but, unlike acetylation and phosphorylation, it does not reduce the charge of the group, leaving a positive +1 net final charge. Methylation marks are deposited and removed by site and residue specific methyltransferases and demethylases. Deimination—also known as citrullination—has been shown to be another way to demethylate residues (Y. Wang et al., 2004). Several methylation groups (up to 3) can be deposited on a single amino acid (Andrew J Bannister & Kouzarides, 2011; Greer & Shi, 2012). Many other histone post-translational modifications were found, among which lysine crotonylation, which is associated with promoters (M. Tan et al., 2011), deimination/citrullination (Y. Wang et al., 2004), β-N-acetylglucosamination (Sakabe, Wang, & Hart, 2010), ADP-ribosylation (Hassa, Haenni, Elser, & Hottiger, 2006), ubiquitylation, sumoylation, tail clipping, proline isomerization, biotinylation, butyrylation, formylation, glutathionylation, 2-hydroxyisobutyrylation, malonylation, hydroxylation, and finally oxidation (Andrew J Bannister & Kouzarides, 2011; Zhao & Garcia, 2015). It was shown that acetylation of histones facilitates access to nucleosomal DNA (D. Y. Lee, Hayes, Pruss, & Wolffe, 1993), and that this effect was even more pronounced on nucleosome arrays (dodecamers), where high levels of acetylation of about 12 acetyl groups per nucleosome leads to decompaction of chromatin and 5-15 times more transcription (Tse, Sera, Wolffe, & Hansen, 1998). Histone modifications are deposited on residues located both on the histone tails and on the histone cores (see structural explanations in
Specific research to dissect the effect of individual modifications focused heavily on the post-translational modification of histone tails (fig. 15), but the histone-fold domains PTMs are catching up (fig. 14).

**Post-translational modifications of the histone-fold domains**

Figure 14: Post-translational modifications on the globular core of the histone octamer. (a) Surface representation of the vertebrate nucleosome core particle (without flexible tails) viewed down the DNA superhelix axis. The functional groups of modified residues are colored according to the type of post-translational modification, with acetylation in green, phosphorylation in red, methylation in blue, sites that can be either acetylated or ubiquitylated in purple, and sites that can be either acetylated or methylated in light blue. The DNA superhelix is transparent light blue, with superhelix locations (0–7) indicated. Ambiguous sites of modification are indicated by curly brackets and dashed lines. (b) View as in a but rotated 90° around the molecular dyad axis (Φ). (c) View as in a but rotated 90° around the horizontal axis looking down at the top of the molecule in a. Representations of nucleosomes were generated using PyMOL (http://www.pymol.org). Adapted from Cosgrove, Boeke, and Wolberger (2004)

Cosgrove, Boeke, and Wolberger (2004) plotted the position of known histone-fold modifications on the surface of the nucleosome and found that out of 27 modifications, twelve are positioned on the nucleosome core and do not interact with DNA, and fifteen are located at the histone-DNA interface (see fig. 14). While most of these modifications were detected by mass spectrometry (Freitas, Sklenar, & Parthun, 2004; L. Zhang, Eugeni, Parthun, & Freitas, 2003) and remain
to be characterized, ongoing work indicates that histone-fold post-translational modifications can directly affect the dynamics of nucleosomes (Lawrence, Daujat, & Schneider, 2016; Tessarz & Kouzarides, 2014).

Iwasaki et al. (2011) crystallized 11 lysine to glutamine mutant nucleosome core particles: H3 K56Q, H3 K64Q, H3 K79Q, H3 K115Q, H3 K122Q, H4 K31Q, H4 K44Q, H4 K59Q, H4 K77Q, H4 K79Q, and H4 K91Q. These mutants mimic the acetylated state of the lysine, but caution has to be observed as mimetic K→G mutations do not always recapitulate the full effect of acetylation (Manohar et al., 2009). They found that while these modifications did not significantly affect the nucleosome structure, they disrupted key interactions, most of them leading to the loss of the residue’s interaction with DNA. Simon et al. (2011) studied the dynamics of some of these acetylated residues using magnetic tweezers (see section 3.2). They concluded that the acetylation of histone-fold domains near the nucleosome dyad (H3 K115ac and H3 K122ac) facilitated the disassembly of nucleosomes, but did not alter DNA breathing and transcription factor binding, and vice versa for the acetylation of residues located near the DNA entry/exit sites (H3 K56ac, H4 K77ac and H4 K79ac). These results indicate that acetylation of the histone-fold domains directly affect nucleosome dynamics.

More in-depth studies were conducted for key residues. The acetylation of H3 K56 was shown to increase DNA breathing seven times (Neumann et al., 2009). In the structure, acetylation of H3 K56 was shown to disrupt the DNA contacts of this residue (Iwasaki et al., 2011). The methylation of H3 R42 was shown to stimulate transcription both in vivo and in vitro by destabilizing the histone-DNA interactions (Casadio et al., 2013; Hainer & Martens, 2011; Hyland et al., 2011). H3 K64ac was shown to regulate nucleosome stability, facilitate nucleosome eviction, and inhibit H3 K64me3 (which is associated with heterochromatin, see Daujat et al., 2009), making it
a transcription-facilitating modification (Di Cerbo et al., 2014). Structurally, the H3 K64Q mutant was shown to disrupt the interactions between the α-helices 1 and 2 of H3, but not DNA interactions (Iwasaki et al., 2011). The phosphorylation of H3 T118 was shown to lower the affinity of the histone octamer for DNA, leading to higher breathing and higher mobility (North et al., 2011). It was suggested it may induce non-canonical nucleosome structures (North et al., 2014).

Interestingly, mutations that mimic (T118E) or inhibit (T118A) this phosphorylation are lethal in S. cerevisiae, and hetero-allelic mutants show defects in transcription and repair, indicating that its apposition is both necessary and tightly regulated (Hyland et al., 2005). The same was observed in Drosophila melanogaster (Graves, Wang, Lagarde, Chen, & Tyler, 2016). Acetylation of H3 K122 and and K115, near the dyad, was shown not to affect DNA breathing, but affect mobility (Manohar et al., 2009). Structurally, these acetylation marks disrupt the interactions of the residues with DNA (Iwasaki et al., 2011). It was also shown that acetylation of H3 K122 is necessary for transcription (Tropberger et al., 2013). In the crystal structure of H4 K91Q, the contact that H4 K91 normally forms with H2B D65 is lost (Iwasaki et al., 2011). While this does not lead to any significant change in the crystal structure, it was shown that H4 K91ac disrupts chromatin assembly (Ye et al., 2005), indicating a weaker tetramer-dimer interface. Finally, core post-translational modifications are not limited to the core histones. For instance, citrullination of H1R54 was shown to disrupt its interaction with DNA (Christophorou et al., 2014).

Not all histone-fold post-translational modifications lead to modified nucleosome dynamics. For example, the tri-methylation of H3 K79 is thought to be more informational than structural. In the crystal structure, it was shown that the lysine 79, when tri-methylated, becomes almost fully solvent-accessible, while not causing any significant structural changes in the nucleosome. It was thus proposed that this modification acts as a “docking site” on the nucleosome for specific
proteins to recognize and bind (Xu Lu et al., 2008). Of note, methylation of H3 K79 is necessary for silencing in *S. cerevisiae* (Ng et al., 2002). It was also shown that the methylation of H2A Q105 prevents its binding by Facilitates access to chromatin (FACT) (Tessarz et al., 2013) and that the acetylation of H3 K56, on top of its effect on DNA breathing, is necessary for binding to chaperones Caf1 and Rtt106, and is found on all newly synthesized histones (C.-C. Chen et al., 2008; Q. Li et al., 2008; Ozdemir, Masumoto, Fitzhjohn, Verreault, & Logie, 2006).

Overall, these findings indicate that histone-fold PTMs can have a direct effect on nucleosome dynamics, but do not have to. Post-translational modification of histone-fold residues can induce breathing, sliding, facilitate disassembly, or prevent or facilitate the binding of other factors, such as silencing factors and chaperones.

**Post-translational modifications of the histone tails**

Histone tails are heavily positively charged and unstructured (Luger et al., 1997a). They are involved in the formation of higher-order chromatin structures (see section 1.8) and in repression of chromatin access (so do the histone-fold domains). Histone tails participate in blocking transcription factor binding (Vettese-Dadey et al., 1994), and removal of histone tails makes transcription factor binding more than 100 times more likely, while increasing breathing by a factor of 5 only (Z. Yang, Zheng, Thiriet, & Hayes, 2005), as was observed by Polach et al. (2000). Removing the histone tails also increases nucleosome sliding (Ferreira, Somers, Webster, Flaus, & Owen-Hughes, 2007; Hamiche et al., 2001). A large body of work by the laboratory of Dr. Jeffrey Hayes helped dissect the role of the individual histone tails (Zheng & Hayes, 2003). The N-terminal tail of H2A binds DNA, around 40 bp away from the dyad (K.-M. Lee & Hayes, 1997). The H2A C-terminal tail was shown to bind near the DNA entry/exit sites in the presence of linker DNA (K.-M. Lee
Since histone tails are involved in repression of chromatin access and higher-order structures, it appears logical that the post-translational modification of their residues modulate these two functions. Trimethylation of H4 K20 was shown to enhance the compaction of higher-order chromatin structures (Xu Lu et al., 2008). On the other hand, the acetylation of H4 K16 completely obliterates higher-order chromatin structure. It does so both by preventing the interaction of the basic tail of H4 with the acidic patch of H2A-H2B dimers of a neighboring nucleosome (R. Zhang, Erler, & Langowski, 2017) and by preventing the action of the remodelers ACF (Shogren-Knaak et al., 2006) and ISWI (Corona, Clapier, Becker, & Tamkun, 2002), which are involved in spac-
ing nucleosomes properly to enhance higher-order structures. Since it prevents the compaction of chromatin, the acetylation of H4 K16 leads to increased transcription (Akhtar & Becker, 2000). In vivo, during the cell cycle phase G2, it was demonstrated that the phosphorylation of H3 S10 leads to the recruitment of deacetylases that remove the acetyl marks on H4 K16, leading to chromatin condensation (Wilkins et al., 2014). However, and contrary to histone-fold PTMs, it appears that most of the histone-tail modifications are involved in factor recruitment, rather than in the direct modulation of nucleosome dynamics (Bowman & Poirier, 2015; Lawrence et al., 2016).

Several classes of proteins recognize histone post-translational modifications. Bromodomain-containing proteins recognize acetylation marks (Mujtaba, Zeng, & Zhou, 2007). Bromodomains are involved in chromatin remodeling, such as in SWI/SNF (Hassan et al., 2002), histone modification, and regulation of transcription (Fujisawa & Filippakopoulos, 2017). Interestingly, some histone acetyltransferases such as Spt-Ada-Gcn5 acetyltransferase (SAGA) also contain a bromodomain, leading to a positive feedback loop of histone acetylation (Hassan et al., 2002). Of note, acetylation of the histone tails increases their propensity to form α-helices (Xiaoying Wang, Moore, Laszczak, & Ausió, 2000), potentially helping form structured recognition sites for further protein binding. Chromodomains, Tudor domains and MBT domains recognize methylation marks (J. Kim et al., 2006). CHD1, a chromatin remodeler, is recruited to active chromatin (marked by the tri-methylation of H3 K4, see Santos-Rosa et al., 2002) via its chromodomains (Flanagan et al., 2005; Pray-Grant, Daniel, Schieltz, Yates III, & Grant, 2005; Sims et al., 2005). The tri-methylation of H3 K4 also prevents binding of Nucleosome remodeling and deacetylase (NuRD), a transcription repressor (Zegerman, Canas, Pappin, & Kouzarides, 2002), and of Suv39h1, which methylate H3 K9 (Nishioka et al., 2002). H3 K9 methylation leads to the binding of HP1 and the formation of constitutive heterochromatin, a transcription inhibiting nucleoprotein complex (Andrew J. Ban-
nister et al., 2001; Lachner, O’Carroll, Rea, Mechtler, & Jenuwein, 2001). Hence, the interplay between the recognition and avoidance of the methylation marks at H3 K4 and H3 K9, separated by only 4 amino acids, is sufficient to define transcriptionally active euchromatin and repressive heterochromatin. The MBT domain of L3MBTL1 binds at mono- and di-methylated H4 K20 and H1 K26 to further compact chromatin (Trojer et al., 2007). Interestingly, HP1 is also recruited by the methylated H1 K26, associating this mark with heterochromatin, but only in the absence of a phosphorylation mark on H1 S27 (Daujat, Zeissler, Waldmann, Happel, & Schneider, 2005). Of note, chromodomains have also been shown to associate with DNA and RNA, adding a layer of complexity to histone PTM recognition and interpretation (Akhtar, Zink, & Becker, 2000; Bouazoune et al., 2002). PHD fingers recognize either acetylation (Lange et al., 2008; Zeng et al., 2010) or methylation (Champagne & Kutateladze, 2009) The Plant homeodomain (PHD) finger of Nucleosome remodeling factor (NURF), a chromatin remodeler, recognizes the tri-methylation of H3 K4. Finally, 14-3-3 domains recognize phosphorylation (Dougherty & Morrison, 2004; Macdonald et al., 2005).

Overall, histone tails post-translational modifications are mostly involved in the recruitment of co-factors, while histone-fold post-translational modifications are mostly involved in the modulation of nucleosome dynamics. Because the factors recruited to nucleosomes are involved in the modulation of nucleosome dynamics, which boils down to the modulation of DNA accessibility, it can be said that all histone post-translational modifications result in the modulation of nucleosome dynamics and DNA accessibility.
1.13 Histone variants

Histone post-translational modifications alter some amino acids of the nucleosome. Heavier relifting of the nucleosome can be achieved by the replacement of core histones with histone variants. Histone variants conserve the histone fold structure and integrate into the octamer, but their tails, and sometimes their core, differ from the canonical histones by several amino acids, up to whole domains. Thus, histone variants provide different binding pockets and nucleosome dynamics compared to the canonical histones. There is currently 25 known histone variants (H2A: 8, H2B: 5, H3: 5, H4: 0, H1: 7) (Draizen et al., 2016). Histone variants mostly differ from the canonical histones by their expression patterns (cell cycle phase, cell type, etc.) and by their tails. Histone variants also get their own set of post-translational modifications, on top of the one shared with the canonical histone. (Henikoff & Smith, 2015). As discussed below, key histone variants include CENP-A, H3.3, H2A.Z, H2A.X, macroH2A (mH2A), and H2A.B (H2A.Bbd). More information on all known histone variants can be found on the HistoneDB (Draizen et al., 2016).

CENP-A

CENP-A is a histone variant that replaces H3 exclusively at centromeres, in all eukaryotes. It is necessary for the assembly of centromeres, kinetochores, and the proper segregation of chromosomes (Amor, Kalitsis, Sumer, & Andy Choo, 2004b). CENP-A was first isolated using the antibodies of patients with an autoimmune disease (Earnshaw & Rothfield, 1985). It can replace H3 in nucleosomes reconstituted in vitro (Yoda et al., 2000) and, in vivo, its assembly into nucleosomes is mediated by the chaperone HJURP (Scm3 in S. cerevisiae) (Dunleavy et al., 2009; Foltz et al., 2009; Stoler et al., 2007). Deposition of CENP-A into centromeric nucleosomes happen during
the G1 phase in human cells and G2 phase in *Schizosaccharomyces pombe* (Lagana et al., 2010; Lando et al., 2012). CENP-A nucleosomes are easier to disassemble than canonical nucleosomes (Conde e Silva et al., 2007). Several non-canonical structures, ranging from right-handed octamer to hemisome, have been proposed for CENP-A nucleosomes, and whether one of them is the right one or all of them are possible remains to be determined (Black & Cleveland, 2011; Henikoff & Furuyama, 2012). Interestingly, CENP-A lacks the lysines 9 and 27 present on H3, which are necessary for the formation of heterochromatin (Amor et al., 2004b). Hence, CENP-A nucleosomes probably block the expansion of heterochromatin domains found around the centromeres (Allshire & Ekwall, 2015).

Unlike the four core histones, CENP-A is rapidly evolving (Henikoff, Ahmad, & Malik, 2001; Malik & Henikoff, 2001). It was proposed that this rapid evolution is meant to match that of the α-satellite DNA repeats that centromeres are usually formed on (Malik & Henikoff, 2001). Indeed, it appears that chromosomes that bind microtubules more tightly are more likely to segregate into the oocyte during gametogenesis. Hence, a tighter bond between the underlying sequence and the histone octamer at the centromere would lead to a larger kinetochore, more microtubule attachment, and a likelier segregation to the oocyte (Henikoff et al., 2001).

CENP-A marks the position of centromeres epigenetically. Evidence that CENP-A is sufficient for the definition of centromeres first came from the formation of neocentromeres (Amor & Choo, 2002; A. W. I. Lo et al., 2001; A. W. Lo et al., 2001). Transgenerational inheritance of mislocalized centromeres that are mitotically stable was observed (Amor et al., 2004a). Accordingly, CENP-A is one of the rare histones conserved in sperm chromosomes (where most histones are replaced by protamines) (Palmer, O’Day, & Margolis, 1990), although this was not observed in *C. elegans* (Gassmann et al., 2012). Note that *C. elegans* has holocentric centromeres (positioned linearly on
the sides of the chromosome, versus a more usual punctum in the middle of the “X”), potentially causing this discrepancy (Maddox, Oegema, Desai, & Cheeseman, 2004). To demonstrate its autonomy in defining centromeres, a CENP-A-LacI chimera was targeted to an array of lac operators, in D. melanogaster. Strikingly, functional centromeres assembled at this location, leading to the formation of functional kinetochores and to chromosome segregation. The centromeres remained at the Lac operators location through several generations, even in the absence of the CENP-A-LacI construct, demonstrating its epigenetic inheritance (Mendiburo, Padeken, Fülöp, Schepers, & Heun, 2011). Since the S. cerevisiae CENP-A protein Cse4p can rescue a CENP-A knockdown in human cells, it is probable that this epigenetic mechanism is conserved in all eukaryotes (Wieland, Orthaus, Ohndorf, Diekmann, & Hemmerich, 2004). Of note, S. cerevisiae is the only known eukaryote with a sequence-defined centromere of only one nucleosome, see (Hegemann & Fleig, 1993). However, previous attempts trying to define inheritable ectopic centromeres using HeLa and CHO cells managed to assemble the centromeres, but did not succeed at establishing complete kinetochores that would have led to inheritance (Van Hooser et al., 2001).

H3.3

Unlike histone H3, H3.3 is integrated in chromatin in a replication-independent manner (Tagami, Ray-Gallet, Almouzni, & Nakatani, 2004), with the chaperones DAXX are HIRA responsible for its deposition in chromatin (Drané, Ouararhni, Depaux, Shuaib, & Hamiche, 2010; Elsässer et al., 2012; Ray-Gallet et al., 2002; Ricketts et al., 2015). Histone octamers reconstituted with H3.3 are more sensitive to salt-dependent disassembly than those assembled with H3, indicating a reduced stability (Jin & Felsenfeld, 2007). H3.3 only differs by H3 by 4 amino acids: serine 31, alanine 87, isoleucine 89, and glycine 90, which are an alanine, a serine, a valine, and a methionine, re-
spectively, in H3 (Franklin & Zweidler, 1977). Residues 87, 89 and 90 are located near the H3-H3 tetramerization interface (Luger et al., 1997a), possibly explaining the difference of stability. Indeed, it was found, in vivo, that H3.3-H4 tetramers split more readily into dimers (Xu et al., 2010). The presence of a serine at residue 31 was shown to inhibit the methylation of H3K27 by ATXR5 on H3.3, limiting H3.3’s capacity for the formation of heterochromatin (Jacob et al., 2014). Interestingly, *S. cerevisiae* only has the H3.3 histone gene, which it uses both as canonical histone H3 and replication-independent histone H3.3 (Szenker, Ray-Gallet, & Almouzni, 2011). *S. cerevisiae* does not form H3K27me-dependent heterochromatin, called facultative heterochromatin, which is usually associated with the repression of tissue-specific genes—*S. cerevisiae* does not have tissues (Saksouk, Simboeck, & Déjardin, 2015). This could indicate that *S. cerevisiae* never had a use for two different histone H3.

Interestingly, H3.3 is necessary for proper transcription and development. In *D. melanogaster*, depletion of H3.3 leads to decreased viability, male infertility and transcriptional defects that are compensated by overexpression of the canonical H3 (Sakai, Schwartz, Goldstein, & Ahmad, 2009). Mutation of H3.3 K4→A leads to normal transcription, but does not rescue gametogenesis, indicating that the methylation of lysine 4 plays a significant role in this process (Hödl & Basler, 2009). In mouse, H3.3 knock-down leads to oocyte death, loss of viability and male sterility (Couldrey, Carlton, Nolan, Colledge, & Evans, 1999; Tang et al., 2015). H3.3 is also thought to play a key role in ageing (Bano, Piazzesi, Salomoni, & Nicotera, 2017). In mouse somatic tissues, H3.3 reaches near saturation levels (Tvardovskiy, Schwämmle, Kempf, Rogowska-Wrzesinska, & Jensen, 2017), and in *C. elegans*, H3.3 positively correlates with lifespan (Piazzesi et al., 2016). H3.3 has also been involved in a wide array of diseases and cancers (Kallappagoudar, Yadav, Lowe, & Partridge, 2015).
Other H3 variants

Other H3 variants are TS H3.4 and H3.5, both testes-specific variants (Schenk, Jenke, Zilbauer, Wirth, & Postberg, 2011; Witt, Albig, & Doenecke, 1996), and H3.X found in testes, brain, and tumors (Wiedemann et al., 2010). It is interesting that all H3 variants have a direct effect on spermatogenesis, potentially indicating a role related to the de-chromatinization of sperm DNA.

H2A.Z

The function and effect on chromatin of the H2A variant H2A.Z are somewhat controversial. Considering its existence in all eukaryotes and cell types, this could indicate a more subtle function than, say, that implied by testis-specificity. H2A.Z is deposited by the chromatin remodeler SWR-1 in a step-wise fashion, leading to transiently asymmetric nucleosomes (Luk et al., 2010). The reverse reaction is mediated by INO80, another chromatin remodeler, which removes exclusively unacetylated H2A.Z (Papamichos-Chronakis, Watanabe, Rando, & Peterson, 2011). Interestingly, the substrate specificity of SWR-1 is disrupted on nucleosomes containing H3 K56ac, leading to the incorporation of either H2A or H2A.Z by the remodeler (Watanabe, Radman-Livaja, Rando, & Peterson, 2013). This set of experiments indicates close proximity between chromatin remodeling and H2A.Z. H2A.Z was shown to facilitate the passage of Pol II through nucleosomes (Santisteban, Hang, & Smith, 2011; Weber, Ramachandran, & Henikoff, 2014), although in vitro reports indicate otherwise (Thakar, Gupta, McAllister, & Zlatanova, 2010). Likewise, different reports indicate that H2A.Z nucleosomes are less stable than canonical nucleosomes (Jin & Felsenfeld, 2007; Suto, Clarkson, Tremethick, & Luger, 2000), but others differ (Ishibashi et al., 2009). Interestingly, and perhaps conclusively, H2A.Z nucleosomes were shown to be more sensitive to temperature.
Studies in *Arabidopsis thaliana* showed that the expression of genes related to high temperature was regulated by H2A.Z. At low temperatures, H2A.Z nucleosomes block transcription, whereas at high temperature, H2A.Z is evicted, although it is unclear whether this is catalyzed or intrinsic disassembly, leading to expression of the high-temperature genes. The same effect of temperature on H2A.Z was observed in budding yeast, indicating a potentially conserved function (Kumar & Wigge, 2010). The idea that H2A.Z’s easier dissociation is associated with gene transcription was also enunciated by H. Zhang, Roberts, and Cairns (2005). They proposed that H2A.Z is used in genes poised for activation because it does not affect repression and eases transcription, while not catalyzing it. Indeed, although it seems to lower the nucleosome barrier to transcription, H2A.Z was shown to inhibit the function of histone modifiers associated with transcription, such as remodelers, methylases and acetyltransferases (B. Li et al., 2005).

In contrast with its transcription-poising activity, H2A.Z was also found to promote the interaction of chromatin with HP1, a heterochromatin-associated protein (J. Y. Fan, Rangasamy, Luger, & Tremethick, 2004), and to mediate the interaction between DNA and the the inner nuclear membrane by interacting with SUN proteins (Gardner et al., 2011). It was also shown to interact with the RNAi-based heterochromatin machinery (Zofall et al., 2009) and to be enriched at Polycomb target genes in embryonic stem (Creyghton et al., 2008). Perhaps helping sorting out these contradictory functions, H2A.Z was found preferentially at the promoter of repressed genes, but H2A.Z acetylated at K14 was enriched at the promoter of active genes. (Millar, Xu, Zhang, & Grunstein, 2006) H2A.Z was also proposed to protect genes from DNA methylation (Zilberman, Coleman-Derr, Ballinger, & Henikoff, 2008).

Deletion of H2A.Z is lethal in *D. melanogaster* and *C. elegans* (Updike & Mango, 2006; van Daal & Elgin, 1992). H2A.Z is necessary for proper mammalian development (Faast et al., 2001).
The C-terminal tail of H2A.Z is involved in its stable association with the nucleosome (Wratting, Thistlethwaite, Harris, Zeef, & Millar, 2012) and is required for development in D. melanogaster (Clarkson, Wells, Gibson, Saint, & Tremethick, 1999). In S. pombe, the deletion H2A.Z or its mutation to make it unacetylable leads to genome instability (H.-S. Kim et al., 2009).

**H2A.X**

H2A.X is primarily associated with DNA double strand breaks, where it is rapidly phosphorylated by ATM and/or DNA-PK on serine 139 and renamed γH2A.X (Burma, Chen, Murphy, Kurimasa, & Chen, 2001; Rogakou, Pilch, Orr, Ivanova, & Bonner, 1998; Stiff et al., 2004; Ward & Chen, 2001). In mammals, γH2A.X forms mega-base pair size foci in a manner of minutes after the DNA breaks (Rogakou, Boon, Redon, & Bonner, 1999). It is proposed that after phosphorylation by ATM/DNA-PK, γH2A.X recruits MDC1, which recruits MRN, itself further activating ATM, leading to a positive feedback loop of H2A.X phosphorylation (Lukas et al., 2004; Stucki et al., 2005; Uziel et al., 2003). The exchange of H2A-H2B dimers with H2A.X-H2B dimers appears to be mediated by the chaperone FACT, and phosphorylation of H2A.X seems to facilitate that exchange (Heo et al., 2008). In S. cerevisiae, the canonical histone H2A possesses the key phosphorylatable residue (serine 129 in this case) of H2A.X, taking its role in DNA repair (Jessica A. Downs, Lowndes, & Jackson, 2000). Because of this and many other inverse relationships, it was proposed that the canonical histone H2A actually evolved from H2A.X, and not the opposite (Talbert & Henikoff, 2010).

H2A.X null mice display growth defect, immune deficiency, radiation sensitivity, and male infertility, as well as DNA repair defect and chromosome instability (Celeste et al., 2002). However, the phosphorylation of H2A.X is not necessary for its function, and it is proposed that phospho-
rylation simply helps with concentrating repair proteins at the double-strand break foci (Celeste et al., 2003; Fink, Imholz, & Thoma, 2007). In addition, while chromatin foci show ATP-dependent local expansion after double-strand break, this expansion is independent of H2A.X (Kruhlak et al., 2006). The phosphorylation of H2A.X was shown to destabilize the nucleosome and inhibit H1 binding (A. Li et al., 2010).

Apart from its role in DNA repair, H2A.X is involved in X chromosome inactivation (Chadwick & Lane, 2005; Fernandez-Capetillo et al., 2003), chromatin regulation during mitosis (McManus & Hendzel, 2005), neural stem cell development (Andäng et al., 2008), cellular senescence maintenance (Rodier et al., 2009), embryonic stem cell development (Banáth et al., 2009), and asymmetric sister chromosome segregation (Charville & Rando, 2013). The multiple functions of H2A.X are further discussed by Turinetto and Giachino (2015).

**macroH2A**

macroH2A (mH2A) is the largest known histone variant—almost three times the size of H2A—due to its large *macro* domain on the C-terminal tail. It is found exclusively in vertebrates and has two variants: mH2A1 and mH2A2 (Gamble & Kraus, 2010). mH2A has been involved with gene silencing. It is concentrated on inactivated X chromosomes (Costanzi & Pehrson, 1998), but only if it is not phosphorylated on serine 137 (Bernstein et al., 2008). It appears that mH2A is displaced from the centrosome, potentially via microtubule transport, to the X chromosome being inactivated (Rasmussen, Mastrangelo, Eden, Pehrson, & Jaenisch, 2000). mH2A also colocalizes with Polycomb, a heterochromatin protein (Buschbeck et al., 2009), interferes with transcription factor binding (Angelov et al., 2003), and inhibits transcription initiation by preventing histone acetylation (Doyen et al., 2006a). The chromatin remodeler SWI/SNF seems to prefer the canoni-
cal H2A over mH2A, while repression-associated remodeler ACF shows no preference, suggesting
the preferential selection of mH2A chromatin for repressive remodeling (Chang et al., 2008). In
relation with its inhibitory functions, mH2A stabilizes the histone octamer. It appears to prefer-
entially form hybrid H2A-mH2A nucleosomes over mH2A-mH2A ones (Chakravarthy & Luger,
2006). Interestingly, mH2A also inhibits H1 binding (Abbott, Chadwick, Thambirajah, & Ausió,
2005) and can protect heterochromatinized genes from silencing (Gamble, Frizzell, Yang, Krish-
nakumar, & Kraus, 2010). mH2A has also been involved in cancer as a tumor suppressor (Kapoor
et al., 2010). Its deletion in mice leads to lipid accumulation in the liver of homozygote females,
but not to the loss of X inactivation or infertility (Boulard et al., 2010).

The macro domain of mH2A was crystallized. It was shown to associate with histone deacetyl-
lases and hence decrease nucleosome acetylation, in accordance with its gene silencing function
(Chakravarthy et al., 2005). In addition, it was shown that macro domains bind ADP-ribose with
high affinity (Karras et al., 2005). As a consequence, mH2A is recruited to sites of Poly(ADP-
ribose) polymerase 1 (PARP1) activation, such as in DNA repair, where it plays a role in the
establishment γH2A.X patterns and inhibit Ku70-Ku80 recruitment (Timinszky et al., 2009). In-
terestingly, it seems to inhibit the enzymatic activity of PARP1, potentially leading to a negative
feedback loop (Nusinow et al., 2007).

H2A.B

H2A.B (also known as H2A.Bbd) is thought to play a role opposite to mH2A and facilitate gene
transcription (Y. Chen, Chen, McEachin, Cavalcoli, & Yu, 2014). Indeed, the inactive X chro-
mosome is fully deficient of H2A.B, while the active one is not (Chadwick & Willard, 2001).
H2A.B-containing histone octamers protect between 118 bp (Bao et al., 2004) and 130 bp (Doyen
et al., 2006b) of DNA. H2A.B was shown to destabilize the nucleosome (Eirín-López, Ishibashi, & Ausió, 2007). H2A.B also lacks an acidic path on its surface that is bound by H4 and necessary for the formation of higher-order chromatin structures (Zhou et al., 2007). H2A.B also exchanges more rapidly than H2A in chromatin (Gautier et al., 2004).

Overall, histone variants provide differential modulation of the several functions of nucleosomes. They can increase or decrease nucleosome dynamic, recruit key sets of factors or inhibit the recruitment of others.

### 1.14 Nucleosome (dis)assembly

The synthesis, assembly and disassembly of histones is highly regulated. Indeed, in the absence of regulation, histones tend to aggregate and cause numerous cell defects (Groth et al., 2005; U. J. Kim, Han, Kayne, & Grunstein, 1988; Prado & Aguilera, 2005). The regulation of histone synthesis has been extensively reviewed by Gunjan, Paik, and Verreault (2005). In vivo and in vitro, nucleosome assembly proceeds through the same steps: first, the H3-H4 tetramer is deposited on DNA, second, the H2A-H2B dimers are added (fig. 17). In both cases, the assembly needs to be regulated to prevent the formation of non-nucleosomal histone-DNA interactions (fig. 16). Indeed, histones are both heavily positively charged and hydrophobic. In high monovalent salt concentration (2M), histones form octamers in the absence of DNA, but otherwise, they fall apart into H2A-H2B dimers and H3-H4 tetramers (Jorcano & Ruiz-Carrillo, 1979). Without proper guidance, tetramers and dimers independently bind to DNA via their positive charges, and potentially, when high enough concentration is reached on DNA, to themselves via their hydrophobic cores, leading to large aggregates. The correct assembly of nucleosomes leads to intermediates (fig. 17).
that may, or not, be relevant in vivo (Zlatanova, Bishop, Victor, Jackson, & van Holde, 2009). For example, H3-H4 tetramers reconstitute in the absence of dimers and form pseudo nucleosomes (Dong & van Holde, 1991), while FACT may mediate the formation of hexasomes during transcription (Belotserkovskaya et al., 2003).

In vivo, the transport and deposition of histones is mediated by chaperones, a class of negatively-charged proteins that prevent the inappropriate interactions of histones. More than 27 chaperones are known across eukaryotes, most of which are conserved from yeast to mammals (Gurard-Levin, Quivy, & Almouzni, 2014). Nucleosome assembly happens through two main pathways that are either DNA synthesis-dependent or-independent. Because the synthesis of the core histones is coupled to replication (Jackson & Chalkley, 1985; Osley, 1991), they are usually incorporated into chromatin through the DNA synthesis-dependent pathway. On the other hand, the histone variants, and in particular, H3.3, are for the most part integrated via the DNA synthesis-independent pathway.

The DNA synthesis-dependent pathway revolves around the chaperone CAF-1 (S. Smith & Stillman, 1989; Verreault, Kaufman, Kobayashi, & Stillman, 1996). H3-H4 dimers are bound by
Asf1 (forming a heterotrimer known as RCAF) and delivered to CAF-1 (English, Maluf, Tripet, Churchill, & Tyler, 2005; Groth et al., 2007; Mello et al., 2002; Jessica K. Tyler et al., 1999). CAF-1 subsequently deposits H3-H4 on the DNA as tetramers in yeast (Winkler, Zhou, Dar, Zhang, & Luger, 2012), and possibly in higher eukaryotes too via dimerization of its p150 subunit (Gurard-Levin et al., 2014; Mattiroli et al., 2017; Quivy, Grandi, & Almouzni, 2001; Sauer et al., 2017). During replication, the histone deposition coupled to the replication complex via the interaction of CAF-1 with PCNA (Ben-Shahar et al., 2009; Gaillard et al., 1996). While it is known that the H2A-H2B dimers are added rapidly after the H3-H4 tetramer, it is unclear which chaperones mediate the transfer. FACT is a prominent contender due to the effect of its depletion on replication speed (Abe et al., 2011) and to its association with the replication fork (Gambus et al., 2006; Okuhara et al., 1999; B. C. Tan, Chien, Hirose, & Lee, 2006; Wittmeyer & Formosa, 1997).

The DNA synthesis-independent pathway can be mediated by HIRA (Ray-Gallet et al., 2002; Ray-Gallet et al., 2011; Tagami et al., 2004) or DAXX (Drané et al., 2010; Goldberg et al., 2010). These pathways are thought to function similarly to the CAF-1 pathway, with the exception that they deposit the histone variant H3.3, while CAF-1 assembles H3.1 (Tagami et al., 2004). In both cases, Asf1 would provide the H3-H4 dimers, and the deposition of the H2A-H2B dimers remains to be clarified (Gurard-Levin et al., 2014).

In vitro, the charges of the histones can be neutralized using anions. The gradual decrease in the concentration of anions allows for the strong interactions between the DNA and histones to establish first, while gradually allowing weaker and weaker interactions (Luger, Rechsteiner, Flaus, Waye, & Richmond, 1997b; Lusser & Kadonaga, 2004). This leads to the deposition of H3-H4 tetramers first, followed by the H2A-H2B dimers. Indeed, H2A-H2B dimers bind DNA stably below 0.6 M NaCl, while the interaction between the H3-H4 tetramer and DNA is stable below 1
M NaCl (Jorcano & Ruiz-Carrillo, 1979). Hence, it became a common procedure to first stabilize histone octamers at 2M NaCl (Eickbush & Moudrianakis, 1978; Jean O. Thomas & Butler, 1977), mix them with DNA, and then gradually decrease the salt concentration to physiological levels to obtain properly reconstituted nucleosomes (Luger et al., 1997b). Apart from salt, RNA and polyanions have been used for nucleosome assembly (T. Nelson, Wiegand, & Brutlag, 1981; Stein, Whitlock, & Bina, 1979). It is also possible to assemble nucleosomes in vitro by using chaperones (Fyodorov & Kadonaga, 2003; Lusser & Kadonaga, 2004). In particular, in vitro reconstitution with NAP-1, a chaperone that binds H2A-H2B and H3-H4 with similar affinities (Andrews, Downing, Brown, Park, & Luger, 2008), but also H1 (Kepert, Mazurkiewicz, Heuvelman, Tóth, & Rippe, 2005), was used to obtain the equilibrium constants of the nucleosome assembly pathway (fig. 17 and Andrews, Chen, Zevin, Stargell, and Luger, 2010; Mazurkiewicz, Kepert, and Rippe, 2006), a measurement that cannot be obtained with salt gradients as the assembly reactions are not reversible in such assays (Thastrom, Gottesfeld, Luger, & Widom, 2004). In all cases, these assays reproduce the order of addition of histones observed in vivo (fig. 17).

The disassembly of nucleosomes is thought to process in reverse order from its assembly. With increasing concentrations of salt, the DNA-H2A/H2B interactions and the dimer-tetramer interactions first release, leading to a potential “butterfly conformation”, where the dimers are still bound to DNA, but not to the tetramer (V. Böhm et al., 2011; Gansen et al., 2009). The H3-H4 tetramer releases from DNA last (Jorcano & Ruiz-Carrillo, 1979). Under force, it was shown that the nucleosome present 3 main barriers: 2 located around 40 bp on each side of the dyad, and the stronger one located at the dyad (Hall et al., 2009). When applying force in a unidirectional manner similar to a motor protein, the nucleosome is lost once the interactions at the dyad are broken, indicating that interactions on only one side of the nucleosome are not strong enough to maintain it (Hall
In vivo, histone dissociation is also chaperoned. Indeed, free histones would lead to aggregation. For example, during transcription, histone disassembly can be mediated by FACT, NAP-1, nucleoplasmin, nucleolin, and ASF-1 (Adkins, Howar, & Tyler, 2004; Angelov et al., 2006; Belotserkovskaya et al., 2003; H. Chen, Li, & Workman, 1994; Schwabish & Struhl, 2006; Walter, Owen-Hughes, Côté, & Workman, 1995). It also appears that the torsional stress applied by the RNA polymerase can lead to eviction of H2A-H2B dimers (Levchenko, Jackson, & Jackson, 2005), although it is unlikely that dimers released this way would remain freely diffusing in vivo, and more likely that chaperones will help the release. Chromatin remodeling complexes have also been shown to disassemble nucleosomes in the reverse order of assembly, in the presence of chaperones. The remodeler RSC was shown, in the presence of NAP-1 and ATP, to disassemble first the H2A-H2B dimers, and then the tetramer (Lorch, Maier-Davis, & Kornberg, 2006). During DNA repair, the chromatin remodeler INO80 is essential for the removal of nucleosomes, while...
both HIRA and CAF-1 will be involved in reassembly (X. Li & Tyler, 2016).

Overall, because of the highly polar and hydrophobic nature of histones, care has to be taken both in vivo and in vitro to favor the formation of nucleosomes and not aggregates (fig. 16). In vivo, the handling of histones is mediated by chaperones, of which Asf1, CAF-1, HIRA, NAP-1, and FACT are key players (Gurard-Levin et al., 2014). In vitro, the histones can be passivated using salts, polyanions, or chaperones (Lusser & Kadonaga, 2004).

### 2 Chromatin condensation

#### 2.1 Condensation of chromosomes

How interphase chromatin morphs into chromosomes during M phase remains to be elucidated (Kschonsak & Haering, 2015). Since nucleosomes do not appear to fold into structures higher than the 10 nm fiber in vivo (see section 1.8), it is probable that either other factors are in charge of the condensation of DNA, or an activation event has to occur to induce the compaction of nucleosomes.

Wilkins et al. (2014) showed that the phosphorylation of H3 S10 can act as this activation event. Indeed, in *S. cerevisiae*, the phosphorylation of H3 S10 leads to the recruitment of the deacetylase Hst2p, which deacetylates H4 K16. In turn, deacetylated H4 K16 can interact with the acidic patch at the surface of a H2A-H2B dimer on a neighboring nucleosome, leading to the compaction of chromatin. Accordingly, it was shown that H3 S10 is required for proper chromosome segregation in *Tetrahymena thermophila* (Y. Wei, Yu, Bowen, Gorovsky, & Allis, 1999) and in *S. pombe* (Mel-lone et al., 2003). However, there is no clear defect from the mutation of H3 S10 to an alanine in *S. cerevisiae* (Dai et al., 2008), implicating that other pathways may function redundantly, or that
this pathway is not necessary to proper chromosome formation. In addition, computer simulations indicate that the nucleosome-mediated compaction of chromosomes would result in a spherical, rather than the observed cylindrical, chromosomes (Marko & Siggia, 1997). Hence, nucleosome compaction is not sufficient to explain the formation of chromosomes. It is likely, however, that the importance of the contribution of histones depends on chromosome length. Indeed, it was shown in *S. cerevisiae* that by increasing the length of a chromosome by 45%, the mutation H3 S10A leads to a loss of the “hypercondensation” of the mutant chromosome (Neurohr et al., 2011). Moreover, *S. cerevisiae* does not form dense, easily visible chromosomes like higher eukaryotes, indicating that its required level of chromosome compaction may be generally lower—probably related to its small genome size—than other eukaryotes (Vas, Andrews, Kirkland Matesky, & Clarke, 2007).

Overall, while nucleosomes contribute to the proper condensation and segregation of chromosomes, they no not appear to be the only players.

Another protein directly involved in the formation of chromosomes is condensin. Indeed, condensin was shown to be involved in the condensation of chromatin in human cells (Lewis & Laemmli, 1982) and in *X. laevis* egg extract (T. Hirano, Kobayashi, & Hirano, 1997). Moreover, in the absence of functional condensin, mitosis proceeds, but is erratic and often leads to DNA bridges or improper chromosome segregation, including in *S. cerevisiae* (T. Hirano & Mitchison, 1994; Saitoh, Goldberg, Wood, & Earnshaw, 1994; Saka et al., 1994; A. V. Strunnikov, Hogan, & Koshland, 1995). Finally, in vitro, chromosome formation was shown to be independent of the presence of histones, arguing for condensin to be the main actor of chromosome condensation (Shintomi et al., 2017). Such chromosomes, however, are more fragile, so the function of nucleosomes may not be chromosome morphogenesis, but its structural reinforcement. Condensin proteins are also present in all kingdoms of life, arguing for their essential role (T. Hirano, 2016).
Consequently, it appears that condensin plays a principal role in the condensation of chromosomes.

2.2 Structural maintenance of chromosomes (SMC) proteins

Condensin is part of the structural maintenance of chromosomes (SMC) family of proteins, along with cohesin and Smc5-6 (Uhlmann, 2016). The first member of the SMC family to be discovered was mukB, the E. coli condensin. In the absence of mukB, cells exhibit chromosome partitioning defect (Niki, Jaffé, Imamura, Ogura, & Hiraga, 1991). Next, Smc1, which was later shown to be part of the cohesin complex (Guacci, Koshland, & Strunnikov, 1997; Losada, Hirano, & Hirano, 1998; Michaelis, Ciosk, & Nasmyth, 1997; A. Tóth et al., 1999), was discovered in S. cerevisiae. Likewise, absence of Smc1 leads to chromosome segregation defects and loss of viability (A V Strunnikov, Larionov, & Koshland, 1993). Finally, subunits of condensin were discovered to share the same functions (T. Hirano et al., 1997; T. Hirano & Mitchison, 1994; Saitoh et al., 1994; Saka et al., 1994; A. V. Strunnikov et al., 1995). Hence, it appeared that SMC proteins are involved in the maintenance and partition of chromosomes.

Further research demonstrated the wide range of functions of the SMC protein family. Cohesin was shown to hold sister chromatids together (Guacci et al., 1997; Michaelis et al., 1997; Uhlmann & Nasmyth, 1998), as well as to organize topologically associating domains (TADs) (Gassler et al., 2017; Rao et al., 2017; Wutz et al., 2017). Condensin was shown to be necessary for proper chromosome condensation and segregation (T. Hirano et al., 1997; T. Hirano & Mitchison, 1994; Saitoh et al., 1994; Saka et al., 1994; A. V. Strunnikov et al., 1995). In additions, condensin is also involved in X chromosome dosage compensation in C. elegans (Chuang, Albertson, & Meyer, 1994), in the establishment of chromosome territories (Bauer, Hartl, & Bosco, 2012), in the pre-
Figure 18: General structure of SMC proteins. All SMC proteins share the same structural elements: a dimer of SMC subunits, which can be a homodimer in prokaryotes or an heterodimer in eukaryotes, an asymmetrically bound kleisin subunit and, finally, accessory factors associated with the kleisin subunit. The Smc5/6, cohesin, and condensin proteins showed are from *S. cerevisiae*. Adapted from Hassler, Shaltiel, and Haering (2018).

vention of somatic homologous pairing (Joyce, Williams, Xie, & Wu, 2012), and in the dispersion of chromocentres (Nishide & Hirano, 2014). Note that all of these functions imply the separation of chromosomes. Bacterial SMC proteins are involved in origin segregation and chromosome separation (Gruber et al., 2014; Xindan Wang, Brandão, Le, Laub, & Rudner, 2017; Xindan Wang, Tang, Riley, & Rudner, 2014). Finally, SMC5-6 is involved in DNA repair (Lehmann et al., 1995). Hence, all rightly-named SMC proteins are involved in the processing and maintenance of chromosomes.

All members of the of the SMC family are multi-subunit complexes that include two long SMC subunits, one kleisin, and accessory factors (fig. 18). Each SMC subunit, ranging from 270 to 350 amino acids (a.a.) (Bürmann et al., 2017), contains a Walker A and Walker B motif at each of its ends. The motifs are separated by coils and a central non-helical domain. The subunits fold on themselves, forming antiparallel coiled-coils on their length and a terminal ATP-binding cassette (ABC)-like ATPase head domain resulting from the assembly of the Walker motifs. At the other end of the folded protein, the non-helical domain dimerizes with that of another SMC subunit.
to form a donut-shaped hinge domain. This results in a V-shaped dimer of about 50nm (D. E. Anderson, Losada, Erickson, & Hirano, 2002; Haering, Löwe, Hochwagen, & Nasmyth, 2002), a length that is highly conserved (Melby, Ciampaglio, Briscoe, & Erickson, 1998). The two ATPase domains of the dimer are linked by a kleisin subunit, creating a topologically closed ring thought to entrap DNA (Cuylen, Metz, & Haering, 2011; Gligoris et al., 2014; Haering, Farcas, Arumugam, Metson, & Nasmyth, 2008; Wilhelm et al., 2015). Interestingly, the kleisin subunit binds the two SMC dimers asymmetrically, even in the symmetrical prokaryotic complex. The N-terminus of the kleisin subunit attaches to the coiled-coil above the ATPase domain of one of the two SMC subunit, while the C-terminal attaches to the bottom of the ATPase domain of the other subunit (Bürmann et al., 2013; Gligoris et al., 2014; Haering et al., 2004). In the presence of ATP, the two ATPase head domains interact by the paired binding of one ATP molecule each, leading to an open ring structure with the two “legs” away from each other, but the feet holding to each other (Diebold-Durand et al., 2017; Lammens, Schele, & Hopfner, 2004). However, in the absence of ATP, the legs collapse and the ATPase heads separate, leading to a rod-shaped structure (Diebold-Durand et al., 2017). The SMC complexes are completed by winged-helix domain (WHD) or “huntingtin, elongation factor 3, protein phosphatase 2A and TOR1 kinase” (HEAT) factors that bind to the kleisin subunit (fig. 18). HEAT proteins are found on the cohesin and condensin subunits (Haering et al., 2002; Hara et al., 2014; Onn, Aono, Hirano, & Hirano, 2007), while WHD proteins are found on the bacterial SMCs and on Smc5/6 (J. J. Palecek & Gruber, 2015; J. Palecek, Vidot, Feng, Doherty, & Lehmann, 2006; Wells, Gligoris, Nasmyth, & Marsh, 2017). Interestingly, the hinge domain, positioned at the other end of the complex, was also shown to interact with these factors, hinting at a possible bent conformation of the SMC proteins (Duan et al., 2009; Eeftens et al., 2016; Murayama & Uhlmann, 2015).
Overall, all SMC proteins from prokaryote to eukaryote—condensin, cohesin, and Smc5-6—share the same structure of a dimer of two long SMC coiled-coil legs with a hinge domain linking the monomers and one ATPase domain at each foot. The 50 nm V-shaped structure is closed topologically by an asymmetrically bound kleisin subunit, itself sporting a variety of HEAT or WHD factors.

2.3 Mechanisms of chromosome morphogenesis

Figure 19: Steps and factors in the morphogenesis of chromosomes. Frog sperm chromatin, which is composed of histones H3/H4 and sperm-specific proteins (SPs), displays a highly compacted, snake-like shape. Nucleoplasmin (Npm) evicts SPs to swell sperm chromatin, and Nap1 deposits H2A (the embryonic variant H2A.X-F in this case) and H2B to assemble nucleosomes, resulting in the formation of ‘banana’-shaped chromatin. It has been proposed that FACT destabilizes and mobilizes the nucleosomes so that they become a good substrate for subsequent actions of topoisomerase IIα (topo IIα) and condensin I (cond I). Topo IIα catalyzes decatenation of intertwined chromatid DNAs, thereby allowing the banana-shaped chromatin to be converted into a ‘cloud’-like morphology. Finally, condensin I drives the formation of single mitotic chromatids with continuous help of FACT and topo IIα. The activation of condensin I requires phosphorylation mediated by Cdk1. Adapted from T. Hirano (2016).

Chromosome morphogenesis was recapitulated in vitro using a set of 6 factors (fig. 19 and Shintomi, Takahashi, and Hirano, 2015) and sperm chromatin from X. laevis. First, nucleoplasmin
evicts the protamines and sperm proteins from the chromatin. Interestingly, *X. laevis* sperm comes with the H3-H4 tetramers, so that only Nap1 and the H2A-H2B dimers are required to assemble nucleosomes. This leads to expanded sperm chromatin, the “banana” morphology. Next, *Facilitates access to chromatin (FACT)* and Topoisomerase II are necessary to expand the chromatin into a cloud-like shape reminiscent of interphase chromatin. Finally, the addition of condensin phosphorylated by Cdk I—this phosphorylation was shown to be required for condensation in vivo (Robellet et al., 2015)—leads to the condensation of the chromatin into its chromatid shape which, would the sister chromatids be linked at the centromere, would give canonical chromosomes. Why *FACT* is necessary for the formation of chromosomes remain an open question, and it was suggested that its function is to open chromatin for the proper function of topoisomerase II and condensin (T. Hirano, 2016; Shintomi et al., 2015). It was also shown by the same group that DNA condensation can proceed in the absence of histones (Shintomi et al., 2017). Using sperm DNA from *M. musculus* which, unlike its *X. laevis* counterpart, is mostly devoid of histones, the authors were able to reproduce chromosome-like structures in *X. laevis* egg extract depleted of the nucleosome-deposition factor Asf1, hence preventing the association of any nucleosomes on the DNA. Consequently, it is likely that condensin, topoisomerase II, and nucleoplasmin at the only factors necessary for DNA condensation, although because the work was performed in extract, the minimal set of proteins is not known. It is to be noted that, although DNA condensation proceeded and that the condensed DNA recreated the expected chromosome shape in the absence of nucleosomes, these chromosomes were not fully condensed and more fragile, arguing for a role of histones in reinforcing the chromosome structure rather than shaping it (Shintomi et al., 2017).

From these experiments, it appears that condensin is the essential factor that transforms the interphase-like DNA “cloud” into mitosis-like DNA rods (fig. 19). It was first suggested that cond-
densin could extrude loops of DNA to achieve its function (Nasmyth, 2001). Computer simulations agreed that, with the help of topoisomerase II, condensin molecules can compact and separate the sister chromatids, with results similar to what is observed during mitosis in vivo, simply by extruding loops of chromatin, randomly (Goloborodko, Imakaev, Marko, & Mirny, 2016). However, one big obstacle to this model was the low ATPase rate of condensin, and generally of SMC proteins (Cobbe & Heck, 2006). Indeed, these rates are several orders of magnitude lower than those of known motor proteins, and hence would not be compatible with a sustained DNA extrusion mechanism (Hassler, Shaltiel, & Haering, 2018). This issue was resolved using single-molecule DNA Curtains with S. cerevisiae condensin (Terakawa et al., 2017). It was demonstrated that condensin indeed is a motor protein, and that albeit its low ATPase activity, it can translocate at around 60 bp/s, with a processivity of more than 10 kbp, suggesting a step size of 16 nm per ATP molecule, and a yet-to-be discovered translocation mechanism. In comparison, usual DNA translocases have a step size of one to few base pairs (1 nm) per ATP molecule (Singleton et al., 2007). Loop extrusion was next demonstrated at the single-molecule level using a similar fluorescence-based technique (Ganji et al., 2018). The authors showed that single S. cerevisiae condensin proteins extrude loops in a unidirectional manner and at a rate that depends on the tension of the DNA molecule. Hence, it was demonstrated theoretically that loop extrusion can lead to the formation of mitotic chromosomes, and experimentally that condensin, which is necessary their formation in vivo, extrudes DNA loops in vitro. It can be concluded that chromosomes are formed in vivo by the extrusion of DNA loops by condensin. This model was also suggested in vivo in prokaryotes (Xindan Wang et al., 2017) and in chicken cell cultures (Gibcus et al., 2018).

How the structure of condensin translates into loop extrusion remains unclear. Several models have been proposed to link the structure of condensin with its loop extrusion activity (fig. 20). First,
Figure 20: Hypothesized loop extrusion mechanisms by condensin. (A) Sequential walking model. (B) DNA pumping model. (C) Extended scrunching model. See main text for details. Note that the sequential walking model makes no assumptions about the SMC coiled coil conformations, whereas the pumping model assumes stiff coiled coils that are under tension when bent open. The extended scrunching model postulates that SMC coiled coils alternate between stiff and relaxed states. Adapted from Hassler, Shaltiel, and Haering (2018).

The sequential walking model stipulates that condensin anchors at one point, for example the hinge, and then “walks” on the DNA, each step being made either by the association of the hinge domain to the ATPase feet, transferring the DNA anchor point each cycle (Duan et al., 2009; Murayama & Uhlmann, 2015), or through the separation of the legs, threading the DNA into the main loop,
with the condensin anchored at the hinge (fig. 20A and Terakawa et al., 2017). This model would explain the large step size and faster loop extrusion on looser DNA observed in single-molecule experiments (Ganji et al., 2018). While this model does not require topological enclosing of the DNA per se, it is not incompatible with it (Cuylen et al., 2011). Second, the DNA pumping model stipulates the existence of two DNA chambers: the larger one formed by the SMC subunits, and a smaller one formed by the kleisin subunit (fig. 20B and Diebold-Durand et al., 2017; Minnen et al., 2016). A DNA loop is formed in the SMC chamber and, upon hydrolysis of ATP, as the legs zip up and the ATPase feet separate, the loop is pushed down into the kleisin “meta-chamber”. DNA would be anchored within the meta chamber, and the process would repeat. In this model, the SMC legs do not define the step size, and the large chamber simply helps thread larger DNA loops in. Hence, this model also predicts larger step sizes on looser DNA. How the DNA loop is threaded into the SMC chamber, however, is unclear. Finally, the extended DNA scrunching model was proposed (fig. 20C and Hassler et al., 2018). In this model, condensin would anchor via the safety belt structured the kleisin and HEAT subunits (Kschonsak et al., 2017). Just like in the sequential walking model, the SMC legs would provide with the step size. DNA would be captured at the hinge upon ATP binding, which would induce a conformational change and bring the hinge domain against the ATPase/kleisin/HEAT subunits. Hydrolysis of ATP would lead to the transfer of the DNA from the hinge to the ATPase domains, where it would bind. This model would explain several biochemical findings related to SMC proteins: the safety belt mechanism from condensin (Kschonsak & Haering, 2015), the importance of the positively charged channel inside the hinge in cohesin (Kurze et al., 2011), and the interaction of the hinge with the accessory subunits in Smc5-6 and cohesin (Duan et al., 2009; Murayama & Uhlmann, 2015).

None of these models, however, have been demonstrated, and the DNA binding sites of con-
densin, as well as the structural dynamics, are part of the missing information. Very few DNA binding sites within condensin are known, with the only one confirmed at the HEAT repeats, forming a “safety belt” with kleisin that entraps DNA (Kschonsak et al., 2017; Piazza et al., 2014). Another putative site is located at the hinge domain (Chiu, Revenkova, & Jessberger, 2004; Griese & Hopfer, 2011; Griese, Witte, & Hopfer, 2010; M. Hirano & Hirano, 2006; Piazza et al., 2014). However, this site seems to favor single-stranded DNA binding over double-stranded DNA binding (which does not disqualify it as a double-stranded DNA binding site). Crystal structures of Rad50, a SMC-related protein, suggested DNA binding sites at the dimerized ATPase domains bound to ATP, and on the coiled-coils next to the ATPase head domains (Liu et al., 2016; Rojowska et al., 2014; Seifert, Lammens, Stoehr, Kessler, & Hopfer, 2016). Because the same structures are found on prokaryotic SMC proteins (Lammens et al., 2004; Woo et al., 2009), it is tempting to extrapolate the sites to all SMC proteins (which models above do), although they remain to be demonstrated. Moreover, very few studies have investigated the conformational changes and structural dynamics of condensin. Indeed, most structural dynamics have been inferred from crystal structures of parts of the SMC proteins and from atomic force microscopy (AFM) and cryogenic electron microscopy (cryoEM). Known structural changes are the dimerization of the ATPase domains upon ATP binding (Lammens et al., 2004), the opening and closing of the hinge domain (Y. Li, Schoeffler, Berger, & Oakley, 2010; Soh et al., 2015; Srinivasan et al., 2018), and the flexibility of the coiled-coil legs and the interaction of the hinge with the ATPase domains (Eeftens et al., 2016). It appears from the angles in the structures that dissociated ATPase domains are only compatible with rod-shaped coiled-coils and a closed hinge, and that the associated ATPase head domains are only compatible with an open hinge and separated, open coiled-coil legs (Hassler et al., 2018). Further studies, notably through single-molecule Forster energy resonance transfer (FRET), will be required to parse
out the conformational changes and mechanisms of loop extrusion by condensin.

Another layer of complexity in parsing out the mechanisms of chromosome morphogenesis lay in the existence of two condensins in most eukaryote, with varying preponderance depending on organisms (fig. 21). While retaining the same SMC subunits, the two condensins differ by their kleisin and accessory factors, resulting in different spatiotemporal regulation and functions. Condensin II is present in the cell nucleus at all times and initiates chromosome compaction in prophase, whereas condensin I is cytoplasmic and only start compaction in prometaphase, after nuclear envelope breakdown (T. Ono, Yamashita, & Hirano, 2013). In addition, condensin II is involved in the axial shortening of chromosome, whereas condensin I handles their lateral compaction (Green et al., 2012; Shintomi & Hirano, 2011). Hence, in eukaryotes, a division of task between condensins is observed, but the importance of the condensins differ between organisms. For example, in *C. elegans*, condensin II only is essential, whereas condensin I only is essential in *Cyanidioschyzon merolae* (T. Hirano, 2016).
These differences may stem from the different types of chromosomes—*C. elegans* has holocentric chromosomes (Maddox et al., 2004)—and genome sizes (fig. 21).

Overall, chromosome morphogenesis from sperm DNA can be recapitulated using 6 factors: histones, nucleoplasmin, Nap-1, topoisomerase II, FACT, and condensin (Shintomi et al., 2015). While histones are not necessary for the formation of chromosomes, they participate in its strength and full compaction (Shintomi et al., 2017). The molecular mechanism through which condensin transforms interphase DNA “clouds” into chromosomes is currently unknown, but most likely relies on the loop extrusion capacity of condensin (Ganji et al., 2018; Goloborodko et al., 2016; Terakawa et al., 2017).

3 Single molecule experimentation

It is very easy to answer many of these fundamental biological questions; *you just look at the thing!*

Feynman (1992)

3.1 The limitations of ensemble techniques

Biology is typically explained with geometrical shapes in a galore of colors linked together by arrows. Each shape represents one molecule, each arrow, one interaction. This pictograph chemistry is the way we think about biology: single molecules interacting with each other, leading to an action as a function of an input, and ultimately to life. However, because molecules are so small, it is difficult to observe them as they interact individually. Hence, scientists have had to devise ways to probe them without seeing them. A very common way is to gather a large amount of molecules
and then probe their properties as a group. The large amount of input molecules leads to a large response to the probe that can be observed more readily. Common examples are spectrometry, where absorption and scattering properties of molecules are probed, or gel electrophoresis, where size and charge can be assessed. While these techniques (colloquially termed “bulk” techniques, and more officially “ensemble” techniques) have produced most of the knowledge we have of biological entities, they present several limitations that can impede the precision of the observations and lead to misinterpretation of the results.

First, ensemble techniques generally have poor temporal resolution. They involve the treatment of large amount of molecules, transferring them from one vessel to another, and this treatment takes time. As a consequence, complex systems have to be develop to obtain temporal resolution below the second, or even below the minute mark. Moreover, because they require handling a large amount of particles, synchronization is often necessary to get temporal measurements with ensemble assays, adding to the complexity. For example, an enzymatic reaction can be synchronized by the precise temporal addition of a substrate. However, because of the large volumes used, proper mixing must be done fast to obtain a signal as uniform as possible. This leads to complex pieces of equipment with advanced microfluidic designs. Hence, ensemble techniques make it hard to observe the behavior of molecules over time.

Second, and more importantly, by design, ensemble techniques average everything, loosing information about the static and dynamic heterogeneities of the population. Because a consequent amount of molecules is needed to obtain the desired signal, ensemble techniques can only inform on the behavior of a population. Let us, for example, measure the speed of the runners of the New York marathon. Sitting in a helicopter, we observe their position, wait 5 minutes, and then observe again. We measure the distance traveled by the main pack of runners, divide it by the time, and get
a velocity. While this measurement is perfectly sensible, it does not contain one the main elements of the marathon, which would be how fast the winner ran. It would also hide the distribution of speeds for each runner, whether some runner went fast at first and then stopped, etc. The distribution of the velocities of the runners is defined by their static and dynamic heterogeneities. The static heterogeneity of the population is defined by the different states the different individuals are found in. For example, the marathon winner probably came with thousands of hours of training, at the peak of his shape, and thus would be in a high state of physical capacity. The average runners may have trained somewhat less and marathons may not be their living, putting them in a lower state of physical capacity. The dynamic heterogeneity of a population can to defined at the changes in the velocity of a single individual. Indeed, each individual probably did not run at the constant velocity obtained from our ensemble measurement. Some went fast at first, and then slowed down, while others did the opposite. A few overzealous runners may have sprinted at the end. How individuals behave over time defines the dynamic heterogeneity of the population, and also plays its role in explaining the velocity. Hence, by observing only the average pack of runners, one lost key information about the population that would explain its velocity. The velocity of the fastest runner may also have been lost by the insensitivity of the measurement. Overall, ensemble measurement of a population leads to simplifications, because it oversees many details about the individuals in the population.

The simplifications of ensemble measurements can be misleading, because we think about biology in term of single molecules, not population averages. Consequently, ensemble measurements are more likely to lead to misinterpretations. While the average velocity of the main pack of runners is a real measurement, that ensemble, averaged value is easy to over interpret. The scientist studying marathons this way could conclude it simply is a very slow method of mass transportation. This
is because the complexity of the individual behaviors in the population is averaged out. Consider the speed of the New York subway. Averaging the distance all trains traveled in 5 minutes gives the average velocity of the subway system. While perfectly relevant, this measurement hides the complexity of the subway system. The trains stop at every station, sometimes someone holds the door, often the train dispatcher is holding it at the station. Moreover, that measurement does not give the actual speed of trains, because it includes the pauses at the stations in the averaged value. By simply measuring the average velocity, one would hide the true behavior of the subway, and think it simply is a slowly moving system. The scientist studying marathons and subways in New York would compare them as two slow means of transportation involving lots of people, potentially thinking they are the same. While one can easily see how this conclusion is wrong, in molecular biology, however, we cannot see so easily, and we are probing the unknown blinded, with mittens over our hands. While great creativity has led to an amazing wealth of knowledge of the living, having the most precise tools for studying what we cannot see and can barely grasp the complexity of is crucial.

Ensemble techniques have allowed to probe the molecular world and have led to tremendous breakthroughs that define our current knowledge of biology. However, ensemble measurements are notoriously blind to dynamic heterogeneity, the dynamic behavior of an individual object of study that leads to its averaged ensemble behavior (the different speeds of an individual runner over time), and static heterogeneity, the context or state of the individual subjects of a population that leads to its averaged behavior (the amount of training before the marathon) (Knight, 2009).
3.2 Overview of single-molecule techniques

To solve these limitations, scientists have devised ways to observe the behavior of individual molecules. This set of techniques is referred to as “single-molecule” techniques. Single-molecule techniques, however, are not content with the behavior of one molecule. To produce meaningful data, the single-molecule biologist collects large data sets of individual molecules and then average them to get insights about the population.

First hand, this may sound counterproductive. Why all of the trouble if in the end, the result is an average? Consider the New York subway example. First, the velocity of every train was measured, and then it was averaged. As it turns out, this is exactly what a single-molecule experiment is. An ensemble experiment would have put all the trains and rails in a same place, started them all at approximately the same time, and then looked at how far that train mass moved from its initial position in a given time. A single-molecule experiment would track the movement of the trains as it happens, train by train. Hence, the single-molecule biologist can separate the time spent stopping, for example at a station, versus the time spent moving between stations. The averaged velocity of the train can hence be refined. Of course, if one just wants to know how fast a train goes from point A to point B, then the ensemble measurement is sufficient. However, the single-molecule experiment has informed that trains pause frequently, with relatively constant dwell times. This extra information is of great help to parse out the function of subways and how they fulfill that function. Likewise, by observing single marathons runners, one can figures that those in a higher initial physical state run faster, and that those who start really fast and in a lower physical state usually end very slow. As such, single-molecule experiments simply provide an added layer of precision by integrating information on the dynamic and static heterogeneities of the population,
and this added precision helps to parse out mechanisms and functions.

Several single-molecule techniques have been devised over the years (Tinoco & Gonzalez, 2011; Zlatanova & van Holde, 2006). While the recent advances of super-resolution microscopy have made in vivo single-molecule studies possible (Sahl, Hell, & Jakobs, 2017), most single-molecule techniques are of biochemical nature and happen in vitro. As well, most single-molecule techniques rely on microscopy. I will present a few of them: fluorescence-based techniques (including Forster energy resonance transfer (FRET)), optical and magnetic tweezers, atomic force microscopy (AFM), and finally, DNA Curtains.

**Fluorescence single-molecule microscopy**

Fluorescence-based techniques exploit fluorophores attached to molecules to observe their behavior. Usually, molecules will be labeled with one or several fluorophores, such as a fluorescent protein, an organic fluorophore or a quantum dot (Specht, Braselmann, & Palmer, 2017). When a fluorophore absorbs a photon at its excitation wavelength, an orbital electron in the fluorophore reaches the excited, singlet state. When the electron relaxes to its ground state, it reemits this energy as a photon of lower energy (longer wavelength) and heat, a phenomenon known as fluorescence. The difference between the absorption wavelength and the emission wavelength is known as Stokes’ shift (Lakowicz, 2006). Hence, an illumination source such as a laser can be used to excite specific fluorophores, and a set of wavelength filters and a camera can be used to record the emitted wavelength. With advanced detectors and fluorophores, optical systems can resolve single molecules above the Rayleigh limit, allowing to probes the interactions and movement of labeled molecules. Many systems exist to resolve single fluorophores (Shashkova & Leake, 2017), and I will here focus on total internal reflection fluorescence (TIRF) microscopy.
Figure 22: (A) Schematic representation of a total internal reflection fluorescence (TIRF) microscopy experiment. Nucleosomes are tethered on a glass surface, with some freely diffusing in solution. The incident light (green) is reflected on at the interface between the glass (blue rectangle) and the solution, leading to the formation of an exponentially decaying evanescent field. Only the nucleosomes close enough to the surface, within the range of the evanescent field, are illuminated (colored nucleosomes versus grey ones). The further away from the surface, the lower the light they receive (faded versus fully colored nucleosomes), leading to a specific enrichment of the signal close to the surface. (B) Effect of the incident angle on the depth of the illumination field in TIRF microscopy. Using refractive indexes of 1.46 (fused silica) and 1.33 (water), the effect of the incident angle on the depth of the evanescent wave was calculated. The critical angle is 65.63°, and wavelength was set to 488 nm.

While conventional epifluorescence microscopy illuminates the whole depth of a sample, TIRF microscopy allows for the observation of a thin optical section, usually of around 100 nm, at the surface of the sample (Daniel Axelrod, 2001). By illuminating a surface at an angle higher than the critical angle, light reflects back and does not enter the sample, a phenomenon called total internal reflection (TIR). The critical angle $\theta_c$ is defined by the refractive indexes in the light path, $n_1$ and $n_2$:

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right)$$

where $n_1$ is the refractive index of the first medium in which the light travels, usually the slide.
(\(n = 1.46\) for fused silica), and \(n_2\) is the refractive index of the medium of the sample, usually water \((n = 1.33)\). Note that TIR can only happen if \(n_2 < n_1\).

When TIR happens, an “evanescent wave” propagates parallel to the surface between the two media (fig. 22). The intensity of the evanescent wave at position \(z\) above the surface of the slide (depth) is given by the equation

\[ I(z) = I_0 e^{-\frac{z}{d}} \]

with

\[ d = \frac{\lambda}{4\pi \sqrt{n_1^2 \sin^2 \theta - n_2^2}} \]

where \(I_0\) is the input light intensity and \(\theta\) is the incident angle. These equations indicate that the illumination field will decrease exponentially with depth. The depth of the illumination field also decreases fast with the incident angle. For example, at the interface between a fused silica slide and water, the critical angle will be around 65.63°, around which the depth of illumination will approximate 6.7 mm. At an incident angle of 66°, the illumination depth approximates 2.5 μm, and at 75°, 100 nm (fig. 22). The wavelength, on the other hand, only has a minor effect on the depth of the evanescent wave.

Because of its thin illumination section, TIRF microscopy provides with a great signal-to-noise ratio for molecules on the illuminated surface. Indeed, all of the molecules more than 100 nm above the surface are not exposed to significant light, and hence do not emit signal. For example, TIRF microscopy has been used to study processes at the cell membrane in vivo (Axelrod, 1981), and in vitro to demonstrate that microtubules are motor proteins (Vale et al., 1996). Study of molecules in-gel, to slow their diffusion, were also conducted (Dickson, Norris, Tzeng, & Moerner, 1996). Ultimately, techniques evolved to immobilize proteins of interest at the surface, and TIRF mi-
Forster energy resonance transfer (FRET)

When two fluorophores are close to each other, the “donor” fluorophore can transfer energy to the “acceptor” one by non-radiative energy transfer. This phenomenon was named Förster Resonance Energy Transfer (FRET), after its discoverer (Förster, 1948, 12). It is also referred to as Fluores-
cent Resonance Energy Transfer, keeping the same acronym. The amount of energy transferred will be proportional to the distance between the donor and acceptor fluorophores. Hence, using FRET, one can measure the distance between two fluorophores. Proteins, nucleic acids, lipids, etc. are labeled at specific locations with fluorophores. By measuring the gain of the acceptor signal and the loss of the donor signal, one can measure their relative distance over time (Roy et al., 2008). The closer the two fluorophores, the higher the transfer to the acceptor, and thus the higher its signal, and the lower the signal of the donor.

FRET is not a single-molecule technique per se. Indeed, it is possible to collect ensemble FRET measurements. However, just like any single-molecule versus ensemble technique, single-molecule FRET increases the precision of ensemble FRET, as discussed above and by Ha et al. (1996). Due to its flexibility and power, single-molecule FRET has been used to study a wide variety of systems including chromatin (Buning & van Noort, 2010) and translation (Tinoco & Gonzalez, 2011).

**Tweezers**

In optical tweezers, or optical trap, a bead of dielectric material, such as glass, is trapped by a light gradient (Ashkin, 1992). By moving the light gradient, one can move the glass bead. A protein or nucleic acid can be tethered between two such beads, or between a bead and a surface. The induced or observed movement of the bead(s) can be used to apply or measure forces on the tethered molecule. For example, DNA, chromatin and nucleosomes have been pulled apart using optical tweezers to determine the forces that keeps them together (Cui & Bustamante, 2000; Hall et al., 2009; S. B. Smith, Cui, & Bustamante, 1996). Just like optical tweezers, magnetic tweezers can study force on single molecules, but instead of trapping beads of dielectric materials, they trap
magnetic beads (Amblard, Yurke, Pargellis, & Leibler, 1996; Crick & Hughes, 1950). Likewise, single-molecules can be attached to these beads and force can be probed and exerted.

**Atomic force microscopy (AFM)**

AFM relies on a cantilever to probe distances or force (Binnig, Quate, & Gerber, 1986). The cantilever has a tip that slides or taps on the surface. As the tip tracks the difference of heights on the surface, the cantilever moves up and down, and these movements are recorded by the reflection of a laser from the top of the cantilever on a detector. Hence, when the surface moves up, the tip moves up, causing the cantilever to move up, which deflects the laser beam, which is captured by the detector. By sliding the tip across a surface, one can get information about its structure down to the sub-nanometer resolution. The cantilever can also be used to exert force. By attaching a molecule one side to the surface and one side to the tip, and controlling the movement of the cantilever, one can use AFM to measure force from 10 to $10^4$ pN (Allen et al., 1993; Florin, Moy, & Gaub, 1994; Zlatanova & van Holde, 2006).

4 DNA Curtains

DNA Curtains is an in vitro single-molecule technique that addresses two common issues of single-molecule studies: the throughput and the biocompatibility of the surface on which molecules are tethered. These issues are addressed by using a lipid bilayer formed on the surface of a glass slide (Granéli, Yeykal, Prasad, & Greene, 2006).

First, lipid bilayers readily prevent DNA and protein binding to glass, hence providing with a biologically relevant way to passivate the surface and prevent non-specific binding (Glasmä-
tar, Larsson, Höök, & Kasemo, 2002). Indeed, in single-molecule studies, non-specific binding can lead to the accumulation of signal on the surface that would prevent the detection of single-molecules. Surface passivation is usually achieved using BSA or PEGylation, the first one usually offering limited results and the second one involving more preparation (Visnapuu, Duzdevich, & C. Greene, 2008a). In order to attach the molecules of interest to the passivated surface, a small amount of biotinylated lipids is added to the lipid stock. After establishment of a bilayer on the surface of a slide, these lipids can bind to streptavidin, which can attach to another biotin conjugated with the molecule of interest. Hence, lipids bilayers can be used to passivate surfaces in a way that is not only bio-compatible, but also bio-relevant, and molecules on interest can be attached specifically on the bilayer itself.

Second lipid bilayers are fluid, allowing for the concentration of molecules against lipid diffusion barriers. A molecule of interest, in our case DNA, attached to a lipid in a lipid bilayer will diffuse with the lipid in the bilayer (Granéli et al., 2006). In the presence of flow, the hydrodynamic force will push the DNA molecule in the bilayer in the direction of the flow. The movement of the DNA molecule will stop when the flow stops or when it hits a lipid barrier, leading to the accumulation of DNA molecules at barriers. Indeed, DNA molecules that were previously randomly diffusing across the surface, potentially long distances away from the barriers, are now all stopping their movement at the same location. Hence, this barrier can be used to concentrate DNA molecules and increase the amount of data points within one field of view. Originally, barriers were made by scratching the surface of the glass slide with a diamond pen. The dip on the surface prevents the bilayer from bridging on both sides, and hence it prevents the diffusion of lipids (Granéli et al., 2006). Subsequently, the barriers were refined by using electron-beam nanolithography (T. Fazio, Visnapuu, Wind, & Greene, 2008; Visnapuu, Fazio, Wind, & Greene, 2008b). In this technique, a
layer of acrylic is deposited on a fused silica slide and barrier patterns are imprinted in it using an electron microscope, similar to how a laser can be used to write on materials. The acrylic exposed to the electron beam is weakened and can be removed with a mild solvent, leaving exposed silica. Subsequently, a layer of metal is deposited by metal evaporation. The metal will coat the whole surface, but it will only bind to the silica where the patterns were previously written, as the rest of the silica is protected by acrylic. Finally, that acrylic layer, with the metal on top, is removed with a strong solvent, leaving only a layer of metal (in this case, chrome) in the shape of the written patterns on the silica slide. Hence, while the original method prevented lipid diffusion by a dip on the surface, this method prevents lipid diffusion by a bump on the surface, with the added advantage that the shape of the bump can be defined down to a 50 nm or less resolution (T. A. Fazio, Visnapuu, Greene, & Wind, 2009; Greene, Wind, Fazio, Gorman, & Visnapuu, 2010).

DNA molecules are attached to the lipids via a biotin ligated at one of their end. Because that end is always the same, all DNA molecules are in the same orientation, meaning that a given sequence on the DNA molecule will be at the same distance from the bilayer for all the DNA molecules. In the case of the lambda phage genomic DNA molecule, which is the one generally used for DNA Curtains, an oligonucleotide conjugated with a biotin is ligated to one of its cosite overhangs. That biotin can then bind to a streptavidin which, with 4 biotin binding sites, can also bind to a biotin on the lipids. Because biotin-streptavidin interactions have a close to covalent strength, the DNA molecules remain attached to the lipids for the whole length of the experiment.

In the presence of buffer flow, DNA molecules align against the plane of the objective, allowing for the visualization of their whole length (fig. 23). Indeed, once against the barriers, all of the hydrodynamic force from the flow is applied onto the length of the DNA molecule, instead of for moving it in the bilayer, leading to its extension with the flow. In the absence of buffer flow, a
Figure 23: Schematics of a DNA Curtains experiment. Bottom: Actual DNA Curtains experiment corresponding to the drawing. The DNA signal is labeled with the intercalating dye YoYoI (depicted in green). Nanopatterns of barriers and pedestals are deposited on a glass/quartz/fused silica slide. Subsequently, a lipids bilayer containing biotinylated lipids is formed on the slide, and DNA molecules are attached by a biotin-streptavidin-biotin link to the lipid bilayer. Note that since the biotin is always added on the same side of the DNA molecule, all DNA molecules are in the same orientation (denoted by the pink sequence on the molecule). Upon addition of buffer flow (blue arrow), the DNA molecules are pushed against the barriers and, since the hydrodynamic force cannot be used for movement, the DNA molecules are forced against the plane of the objective, allowing for the visualization of their whole length. In the presence of anti-digoxigenin on the pedestals, the DNA molecules can attach on them, leading to “double tethered” molecules that remain aligned with the plane of the objective even in the absence of buffer flow.

DNA molecules retracts, since no force stretches it, and it returns to Brownian diffusion. Because DNA Curtains are based on TIRF microscopy, in the absence of flow, a DNA molecule appears as a spot of relatively faint signal. Indeed, the lambda phage DNA molecule is about 16 μm long, and in the absence of buffer flow, it is allowed to explore the whole 3D space around it. Consequently, with an illumination depth of about 100 nm, most of the DNA sequence is not visible.
To observe the DNA along its length even in the absence of buffer flow, DNA Curtains use pedestals to double-tether the DNA (fig. 23). Pedestals are nanolithographed patterns strategically positioned three quarter of the length of the lambda phage DNA molecule away from the barriers (12 μm). Just like the barriers, pedestals are made of chrome, but they expose a larger surface on which proteins can readily adsorb. Anti-digoxigenin antibodies are adsorbed on the pedestals, and the second cos overhang of the lambda phage DNA molecule (the first one being attached to a biotinylated oligonucleotide) is ligated to an oligonucleotide conjugated with a digoxigenin. When buffer flow is applied on the DNA, its digoxigenin will hover the anti-digoxigenin-infused pedestal, and an interaction may happen. When the interaction happens, the DNA molecule is tethered on both ends: one to the lipid, and one to the pedestal. Hence, it cannot retract and remains visible along its length even in the absence of buffer flow. Because pedestals are located at 3/4 of the length of the DNA molecule, it is not stretched and can allow for conformational changes.

![Diagram of DNA Curtains flowcell assembly](image)

**Figure 24:** Schematics of the flowcell assembly for DNA Curtains. A fused silica slide with two drilled holes is nanolithographed. A flow chamber is carved in double-sided tape and placed around the holes on the slide. A coverslip is added to close the flow chamber and the slide is heated to melt the tape and seal the chamber. Finally, two ports are glued onto the holes to attach syringes and obtain a air-free flow chamber.

Finally, to be able to assemble the DNA Curtains experiment and apply buffer flow, DNA Curtains are assembled in flow chambers (fig. 24 and Ma, Steinfeld, and Greene, 2017). Two holes are drilled on a fused silica slide, on which the barriers and pedestals patterns are then nanolithographed. A flow chamber is carved in double-sided tape and placed around the holes on the slide. A coverslip is added to close the flow chamber and the slide is heated to melt the tape and
seal the chamber. Finally, two ports are glued onto the holes to attach syringes and obtain an air-free flow chamber. This construction allows to push buffer up to about 2.5 mL/min. Using the flow cell, a lipid bilayer can be deposited, followed by the incubation of anti-digoxigenin to adhere to the pedestal. The remaining exposed surfaces on the pedestals and slide are coated with bovine serum albumin (BSA), and streptavidin is incubated to bind the biotin on the lipids. The excess streptavidin is flushed out, and the lambda phage DNA functionalized with a biotin and a digoxigenin is added to the flow cell. Finally, if double-tethering is required, buffer flow is applied to stretch the DNA molecules so that their digoxigenin-labeled end hovers and attach to the pedestals. The DNA Curtains experiment can then be used to probe DNA-protein interactions. Notably, DNA Curtains have contributed to our understanding of CRISPR (Redding et al., 2015; Sternberg, Redding, Jinek, Greene, & Doudna, 2014), replication (Duzdevich et al., 2015), telomeres (Erdel et al., 01 03, 2017), nucleosomes (Visnapuu & Greene, 2009), DNA condensation (Terakawa et al., 2017), DNA repair (J. Y. Lee et al., 2015; Qi et al., 2015), and phase-transition in heterochromatin (Larson et al., 2017), among others.

Overall, DNA Curtains allow for the simultaneous visualization of hundreds of DNA molecules in real-time at the single-molecule level. They offer high throughput and bio-relevant surface passivation.
Chapter 1

Establishment of a protocol for the study of nucleosomes using DNA Curtains

1.1 Single-molecule studies of nucleosomes

Single-molecule studies have been linked with the nucleosome since its inception. Indeed, a large body of literature that led to the understanding of chromatin as a polymer of nucleosomes came from electron microscopy, with which single molecules of DNA and single nucleosomes were observed (section 1.1 and A. L. Olins and Olins, 1974; Oudet et al., 1975). Likewise, single-molecule studies of chromatin structure were performed with atomic force microscopy (AFM) (Allen et al., 1993).

After Widom’s work on nucleosome breathing (see section 1.10), much of the field turned towards single-molecule Forster energy resonance transfer (FRET) to directly demonstrate the transient release of DNA. The first observations of breathing with single-molecule FRET failed to account for the blinking of the acceptor dye. As a consequence, the authors overestimated the amount of breathing (Tomschik, Zheng, van Holde, Zlatanova, & Leuba, 2005). Subsequently, an extremely carefully controlled experiment was published by Koopmans, Brehm, Logie, Schmidt, and van Noort (2007). First, they noticed that while their reconstitution efficiency was north of
85%, they only observed FRET on about 10% of their surface-immobilized molecules. They narrowed this discrepancy down to the surface-induced dissociation of nucleosomes, although they passivated their surface with poly(ethylene glycol) (PEG), emphasizing the importance of surface passivation in single-molecule studies. Second, on top of pointing out the overestimation of the values from the previous study, they showed that 95% of all nucleosomes do not breath, with the remaining 5% in the same order of magnitude as what was published by G. Li et al. (2004), G. Li and Widom (2004). The results were then reproduced by Tomschik, van Holde, and Zlatanova (2009), in conjunction with a correction letter for their first article. Finally, S. Wei, Falk, Black, and Lee (2015) combined single-molecule FRET with advanced data analysis to show that nucleosomes breath at the millisecond timescale, an order of magnitude less than previously reported. Hence, single-molecule FRET has provided with the direct demonstration of nucleosome breathing, and its refinement to the millisecond resolution. Additionally, T. T. Ngo et al. (2015) used single-molecule FRET coupled with optical tweezers to demonstrate the effect of sequence flexibility on nucleosomes, and that asymmetric sequences can lead to nucleosomes being easier to evict from the stiffer side than the other, more flexible one.

Single-molecule experiments also provided a wealth of information on how nucleosomes are handled by other factors. Blosser, Yang, Stone, Narlikar, and Zhuang (2009), Deindl et al. (2013) used single-molecule FRET to show that nucleosome remodeling by ACF and ISWI happens in steps, a measurement that would have been hard to make with ensemble techniques, as the non-synchronous steps of each complex would have averaged out. Shundrovsky, Smith, Lis, Peterson, and Wang (2006) used optical tweezers to study nucleosome remodeling by Switch/Sucrose Non-Fermentable (SWI/SNF), obtaining an average of 28 bp displacement per remodelling event, with a catalytic efficiency of 0.4 (min nM)^{-1}. For the same remodeler, using the same tools, Yongli
Zhang et al. (2006), showed a nucleosome translocation rate of 13 bp/s, as well as the formation of loops by the remodeler, hinting at its intrinsic remodeling mechanisms. All of these measurements demonstrated that chromatin remodelers are bidirectional. In a similar fashion, Mack, Schlingman, Ilagan, Regan, and Mochrie (2012) used optical tweezers to study SIN (SWI/SNF-independent) mutant nucleosomes. The SIN group of nucleosome mutations, such as the H4 R45H they studied, was shown to rescue the depletion of SWI/SNF in *S. cerevisiae* (Kruger et al., 1995; Muthurajan et al., 2004). Hence, by measuring the binding strength difference between wild-type (WT) and SIN mutant nucleosomes, they were able to estimate the transcriptional barrier of nucleosomes, which is normally relieved by SWI/SNF, to $5\ k_B T$. High resolution single-molecule force and FRET measurements also measured the steps in nucleosome dissociation (Gansen et al., 2009; Hall et al., 2009), which suggest how an enzyme such as an RNA polymerase would traverse a nucleosome. In addition, Hodges, Bintu, Lubkowska, Kashlev, and Bustamante (2009) directly measured transcription by Pol II through a nucleosome at the single-molecule level using optical tweezers, showing how nucleosomes generally inhibit Pol II and that the core histones, instead of being evicted by the polymerase, probably are transferred behind it. Such loop-mediated recapture was also suggested in the absence of any polymerase, simply by unzipping the DNA using optical tweezers (Brennan, Forties, Patel, & Wang, 2016).

Single-molecule techniques have also been used to study chromatin. Chromatin assembly and disassembly in *X. laevis* egg extract was probed using optical tweezers, demonstrating the importance of ATP in chromatin disassembly, as well as the forces involved in assembly (Bennink et al., 2001; Yan et al., 2007). The same tools were used to determine the forces involved in nucleosome-nucleosome interactions in chromatin (Cui & Bustamante, 2000). Kruithof et al. (2009) also used single-molecule force measurements to show that the 30 nm fiber adopts a solenoid conformation.
Kilic et al. (2018) used single-molecule FRET to show the dynamics of nucleosome interactions in chromatin at the micro-second scale. Interestingly, the heterochromatin protein HP1α stabilizes these normally short-lived interactions, but is not enough to prevent transcription factor binding.

Other single-molecule techniques, such as cryogenic electron microscopy (cryoEM), have been used to study nucleosomes (Wilson & Costa, 2017, Pt 6). In-depth reviews of the studies of nucleosomes at the single-molecule level have been written (Buning & van Noort, 2010; Killian, Li, Sheinin, & Wang, 2012; Ordu, Lusser, & Dekker, 2016). Overall, single-molecule studies of nucleosomes have refined and expanded our understanding of their mechanisms, dynamics and interactions with DNA-based processes. Many of these observations, such as the underlying inhibitory dynamics of nucleosomes, would have been difficult or impossible to probe with ensemble techniques.

1.2 Study of nucleosomes using DNA Curtains

DNA Curtains have the potential to address previously unexplored questions about nucleosome biology. While DNA Curtains do not have the structural and biophysical resolution of FRET, tweezers, or cryoEM, they are not limited to single nucleosomes and are especially adapted to resolve the outcome of interactions between nucleosomes and other factors.

Because DNA Curtains allow for the visualization of long DNA molecules in all their length, they are especially suited for the study of several nucleosomes. For example, DNA Curtains make it easy to measure the number of nucleosomes per DNA molecules, their position on the sequence, their number of subunits, and their half-life, as well as to measure the relations between these observations. Interesting studies of the effect of post-translational modifications on motility, stability
and sequence preference of nucleosomes could be conceived. In addition, the effect of histone chaperones or chromatin remodelers on these parameters could be investigated. For example, it is still relatively unclear how chromatin remodelers evenly space nucleosomes. Nucleosome Curtains could be used to directly observe how remodelers move nucleosomes. Chromatin Curtains, made with domains of tightly interspaced chromatin or with fully chromatinized DNA, could also be established.

Virtually any DNA-based process can be studied with the presence of nucleosomes. It is generally unclear how nucleosomes affect DNA-based processes, and how DNA-based processes handle nucleosomes. A key remaining question in all DNA-based processes is: what happens to the nucleosome when it is displaced for DNA access, for example, during replication, transcription, and repair. While the question has been extensively studied in the context of transcription, it remains enigmatic (Teves, Weber, & Henikoff, 2014). The study of such collisions, with one or several nucleosomes, is feasible using Nucleosome Curtains, and would allow to track the outcome for both the polymerase/helicase/etc. and the nucleosome in real-time, at the single-molecule level, whereas these studies usually rely on ensemble measurements and gels. Due to their flexibility in the density of nucleosomes, Nucleosome Curtains could also be used to study the preferential binding of proteins to DNA versus nucleosomes/chromatin. In short, there is an immense quantity of questions that can be addressed using Nucleosome Curtains.

Visnapuu and Greene (2009) have previously studied nucleosomes using DNA Curtains. In addition, J. Y. Lee and Greene (2011) have published a protocol for the study of nucleosomes using DNA Curtains. Visnapuu and Greene (2009) have studied nucleosomes using FLAG-tagged histones. The histone octamers with a FLAG-tag on one of the four core histones were first reconstituted into nucleosomes on λ phage genomic DNA, assembled into a DNA Curtains experiment,
and then labeled in situ with a quantum dot-labeled anti-FLAG antibody ("SiteClick Qdot 705 Anti- 
body Labeling Kit - Thermo Fisher Scientific," n.d.). The excess fluorescent labels were flushed 
out after the reaction. Using this protocol, they were able to determine the preference of nucle-
osomes for GC-rich sequences, as well as to compare the accuracy of algorithms predicting the 
sequence preference of nucleosomes.

I first replicated the same method. However, several issues became apparent using this pro-
tocol. First, the in situ labeling with quantum dots leads to a poor labeling efficiency. Indeed, 
nucleosome concentration in a DNA Curtains experiment is remarkably low. For example, if 100 
DNA molecules are present in a field of view, this would lead to a DNA concentration of about 
30 fM. Considering 5 nucleosomes per DNA molecule, with two tags per nucleosomes, this is 
in the order of 300 fM of reagent. Likewise, the concentration of label has to be limited to tens 
of nanomolars, because high concentrations of free quantum dot labels lead to non-reversible in-
teractions with the surface and results in poor signal-to-noise ratio. Overall, this leads to limited 
labeling efficiency. Second, interactions between FLAG-tags and antibodies have a limited half-
life of about 45 minutes (S. Redding, personal communication, and own measurement, data not 
shown). As a consequence, some experiments requiring longer incubation periods will result in 
low amounts of labeled nucleosomes, both because the labeling efficiency was initially low and 
because the fluorophores will have released from the histones. Finally, quantum dot labels are big. 
The diameter of a quantum dot is estimated to be around 20 nm, twice that of a nucleosome. In 
addition, antibodies have a length of about 15 nm. Hence, the total size of the label is up to 35 nm, 
more that three times the size of a nucleosome. The addition of a quantum dot in the initial steps 
of the assembly of the replication complex inhibits it (D. Duzdevich, personal communication). 
Hence, it is likely that the addition of a quantum dot on a nucleosome would hamper processes that
interact with it. Most notably, a complex like condensin or cohesin may not be able to slide through a nucleosome in the presence of a quantum dot (Stigler, Çamdere, Koshland, & Greene, 2016).

To correct these drawbacks, I developed a protocol using organic fluorescent dyes instead of quantum dots. Organic dyes are small molecules of less than 1 kDa that can be conjugated covalently to proteins. Unlike quantum dots, which display fast blinking (Nirmal et al., 1996) and no photobleaching, organic dyes blink and photobleach, but these issues can be mitigated using specific chemicals (Dave, Terry, Munro, & Blanchard, 2009; Rasnik et al., 2006). Albeit the limitations of organic dyes as fluorophores, I decided to attach dyes covalently to the nucleosomes for the following reasons:

• Virtually infinite half-life of the interaction, so that the fluorophore will not be lost before or during the experiment, except for photobleaching.

• As a consequence of the long half-life, the excess fluorophore can be removed easily before the experiment without compromising the labeling efficiency.

• A bulky interaction module is not necessary, making the dimension of the labeled nucleosomes closer to wild-type.

• Photobleaching can be used to quantify the number of subunits in a nucleosome.

Many labeling techniques are available (Sahoo, 2012). I opted for maleimide conjugation because of its efficiency, specificity and widespread use. Indeed, histones contain many lysines, which would lead to several labels per histone using NHS-ester conjugation, an undesired outcome. Maleimide is specific to cysteine residues, and because yeast histones do not contain cysteines, it is easy to mutagenize them to replace a serine or threonine with a cysteine, which have a similar size and polarity. Also, just like NHS-ester conjugation, maleimide conjugation leads to good
labeling efficiency on solvent-exposed residues. Finally, many groups have tested and published
cysteine-mutant histones, making it easy to select labeling sites without having to heavily control
the consequences of the mutation.

It is common to use histones from *X. laevis* to assemble nucleosomes, because they have been
extensively characterized, crystallized and controlled (Luger et al., 1997a; Luger et al., 1997b).
However, I chose to use histones from *S. cerevisiae*. Histones are extremely conserved across
eukaryotes, and it is widely accepted that the few sequence differences are unlikely to result in
strong biochemical differences. As a consequence, in vitro, *X. laevis* nucleosomes are readily
mixed with human or yeast proteins. However, *S. cerevisiae* histones are the most divergent across
eukaryotes (see section 1.6 and White et al., 2001), and they have been shown to be less stable than
their metazoan counterparts (K. Tóth et al., 2013). Since the budding yeast is a unicellular organism,
with most of its genome expressed, it seems logical that its nucleosomes would intrinsically provide
a more open chromatin structure. In addition, *S. cerevisiae* only has three histone variants: H2A.Z,
CENP-A and H1. As a consequence, *S. cerevisiae* histones may need to wear several hats. For
example, *S. cerevisiae* H2A contains the tail sequence of H2A.X that is phosphorylated during DNA
repair. Moreover, most of the proteins we work with in the laboratory are from *S. cerevisiae*, and
since the processes we study imply the direct handling of chromatin, it is possible that these proteins
have evolved specificities for *S. cerevisiae* histones. Consequently, I decided to use histones from
*S. cerevisiae*.

This train of thoughts was confirmed by two studies published after I had started the project.
In a first study, it was shown that some of the divergent amino acids of *S. cerevisiae* histone H3 are
necessary for normal growth rates (McBurney et al., 2016). These experiments confirmed that the
divergence of *S. cerevisiae* histones provides a function and hints that histones that do not share
these divergences are improper replacements. In a second study, all core histone genes (but not the variants) from *S. cerevisiae* were replaced with *H. sapiens* histones (Truong & Boeke, 2017). Out of $10^7$ cells, 8 ended up forming colonies, only starting to appear on plates after 20 days. They presented severe growth defect and aneuploidy. Interestingly, the “humanized” yeasts were also slower to adapt to new environments, like switching from glucose to galactose, indicating slower activation of their transcriptional programs, due to the less permissive nature of human histones. This report confirms that *H. sapiens* histones, and probably all others too, are a poor replacement for *S. cerevisiae* histones. Of note, about 50% of human proteins can replace their *S. cerevisiae* homolog (Kachroo et al., 2015), confirming that this assay is relevant.

While the data presented below was collected with *S. cerevisiae* histones, the protocol can easily be adapted to histones from other species, as the key complications from the usage of histones on DNA Curtains are independent from the species. This method has already been used in projects involving the study of cohesin (Stigler et al., 2016), condensin (see chapter 2), DNA repair (Xue, C et al., in preparation), replication (data not shown), and for developing FRET on DNA Curtains (J. Stigler, personal communication). It is easily adaptable and provides with a solid platform for the study of nucleosomes and chromatin using DNA Curtains.

### 1.3 Selection of cysteine mutation sites for labeling

To label nucleosomes fluorescently, I first needed to append a dye on the histones. To do so, I selected potential amino acids to mutate into cysteines. In order to observe every individual histone in the nucleosome, I selected a site in each of the four core histones, for example to measure H2A-H2B dimer versus H3-H4 tetramer half-lives or fates after collisions.
My first strategy was to use mutations previously characterized in the literature. Most work on dye-labeled cysteine mutant nucleosomes was done for FRET using *X. laevis* histones. Hence, I aligned the sequences of *X. laevis* and *S. cerevisiae* histones using Clustal Omega (Sievers et al., 2011) to be able to mutate the equivalent amino acids. I selected four cysteine mutants, one for each histone:

- H2A S47C, the equivalent of *X. laevis* H2A A45C published by Kassabov, Henry, Zofall, Tsukiyama, and Bartholomew (2002)
- H3 V36C, from H3 V35C in *X. laevis* (Zheng et al., 2005)
- H4 S48C, as published by Brogaard, Xi, Wang, and Widom (2012). Note that the authors call it ”H4 S47C” because they ignore the first methionine, which is usually removed in *S. cerevisiae*, while I consider the whole genetic sequence.

The positions of the mutated sites are shown in fig. 1.2.

While I ended up using these mutants (except for H3 V36C, which never expressed), I also designed another set of mutants at a later stage of the project. Indeed, the position of these cysteines is optimized to be close to the DNA to allow for FRET or hydroxyl radical footprinting, which is not to our interest. My second selection of residues was based on the following criteria:

- Solvent exposure. For proper labeling, it is imperative that the residue be easily accessible from the solvent and not buried within the structure, like H4 S48 is (see fig. 1.2).
- No known growth defect when mutated, as probed by the HistoneHits database (Huang et al., 2009).
- No to minimal phenotype when mutated, as probed by the HistoneHits database. In particular,
no DNA repair or silencing defect.

- Not too close to a residue that would cause defects when mutated. Indeed, the fluorophore may cause steric hindrance or charge repulsion to the factors binding or editing the nearby amino acids.
- Ideally, not a conserved residue, the rationale being that conserved residues are conserved for a reason (although divergent residues can also be divergent for a reason, as observed by McBurney et al. (2016)).
- No known importance for the function of the cell or of the nucleosome in the literature.

As it turns out, these criteria are pretty stringent. Histones have been highly optimized through evolution—probably because at their current size of about 130 amino acids (a.a.) each, they already more or less equal DNA in mass, making it advantageous to get rid of the superfluous amino acids—and most mutations lead to a phenotype, especially in H3 and H4. I selected one amino acid from the tail and one amino acid from the core domains for each histone:

- H2A S11C and T25C
- H2B S27C and S126C or S127C
- H3 S22C and S136C
- H4 S64C (no suitable tail residue could be found)

Looking back, I think the easiest, safest and most efficient way to label histones would be to add a cysteine at the beginning of the N-terminal tail, right after the methionine. Indeed, N-terminal tails are unstructured and exposed to solvent, and by adding the cysteine instead of mutating another residue, there is less risk of disrupting an existing function or interaction.

I have mutated all four histones from the initial site selection using Inverse PCR mutagenesis
Purification of recombinant histone octamers

Using the mutated plasmids, I proceeded to purify the recombinant *S. cerevisiae* histones in *E. coli* following the protocols from J. Y. Lee and Greene (2011), Luger et al. (1997b).

These protocols are based on three key properties of the histones: (i) their intrinsic hydrophobicity that causes them to aggregate, (ii) their high amount of positively charged amino acids, and (iii) the ability of the histone-fold domains to refold reproducibly from a denatured state into an octamer. The protocol has three main steps: the purification of histones, the purification of histone octamers, and the assembly of nucleosomes. In the case of dye-labeled nucleosomes, an extra labeling step is inserted before the nucleosome assembly (see fig. 1.1).

First, histones are purified. They are expressed individually in *E. coli*, in which they will form aggregates. Because histones contain rare codons, an *E. coli* strain containing a pRARE plasmid, such as BL21 Rosetta cells, is recommended. The aggregates are isolated by centrifugation from the supernatant and solubilized in urea, a non-charged chaotropic agent. This provides with a first layer of purification. Next, the denatured histones are purified on a cation-exchange chromatography column. Their high basicity, matched by few proteins, provides another layer of purification. To help with purification, an anion-exchange column is placed before the cation exchange column, and removed before elution. Typical elution profile and gels of fractions for all four histones are shown in fig. 1.3. Note that the presence of cysteines does not affect the elution profile. Finally, the individual histones are lyophilized and the resulting powder is stored at -20°C.

Second, histone octamers are assembled and purified. Powder from each histone (H2A, H2B,
H3, H4) is denatured in a strong chaotropic agent (usually guanidium). An equimolar ratio of each denatured histone is subsequently mixed and slowly dialyzed against 2M monovalent salt (usually NaCl or KCl), leading to the slow refolding of the histone octamer. The high salt concentration helps maintain the histones as one octameric complex (Eickbush & Moudrianakis, 1978). Finally, the octamers are separated from aggregates, dimers, and tetramers by size exclusion chromatography (fig. 1.4). This purification step helps get rid of aggregates and ensures the proper stoechiometry of histones. Indeed, if all octamers, tetramers and dimers are kept together, it is likely, due to aggregation during refolding, that their ratio will not remain equimolar. Hence, one could obtain more tetramers than dimer, which would not lead to the formation of proper nucleosomes. Consequently, either the dimers and tetramers are purified individually, either the octamers are purified as a whole. Note that an octamer purification will yield dimers and tetramers, too, which could be used for reconstitution of nucleosomes. Overall, I have purified 6 different octamers construct: WT, H2A-3xFLAG, H2B-3xFLAG (both as published by Visnapuu and Greene (2009)), H2A S47C, H2B T119C, and H4 S48C (fig. 1.5).

If cysteine-mutant histone octamers were assembled, they can now be labeled. I have been labeling whole octamers, rather than single histones, because it is more convenient to try different dyes and optimize the labeling reaction. Indeed, one octamer preparation can be splitted into smaller aliquots and labeled, which is easier than running several octamer assembly reactions in parallel. In addition, this strategy helps decrease the time handling the dye-labeled octamer, and consequently decreases exposure to light and photobleaching of the stock. However, it is also possible to first label the histones, and then reconstitute the octamers. This second strategy has the added benefit that several dyes can be assembled in the same octamer, for example to label dimers and tetramers, and that the excess dye can be separated from the octamers during size-exclusion chromatography,
hence hitting two birds out of one stone. To label the cysteine-mutants, they are first incubated with tris(2-carboxyethyl)phosphine (TCEP) to break the disulfide bonds. The TCEP and other reducing agents in the buffer are then removed—TCEP was shown to inhibit labeling reactions (Tyagarajan, Pretzer, & Wiktorowicz, 2003)—and the protein is incubated with the maleimide-conjugated dye at a 5:1 to 10:1 maleimide:cysteine molar ratio. Once the labeling has proceeded, the excess dye is removed. All clean-up steps are achieved via spin-concentration, by rinsing 5+ times with buffer (Nanda & Lorsch, 2014), but other methods, such as size exclusion chromatography, may result in better separation. Overall, the reactions show great specificity (fig. 1.6). In general, H4 S48C octamers label less well than its dimer counterparts, and H2B T119C octamers seem to label marginally better than H2A S47C. This can be explained by their relative solvent exposure, with H2B being the most exposed and H4 the least (fig. 1.2).

1.5 Purification of λ DNA

The genome of the bacteriophage Lambda measures 48.5 kbp, and can readily be obtained commercially or by infecting E. coli with the phages. In addition, λ phage DNA is packaged in its linear form by the phage, leaving two 12 bp cohesive ends that are useful for adding specific groups, such as biotin, digoxigenin, or fluorescent dyes, via functionalized oligonucleotide ligation.

To assemble λ for DNA Curtains experiments, a biotin is added on one side, to attach the molecule to the lipids surface, and a digoxigenin is added on the other side, to interact with the anti-digoxigenin antibody on the nanolithographed pedestals. This is achieve by the ligation of functionalized oligonucleotides in a large excess to λ DNA (about 1:100 λ:oligonucleotide, molar ratio). To remove the excess of functionalized oligonucleotides from the DNA, a previous purifi-
cation protocol of λ DNA used size-exclusion chromatography. However, this protocol was slow and labor intensive, as the media bed often needed to be repacked. In addition, the resulting DNA was heavily diluted and its integrity was often compromised.

To correct these flaws, I established a simple purification protocol based on the PEG precipitation of DNA (Lis & Schleif, 1975). Simply, because the oligonucleotides are very short (12 bp) and the λ phage genomic DNA is very long (48.5 kbp), they can easily be separated by size-specific PEG DNA precipitation. The λ DNA is pelleted by centrifugation, while the short oligonucleotides do not precipitate. The pellet is subsequently washed and resuspended in the desired volume of buffer. The precipitation efficiency is close to 100%, hence, the DNA can be resuspended to the approximate desired concentration (see fig. 1.7 for a representative purification). The whole purification protocol can happen in less than a day.

1.6 Reconstitution of nucleosomes

To assemble nucleosomes, histone octamers (or a 2 to 1 ratio of H2A-H2B dimers and H3-H4 tetramers) are mixed with DNA in high salt, and the salt is slowly dialyzed away. The ratio of octamers to DNA can be modulated to provide the proper amount of nucleosomes per DNA molecule. However, as with all reactions, the assembly efficiency is not absolute, and a higher molar ratio of octamers to DNA is generally advisable to obtain the required amount of nucleosomes. It appears that the reaction assembly efficiency saturates around 1 μM of histone octamers (fig. 1.14). Hence, at 1 μM, a 1:1 molar ratio of octamer:DNA should give close to 1 nucleosome per DNA molecule, ignoring the imprecision in the concentration measurements. However, below 1 μM, a ratio higher than 1:1 will be required to achieve 1 nucleosome per DNA molecule.
Several methods can be used to dialyze salt, resulting in different dynamics of salt-concentration over time. In the linear method, salt-free buffer is added continuously as buffer is removed from the dialysis vessel, leading to a linearly decreasing salt gradient, with a constant volume of buffer. In the step gradient method, dialysis is performed in successive incubations with buffers of decreasing salt concentrations. This leads to a smoothed step-function salt dialysis, similar to a series of stacked exponential decays. Finally, in the exponential decay method, salt-free buffer is added continuously, but unlike the linear method, no buffer is removed from the dialysis vessel, leading to a gradually slowing salt dialysis. This method has the advantage of simplicity and of giving more time for histones and DNA to equilibrate at lower salt concentrations, where dynamics are expected to be slower. I have not noticed a difference between the three methods. It is also possible to reconstitute nucleosomes using the chaperone Nap1 (Lusser & Kadonaga, 2004). This strategy works equally well for Nucleosome Curtains.

It was proposed that the addition of glycerol and surfactant to the assembly buffer helps make the reaction reversible (Thastrom et al., 2004). I have tried a buffer similar to that published (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0, 10% glycerol, 0.1% NP-40) and found no difference with regular buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0). Hence, for the sake of simplicity, I do not recommend this buffer. Its main drawback is that permanent marker dissolves in it, leading to the loss of all labels on reconstitution reaction tubes.

Overall, all existing nucleosome reconstitution methods are proper for Nucleosome Curtains.
1.7 Selection of organic dyes for labeling nucleosomes

Before this project, organic fluorophores have not been extensively used with DNA Curtains. The first step to observe dye-labeled nucleosomes on DNA Curtains was hence to select an organic fluorophore.

Organic fluorophores differ widely in their qualities. Key qualities for DNA Curtains experiments are brightness and photostability. In addition, they must not impede the formation of nucleosomes and must be available to order with a maleimide tag. I first screened for their fluorescent properties by labeling octamers, assembling Nucleosome Curtains, and observing the behavior of the dye.

Because we have lasers emitting at 488 nm, 532 nm, 565 nm and 647 nm, I screened for the following dyes in the matching channels: Atto-488, Alexa-488, Atto-532, Dyomics-530, Dyomics-549P1, Atto-565, Dyomics-634, Alexa-647, and Atto-647N. Overall, it appeared that the longer the wavelength, the more stable the dye. By far, fluorophores absorbing in the 647 nm range outcompete those in the 488 nm range when it comes to photostability. Indeed, the 488 dyes, of which Alexa-488 was the best in my hands, would photobleach after a few seconds. In comparison, Atto-647N can be observed for minutes before the signal is lost. As such, I have not used 488 nm dyes for this protocol. Out of the dyes screened, I recommend using, in order, Atto-647N, Atto-565, Dyomics-530, and Alexa-488 (if no other option is available).
1.8 **Microccocal nuclease (MNase) treatment of nucleosomal DNA gives expected protection**

*Microccocal nuclease (MNase)* was used early on to define how much DNA histones protect and to demonstrate the polymeric nature of chromatin (Noll, 1974). *MNase* is a weak endonuclease with slight sequence preference for the 5’ end of A/T base pairs (Nikitina et al., 2013; Pauli, Seebeck, & Braun, 1982). *MNase* does not readily cut the DNA in contact with the histone octamer, but can cut the linker DNA around the histones. Using *MNase* digestion of chromatin, it was shown that nucleosomes protect about 150bp of DNA (this number varies between authors and with time, see Nikitina et al. (2013), Noll (1974)), while non-nucleosomal histones and assembly intermediates do not. For example, the H3-H4 tetramer protects 73bp (Dong & van Holde, 1991), and the hexasome protects 110bp (Arimura, Tachiwana, Oda, Sato, & Kurumizaka, 2012). I am not aware of the H2A-H2B dimers giving any protection on their own. At lower concentrations of *MNase*, not all linkers are cut, leading to a ladder of protected sequences. Depending on the length of the linker DNA, protected fragments are observed every multiple of 150bp to 200bp (Noll, 1974).

*MNase* has become the gold standard for determining the structure of nucleosomes. Consequently, I proceeded to test my nucleosomes reconstitutions by *MNase* digestion. First, I titrated nucleosomes and *MNase* to test for ideal conditions (fig. 1.8). As can be observed, laddering can be reproduced at higher nucleosome to DNA ratios. In the absence of histones, the λ DNA is fully digested to fragments of less than 100 bp, indicating that the protection comes from the octamers, and not from the DNA itself. Next, I tested the nucleosomes assembled with cysteine-mutant octamers, but unlabeled, to confirm that the mutations alone did not lead to any defect. As expected,
the mutant octamers provided the same amount of protection as the WT ones, around 150 bp, confirming that they are nucleosomes, and not tetramers or hexasomes (fig. 1.9). Finally, I tested the labeled nucleosomes assembled for Nucleosome Curtains with all three cysteine-mutant octamers labeled with Dyomics 530 (fig. 1.10). As expected, they also protected about 147bp of DNA, confirming that they are full nucleosomes and that the dye does not cause defects. While the intensity of the protected band varies, this intensity is true for the whole lane, with the sub-100 bp degradation product varying in the same amount. Hence, the difference is probably due to the variability in the amount of nucleosomal DNA used, and not to a defect in the nucleosomes themselves. It is possible that some constructs are most likely to aggregate during reconstitution, leading to a lower final amount of DNA. The fact that bands of lower intensity are also at lower positions is a consequence of the DNA to MNase ratio: higher concentrations of MNase will have more opportunities to cut DNA within the nucleosome (fig. 1.8). To confirm the absence of defect from dye-labeled nucleosomes, the MNase digestion was performed at several reconstitution ratios of DNA:octamer with H2A S47C DY530 (fig. 2.15). As can be observed, protection is similar to WT, confirming that dye-labeled octamers form proper nucleosomes and that the addition of a dye does not affect the nucleosome.

1.9 **Electrophoretic mobility shift assay (EMSA) of nucleosomes**

A second method to test the formation of nucleosomes is by EMSA. Simply, nucleosomes are reconstituted on a short DNA fragment (usually 147 bp) that can be dye-labeled (see appendix E). The nucleosome core particle (NCP) are then resolved on a 5% polyacrylimide native gel and imaged for DNA and protein content. Most often, the Widom 601 sequence (see introduction
section 1.9) is used for sequence, because of its high binding affinity.

I have purified 147bp 601 sequences, with or without a dye (fig. 1.11). Subsequently, I have reconstituted nucleosomes with them and resolved them by EMSA. As can be seen in fig. 1.12, NCPs are reconstituted and shift at the same size as MNase products, mutually confirming that they are nucleosomes. In addition, it can be observed that several bands appear at the 2:1 DNA:octamer molar ratio, and that in all cases, not all the DNA is formed into nucleosomes, indicating that the reconstitution reaction is not fully efficient.

At first, I thought that the several shifted bands observed at the 2:1 ratio and, to a lesser extent, in the 1:1 ratio were shifted nucleosomes that were not perfectly centered on the 147 bp sequence. However, heat-shifting them as recommended in the literature (Luger, Rechsteiner, & Richmond, 1999) turned out to be ineffective. Deeper inquiry lead me to realize, with the help of F. Erdel and D. Duzdevich, that these products are not nucleosomes but non-nucleosomal histones. Indeed, at the concentrations used, not all histones become nucleosomes (figs. 1.12 and 1.14 and text below). These unbound histones weakly interact with the DNA in a non-nucleosomal manner. They can easily be chased off by competitor DNA (fig. 1.13). In the absence of competitor DNA, bands other than the octamer form, but in the presence of excess large fragments of competitor DNA, they disappear, indicating that they are not nucleosomes, but histones binding to the DNA non-specifically. Using labeled histones, it can be shown that these bands correspond to H2A-H2B dimers or H3-H4 tetramers, but not octamers (F. Erdel, personal communication). Hence, non-nucleosomal histone subspecies interact weakly with the DNA, shifting it, but these non-nucleosomal histones can be absorbed by competitor DNA.

To address the efficiency of the reconstitution reactions, I first titrate the concentration of histone octamers and DNA at a 1:1 ratio (fig. 1.14). As can be seen, the reaction appears to saturate
around 1 μM of reagents. This titration indicates that at the concentrations used for Nucleosome Curtains, most of the histones will not become nucleosomes. Shifting the concentration of λ DNA to the right range to obtain the desired low number of nucleosomes per DNA molecules and 100% reconstitution efficiency is not feasible due to the high viscosity of λ DNA. In addition to μM range concentrations, the ratio of octamers to DNA has to be slightly shifted to accommodate for the concentration measurement error (fig. 1.15).

EMSA of dye-labeled histone octamers also demonstrate that dye-labeled octamers are just as efficient at forming nucleosomes, and that the non-nucleosomal histones, but not the nucleosomal histones, bind the excess competitor DNA (fig. 1.16). This result indicates that in the presence of high concentrations of competitor DNA, nucleosomes remain stable, but non-nucleosomal histones rapidly exchange.

Overall, the EMSAs have demonstrated that the reconstitution reactions form proper nucleosomes, that the nucleosome reaction reaches maximum efficiency around 1 μM of histone octamers, that dye-labeled and WT histone octamers both form nucleosomes, and that non-nucleosomal histone subspecies bind to free DNA and can easily be removed from the DNA of interest with competitor DNA.

1.10 Assembly of Nucleosome Curtains

Assembly of a Nucleosome Curtains is in all steps similar to the assembly of a DNA Curtains (Greene et al., 2010; Ma et al., 2017), except that the DNA cannot be incubated in the flow cell in the presence of salts (see section 1.11).

First, a flow cell is prepared. Two holes are drilled into a fused silica slide. That slide is cleaned
and nanopatterns are lithographed on it using electron beam lithography and chrome deposition (see section 4 and T. Fazio et al., 2008). The patterned slide is then cleaned and assembled into a flow cell by forming a channel between the two holes with double-sided tape. The tape is sandwiched between the slide and a coverslip and melted at 180°C to seal the flow chamber. Finally, ports are mounted on the holes using hot glue to allow for the airtight connection of tubing.

Second, the DNA Curtains experiment is prepared. The flow cell is connected to syringe ports and then air is removed from the flow chamber with water. The flow cell is rinsed with buffer and a lipid bilayer is deposited on the silica surface. This lipid bilayer contains biotinylated lipids, allowing for the binding of DNA via biotin-streptavidin-biotin interactions. If tethering of the DNA molecules at both ends (“double-tethering”) is desired, anti-digoxigenin antibody is incubated, to subsequently interact with the digoxigenin at the end of the DNA molecules. The antibody adsorbs on the exposed chrome surface, on which lipids cannot form a membrane. The flow cell is then passivated with buffer containing bovine serum albumin (BSA) and streptavidin is added to bind the biotinylated lipids. Finally, the reconstituted nucleosomal DNA is added.

Last, the DNA Curtains experiment is transferred to a total internal reflection fluorescence (TIRF) microscope with a syringe pump. Buffer is slowly injected through the chamber to concentrate DNA molecules at the barriers, and the flow cell is incubated in the absence of flow for the DNA molecules to spread against the barriers by diffusion. If double-tethering is desired, the DNA is pushed at higher buffer flow so that its digoxigenin-labeled end aligns with the pedestals. To remove the non-nucleosomal histones, salt buffer is injected with a fast flow rate, to minimize the probability that the histones interact with the surface of the flow cell, or salmon sperm DNA is injected over several milliliters of buffer (see section 1.11). While salmon sperm will lead to less histone binding to the surfaces, it may also itself interact with the surface. This can be an issue, as
it will titrate DNA-binding proteins off the λ DNA, and if these proteins are labeled, it will result in high background noise and low signal-to-noise ratio. Finally, the desired proteins or measurements can be made. Representative examples of Nucleosome Curtains for all three mutants are shown in fig. 1.17.

1.11 Histones readily adhere to glass surfaces in the presence of salts

The first thing I observed in Nucleosome Curtains assays was the amount of noise on the surface of the flow cell (fig. 1.18, left panel). It appeared to me that the fluorophores may interact with the lipid bilayer (L. D. Hughes, Rawle, & Boxer, 2014). However, adding each of the organic dyes in the absence of histones to the DNA Curtains experiment did not result in significant binding. Atto-647N exhibited the highest surface binding, but it was still less than in the presence of histones. However, adding dye-labeled octamers in the flow cell led to consequent surface binding, indicating that the histones, rather than the dyes, were responsible for the surface interactions. This can be explained by the fact that histones are extremely positively charged, and glass is negatively charged. In addition, it is possible that the hydrophobic core helps the histones solubilize in the lipids bilayer to reach the glass surface. The presence of histones on the surface of the flowcell also reduces the fluidity of the membrane, resulting in the poor quality of the DNA Curtains experiment. Hence, photobleaching the surface before an experiment, which would remove all of the background signal, would increase the signal-to-noise ratio but would not rescue the quality of the surface, and as such, did not turn out to be a sufficient solution. To mitigate the surface binding, I tried to increase the
concentration of BSA in the flow cell, to compete for the non-specific binding sites, to no avail. Next, I tried to add salt, competitor DNA, or surfactants in the flow cell, such as NP-40, which did not help neither. I figured that maybe chaperones would help prevent histone binding to the surface. However, adding an excess of NAP-1 did not help. I next hypothesized that the best method to prevent the histones from binding to the surface was to prevent them from ever entering the flow cell. Hence, I tried to separate the free histones from the nucleosomes. This solution was tricky, because very small amounts (30 μL) of material are reconstituted at once, and their shelf life is limited to about a week, making the purification of big batches costly in time and material. I tried to PEG-precipitate the DNA and to resuspend it afterwards in a histone-free buffer. It helped, but not sufficiently. I next tried to separate the DNA from the free histones on a Superose 6 size-exclusion column. Unexpectedly, the column did not succeed at separating the lambda DNA, which eluted in all fractions. I suspect that the nucleosomes may have interacted with the media, preventing proper migration and separation of the DNA. Finally, I tried to run the sample on a DEAE column, a cation-exchanger. The hypothesis was that free histones, which are heavily charged, would interact with the media and get stuck, while nucleosomal DNA, which is heavily negatively charged, would end up in the void, or elute more readily than free histones. While this solution first gave promising results, it ultimately failed, as it was not reproducible. I suspect that the histones may have bound to the column and saturated it over time, somehow never eluting.

In the meantime, my colleague Daniel Duzdevich tried to absorb the free histones using glass beads, figuring that since the histones bind to the glass, one may as well adsorb them all on glass first, and then collect the cleaned up nucleosomal DNA. Since we also observe DNA sticking to the glass surface, and that techniques such as DNA combing are entirely based on DNA sticking to glass, I assumed this solution would not work. As it turns out, the glass beads did not help
much, although they did help sufficiently to make the Nucleosome Curtains experiment with low nucleosome concentrations usable. However, what was remarkable was that the glass beads would adsorb more free histones in the presence of salt than in the absence of salt. At first, this defied our understanding of biochemistry. The higher the salt, the less likely the interactions, as salts compete with the histones for the glass surface. However, I ultimately resolved the problem with the following hypothesis: histone are polycations and DNA is a polyanion. In the absence of salts, free histones readily interact with the DNA through weak polar interactions, and nothing but other histones compete for the same sites. As a consequence, if the number of DNA binding sites is superior or equal to the number of histones, all histones will be bound to the DNA at equilibrium. It is expected that the off-rate of such interactions is very low, because nothing competes for the same sites, and histones contain many lysines and histidines, making it unlikely that all the contacts would release at once. However, upon addition of salt, histones and cations compete for the negative charges on the DNA. Hence, the off-rate of histones on DNA is going to increase, while their on-rate is going to decrease. In the presence of a glass flow cell, or of glass beads, histones that are freely exchanging between DNA and solution may interact with the glass. Once the histone binds the glass, it does not come off (unless very harsh conditions unsuitable for an experimental setup, but suitable for cleaning, are used). The glass acts as a “histone sink”, potentially due to its high concentration of negative charges. This results to the concentration, over time, of histones on the glass, and this concentration is catalyzed by the salts, which speed up the exchange of histones between the DNA and solution.

With that hypothesis in mind, a simple solution arose to fix the surface binding issues: inject the nucleosomal DNA in the flow cell in the absence of salts. Indeed, this solution prove to be the right one: in the absence of salt, the surface binding was minimal (fig. 1.18). As expected,
in the absence of salt, the DNA was labeled along all its length with histone signal. However, upon addition of salt, this signal is redistributed on the surface, and, if flow is sufficiently fast, flushed out of the flow cell (fig. 1.19). 50mM of sodium chloride is sufficient to compete the non-nucleosomal histones off the DNA. The presence of dyes on the surface can be fully avoided by competing the non-nucleosomal histones off the λ DNA with salmon sperm DNA, in the absence of salt. In that case, salmon sperm DNA acts as the competing anion and gradually depletes λ DNA of non-nucleosomal histones. Interestingly, this effect was recapitulated in gels, as explained in section 1.9. As a consequence, the main difference in the protocol for Nucleosome Curtains versus DNA Curtains, on top of using nucleosomal DNA instead of naked DNA, is to inject the DNA in the absence of salt, and subsequently clean it with a rapid injection of either 50 mM or more monovalent salt or 0.5 g/L sonicated salmon sperm DNA.

1.12 **Nucleosomes are stable at low concentration for long periods of time**

It has often been proposed that nucleosomes are not stable at low concentration (Claudet, Angelov, Bouvet, Dimitrov, & Bednar, 2005; Eslami-Mossallam et al., 2016; Godde & Wolffe, 1995; T. A. Hagerman et al., 2009; Thastrom et al., 2004). This would be an issue for Nucleosome Curtains, where concentrations of nucleosomes would range from the femtomolar to the picomolar (see section 1.2). Hence, I decided to test the stability of nucleosomes in Nucleosome Curtains.

First, I tested their stability at higher concentration (about 0.8 μM) over an hour by EMSA. I tested the stability of nucleosomes in two buffers: the CC buffer (10% glycerol, 10mM Tris-HCl
pH8.0, 1mM EDTA pH8.0, 0.1% NP-40), in which nucleosomes are supposedly stable and can freely exchange (Thastrom et al., 2004), and the LHR buffer (30mM Tris-Acetate pH8.0, 50mM KCl, 2.5mM ATP, 1mM MgAc2, 0.2mg/mL BSA), which represents relatively physiological conditions. The half-life of the NCPs in both buffers exceeded the thousands of minutes, far larger than the measurement window (fig. 1.20). Hence, it can be said that nucleosomes at high concentrations are stable and do not significantly fall apart within the experimental window.

Second, I tested the stability of nucleosomes on Nucleosomes Curtains. This time, I used CAB buffer (40mM Tris-Hcl pH7.5, 0.5g/L BSA, 125mM NaCl, 5mM MgCl2, 1mM DTT, 4 mM ATP), as required for the condensin experiments (see chapter 2). Once again, the half-life was so long that it could be properly estimated (fig. 2.21). Hence, it appears that nucleosomes are stable in Nucleosome Curtains, albeit their extremely low concentration.

Concentration-dependent stability would imply fast exchange rates between the bound complex and the pool of free molecules. At lower concentrations, the assembly would be less likely, while the disassembly is independent of concentration. Hence, the apparent stability would decrease as the pool of molecules transfers to the unbound state. However, this effect would only happen if the $k_{off}$ is high. However, nucleosomes are known to be extremely stable, requiring more than 0.6 M of salts to start disassembling the dimers (Ruiz-Carrillo & Jorcano, 1979). In consequence, it is unlikely that nucleosomes would be destabilized at low concentrations. Why many authors reported instability at low concentration remains to be clarified. My hypothesis is that they may have been looking at the fast exchanging non-nucleosomal histone subspecies, instead of at the nucleosomes. In all cases, nucleosomes are stable for hours in the context of Nucleosome Curtains.
1.13 Nucleosomes prefer GC-rich regions

It was previously shown that nucleosomes prefer GC-rich regions (Chua et al., 2012; Visnapuu & Greene, 2009). Lambda phage genomic DNA has the interesting property of being polar with respect to sequence composition: one side is GC-rich, while the other side is AT-rich. It is hence easy to estimate the general sequence preference of a protein: it suffices to look where it preferentially binds on the λ DNA molecule. I hence proceeded to confirm that dye-labeled nucleosomes followed the same pattern by recording the position of individual nucleosomes on a Nucleosome Curtains. In order to avoid DNA molecules that are not full length, which would bias the distribution towards the side closer to the barrier, the DNA was double-tethered. I was able to reproduce the data from Visnapuu and Greene (2009) and show that dye-labeled nucleosomes have a slight preference for the GC-rich sequence (fig. 1.21).

1.14 Photobleaching steps increase with increased reconstitution molar ratio

One property of organic fluorophores is that they photobleach. When exposed to photons, they can enter a “triplet state” that prevents any further absorption and emission until the triplet state is relaxed. This is a disadvantage because it means that the fluorophore will randomly stop emitting signal in a manner dependent on light exposure. The higher the amount of photons, the more likely the photobleaching. On the other hand, photobleaching allows to count for the number of fluorophores in a pixel. Because fluorophores emit quanta of light, every loss of a fluorophore
leads to a discreet loss of signal intensity. Hence, by photobleaching a region of interest, one can learn how many fluorophores were in it.

Counting photobleaching steps is of special interest in the case of dye-labeled nucleosomes, because they have two cysteine-mutant histones per complex and, hence, a single nucleosome should give two photobleaching steps. On the other hand, an incomplete nucleosome or an aggregate would give one or lots of photobleaching steps, respectively. Hence, counting the number of photobleaching steps gives us information on the quality of the nucleosomes. It is important to note, however, that the labeling efficiency of nucleosomes is never 100%, and thus that if cannot be expected to find only one population of dual-labeled nucleosomes, but rather a distribution. The following elements will affect the number of photobleaching steps observed:

- The integrity of the nucleosome (one or two labeled subunits, or aggregates)
- The number of nucleosomes in the pixel
- The labeling efficiency
- The labeling specificity (the number of non-specifically bound dyes on the nucleosome)
- The amount of photobleaching prior to the measurement

Hence, while photobleaching steps can be informative, their interpretation must be done with proper controls.

I have quantified the photobleaching steps for H2A S47C nucleosomes labeled with Dyomics 530 at two reconstitution ratios: 1:40 and 1:60, DNA:octamer (fig. 1.22). Interestingly, the number of photobleaching steps increases with the reconstitution ratio. This could be explained several ways. First, it is possible that at higher reconstitution ratios, several nucleosomes occupy the same pixel. Indeed, the number of nucleosomes per DNA molecules is bimodal and can go up to an av-
verage of eight in the 1:60 dataset (fig. 1.23). In addition, this hypothesis would explain the number of 3+ steps photobleaching events. Second, it is possible that the concentration of histone octamers is too low at the 1:40 ratio, and that incomplete nucleosomes (hexasomes) are formed. At a ratio of 1:40, the octamer concentration during reconstitution is about 67 nM, and 100 nM at a ratio of 1:60. However, the equilibrium constant of the reaction between hexasome and nucleosome has been estimated at about 2.7 nM (Rippe et al., 2008). At 67 nM, one would only expect the concentration of hexasomes to be around 12 nM, or about 18% of the initial octamers, while at 100 nM, the concentration of hexasomes would be 15 nM, about 15% of the initial octamers (the rest forming nucleosomes). With \( O_0 \), the initial concentration of octamers, \( H \), the concentration of hexasomes at equilibrium, \( N \), the concentration of nucleosomes at equilibrium, and \( D \), the concentration of dimers at equilibrium:

\[
D = H
\]

\[
N + H = O_0
\]

\[
N = O_0 - H
\]

\[
K_D = \frac{H \cdot D}{N} = 2.7 \times 10^{-9} \text{nM}
\]

\[
K_D = \frac{H \cdot H}{O_0 - H}
\]

\[
K_D \cdot (O_0 - H) = H^2
\]

\[
0 = H^2 - K_D \cdot (O_0 - H)
\]

\[
H = \frac{-K_D \pm \sqrt{K_D^2 - 4 \cdot K_D \cdot O_0}}{2}
\]
\[ O_0 = 67 \times 10^{-9} \text{nM} \quad O_0 = 100 \times 10^{-9} \text{nM} \]

\[ H \approx 12 \times 10^{-9} \text{nM} \quad H \approx 15 \times 10^{-9} \text{nM} \]

\[ \frac{H}{O_0} \approx 0.18 \quad \frac{H}{O_0} \approx 0.15 \]

Hence, the 3% difference is unlikely to cause the large shift observed between the photobleaching distributions observed at molar ratios 1:40 and 1:60. Note, however, that this is assuming that the reconstitutions reach equilibrium, which is not guaranteed. Third, it is possible that the amount of photobleaching prior to the capture of the photobleaching steps was greater in the 1:40 dataset, and that this shifted the distribution down.

More experiments will be required to figure out this discrepancy. However, it can be said that a good amount of nucleosomes follow a two step photobleaching pattern as expected, and it is probable that the high amount of single steps is due to incomplete labeling or photobleaching. Importantly, the nucleosomes do not exhibit a significant amount of more than two steps photobleaching events (in the case of the 1:40), indicating that we are not looking at aggregates.

### 1.15 Positioning nucleosomes on a 601 sequence

It is of interest to be able to position a nucleosome at a specific position on the DNA. The best way to do so is to use a strong positioning sequence, such as the Widom 601 sequence (Thåström et al., 1999). However, P. T. Lowary and Widom (1997) published that the probability \( p \) of an histone octamer binding to one of the \( N \) given positioning sites with a difference of binding energy
\[ \Delta \Delta G_{HO} \] in a DNA molecule of length \( L \) is given by the equation:

\[
p = \frac{1}{1 + \left( \frac{L-146-N}{N} \right) e^{-\frac{\Delta \Delta G_{HO}}{RT}}}
\]

For a 601 sequence, the \( \Delta \Delta G_{HO} \) versus random sequences is -3.45 kcal/mol (J. Widom, 2001). With the gas constant \( R = 1.9858775 \times 10^{-3} \) kcal/(K mol) and a temperature \( T = 277 \) K for reconstitutions (4°C), the probability that an octamer chooses a 601 sequence rather than any other random sequence is 7% for 7 repeats of the 601 sequence in a λ phage genomic DNA construct. Hence, the addition of 601 sites in λ DNA is unlikely to lead to a strong enrichment. However, previous students had managed to obtain a strong enrichment at the 601 sites using an array of 7 or 13 601 repeats in λ DNA (E.C. Greene, personal communication).

I spent a significant amount of time trying to reconstitute nucleosomes on 601 repeats in the same lambda DNA constructs (λ-7x601 and λ-13x601) as used previously. However, I would never obtain significant enrichment, while I could obtain enrichment on a plasmid with 19 repeats of the 601 sequence, as measured by MNase assay (fig. 1.24). Indeed, the 601 repeats are spaced by a linker 50 bp, leading to different laddering than the random sequences, which give a linker of close to 0 bp. Hence, 601 repeats ladder every 200bp, while WT repeats ladder every 150 bp. The λ-7x601 and -13x601 DNA constructs laddered every 150 bp, so it did not give the same enrichment, confirming that there was an issue with these constructs (fig. 1.25).

I managed to obtain enrichment at a 7x601 site positioned in the middle of the λ DNA molecule after building a new λ construct (fig. 1.26). I actually had to build several constructs for one to work. I believe that recombination happens during the production of λ DNA, leading to poor reproducibility in the number of sites from one molecule to another and one preparation to another. I did not
confirm the spacing of the nucleosomes by MNase assay. Indeed, at higher nucleosome concentrations, this DNA became unusable on a Nucleosome Curtains, because nucleosomes would bind to the surface, even in the absence of salt. I confirmed that the issue came from the purification of the λ DNA construct, but could not isolate the problematic step or compound causing nucleosome binding. The problem was reproducible across several DNA purifications.

Albeit the DNA purification issues, I have shown that it is possible to position nucleosomes on a Nucleosome Curtains experiment. Interestingly, the specificity of the positioning is higher than expected theoretically (P. T. Lowary & Widom, 1997), indicating that either the 601 sequence exhibit higher affinity than previously thought, or the reconstitutions are not achieving equilibrium, or the theory has to be revisited.

1.16 Conclusion

I have presented the methodology and controls for Nucleosome Curtains experiments. In short, recombinant, cysteine-mutant histone octamers are purified, labeled with a fluorophore, such as Atto 647N, Atto 565 and Dyomics 530, and reconstituted on λ DNA prepared for the experiment by ligation with biotinylated and digoxigenylated oligonucleotides. The nucleosomal DNA is used for the Nucleosome Curtains experiment just like in a regular DNA Curtains experiment, except that it is incubated in the flow cell in the absence of salt. Indeed, in the presence of salt, the non-nucleosomal histones freely exchange and bind irreversibly to the surface, leading to a poor signal-to-noise ratio, as well as a poor surface. Once the DNA has incubated, the excess non-nucleosomal histones can simply be flushed out by flowing more than 50 mM of salt in the flow cell at a fast flow rate (1 mL/min), to minimize the interaction of the histones with the surface. EMSAs have
confirmed that nucleosomes are stable in the presence of competitor, such as salt or salmon sperm DNA, but not non-nucleosomal histone subspecies.

Dye-labeled nucleosomes protect the expected 150 bp of DNA in MNase assays and shift DNA just like WT nucleosomes in EMSA. Hence, they are proper nucleosomes and ideal substrates for single-molecule experiments. Indeed, unlike QDot-labeled nucleosomes, they labeling efficiency is high, their interaction with the label is covalent and hence stable, and their label is small compared to a nucleosome. Dye-labeled nucleosomes are stable for hours in a Nucleosome Curtains experiments. Finally, Nucleosome Curtains can be used with 601 sequences to position nucleosomes at specific positions on the DNA. This capacity opens a whole avenue of experiments with chromatin domains.

Overall, Nucleosome Curtains present great flexibility and stability, are heavily reproducible, and have already been used in several projects. I will now present their use in the context of DNA condensation by condensin.
1.17 Figures

Figure 1.1: Steps in the purification of histone octamers. Refer to section 1.4.
Figure 1.2: Position of the cysteine mutations in the nucleosome. The mutated sites are shown as purple spheres. H2A is colored in yellow, H2B in red, H3 in cyan and H4 in green. H3V36C is not shown because it is not seen in the structure (it is positioned in the unstructured tail) and because it did not express. Mutation sites have been framed on one of the two subunits to ease the localization. Left: view from the front of the nucleosome. Right: view from the side of the nucleosome, a rightwards 90° rotation against the vertical axis from the position shown left. Structure from White, Suto, and Luger (2001).
Figure 1.3: Elution profile for all four core histones. The elution profiles obtained from the Unicorn software (GE Healthcare Akta Systems) are shown on top, and the fractions resolved on 15% poly-acrylimide gels and stained with Coomassie Brilliant Blue are shown at the bottom for histones (A) H2A, (B) H2B, (C) H3, and (4) H4.
Figure 1.4: Elution profile of the histone octamers with cysteine mutant H2A S47C. The elution profile obtained from the Unicorn software (GE Healthcare Akta Systems) is shown on top, and the fractions resolved on a 15% poly-acrylimide gel and stained with Coomassie Brilliant Blue is shown at the bottom. All four histones are resolved. Numbers at the bottom of the gel indicate the fraction number. Each peak in the chromatogram is identified with its content.
Figure 1.5: 15% SDS-PAGE of different purified histone octamers. Coomassie blue staining. Each histone is indicated on the right, while the mutated histone in the octamer is indicated at the bottom. Note that H2A-3xFLAG octamers seem to have some degradation products.
Figure 1.6: 15% SDS-PAGE gel of dye-labeled histone octamers. Octamers were labeled with the fluorophore Dyomics 530. The leftmost panel is the Coomassie blue staining, the middle panel is the Dyomics 530 fluorescent signal when illuminated with a 532 nm laser, and the rightmost panel is the overlay of the two previous panels. The mutated histone in the octamer is indicated at the bottom. H4 S48C octamers have a reproducibly lower labeling efficiency, probably due to the position of the cysteine, slightly buried within the nucleosome (see fig. 1.2).
Figure 1.7: 0.75% agarose gel of the purification of λ DNA by PEG precipitation. Two ligations of λ phage genomic are shown. One with the biotin on the “right” side and the digoxigenin on the “left” side (BioRDigL) of the DNA molecule, and the second, the opposite (BioLDigR). The DNA was separated from the short oligonucleotides by precipitation with 10% PEG8000, and the resulting pellet was resuspended to a calculated final concentration of 0.5g/L. The actual concentration, measured with a ThermoFischer NanoDrop spectrophotometer, is shown at the bottom, and is within error, indicating a recovery of around 100%. The left side of the gel is the resuspended pellets, while the right side is the supernatants (“Sup.”). The oligonucleotides are seen exclusively in the supernats, where very little λ DNA is left. The lambda DNA was obtained from NEB (product number N3011S).
Figure 1.8: Titration of MNase and nucleosomes on λ DNA. WT histone octamers were reconstituted on λ DNA at the given ratios. MNase assays were performed at different concentrations of MNase. The higher the concentration of nucleosomes, the more likely the protection of nucleosome repeats (laddering). On the other hand, the higher the MNase concentration, the shorter the protected sequence, as the enzyme is more likely to profit from nucleosome breathing and access the sequences at the edges of the nucleosomes.
Figure 1.9: MNase digestion of nucleosomes reconstituted with WT, FLAG-tagged, and cysteine-mutant histone octamers. Nucleosomes were reconstituted on λ DNA. Several reconstitution ratios were used in order to see whether laddering increases with higher concentrations (Noll, 1974). The top gel is the reaction in the absence of MNase (as indicated in the right column), whereas the bottom gel is in the presence of 1 U/μL MNase. Slight increase in protection is observed as the molar ratio of DNA to increases. However, these ratios do not go high enough to saturate the DNA and cause serious laddering, except for a slight increase in the di-nucleosome band at 300bp (which indicates the absence of linker DNA). H4 S48C seems to increase faster and blur at a molar ratio of 1:200, indicating potential aggregation or defect at higher concentration. Differences in intensities can be explained by the uncertainty in the stock concentration of the nucleosomes and in the MNase digestion assays. The deprotonized DNA fragments were separated on a 2% agarose gel and stained with Sybr Safe.
Figure 1.10: MNase digestion of nucleosomes reconstituted with dye-labeled octamers on λ DNA. WT, H2A S47C, H2B T119C, and H4 S48C histone octamers labeled with Dyomics 530 were used to reconstitute nucleosomes on λ DNA. MNase digestion with either 0 (control) or 1 U/μL of enzyme were prepared. The deprotonized DNA fragments were separated on a 2% agarose gel and stained with Sybr Safe. While different octamers present different protection, both in term of band intensity and protected sequence, this can be explained by the varying input amounts. Refer to text for more details.
Figure 1.11: 2% agarose gel of purified 147bp 601 DNA fragments using the protocol described in appendix E. The 601 fragment labeled with Atto565 is shifted up slightly due to the charge and mass of the dye.
Figure 1.12: EMSA of NCPs and MNase product. Nucleosomes were reconstituted on 147 bp 601 sequences at different ratios, as well as on λ DNA. The λ nucleosomal DNA was digested with microccocal nuclease, and the result of the digestion was resolved along the NCPs. All products migrate at the same positions, indicating that NCPs and λ DNA nucleosomes are indeed full size nucleosomes. The left panel is the gel stained with the DNA stain Sybr Safe (Invitrogen), while the right panel is the same gel with the protein content labeled by silver stain. Interestingly, unprotected 150 bp sequences are observed in the MNase assay, indicating that nucleosomes felt apart during the procedure. The extra bands shifted in the 2:1 NCP ratio are not nucleosomal products (see text). The U-shape of the bands is due to excessive amounts of loaded material.
Figure 1.13: Competition of non-nucleosomal histones on an EMSA. NCPs were reconstituted with Atto565-labeled 147 bp 601 DNA molecules at a molar ratio of 1:1 with WT octamers. Reconstituted nucleosomes were incubated shortly in the buffers indicated below the gels and resolved on a 5% polyacrylimide native gel. As can be observed, the addition of salmon sperm DNA is extremely efficient to remove the bands located above the shifted NCPs, while the NCP signal is not affected by the salmon sperm DNA. This result indicates that the extra bands are not nucleosomes, but non-nucleosomal histones weakly bound to DNA that readily exchange. Interestingly, salt competition did not clear the additional bands. It could be because the high salt concentration destabilizes the nucleosomes, accelerating the exchange between nucleosomal and non-nucleosomal histones, or because the salt does not enter the gel, and hence as the reaction enters the gel, non-nucleosomal histone species are reassembled.
Figure 1.14: Titration of histone octamers on dye-labeled 147 bp 601 DNA. Nucleosomes were reconstituted with WT histone octamers and 147 bp 601 DNA at a 1:1 molar ratio, at different concentrations. The fraction of total DNA shifted was measured and plotted as a function of concentration. The reconstitution reaction seems to saturate around 1 μM. Inset: the gel used for the titration (DNA signal only is shown). The shaded area represents the concentration ranges used for reconstitutions for Nucleosomes Curtains, indicating that the majority of the histone octamers will not reconstitute into nucleosomes in these assays.
Figure 1.15: Screen for the optimal reconstitution ratio for the assembly of NCPs. Nucleosomes were reconstituted in the micromolar range with Atto565-labeled 147 bp 601 sequences and WT histone octamers. The R ratio is the molar ratio of octamer to DNA. This ratio is indicative, because of the usual error in the estimation of concentrations. The optimal R ratio to get rid of non-nucleosomal histones and minimize the amount of free DNA was 1.2 in this case. Left: DNA signal, labeled with Sybr Safe. Right: DNA signal from the 601 DNA (Atto565).
Figure 1.16: EMSA of NCPs reconstituted with dye-labeled nucleosomes. Nucleosomes were reconstituted with Atto565-labeled 147 bp 601 DNA and Alexa488-labeled H2A S47C octamers at different ratios of octamer to DNA. The reconstitutions were chased, or not, with 1 g/L of salmon sperm DNA. All three panels are the same gel with different signal shown. Left: DNA signal (with histone signal bleed-through), center: 601 DNA signal from the Atto565 dye, right: histone signal from the Alexa488 dye. As can be observed, the Alexa488 and Atto565 signal overlap, indicating properly formed NCPs. Addition of competitor DNA shifts the non-nucleosomal histones up, demonstrating that they exchange fast and are not stable, unlike the nucleosomal histones that remain bound.

Figure 1.17: Examples of Nucleosome Curtains with all three cystein-mutant histone octamers (H2A S47C, H2B T119C, H4 S48C) labeled with Dyomics 530 (DY530). Top panel is the curtain in the absence of flow, with the DNA retracted to the barrier, hence not visible, while bottom panel is in the presence of 1 mL / min flow rate. The absence of signal in the absence of flow demonstrate that the histones are bound to the DNA only. The only label present is that of the fluorophore Dyomics 530 attached to the octamer by the cysteine mutant indicated at the top. DNA was washed with sonicated salmon sperm DNA (H2A S47C DY530 and H2B T119C DY530) or salt (H4 S48C DY 530) prior to imaging.
Figure 1.18: Representative examples of surface binding of histones as a function of salt. Nucleosomes were reconstituted on λ DNA at a 1:50 ratio with H2A S47C octamers labeled with Dyomics 530. In the left panel, the DNA was incubated in the flow cell in the presence of salt. In the right panel, the DNA was incubated in the flow cell in the absence of salt. The intensity of both images was adjusted to a range of 300 A.U. to 7000 A.U. The sole label is the Dyomics 530 present on H2A S47C.
Figure 1.19: Effect of salt on non-nucleosomal histones on DNA. Nucleosomes were reconstituted on \(\lambda\) DNA using H2B T119C histone octamers labeled with Dyomics 530. Nucleosomal DNA was injected in the absence of salt and imaged. Subsequently, 300 mM of potassium glutamate were injected in the flow cell, and the DNA was imaged. In the top panel, before the injection of salt, the DNA is cover with histones. In the bottom panel, in the presence of salt, only a small fraction of the histones, those that formed nucleosomes, remain. The intensity of the images have been modulated for readability. The range on the top image goes from 3,000 A.U. to 50,000 A.U., while that of the bottom image goes from 700 A.U. to 3,000 A.U. This difference means that the brightest pixel in the top image is actually 16.7 times brighter than the brightest pixel in the bottom image, illustrating the amount of non-nucleosomal histones on the DNA. The sole label is the Dyomics 530 present on H2B T119C.
Figure 1.20: Nucleosome stability at high concentration. Nucleosomes were reconstituted on an Atto565-labeled 147bp 601 DNA sequence with WT histone octamers at a concentration of 0.79 μM. They were incubated in either CC buffer (10% glycerol, 10mM Tris-HCl pH8.0, 1mM EDTA pH8.0, 0.1% NP-40) or LHR buffer (30mM Tris-Acetate pH8.0, 50mM KCl, 2.5mM ATP, 1mM MgAc2, 0.2mg/mL BSA) for up to an hour and resolved on a EMSA assay. The ratio of shifted nucleosomes is plotted as a function of time and buffer. While the half-life is too long to be precisely estimated, it can be said that the nucleosomes remain stable and do not significantly fall apart within an hour.
Figure 1.21: Position distribution of nucleosomes on λ DNA. Dye-labeled nucleosomes were assembled on λ DNA and imaged on a Nucleosome Curtains experiment. The position of the individual nucleosomes was recorded and plotted as a histogram as a function of the position on the DNA. In addition, the GC-richness of the λ DNA in that pixel is plotted. It can be seen that nucleosomes exhibit a slight preference for GC-rich sequences. The empty bins on both sides of the histogram at due to the poor signal to noise ratio at the barriers and pedestals, making it impossible to collect data in their vicinity.
Figure 1.22: Effect of nucleosome reconstitution ratio on photobleaching steps. The number of photobleaching steps per puncta was measured for reconstitution ratios of 1:40 and 1:60. An example of two-steps photobleaching is shown as an inset on the left plot. The average number of photobleaching steps shifts with the reconstitution ratio.
Figure 1.23: Number of nucleosomes as a function of reconstitution ratio. Dye-labeled nucleosomes were reconstituted at the ratios of DNA to octamer of 1:40, 1:60 and 1:80. The number of puncta per DNA molecule was measured and plotted as a swarm plot. The mean number of nucleosomes per DNA molecule and its 95% confidence interval are plotted in black. Note that the distribution is bimodal at 1:60, potentially indicating several DNA molecules per well. Indeed, the mean of the second distribution, around 8, is about twice that of the first distribution, around 3, indicating that 2 or more DNA molecules are overlayed in the higher distribution.
Figure 1.24: MNase digestion of 19x601 plasmids. Plasmids with (pUC18-19x601) or without (Control) 19 repeats of the 601 sequence spaced by 50 bp was used to reconstitute nucleosomes at different ratios. Because the 601 repeats are spaced by a linker of 50 bp, it is expected that they form a ladder with repeats of 200 bp, instead of repeats of 150 bp like on non-positioning sequences. It can be observed that this is what happens on the 601-containing plasmid, but not on the control plasmid, indicating that the nucleosomes are positioned on the 601 sequences. Repeats of 150bp can also be observed on the 601 plasmids, probably due to the binding of non-601 sequences on the rest of the plasmid.
Figure 1.25: MNase assay on λ 601 constructs. Nucleosomes were reconstituted with WT octamers on λ DNA at several ratios. Three lambda DNA constructs were used: WT, 7x601, and 13x601. It is expected that the 601 arrays lead to a spacing of 50 bp between the nucleosomes, and hence that the laddering with 200bp repeats above the single 150bp band. However, all constructs behave just like wild-type, indicating that there is no preferential binding to the 601 sites.
Figure 1.26: Nucleosome Curtains of dye-labeled histones on a 7x601 λ DNA construct. H2B T119C DY530 nucleosomes were reconstituted at a molar ratio of 50:1 octamer:DNA on λ DNA containing 7 601 repeats in its center. In the absence of flow, the DNA retracts to the barrier and no signal is seen. In the presence of flow, the DNA is extended and the nucleosome signal appears concentrated at the 601 sites.
Chapter 2

Single-molecule nucleosome curtains reveal nucleosomes are not a barrier to loop extrusion by condensin

2.1 Introduction

Since its first publication in 1882 (Flemming, 1882), our knowledge of chromosomes, chromatin and mitosis has advanced tremendously, but it remains incomplete (Alberts, 2008). The exact steps, factors and functions required to assemble chromosomes remain elusive. This work focuses on two of chromosomes’ most abundant proteins: histones and condensin.

2.1.1 Condensin is necessary for the proper formation of chromosomes

Condensin was shown to be involved in proper chromosome segregation. In the absence of functional condensin, mitosis proceeds, but is erratic and often leads to DNA bridges or improper chromosome segregation (T. Hirano & Mitchison, 1994; Saitoh et al., 1994; Saka et al., 1994; A. V. Strunnikov et al., 1995). It was first pointed in human cells (Lewis & Laemmli, 1982) and later shown in *X. laevis* egg extract (T. Hirano et al., 1997) that condensin is involved in the condensa-
tion of chromatin, hence its name. Interestingly, compaction mechanisms are redundant (Vagnarelli et al., 2006), and DNA bridges formed in the absence of condensin usually get resolved in a mechanism that is dependent on Aurora B (Mora-Bermúdez, Gerlich, & Ellenberg, 2007) and could involve the kinesin KIF4A (Mazumdar, Sundareshan, & Misteli, 2004). However, the exact way by which condensin condenses chromatin into chromosomes, and how this prevents DNA bridges, is only starting to get clarified.

2.1.2 Condensin is a large, ring-shaped ATPase protein

Condensin is part of the structural maintenance of chromosomes (SMC) protein family, which also includes the cohesin complex. SMC proteins are large multimers of more than 1,000 amino acids, most of it coming from a heterodimer of SMC subunits. Each subunit contains a Walker A and Walker B motif at each of its ends. The motifs are separated by coils and a central non-helical domain. The subunits fold on themselves, forming antiparallel coiled-coils on its length and a terminal ATP-binding cassette (ABC)-like ATPase head domain resulting from the assembly of the Walker motifs. At the other end of the folded protein, the non-helical domain dimerizes with that of another SMC subunit to form a donut-shaped hinge domain. This results in a V-shaped heterodimer about 50 nm long. The protein is complemented by a kleisin, which bridges the two ATPase head domains and closes it topologically (T. Hirano, 2006). In budding yeast, the condensin complex is composed of the SMC subunits Smc2 and Smc4, the kleisin Brn1 and the HEAT-repeat proteins Ycs4 and Ycg1 (Piazza et al., 2014).
2.1.3 Condensin binds to DNA

Condensin has been shown to bind DNA via its HEAT-repeat subunits, themselves interacting with the kleisin subunit. Binding of the HEAT-repeat subunits to DNA stimulates the ATPase activity of the SMC subunits (Piazza et al., 2014). Interestingly, the presence of a functional hinge domain is also necessary for proper DNA binding, indicating either long range interaction between the domains (M. Hirano & Hirano, 2006), topological enclosing of the DNA (Cuylen et al., 2011), or both.

Crystal structures of the S. cerevisiae kleisin and HEAT proteins Brn1 and Ycg1 interacting with DNA have demonstrated the formation of a “safety-belt” mechanism to entrap DNA (Kschonsak et al., 2017). In the structure, the two proteins form loose interactions with the DNA backbone, and the stability of the interaction comes from the stable entrapment of DNA by a flexible (and hence unresolved in the structure) loop of the kleisin Brn1. This work also indicates that the other HEAT subunit of the S. cerevisiae condensin, Ycs4, has little, if any, impact on the DNA binding capacity of condensin. It remains unclear whether this DNA binding domain is the only one, or if more subunits interact with DNA (Griese et al., 2010; Kinoshita, Kobayashi, & Hirano, 2015).

In their topological studies of DNA entrapment, Cuylen et al. (2011) cut the kleisin subunit to show that DNA was topologically enclosed, and propose that the DNA gets trapped between the SMC subunits. In the light of the crystal structure, it is possible that the DNA never makes it to the large condensin ring, and that it remains entrapped in the kleisin-HEAT ring. Overall, the exact interactions of condensin with DNA and how they relate to its looping and motor activity remains to be defined.
2.1.4 Simulations show that DNA looping is sufficient for chromosome formation

Because the main phenotype out of condensin’s deletion is DNA bridges, it is supposed that the main role of condensin is to separate sister chromatids, which are concatenated post replication. It has been proposed that condensin could extrude loops of DNA to fulfill this function (Nasmyth, 2001).

Computer simulations agreed with that hypothesis, showing that a simple motor protein that extrudes loops of DNA is sufficient to reproduce the chromatin changes from interphase to mitosis (Alipour & Marko, 2012; Goloborodko et al., 2016). In the simulation, the authors defined chromatin as a multimer (5000 subunits, 30 Mbp) of 600 bp, 10 nm diameter monomers (roughly equivalent to 3 nucleosomes). They modeled condensin as a motor protein with that can bridge these chromatin monomer, bringing two monomers—one per SMC subunit— together at each step, hence forming a loop in between the SMC monomers. The condensin molecules stop extruding loops when they collide, and are allowed to randomly bind and unbind from DNA. Strikingly, this simulation showed that intertwined, randomly coiled chromosomes could be transformed into prophase-like chromosomes (Goloborodko et al., 2016).

Interestingly, another computer simulation showed that condensin could condense chromosomes by simply stabilizing its random interactions (Cheng et al., 2015). This model was also able to explain how chromosomes separate. Here, condensin would simply act as a “weak glue” between contact points. Every time two condensin molecules bound to DNA would interact, simply by diffusion of the DNA molecule, they would transiently stabilize that interaction. Over time, the intramolecular interactions become more likely, and the DNA molecules separate. This simple
and elegant model would satisfy Occam’s razor. At that point, however, it was unclear whether condensin had any loop-extruding or interaction-stabilizing capacity, and further experimental evidence was necessary to prove or disprove the models.

2.1.5 Condensin extrudes loops of DNA

A first evidence came from our single-molecule experiments, showing that condensin was a motor protein that translocates on DNA in an ATP-dependent fashion. Not only does condensin translocate on DNA, but it can also translocate on DNA while holding another piece of DNA, indicating a simple looping model by which condensin would grab a piece of DNA and drag it in one direction (Terakawa et al., 2017). This model is interesting considering that this would imply at least two DNA binding sites, while only one has been confirmed in vitro.

The definitive proof of condensin’s loop extrusion capacity came from subsequent single-molecule studies, where the authors showed that condensin forms DNA loops in an ATP-dependent manner (Ganji et al., 2018). However, unlike in the computer simulation model (Goloborodko et al., 2016), and much like in the previous single-molecule experiments, condensin would only extrude loops in an asymmetric manner (Ganji et al., 2018).

Hi-C studies in vivo have demonstrated that condensin is necessary for the proper folding of chromosomes (Gibcus et al., 2018; Kakui, Rabinowitz, Barry, & Uhlmann, 2017). Condensin I and II (see section 2), or the single condensin in S. pombe, orchestrate the folding of chromosomes by forming bigger and bigger loops, potentially in a spiral staircase conformation.

All of this indicates that condensin can extrude loops of DNA, hence untangling sister chromatids and forming prophase chromosomes. The way condensin prevents the formation of DNA
bridges during mitosis is getting clarified (van Ruiten & Rowland, 2018).

2.1.6 Histone chaperones are necessary for the formation of chromosomes in vitro

It remains unclear whether condensin can extrude loops of chromatin. In vitro studies with *X. laevis* egg extract have shown that chromosome-like structures can be assembled with just six factors: the core histones, three histone chaperones, topoisomerase II and condensin I (Shintomi et al., 2015). It is unclear whether the histone chaperones are necessary only to assemble nucleosomes, or also to disassemble them so that condensin can form loops on the DNA substrate. Remarkably, the authors were also able to demonstrate that nucleosomes are not necessary for DNA condensation, and hence that condensin probably both untangles and compact DNA fibers (Shintomi et al., 2017).

To figure out whether condensin can condense DNA in the presence of nucleosomes, we studied the effect of nucleosomes on loop extrusion by *S. cerevisiae* condensin at the single-molecule level. First we developed a DNA Curtains technique to observe DNA condensation and loop extrusion by condensin. Second, we developed a technique to observe *S. cerevisiae* nucleosomes at the single-molecule level using DNA Curtains. Finally, we combined both techniques to study the effect of nucleosomes on loop extrusion by condensin.
2.2 Results

2.2.1 Single molecule observation of condensin reveals its DNA condensation and looping activity

We first developed a system for the study of DNA condensation by condensin. Building on our previous publication, we used recombinant condensin from *S. cerevisiae* (Terakawa et al., 2017). We labeled the end of the DNA molecule with a quantum dot via CRISPR-Cas9 (Sternberg et al., 2014) (fig. 2.1) and continuously injected unlabeled condensin at a flow rate of 0.1mL/min, which is strong enough to align the DNA molecules with the observation field, but not strong enough to disrupt the interaction between condensin and DNA (fig. 2.2). We observed that, upon the addition of wild-type (WT) condensin, the DNA molecule would compact in the presence of ATP, but not in the presence of ATPγS or using an ATPase-dead condensin that are mutated in the γ-phosphate switch loops (Q-loops) of the SMC2 and SMC4 subunits (fig. 2.3). This indicates that the compaction process depends on the ATPase activity of condensin. We measured the compaction velocity and found an average rate of 0.42kbp/s (fig. 2.4). Interestingly, two velocity peaks were found, much like what was observed for the motor activity of condensin (Terakawa et al., 2017). Overall, this indicates that condensin exerts some compaction action on DNA that is dependent on ATP.

We then labeled both condensin and the DNA with QDot605 and 705, respectively. We observed three different scenario regarding the directionality of DNA compaction versus that of condensin (fig. 2.5). In one case, both DNA and condensin would move, indicating that either condensin is translocating in one direction while looping DNA behind it, or condensin is looping DNA
in both directions (symmetric compaction). In a second case, condensin would remain mostly static, while the DNA would move towards it, indicating that condensin loops DNA from its rear side (asymmetric compaction). Finally, in the third case, condensin would loop DNA towards the front, while the DNA in the back would only move at the same speed as condensin, indicating the absence of compaction in the backward direction (asymmetric compaction). When quantified, we observed that the majority (66%) of compaction events are asymmetric (fig. 2.6). This is in good accordance with our previously published model, where condensin grabs a piece of DNA and compacts in by translocating along its length, hence looping it (Terakawa et al., 2017).

It was previously published that condensin can compact DNA by looping it (Ganji et al., 2018). To confirm that what we observed was DNA looping, we repeated our same assay but labeled the whole DNA molecule with the intercalating dye YoYoI, instead of only labeling its end with a QDot. Strikingly, we observed the formation of DNA puncta after the injection of condensin, but not before (fig. 2.7). When analyzed over time, these puncta would move linearly, carrying the DNA molecule with them (fig. 2.8). We identified this as the same DNA condensation event as observed with the QDot-labeled DNA. While the majority of the condensation events happened with one punctum, occasional 2 and 3 puncta per DNA molecules were observed during condensation (fig. 2.10). We then confirmed that condensin was responsible for the apparition of puncta and the condensation of the DNA. We measured the number of condensation events 30 minutes before and after adding condensin, and no compaction could be observed in the absence of condensin (fig. 2.9).

Our data suggest that the observed DNA puncta are loops of DNA generated by the action of condensin. A lengthening piece of DNA can sometimes be observed behind the punctum. However, this DNA is usually further condensed, and the resulting loop is observed as a punctum. Stronger buffer flow can lead to the stretching of the punctum, indicating that it is a loop of DNA. Note that
we do not know how many loops and condensin molecules are present per punctum, but know that several loops can be present in one punctum (fig. 2.11). Overall, these results indicate that we have a functional assay to observe the formation of DNA loops by condensin.

### 2.2.2 Condensin associates with DNA loops

We then checked the association of these loops with condensin. We labeled both condensin and the DNA and, in accordance with what was previously published (Ganji et al., 2018), we observed that condensin colocalizes with the loop (fig. 2.12). When quantifying the colocalization (fig. 2.13), we observed that the majority of DNA puncta (80%) did not colocalizes with condensin. The discrepancy between this result and that of fig. 2.9 can be explained by the poor labeling efficiency of condensin with the anti-HA-conjugated QDot. All puncta are most likely associated with condensin, but only a minority of condensin molecules is labeled. It is also interesting to note that not all condensin molecules lead to the formation of DNA puncta. The majority of condensin molecules does not form a visible punctum (58%). This can be explained by the possibility that the puncta are smaller than our detection limit, and that many condensin molecules bind on DNA but remain static (data not shown). It is also possible that condensin has two modes of action. One where it is actively extruding loops, and one where is it only moving on DNA.

### 2.2.3 Assembly of nucleosomal DNA Curtains

It is unclear whether condensin can actively extrude loops of DNA in the presence of nucleosomes. To test the effect of nucleosomes on condensin, we next proceeded to to assemble nucleosomes for DNA Curtains. First, we purified histone octamers using the protocol described in J. Y. Lee and
Greene (2011). We purified every individual histone independently, including the cysteine-mutant histones H2A-S47C, the S. cerevisiae equivalent of the H2A-A45C mutant described previously using X. laevis histones (Kassabov et al., 2002). We then assembled the histones into octamers and labeled the cysteine-mutant octamers with the maleimide conjugated fluorescent dye Dyomics 530. These histones display the expected mass weight and the labeling is very specific, with only a small amount of the signal (5%) colocalizing with H4 instead of H2B (fig. 2.14).

The purified histone octamers were reconstituted into nucleosomes with lambda genomic DNA using salt dialysis (J. Y. Lee & Greene, 2011). The reconstituted nucleosomes, both with WT and H2AS47CDY530 octamers, showed proper protection of about 150bp on a microccocal nuclease (MNase) assay, with light amount of laddering at higher concentrations (fig. 2.15). Laddering, obtained from the protection of multiple consecutive nucleosomes, is expected when nucleosomes are close enough to prevent MNase digestion in between each nucleosome (Noll, 1974). Note that the concentration of nucleosomes is kept relatively low on purpose. We are aiming to observe single nucleosomes on the DNA Curtains experiments, and as such laddering, indicating high density and proximity of nucleosomes, is not desired.

We next measured the properties of the nucleosomes on DNA Curtains. As expected, each DNA molecule was bound by several individual nucleosomes (fig. 2.16). To confirm that the nucleosomes were bound to the DNA molecules, and not to the surface, buffer flow was applied and released. In the presence of buffer flow, DNA molecules are extended against the plane of the objective, allowing for the visualization of their whole length, and of the bound nucleosomes. Once buffer flow is released, the hydrodynamic force applied on the DNA molecules is lost, and the DNA molecules relax and go back to 3D diffusion, hence disappearing from the field of view. In this situation, the nucleosomes disappear with the DNA molecules, confirming that they are not bound
to the surface, but to the DNA (fig. 2.17). As expected, the number of nucleosomes increased as a function of the DNA:Octamer reconstitution ratio (fig. 2.18). Because the number of single nucleosomes per pixel quickly exceeded 1, we relied on signal intensity to quantify the increase at higher ratios (fig. 2.19). In addition, the nucleosomes followed the previously published (Visnapuu & Greene, 2009) sequence preference (fig. 2.20). We tested the stability of the nucleosomes over time in the conditions of the experiment, and measured a half-life far longer than 100 minutes, although we refrained from estimating it, since it exceeded the observation window (fig. 2.21). Hence, nucleosomes are stable within our experimental window Finally, to we measured the photobleaching steps of the nucleosomes to confirm that they contained the proper number of subunits (fig. 2.22). Indeed, a proper nucleosome with 2 subunits of each histone and one fluorophore per labeled subunit should exhibit 2 photobleaching steps. The majority of the nucleosomes displayed 1 or 2 steps, consistent with 2 subunits and incomplete labeling. Few of them presented more than 2 steps, and those are potentially nucleosomes with H4 labeled (5% of all signal, as shown in fig. 2.14) or several nucleosomes in the same pixel. It is possible that some of the single-labeled nucleosomes are actually hexasomes, but the MNase assay (fig. 2.15) seems to indicate that this would be a minor population, as no hexasome protection band is visible (Arimura et al., 2012).

Overall, these results indicate that I have developed a working protocol for studying nucleosomes at the single-molecule level.

### 2.2.4 Nucleosomes do not affect condensation velocity

To observe the effect of nucleosomes on DNA condensation, we first reconstituted wild-type nucleosomes on lambda phage genomic DNA, and then used that DNA to observe condensation with
5 nM of unlabeled condensin injected at 0.1 mL/min (fig. 2.23). Interestingly, the condensation velocity remains fairly constant against nucleosome density (fig. 2.24). The very weak, positive slope (0.008 μm/s) could be explained by the intrinsic compaction of DNA by nucleosomes. Indeed, nucleosomes on their own provide with a compaction factor of about 4.4 fold (Richmond & Davey, 2003). Hence, more nucleosomes on the DNA molecule would cause more compaction, resulting in more DNA base pairs per μm. It is, however, hard to estimate the contribution of nucleosomes to compaction in each DNA molecule, as the number of nucleosomes per DNA molecule can vary greatly (fig. 2.18), and because nucleosomes are not labeled in this experiment. Hence, we opted to plot the data in μm per second, rather than base pairs per second, to avoid adding the uncertainty of the nucleosome compaction factor to the data. In all cases, the absence of a negative slope indicates that nucleosomes have very little effect on condensin. How can condensin not be affected by structures that are roadblocks for most of the motor proteins (Clapier et al., 2017; Gurard-Levin et al., 2014; Hauer & Gasser, 2017; Teves et al., 2014)? We hypothesized that either nucleosomes are actively evicted by condensin, or condensin forms loops in between nucleosomes, or the condensation activity is independent of the width and conformation of the DNA substrate (fig. 2.25).

### 2.2.5 Condensin does not affect nucleosome organization

First, we tested whether condensin evicts nucleosomes. We reconstituted nucleosomes with a dye on the engineered H2A S47C site (H2A S47C DY530, fig. 2.16) and used them for our DNA Curtains experiments. We injected 5 nM of condensin, and let the reaction proceed to equilibrium for 10 minutes. Then, buffer was pushed at 0.1 ml/min, so that the DNA would align with the obser-
vation field, but the interactions between condensin molecules and the DNA molecules would not be disrupted. The compaction state of the DNA was recorded, and 0.5 M of NaCl was flowed in to release the interactions between condensin and DNA. As can be observed, the fully compacted DNA can be decompacted by the addition of 0.5 M salt, and while this affects the stability of the nucleosomes, most of the signal seems to remain bound (fig. 2.26). To make sure that the compaction was not due to the higher concentration of proteins, we repeated the experiment with an ATPase-deficient (Q-loop mutant) condensin, as well as with the non-hydrolysable ATP substitute ATPγS, and both show no compaction compared to the experiment with ATP (fig 2.27). We quantified the effect of condensin and induced decompaction on several reconstitution ratios (fig. 2.28). As can be seen, the drop of nucleosome signal before and after induced decompaction is not strikingly different, indicating that condensin does not affect the stability of nucleosome, and hence barring the possibility that condensin proceeds on nucleosomal DNA by evicting nucleosomes. To confirm this, we observed the compaction of nucleosomal DNA by 1 nM of unlabeled condensin. At lower concentrations of condensin, DNA loops spontaneously decompact, probably due to the release of the condensin molecule from at least one attachment site on the DNA. We found that during these spontaneous decompaction events, when the DNA is released, nucleosomes were positioned identically to before compaction (fig. 2.29). This further reinforce the idea that condensin does not remodel or evict nucleosomes during DNA looping (fig. 2.25C).

### 2.2.6 Condensin assembles loops containing nucleosomes

Next, we addressed whether condensin extrudes loops of DNA between or with nucleosomes (fig. 2.25A-B) To do so, we tracked the loops as well as the nucleosomes, and observed whether
loops would be formed over or around nucleosomes. We injected 1 nM of condensin at 0.1 mL/min for 30 mins, and recorded both DNA and nucleosome signal. Astoundingly, upon collision with nucleosomes, DNA loops would continue growing, integrating the nucleosomes inside of it (fig. 2.30). When quantified, 80% of all collisions result in continued compaction, with the integration of the nucleosome into the DNA loop (fig. 2.31). The second likeliest outcome would be loop release upon collision, occurring in less than 14% of all cases.

Next, we quantified the velocity and processivity of DNA loops colliding with a nucleosome versus those on naked DNA. As expected from fig. 2.24, the act of colliding with a nucleosome does not lead to a significant difference in either velocity (fig. 2.32, KS-statistic = 0.08017) and processivity (fig. 2.33, p-value = 0.63 on a logrank test). This result indicates that nucleosomes do not affect condensin, which can directly form loops of nucleosomal DNA (fig. 2.25B).

To confirm that the nucleosomes did not slow down loop extrusion, we also measured the dwell time at collision, for collisions leading to the integration of the nucleosome in the loop (fig. 2.34). The survival distribution exhibited features of a sum of two exponentials and was fitted as such. Around 85% of all collisions are resolved in less than 5 s, which is our time resolution limit, and hence cannot be estimated. The 15% left can be plotted to a single exponential with a half-life of about 65 s. This second population may come from a different conformation of condensin, or from an experimental artifact. For example, it is possible that the size of a condensin ring varies over time, or from molecule to molecule, potentially switching from open to closed or bent (Hassler et al., 2018). If a condensin with a closed or bent ring encounters a nucleosome, it is possible that it cannot integrate it into the loop, the ring being too small for a nucleosome to go through. Only when the ring opens would the condensin be able to extrude past the nucleosome, leading to a longer half-life at the site of the nucleosome. Alternatively, it is possible that transient interactions
between the nucleosome and the surface, or other experimental artifacts, lead to this population of slowly-resolving interactions.

2.3 Discussion

Overall, our study shows that condensin forms loops to compact DNA, that it can compact nucleosomal DNA without altering the binding or spatial distribution of the nucleosomes, and that it can integrate nucleosomes in DNA loops at the same velocity and processivity as loops of naked DNA. These exciting results imply a new model for the way the motor domain of condensin proceeds to loop DNA.

2.3.1 Condensin forms loops to condense DNA

In agreement with what was previously published (Ganji et al., 2018), our results indicate that condensin compacts DNA in a unidirectional manner (fig. 2.6) by forming a loop (fig. 2.8). In our system, loops appear like puncta, although they occasionally stretch with the buffer flow (fig. 2.11). We believe there is two reasons for this. First, we use a very low flow rate, which may not be amenable for stretching small loops. Second, as the loop gets bigger, more condensin molecules may load on it and compact it, forming loops in the loop and hence stabilizing its “puncta” appearance (fig. 2.11). Interestingly, it was shown that condensin forms clusters on chromosomes, which would agree with that observation (Sun, Biggs, Hornick, & Marko, 2018).
2.3.2 Condensin condenses nucleosomal DNA

Importantly, our research shows that condensin can compact DNA at a velocity independent of the concentration of nucleosomes (fig. 2.24). This is an intriguing result that can be explained by several hypotheses: either the compaction activity of condensin is unaffected by roadblocks, for example because it can easily avoid them or step over them, or condensin has some remodeling activity, and hence can push away or evict the nucleosomes in its way, or condensin cannot handle nucleosomes and compacts DNA around them (fig. 2.25). Our data supports a model in which condensin can readily integrate nucleosomes into the compacted DNA.

First, condensin does not affect the position or presence of nucleosomes on DNA. Indeed, in our experiments, nucleosomes are not lost or displaced during DNA condensation (figs. 2.28 and 2.29). It remains possible that condensin slightly remodels nucleosomes, for example by catalyzing the dissociation of the H2A-H2B dimers interface with the tetramers (Andrews & Luger, 2011), allowing them to linearize during the passage of condensin. However, such an event would use some of the energy used by condensin and result in a slow down of DNA compaction, which we do not observe, arguing against this possibility. Moreover, because the histones would still be bound to the DNA during the passage of condensin, a relatively large active site would still be required if the motor is tracking the DNA backbone. Hence, we believe condensin’s remodeling activity, if it exists, is not a contender for explaining the lack of effect of nucleosomes on condensin’s compaction and motor activity.

Second, condensin does form loops with nucleosomal DNA, as can be observed from our dual-labeled experiments (fig. 2.30). This excludes the hypothesis that condensin forms loops exclusively with naked DNA. The vast majority of collisions between a DNA loop and a nucleosome...
results in a seemingly pause-less (at least, below our time resolution) integration of the nucleosome into the loop (figs. 2.31 and 2.34). While our time resolution does not allow to quantify the dwell time of the majority of the collisions, 15% of all integration events behave with a half-life of 65s. It is unclear whether this is a consequence of different condensin or nucleosome conformations, or because of an experimental artifact (misfolding, aggregation, etc.) In all cases, because increasing amounts of nucleosomes do not slow down overall DNA compaction by condensin (fig. 2.24), it appears that a longer pause by 15% of condensin molecules can be rescued by the other 85%. It is also important to note that a single loop can integrate several nucleosomes (fig. 2.29), implying that condensin can proceed on a low density of nucleosomes or on chromatin.

Overall, our data indicate that nucleosomes are not obstacles for condensin, and that condensin can extrude loops of DNA containing nucleosomes.

### 2.3.3 Mechanism of loop extrusion through a nucleosome by condensin

It is currently unclear how the DNA binding, loop extrusion, and ATPase activities of condensin are orchestrated (Hassler et al., 2018). The fact that the loop extrusion activity of condensin is not slowed down by nucleosomes can indicate that either the enzyme can accomodate a nucleosome into its active site, or that the nucleosome never encounters the active site. Two possible models are depicted in fig. 2.35.

If condensin topologically entraps DNA (Cuylen et al., 2011), then nucleosomes would have to go through the condensin ring (fig. 2.35A). If the ring is formed by the SMC subunits, the diameter of about 30 nm—considering each straight subunit is 50 nm long (Haering et al., 2002), a ring thus formed would have a diameter of 30 nm—could easily accomodate the 10nm nucleosomes, but
larger, higher-order structures could potentially prevent the loop extrusion activity of condensin. It is also possible that the topological ring is not formed by the SMC subunits. In the crystal structure of the DNA binding domain of condensin located on the “huntingtin, elongation factor 3, protein phosphatase 2A and TOR1 kinase” (HEAT) and kleisin subunits, Kschonsak et al. (2017) show that condensin “buckles up” DNA with a flexible linker. In budding yeast, this flexible linker is longer than in other species. It would be interesting to see whether reducing the size of this linker affects the outcome of collisions between nucleosomes and condensin.

Alternatively, it is possible that nucleosomes never encounter the active site of condensin, because the molecule does not entrap DNA topologically (fig. 2.35B). As depicted, condensin could use its kleisin-HEAT subunits to anchor to the DNA and its SMC subunits to “inchworm” its way on the DNA. Indeed, SMC subunits are flexible and the hinge domain was shown to make contacts with the HEAT subunits (Duan et al., 2009; Eeftens et al., 2016; Murayama & Uhlmann, 2015). In this model, condensin would be unaffected by the conformation of the DNA (heterochromatin, centromere, etc.), because it could slide and anchor around or into them.

Overall, our studies show that condensin extrudes loops of DNA, and that it does it in the context of nucleosomes as well as on naked DNA. Nucleosomes are not an obstacle for condensin and are readily integrated into growing loops of DNA being extruded by condensin. How exactly condensin assembles loops such that roadblocks are ignored, and how big of a roadblock it can ignore, remain to be determined.
2.4 Methods

Purification of histones

Escherichia coli strain Rosetta-2 were transformed with a pet11a plasmid containing one of the following S. cerevisiae genes: HTA1, HTB1, HHT1, HHF2 or one of the cysteine-mutant histones described above. 6L of bacteria were grown at 37°C in LB to O.D.600 = 0.6, expression was induced with 0.45 mM IPTG and the cells were harvested after 3h of expression. They were pelleted at 4000xg, 10 mins, 4°C, resuspended in 80 mL of Tris-Sucrose (50mM Tris-HCl pH8.0, 10% w/v sucrose, 1mM benzamidine) and frozen at -80°C overnight. The cells were then mixed with 10 mL of inclusion buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol, 1 mM benzamidine, 1% w/v Triton X-100) and sonicated for 3 mins at 60% power output, in sequences of 10 seconds on, 50 seconds off. The inclusion bodies were harvested by centrifugation at 20,000 xg for 20 mins, 4°C, and rinsed 4 times with inclusion buffer, the last two times in the absence of Triton X-100. Rinsing was done by resuspending the pellets in inclusion buffer, and centrifuging them afterwards. The inclusion bodies were then solubilized in 30 mL of unfolding buffer (20 mM Tris-HCl pH 7.5, 7 M guanidium-HCl, 10 mM DTT) by shaking them at room temperature for 1 h. They were spun down at 20,000 xg for 20 mins, 4°C, to pellet cell debris. The supernatant, containing the histones, was dialyzed against 1 L of urea buffer (10 mM Tris-HCl pH 8.0, 7 M urea, 1m M EDTA pH 8.0, 5 mM β-mercaptoethanol and either 100 mM NaCl for H2A and H2B or 200 mM NaCl for H3 or H4) in a 3,500 MWCO dialysis tubing overnight at 4°C, plus two more buffer changes every 2 h. Urea buffer was deionized using AG 501-X8 resin (#1437424, Bio-Rad) for several hours at room temperature or overnight at 4°C. Us-
ing an Akta FPLC system (GE Healthcare), dialyzed histones were then injected at 1 mL/min onto 80 mL of Q sepharose followed by 80 mL of SP sepharose ion exchange media (GE Healthcare) pre-equilibrated with urea buffer. The media was rinsed with more than 150 mL of urea buffer, the Q sepharose media was removed, and the histones bound to the SP sepharose media were eluted at 0.4 mL/min using a salt gradient to 1 M (in urea buffer) over 115 mL. Fractions were analyzed by SDS-Page gel and the ones containing the histone were merged and dialyzed in dialysis buffer (10 mM Tris-HCl pH 8.0, 5 mM β-mercaptoethanol) using some 3,500 MWCO dialysis tubing. The first dialysis was conducted overnight at 4°C, followed by at least 3 buffer changes at 2 h intervals, the 2 last ones in the absence of β-mercaptoethanol. Finally, the histones were snap-frozen in liquid nitrogen and lyophilized in a Labconco FreeZone 1 lyophilizer for at least 24h. Histone powder was stored at -20°C for subsequent use.

Assembly and purification of octamers

Around 5 mg of powder from each histone H2A, H2B, H3 and H4 (with at most one of the cysteine mutants) were dissolved in 1 mL of unfolding buffer (20 mM Tris-HCl pH 7.5, 7 M guanidium-HCl, 10 mM DTT) each, for 1 h at room temperature with agitation. The histone concentrations were measured using a Nanodrop (Thermo Fisher). Histones were mixed at equimolar ratio and the mixture was diluted to 1 g/L with unfolding buffer. They were subsequently dialyzed in TEB2000 (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 5 mM β-mercaptoethanol, 2 M NaCl) for 48 h with at least 4 buffer changes and concentrated to 1 mL using a 10,000 MWCO spin concentration device (Vivaspin 6, GE Healthcare). The octamers were purified at a flow rate of 0.2 mL/min on a Superdex S-200 16/60 size exclusion column (GE Healthcare) pre-equilibrated with TEB2000.
The fractions were resolved on SDS-PAGE and the ones containing the octamers were merged and concentrated using a Vivaspin 6 10,000 MWCO spin-concentration device until the concentration of octamers was superior to 20 µM. Finally, half a volume of glycerol was added to the octamers (final concentration of 50% glycerol) and they were stored at -20°C for subsequent use.

Assembly of nucleosomes

Lambda BioLDigR DNA was mixed with histone octamers at a final concentration of DNA of 1.6 nM and a molar ratio of 1:40 DNA:octamer. The DNA was first mixed with TEB2000 and NaCl to reach a final concentration of NaCl of 2 M, and the octamers were added afterwards. The reaction was transferred into a dialysis device made with the cap of a PCR tube and a piece of 10,000 MWCO SnakeSkin dialysis tubing (#68100, Thermo Fisher). The dialysis device was inserted on a floater and incubated in a beaker with 100 mL of TEB2000, with agitation. 40 mL of TE (10mM Tris-HCl pH8.0, 1mM EDTA pH8.0) were added to the beaker at a rate of 0.5mL/min using a peristaltic pump, and the dialysis device was then incubated in 100 mL of TE for at least 2h, with agitation. Finally, the dialysis membrane was pierced, the nucleosomal DNA was collected in an eppendorf and spun down at 4°C, 15,000 xg for 5 mins, and kept at 4°C for subsequent use.

Nucleosome curtains

Nucleosome curtains were assembled as DNA curtains, described previously (Greene et al., 2010). All buffers were devoid of salts, to prevent the exchange of histones, and devoid of magnesium, to prevent the compaction of DNA by nucleosomes. Before the actual experiment, working buffer containing salt and magnesium is injected at 1 mL/min for 3 mL to flush out the non-nucleosomal
histones and prevent their binding to the surface.

**Single-molecule DNA condensation experiments**

A DNA Curtains or Nucleosome Curtains experiment was prepared (see Greene et al., 2010 and above). The flow cell was washed with 3 mL of condensin-ATP buffer (Condensin ATP Buffer (CAB, 40mM Tris-HCl pH 7.5, 0.5g/L BSA, 125 mM NaCl, 5mM MgCl$_2$, 1mM DTT, 4 mM ATP) containing 0.5 nM YoYoI when needed. The flowcell was incubated in CAB buffer for 5 minutes, and then condensin at concentrations of 1 or 5 nM was injected at 0.1 mL/min for 30 mins.

**Quantification of condensation events**

A single-molecule DNA condensation experiment (see above) was prepared. Kymograms of the DNA molecules were collected and the number of kymograms showing wedges and bright spots was counted by hand. The data was plotted using Seaborn (Waskom et al., 2018).

**Quantification of the number of puncta per condensation event**

A single-molecule DNA condensation experiment (see above) was prepared. Kymograms of the DNA molecules were collected and the number of resolvable puncta for each compaction event was recorded. The data was plotted using Seaborn (Waskom et al., 2018).

**Quantification of colocalization between labeled condensin and DNA loops**

A single-molecule DNA condensation experiment (see above) was prepared, but condensin was preincubated at 1 μM for 5 mins with 3 μM of Qdot-705-labeled anti-FLAG antibody, on ice.
The labeled condensin was diluted to 1 nM and the experiment proceeded as described above. Kymograms of the DNA molecules were collected and the events of labeled condensin only, labeled condensin with a DNA punctum, or a DNA punctum only were recorded. The data was plotted using Seaborn (Waskom et al., 2018).

**SDS-PAGE of dye-labeled histone octamers**

A 15% SDS polyacrylimide gel was prepared. Around 15 pmol of histone octamer was loaded in each lane. The gel was resolved for 1h20 at room temperature, 160 V. The signal from the dyes was collected using a GE Typhoon FLA9000 laser scanner at 532 nm. The gel was then stained with Coomassie Brilliant Blue and digitized with a scanner.

**Flow on/flow off experiment**

A nucleosome curtains experiment was prepared as described above, using a reconstitution molar ratio of 1:40 (DNA:histone octamer). DNA was left to diffuse, a barrier with high density of DNA was picked, and buffer flow was applied at a rate of 1mL/min. The flow was cut off using a stopcock valve for 5 s, and then released to attest of the presence of nucleosomes on DNA molecules exclusively.

**Quantification of the number of nucleosomes and nucleosome signal per DNA molecule**

A Nucleosome Curtains experiment was prepared with labeled nucleosomes. DNA was extended with a flow rate of 1 mL/min. Subsequently, kymograms were generated, and the number of nu-
cleosome spots per DNA molecule was quantified by hand. Alternatively, the total signal on the DNA molecule was integrated. The data was plotted using Seaborn (Waskom et al., 2018).

**Position distribution of nucleosomes on lambda DNA**

A Nucleosome Curtains experiment was prepared with labeled nucleosomes. DNA was extended with a flow rate of 1 mL/min. The signal over a whole barrier was integrated as a function of the distance from the barrier using FiJi (Schindelin et al., 2012). To quantify the AT-richness of the λDNA, the sequence was obtained from NEB, and the number of A/T base pairs per 1000 bp was counted. The data was plotted using Seaborn (Waskom et al., 2018).

**Survival distribution of nucleosomes in CAB**

A Nucleosome Curtains experiment was prepared with labeled nucleosomes reconstituted on DNA at a 1:60 molar ratio. The DNA was lightly extended with 0.1 mL/min flow in condensin-ATP buffer to mimick the conditions of a condensin experiment. Frames were recorded every 60s. The lifetime of every trace was recorded by hand and plotted as a survival distribution using Lifelines (Davidson-Pilon et al., 2018). The error was calculated by bootstrap. The data was plotted using Matplotlib (Caswell et al., 2018).

**Calculation of photobleaching steps**

A Nucleosome Curtains experiment was prepared with Dt-530 labeled nucleosomes reconstituted on DNA at a 1:40 molar ratio. The DNA was extended at 1 mL/min and imaged at 5 frames/second until most of the nucleosome signal had disappeared. Kymograms of the DNA molecules were gen-
erated and the intensity profile of the molecules was generated using cats (C. Moevus, 2018). The number of steps in the intensity profile was counted by hand and plotted using Seaborn (Waskom et al., 2018).

**Calculation of compaction velocity and processivity**

Single-molecule DNA condensation experiments were performed as described above. Kymograms of the DNA molecules were generated and each compaction event was tracked by hand by selecting the points that define the compaction lines. To measure velocity, the slopes were calculated in between each inflexion point and recorded. For processivity, the horizontal distance between the points for all segment with a negative slope (going up) was added, for each continuous compaction event. The data was plotted using Seaborn (Waskom et al., 2018).

**Quantification of the outcome of collisions**

Single-molecule DNA condensation experiments were performed with Atto-647N-labeled H2A S47C nucleosomes reconstituted on lambda DNA at a 1:60 ratio. Kymograms were generated and the outcome of analyzable collisions was classified by hand. The data was plotted using Matplotlib (Caswell et al., 2018).
2.5 Figures

Figure 2.1: Presentation of the dCas9 system to study DNA condensation. (A) Representation of a DNA Curtains experiment with the DNA labeled at its end using QDot705-Cas9. (B) Actual image of a DNA Curtains barrier where DNA molecules are extended and dCas9 molecules are bound. The DNA molecules are not labeled and thus invisible. The dCas9 molecules labeled by Qdot 705 (magenta) are supposed to bind 6.4 kbp away from the DNA end.
Figure 2.2: Schematic drawing of the compaction experiment. A DNA Curtains flowcell was prepared as previously described (Greene, Wind, Fazio, Gorman, & Visnapuu, 2010) and condensin was continuously injected at 0.1 mL/min for the duration of the experiment. In the presence of buffer flow (blue arrow), the DNA is extended in its length against the surface, making it visible under the objective. When condensin binds to DNA, it starts to compact it, presumably via loop extrusion, resulting in the apparent shortening of the DNA molecule, and, if DNA is stained with an intercalating dye, in the formation of a bright punctum at the position of the forming DNA loop.
Figure 2.3: Kymograms of DNA compaction by condensin as a function of ATP hydrolysis. The change of position of dCas9 labeled by QdotTM 705 (magenta) depends on the injected condensin (wild type or ATPase mutant [SMC2 Q147L, SMC4 Q302L]) and on the substrate (ATP or ATPγS) molecules. Wild type condensin and ATP molecules were injected in the top panel. Mutant condensin and ATP molecules were injected in the middle panel. Wild type condensin and ATPγS molecules were injected in the bottom panel.
Figure 2.4: Histogram of compaction velocity of deca9-QDot labeled DNA by condensin. The plot was fitted by sum (black) of two Gaussian distributions (blue and red, respectively). The means of the red and blue distributions are $0.4 \pm 0.2$ [kbp/sec] and $0.8 \pm 0.3$ [kbp/sec], respectively.
Figure 2.5: Three representative kymograms showing change of the position of condensin labeled by Qdot™ 605 (green) and dCas9 labeled by Qdot™ 705 (magenta) after injecting condensin and ATP molecules. The distance between barrier and the condensin molecule and its time derivative are defined as a and Va, respectively. Similarly, the distance between condensin and dCas9 molecules and its time derivative are defined as b and Vb, respectively. The top kymogram is the example in which both Va and Vb have finite values as the DNA molecule is compacting. The middle (bottom) kymogram is the example in which Va (Vb) is almost zero and Vb (Va) has a finite value. The right-side values represent the number of events per the number of all the compaction events. The white bars represent length and time scales, respectively.
Figure 2.6: The plot showing $V_a$ versus $V_b$. This plot shows that asymmetric compactions ($V_a > 0.2$, $V_b < 0.2$ or $V_a < 0.2$, $V_b > 0.2$) were realized in the 66% of the compaction events.

Figure 2.7: Barrier view of the compaction of YoYoI-labeled DNA by condensin. DNA molecules labeled with 0.5 nM YoYoI were observed at 0.1 mL/min (A) before and (B) after the addition of 1nM unlabeled condensin. The addition of condensin results in the formation of puncta on the DNA molecules, as well as their shortening, as depicted in fig. 2.2.
Figure 2.8: Kymogram of a single DNA molecule compacted by condensin. A YoYoI-labeled DNA molecule, like those showed in fig. 2.7, is tracked over time in the presence of 1 nM unlabeled condensin. Compaction events, depicted by green arrows, lead to the the apparition of increasingly bright puncta. Upon release (red arrows), the DNA can extend slowly (first red arrow) or fast (second red arrow). Loops can also move in the absence of DNA compaction (blue arrow), as observed from the constant length of the DNA molecule and constant intensity of the punctum, albeit its movement.

Figure 2.9: Quantification of the fraction of DNA molecules compacted in the presence or absence of condensin. Experiments as described in fig. 2.2 were observed for 30 minutes in the absence or presence of 1 nM condensin and quantified. In the absence of condensin, no compaction is observed, whereas in the presence of 1 nM of unlabeled condensin, almost all of the DNA becomes compacted.
Figure 2.10: Quantification of the number of puncta per compaction events. Experiments were performed for 30 minutes in the presence of 0.5 nM YoYoI and 1 nM condensin, as shown in fig. 2.2, and the number of puncta participating in each compaction event was recorded. Most compaction events only happen with 1 punctum. The number of condensin molecules per punctum is unknown.
Figure 2.11: Kymogram of formation and compaction of DNA loops by condensin. Experiments were performed for 30 minutes in the presence of 0.5 nM YoYoI and 1 nM condensin, as shown in fig. 2.2. Condensin forms loops of DNA, as can be seen from the trailing DNA signal that gets longer as compaction happens (green arrow). The extension is observed because the buffer flow extends the DNA. Interestingly, the loop is further condensed, potentially by (an)other condensin(s) (third green arrow). As the loop compacts, it gets smaller and smaller until the flow is not strong enough to extend it. Upon addition of stronger flow, the loop is extended further, demonstrating that it still is a loop.
Figure 2.12: Kymogram of YoYoI-labeled DNA molecule (green) with QDot705-labeled condensin (magenta). 1 nM of condensin was injected at 0.1 mL/min for 30 mins. The DNA punctum colocalizes with condensin, demonstrating that condensin is responsible for the DNA compaction and the formation of the punctum.
Figure 2.13: Quantification of the colocalization between condensin and DNA puncta. The number of DNA puncta, condensin signal, and their overlap was recorded over a period of 30 mins and normalized per DNA molecule. The majority of DNA puncta does not colocalize with condensin, and the majority of labeled condensin does not colocalize with DNA puncta. On average, 1.11 puncta and 0.53 condensin are observed per DNA molecules, with only 0.22 of these colocalizing.
Figure 2.14: SDS PAGE of WT and H2AS47C histone octamers. The proteins were stained using a Coomassie Protein Stain solution, and the dye signal was recorded at 532nm. The percentage of labeling specificity was quantified using Fiji (Schindelin et al., 2012).

Figure 2.15: MNase assay of reconstituted nucleosomes. Nucleosomes reconstituted with either WT or H2AS47CDY530 octamers were incubated for 30 minutes with or without MNase. The reactions were quenched with EGTA, SDS and Proteinase K and subsequently extracted with phenol:chloroform. The DNA fraction was run on a 2% agarose gel for 1 hour.
Figure 2.16: DNA Curtains experiment with H2A S47C DY530 nucleosomes. Nucleosomes are illuminated with a 40mW 532nm laser using total internal reflection fluorescence (TIRF) microscopy. Every "line” is a DNA molecule, while every signal punctum is a nucleosome.
Figure 2.17: Kymogram of a flow on/flow off on a nucleosome-reconstituted DNA molecule. (A) Representation of the data below. When the flow is on, the DNA is extended, and the nucleosome signal is visible. In the absence of flow, the DNA retracts, and the nucleosome signal is at the barrier. This experiment demonstrates that nucleosomes are bound to DNA, and not to the surface. Arrows represent the histone signal. (B) Actual kymogram of a single DNA molecule from the field of view shown in fig. 2.16. Only nucleosomes are tagged, DNA is unlabeled. With flow on at a rate of 1mL/min, the DNA molecule is extended against the barrier, lining up with the plane of the objective, allowing to see it in its full length. Subsequently, flow is temporarily removed, and the DNA molecule returns to its loose conformation against the barrier, causing the disappearance of the nucleosome signal, and hence indicating that all the nucleosomes are bound to the DNA molecule, and not to the surface. When flow is restored to a rate of 1mL/min, the nucleosome signal reappears at the exact same positions, indicating that this process is reversible, as expected.
Figure 2.18: Number of nucleosomes per DNA molecule as a function of the reconstitution ratio. Nucleosomes were reconstituted on lambda phage genomic DNA (48.5kbp) at different ratios of DNA to histone octamers using histone octamers labeled on H2A S47C with DY530. The number of nucleosomes per DNA molecule was counted using a DNA curtains experiment, in condensin-ATP buffer, in the presence of 1mL/min flow. The resulting distributions were plotted as box plots.
Figure 2.19: Nucleosomes signal per DNA molecule as a function of the reconstitution ratio. Nucleosomes were reconstituted at increasing molar ratios of DNA to histone octamers, and the number of nucleosomes per DNA molecule was measured. As expected, a higher ratio towards octamers leads to more nucleosomes per DNA molecule.
Figure 2.20: Nucleosome binding as function of AT-richness of the underlying sequence. Signal from fig. 2.16 was plotted as a function of the distance from the barrier using Fiji (Schindelin et al., 2012), which equates to position on the DNA molecule with a ratio of 1 pixel = 1 kb. Dye-labeled nucleosomes present a preference for GC-rich regions of the phage lambda genomic DNA, probably due to higher poly(dA-dT) content of the AT-rich region (Segal & Widom, 2009). Blue: nucleosome signal, orange: AT-richness of the underlying DNA sequence.
Figure 2.21: Survival probability of nucleosomes in CAB buffer. Nucleosomes were reconstituted on lambda phage genomic DNA (48.5 kbp) at a 1 to 60 ratio of DNA to histone octamers using histone octamers labeled on H2AS47C with DY530. The DNA was lightly extended with 0.1 mL/min flow in condensin-ATP buffer to mimick the conditions of a condensin experiment. Frames were recorded every 60s. The lifetime of every trace was recorded by hand and plotted as a survival distribution. The error was calculated by bootstrap. More than 90% of the signal nucleosomes remained after 100 mins, indicating that nucleosomes are extremely stable even at low concentrations, in the presence of salts and magnesium. The observed loss of signal may have been a consequence of photobleaching rather than histone dissociation, although the photobleaching rate was not measured.
Figure 2.22: Photobleaching steps of H2A S47C DY530 labeled nucleosomes. Lambda DNA was reconstituted at a ratio of 1 to 40 with octamers labeled on H2AS47C with DY530. Nucleosomes were observed in condensin-ATP buffer with 1mL/min flow rate. Only the nucleosomes were labeled, DNA was unstained. The field of view was exposed to a 40mW laser until most of the signal had photobleached, and photobleaching steps for each individual spot were counted by hand. Most of the spots displayed one or two photobleaching step, consistent with an incompletely labeled set of histone octamers. Few other nucleosomes displayed a 2 or 3 steps photobleaching behavior, potentially due to the low probability of labeling on H4 (see fig. 2.14 or the presence of multiple nucleosomes in the same pixel. The inset shows a representative example of a 2-steps photobleaching behavior exhibited by one of the spots analyzed.
Figure 2.23: Kymogram of the condensation of nucleosomal DNA by 5 nM condensin. Nucleosomes were reconstituted on lambda phage genomic DNA at a 1:80 molar ratio with WT histone octamer. A nucleosome curtains experiment was prepared, and the DNA is labeled with 0.5nM YoYoI. The nucleosomes are not labeled. 5 nM of condensin in CAB buffer were injected at 0.1 mL/min for 30 mins. The beginning of DNA compaction is suggested by the green arrow.
Figure 2.24: Condensation velocity as a function of nucleosome concentration. The velocity of DNA condensation by 5nM of condensin injected at 0.1mL/min is plotted as a function of different DNA:octamer reconstitution ratios, a higher ratio giving a higher number of nucleosomes per DNA molecule (see figs. 2.18 and 2.19). Velocities were tracked from kymograms by hand, by marking the beginning and end of the condensation events and calculating the slopes from these points. The value in pixel per frames was then converted into μm per seconds. Each velocity measured is shown as one point in the colored swarm plots. The average value for each ratio is shown in black. The slope was fitted with a linear curve and the equation is given in the top right corner.
Figure 2.25: Possible outcomes of the interaction between condensin and nucleosomes during loop extrusion. Upon collision with a nucleosome, (A) loop extrusion by condensin could be inhibited, requiring, for example, the histone chaperone *Facilitates access to chromatin* (FACT) to relieve the nucleosome barrier (Shintomi, Takahashi, & Hirano, 2015), (B) the nucleosome could be integrated into the loop, for example because the ring of condensin is large enough to integrate a nucleosome, (C) the nucleosome could be disassembled, and with the loop extrusion proceeding past, suggesting that condensin exhibits nucleosome remodeling functions.
Figure 2.26: Nucleosomes remain bound through complete and reversible compaction of nucleosomal DNA by condensin. In either control (left) or +condensin (right) experiments, DY530-labels on H2A S47C-containing nucleosomes assembled in DNA molecules (DNA:octamer = 1:80) are first extended and visualized in 1 mL/min flow (top). Following either a mock injection of reaction buffer only, or an injection of 5 nM condensin in reaction buffer, and a 10-minute incubation at room temperature in the absence of flow, DNA is subject to 0.1 mL/min flow and visualized (middle). DNA molecules in control experiments are extended by flow (left), whereas those incubated with condensin have been fully compacted and resistant to flow extension (right). Finally, DNA is washed for 5 minutes with buffer containing 500 mM NaCl at 1 mL/min flow rate (bottom). Nucleosomal DNA incubated without condensin (left) appears similar to pre-injection state, with a slight loss in signal. DNA previously incubated with and compacted by condensin (right) is also reversibly extended, with a similar loss in signal.
Figure 2.27: Compaction of nucleosomal DNA is dependent on condensin ATPase activity. In either Q-loop mutant (left) or ATPγS (right) negative control experiments, DY530-labels on H2A S47C-containing nucleosomes assembled in DNA molecules (DNA:octamer = 1:80) are first extended and visualized in 1 mL/min flow. Following injection of either 5 nM condensin Q-loop mutant in the presence of 4 mM ATP (left), or an injection of 5 nM wildtype condensin in the presence of 4 mM ATPγS (right), and a 10-minute incubation at room temperature in the absence of flow, DNA is subject to 0.1 mL/min flow and visualized. DNA molecules in both control experiments are extended by flow. As a positive control, the flow cell is washed with buffer before an injection of 5 nM wild type condensin in the presence of 4 mM ATP. After incubating the reaction for 10 minutes at room temperature, both control experiments showed complete compaction under 0.1 mL/min flow. Finally, DNA is washed for 5 minutes with buffer containing 500 mM NaCl at 1 mL/min flow rate. DNA previously compacted by condensin is again reversibly extended. Scale bar = 5 μm
Figure 2.28: Quantification of H2A S47C DY530 signal intensity before compaction and after induced decompaction. H2A S47C DY530 signal from single molecules of nucleosomal DNA is quantified at the pre-injection (panel C, top) and post-salt wash (panel C, bottom) steps, averaged, and normalized to the pre-injection level. Signal intensities at these two stages, as a result of the combined effects of incubation in the absence or presence of condensin and salt wash, is plotted as percentages of pre-injection intensities. Error bars represent standard deviations from 2-3 repeat experiments.
Figure 2.29: Kymogram of the spontaneous release of H2A S47C DY530 nucleosomal DNA by condensin. Nucleosomes were reconstituted at a 1:60 ratio (DNA:Octamer) with octamers labeled on H2A S47C using DY530. The visible signal comes exclusively from the nucleosomes, labeled with an asterisk, as both DNA and condensin are unlabeled. Condensin is injected in the flow cell at a concentration of 1 nM, at a flow rate of 0.1 mL/min in condensin-ATP buffer. This flow rate lightly extends the DNA but does not prevent condensin's action. Green arrows indicate the beginning of loop formation by condensin while a red arrow indicates the release of a loop. When loops form, the DNA is seen as getting shorter, as its content is integrated into the loop. One can see that the nucleosomes are readily integrated into the loops formed by condensin, and that once the loops are released, they are found at the exact same position as they were, indicating that condensin does not remodel or evict nucleosomes while forming loops.
Figure 2.30: Kymogram of continued compaction over a nucleosome by a loop of DNA being extruded by condensin. Nucleosomes labeled on H2A S47C with ATTO647N (magenta) were reconstituted at a ratio of 1:60 on lambda phage genomic DNA labeled with YoYoI (green). A nucleosome curtains experiment was prepared. Movies were recorded at 1 frame per 5 seconds, in condensin-ATP buffer with 0.5 nM Yoyo-1 and 1 nM unlabeled condensin. axis. The same kymogram is shown in the DNA and nucleosome channels, and then as a merge of the two channels. DNA condensation does not pause at nucleosomes, demonstrating that nucleosomes are not a barrier to DNA condensation by condensin. All scale bars indicate 50s in the horizontal axis and 3 µm in the vertical.
Figure 2.31: Quantification of the outcomes of collisions between a DNA loop extruded by condensin and a nucleosome from the same data sets as fig. 2.30. The vast majority of collisions results in bypass, where nucleosomes are integrated into the loop.
Figure 2.32: Comparison of the velocity of loop extrusion in the presence (orange) and absence (blue) of nucleosomes indicates no statistical difference in the distributions, as tested by a two-sample Kolmogorov–Smirnov test with a significance level of 0.05 (KS-statistic = 0.08017). The mean ± standard deviation of the distribution of velocities are (234 ± 214) bp/s in the absence of nucleosomes and (269 ± 233) bp/s in the presence of nucleosomes.
Figure 2.33: Comparison of processivities between loops extruded in the presence (orange) and absence (blue) of nucleosomes. Median processivities under these two conditions are essentially identical (10,750 bp) and the difference is not statistically significant, as tested by a logrank test (p-value = 0.63).
Figure 2.34: Survival distribution of DNA loops integrating a nucleosome. The dwell time of all loops at a nucleosome that resulted in the integration of the nucleosome in the loop were measured and plotted as a survival distribution. Around 85% of all collisions lasted for less than 5s, our time resolution. Because it is under our time resolution, the half-life of the interaction cannot be measured. The other population, around 15% of the collisions, displays a half-life of 65s. The data was fitted to a sum of exponential decays.
Figure 2.35: Potential mechanisms of nucleosome bypass during loop extrusion by condensin. (A) If condensin forms a topologically closed loop around DNA, nucleosomes would have to go through the active site, potentially in between the SMC subunits, as depicted here. This model implies that maximum size of an obstacle condensin can extrude a loop through is defined by the size of the ring. (B) Alternatively, if condensin does not entraps DNA, then the enzyme can extrude DNA loops independently of the diameter of the chromatin. One such model, depicted here, would happen by the “inchworming” of the SMC subunits along the DNA.
Conclusion

My thesis was centered around the development of Nucleosome Curtains for the high-throughput study of nucleosomes at the single-molecule level. In addition, I have applied Nucleosome Curtains to study the effect of nucleosomes on loop extrusion by condensin.

Nucleosomes are the basic unit of chromatin. They are composed of an octamer of histone proteins—(H2A-H2B)2(H3-H4)2—around which wrap 147 bp of DNA. Since the discovery of nucleosomes in 1975 (Oudet et al., 1975), a tremendous amount of work has been done to better understand nucleosomes and their effect on DNA-based processes. The nucleosomes were crystallized, and their subunits, post-translational modifications, dynamics and regulation were parsed out. Because nucleosomes directly protect DNA, most DNA-based processes have had to develop mechanisms to handle them, or to exploit their inherent dynamics. Indeed, nucleosomes are generally inhibitory to DNA access. As such, a large body of literature was dedicated to the study of the interactions between nucleosomes and transcription, DNA repair, and replication. However, the tools used to study nucleosomes showed their limitations. Indeed, traditional ensemble tools, such as gel electrophoresis or fluorimetry, cannot easily resolve the behavior of single-molecule, and hence are prone to loose the information about the static and dynamic heterogeneities in a population due to averaging. Static and dynamic heterogeneities, namely the different states of a molecule in a population and their fluctuations over time, can explain the distribution of a popula-
tion, providing with an extra layer of precision. Sensitive single-molecule experimental tools were used to solve the limitations of ensemble techniques. For example, single-molecule Forster energy resonance transfer (FRET) was used to demonstrate transient site exposure in nucleosomes, which are random and hence hard to synchronize and resolve temporarily using ensemble methods.

Single-molecule techniques have their own flaws, one of which is their relatively low throughput. Another one, more specific to techniques based on total internal reflection fluorescence (TIRF) microscopy, is the interactions of the molecules of interest with the surface. Indeed, it was shown that even passivated surfaces will affect the integrity of nucleosomes (Koopmans et al., 2007). To solve these issues, DNA Curtains rely on a lipid bilayer, which is biologically relevant, can passivate a surface, and can be used to concentrate DNA molecules against defined barriers. DNA Curtains is a single-molecule, TIRF microscopy-based technique that allows for the observation of hundreds of DNA molecules in real-time, at the single-molecule level, by concentrating DNA molecules under the objective and passivating surfaces with lipids a biologically-relevant substrate. DNA Curtains are especially adapted to study the interactions and movement of proteins on DNA. Hence, it is a great platform to study nucleosomes and their interactions with DNA-based processes.

I first implemented a protocol to study nucleosomes using DNA Curtains (chapter 1). While nucleosomes were previously studied using DNA Curtains, the protocol relied on labeling with quantum dots, which are bulky and can inhibit interactions due to steric hindrance. In addition, quantum dots were attached to nucleosomes using a FLAG-tag antibody, which led to low labeling efficiency. To solve these issues, I labeled recombinant nucleosomes covalently with small organic dyes using maleimide conjugation. Organic dyes have a mass weight of about 1/100th of an histone octamer. I selected the dyes Atto-647N, Atto-565, and Dyomics-530 as the fluorophores with the best properties for DNA Curtains experiments. For labeling nucleosomes from S. cerevisiae with
maleimide-conjugated dyes, I mutated amino acids into cysteines at the sites H2A S47, H2B T119, and H4 S48. Dye-labeled *S. cerevisiae* histones protect the expected 147 bp (figs. 1.10 and 2.15) and give an amount of photobleaching steps consistent with full nucleosomes but incomplete labeling (fig. 2.22). On a Nucleosome Curtains experiment, they appear as single puncta (figs. 1.17 and 2.16). Nucleosome Curtains are assembled exactly like DNA Curtains, except that nucleosomal DNA is used instead of naked DNA, and that the nucleosomal DNA has to be incubated in the flow chamber in the absence of salts. Indeed, in the presence of salts, non-nucleosomal histones can freely exchange and bind non-reversibly to the surface, leading to poor signal-to-noise ratios (fig. 1.18). Overall, Nucleosomal Curtains is a stable, working technique that enables the study of nucleosomes and their interactions with proteins in real time, at the single molecule level, with a high-throughput.

Second, I used the Nucleosome Curtains technique to study the effect of nucleosomes on DNA condensation by condensin (chapter 2). During mitosis, interphase chromatin morphs into compact chromosomes. This process is dependent on condensin, which was shown to form chromosomes by extruding loops of DNA (Ganji et al., 2018; Gibcus et al., 2018; Goloborodko et al., 2016; Terakawa et al., 2017). However, it was unclear whether this process could proceed in the presence of nucleosomes. Indeed, the morphogenesis of chromosomes can be reproduced in vitro from *X. laevis* sperm DNA with just 6 factors: nucleoplasmin (to remove sperm proteins), histones, Nap-1 and *Facilitates access to chromatin* (FACT) (histone chaperones), topoisomerase II, and condensin (Shintomi et al., 2015). It was shown that the absence of FACT prevents chromatin condensation, and it was proposed that nucleosomes have to be handled by the chaperone to allow for the proper function of condensin. Hence, the hypothesis was that nucleosomes inhibit loop extrusion by condensin. To test this hypothesis, DNA condensation by condensin was first observed using DNA Curtains
(fig. 2.8). Next, the experiment was recapitulated using unlabeled nucleosome curtains, showing that the speed of DNA condensation by condensin is not affected by nucleosomes (fig. 2.24). Finally, the direct interaction between compacting DNA and nucleosomes was observed, showing that 80% of the time, nucleosomes are readily compacted by condensin, without a visible slowdown (figs. 2.30 to 2.32). Hence, Nucleosome Curtains have revealed that nucleosomes are not a barrier to DNA condensation by condensin, and that condensin can readily condense nucleosomal DNA.

In conclusion, my thesis had led to the development of a new single-molecule technique called Nucleosome Curtains for the high-throughput study of nucleosomes in real time, at the single-molecule level. This technique was applied to the study of DNA condensation by condensin, showing that condensin is not affected by the presence of nucleosomes. Nucleosome Curtains is a promising technique for the study of nucleosome and chromatin-based processes. It has already been applied to study DNA condensation, DNA repair (C. Xue et al., submitted), cohesin (Stigler et al., 2016), and replication (data not shown).
Future Directions

Nucleosome Curtains is a powerful new tool for the study of nucleosomes at the single-molecule resolution. Future work can address the discrepancy observed between the protection in micrococcal nuclease assays and the photobleaching steps at the single molecule level (see section 1.14). Notably, using a more efficient protocol for the separation of unconjugated dyes, such as using small size exclusion spin columns, will help estimate precisely the labeling efficiency and the number of subunits per nucleosomes. In addition, it will be interesting to label the nucleosomes both on the H3-H4 tetramers and H2A-H2B dimers using different dyes. This could be achieved by purifying cysteine mutants dimers and tetramers separately, labeling them individually, and then using both labeled subunits to reconstitute nucleosomes. Alternatively, single histones could be labeled, and then octamers could be assembled with the two labeled histones at once. Finally, another exciting avenue to explore with Nucleosome Curtains is the formation of chromatin domains. Indeed, Nucleosome Curtains are usually composed of nucleosomes at low density, so that single nucleosomes can be observed. However, in vivo, nucleosomes are packed tightly, potentially leading to an added challenge for DNA-based processes as compared to single nucleosomes. The complete saturation of DNA with nucleosomes in a Nucleosome Curtains experiment results in the tight compaction of the DNA, to an extent where the study of proteins on DNA becomes hard because the DNA is too compacted to observe movement. To observe chromatin in Nucleosome Curtains, a domain of
repeated strong positioning sequences, such as the 601 sequence, could be inserted in the lambda phage genomic DNA to force nucleosomes to position close to each other. To saturate the repeats, one could incubate 1 μM of histones with 1 μM of 601 sequences, some of which would be inserted in the λ DNA, and the rest supplied as 147 bp molecules. Since nucleosome deposition approximates 100% at 1 μM, most of the 601 sequences would form nucleosomes, and because the number of histone octamer equals the number of 601 sequences, close to none of the non-601 sequences would form nucleosomes, leading to clean, saturated chromatin domains. Alternatively, chromatin remodelers such as ACF or Switch/Sucrose Non-Fermentable (SWI/SNF) could be used to evenly space nucleosomes, although the predictability of the size and position of the chromatin domains with this technique remains to be clarified.

The application of Nucleosome Curtains to study condensin has shown that nucleosomes have no visible effect on DNA condensation. This discovery implies a new mechanism for the motor activity of condensin. Indeed, condensin only hydrolyzes about 2 ATP molecules per second, versus hundreds to thousands for usual DNA translocases (Hassler et al., 2018). Hence, potential models need to accommodate a large pore size into which a nucleosome can go through, and a large step size that explains the low ATP hydrolysis. This could be explained by the topological enclosing of DNA into the Smc2-Smc4 ring, which if perfectly circular would have a diameter of about 30 nm. Alternatively, this could be explained by the inchworming of condensin on DNA, using its DNA binding site in the hinge domain (fig. 2.35). To separate these models, it would be interesting to see whether there is a size limit to the DNA roadblocks during DNA condensation. One could label nucleosomes with a quantum dot, or with several quantum dots, and see whether condensin can still go through. If not, it is highly likely that condensin encloses DNA topologically, favoring the first model. In addition, further studies will be necessary to link the conformational changes
of condensin with its DNA condensation activities. Here, the usage of Forster energy resonance transfer (FRET), especially in combination with DNA Curtains, would prove very useful. For example, one could label the hinge domain and the HEAT or kleisin subunits of condensin with a donor and acceptor dye, and see whether the motor activity of condensin leads to cycles of proximity and distance between the dyes, and whether the velocity of condensin correlates with the transition rate between the distances. Likewise, FRET pairs located on each Smc subunit at the hinge domain would help determine whether the hinge opens, and how this opening correlates with DNA condensation velocity. Another interesting experiment will consist in confirming the result on single nucleosomes with domains of chromatin. Indeed, one could imagine that repeated, dense nucleosomes arrays could impede condensin binding to DNA. In addition, with large enough arrays, the effect of higher-order nucleosome structures, such as the 30 nm fiber, could be assessed. Finally, another area of research, potentially beyond the scope of DNA and Nucleosome Curtains, will be to address the discrepancy between DNA condensation time in vivo and in vitro. Indeed, compaction can happen at up to 1.5 kb/s in vitro (Ganji et al., 2018). On the other hand, in *C. elegans*, condensation takes about 8 minutes (Maddox, Portier, Desai, & Oegema, 2006). At 1.5 kb/s, if condensation was fully processive and happened linearly, only 140 condensin molecules would be necessary to condense the genome in 8 minutes. However, tens to hundreds of thousands of condensin molecules are present in eukaryotic cells during mitosis (Walther et al., 2018), which implies that each condensin probably proceeds at slower speeds. It could be that because DNA condensation is random and distributive, a lot of condensation events are “lost”, with loops constantly releasing and reforming, leading to a slower apparent velocity, while the velocity of single condensin molecules is in par with the measurements in vitro. Alternatively, it could be that barriers control the condensation of DNA, and hence that condensation velocity and/or processivity are
limited in vivo. A good first experiment would be to properly time DNA condensation simulations using the condensation velocity and processivity obtained in vitro and the number of condensin molecules calculated in vivo. If condensation happens faster in the simulations than in vivo, it would indicate that further research is necessary to understand the spatio-temporal regulation of DNA condensation.
References


218


Charville, G. W., & Rando, T. A. (2013, August 1). The mortal strand hypothesis: Non-random chromosome inheritance and the biased segregation of damaged DNA. *Segmenting DNA asymmetrically during cell division, 24*(8), 653–660. doi:10/gfgq2f


224


Davey, C. A., Sargent, D. F., Lugker, K., Maeder, A. W., & Richmond, T. J. (2002, June 21). Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution††We dedicate this paper to the memory of Max Perutz who was particularly inspirational and supportive to T.J.R. in the early stages of this study. Journal of Molecular Biology, 319(5), 1097–1113. doi:10/frhjqr


Finch, J. T., & Klug, A. (1976, June 1). Solenoidal model for superstructure in chromatin. *Proceedings of the National Academy of Sciences*, 73(6), 1897. doi:10/fwgw5n


229


Georgieva, M., Roguev, A., Balashev, K., Zlatanova, J., & Miloshev, G. (2012, May 1). Hho1p, the linker histone of Saccharomyces cerevisiae, is important for the proper chromatin organization in vivo. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 1819*(5), 366–374. doi:10/fx52t4


243


252


254


255


SiteClick Qdot 705 Antibody Labeling Kit - Thermo Fisher Scientific. (n.d.).


Thoma, F., Koller, T., & Klug, A. (1979, November 1). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. The Journal of Cell Biology, 83(2), 403–427. pmid: 387806


261


263


Appendix: Protocols

All protocols can be found in section 2.4. Additional or detailed protocols are given here.

A Inverse PCR Mutagenesis

Background

This protocol is used to mutate base pairs using a PCR reaction. I originally obtained it from Bryan Gibbs. In short, two primers amplify away from the mutated base, amplifying the whole plasmid (hence it may prove problematic for large plasmids, and one should use cut sequences). The first primer contains the mutated bases, while the other starts at the base right before the mutated ones, on the opposite strand. Note that you could also use this protocol to insert a sequence, following the same principles.

Steps

1. Design the primers. The primers will amplify in opposite directions from the mutated site. The top primer will contain the mutated base(s), while the bottom primer will start right before the mutated base(s). For example, on the following sequence, to mutate the taa
codon to caa in:

\[\ldots ATGACCATGATTACGCAAGCGCGtaaTTAACCCTCACTAAAGGAACAAAAGCT\ldots \]

\[\ldots TACTGGTACTAATGCGGTTCGCGCattAATTGGGAGTGATTTCCCTTGTTTTCGA\ldots \]
One would design the following primers:

5'p-caaTTAACCTCACTAAAGGGAAC-3'
3'-GTACTAATGCGGTTCGCGC-5'

which would align as follow

5'p-caaTTAACCTCACTAAAGGGAAC-3'
...ATGACCATGATTACGCAAGCGGCTaaTTAACCTCACTAAAGGGAACAAAAGCT...
...TACTGGTACTAATGCGGTTCGCGCattAATTGGGAGTGATTTCCCTTGGGTTCGA...
3'-GTACTAATGCGGTTCGCGC-5'

The top primer needs to be phosphorylated on its 5', but not the bottom one.

2. Run the PCR reaction with a polymerase processive enough for the length of your plasmid. Q5, pHusion, or LongAmp may be appropriate.

3. After the PCR, ligate the reaction using T4 DNA ligase. Simply add ligase buffer and ligase to 10 μL of PCR reaction. This will circularize the newly mutagenized plasmids. I like to let the ligation run at room temperature, overnight.

4. Digest the ligation product with DpnI. DpnI will digest the methylated original plasmids but not the newly-synthesized ones. Let the reaction incubate for 1 h at 37°C.

5. Transform in DH5α or your favorite plasmid-amplification strain.

6. Pick and sequence the colonies to screen for positive hits.

B Labeling of histone octamers with maleimide-conjugated dyes

Goal

This protocol is used to label histone octamers with a dye. The labeling depends on the reaction of maleimides with free thiol groups, such as the one of cysteines. The octamers must hence be reconstituted with one of the histones mutated to contain a solvent-accessible cysteine. Yeast histones do
not contain cysteines. Metazoan histones contain a cysteine on H3, buried at the H3-H3 interface. Rationale: The whole labeling procedure happens in only one spin concentration device to lose as little histone octamers as possible. This protocol can hence be coupled with the last concentration step of histone octamers purification, if you have an approximate idea of the molarity of your purified histone octamers post-column. This protocol was developed because many small chemicals need to be added, and then removed, on the same sample. For example, β-mercaptoethanol is necessary to prevent histone octamers from forming disulfite bonds, but it reacts with maleimide and hence quenches the labeling reaction. TCEP breaks down disulfite bonds very well and fast, but it inhibits maleimide reaction with free cysteines. Finally, the excess dye must be removed to prevent side reactions, interaction with the lipid bilayer and to quantify the Degree Of Labeling (DOL).

Considerations

- Octamers crash below 2M salts, which is the reason for high salts. For any other protein, you can change that.
- B-mercaptoethanol and TCEP are preferred over DTT because its absorption peak interferes with the quantification of the histone octamers concentration at 280nm.
- You can replace β-mercaptoethanol by TCEP in every step involving β-mercaptoethanol. TE2000β is used for historical reasons. TCEP is better.
- A former version of the protocol included passivation of the column with BSA. The advantage of having (barely) more octamers out of the column was outweighed by the added imprecision in the measurement of the octamer concentration (50

Material

- 10,000 MWCO spin concentration column (Amicon Ultra 0.5mL 10,000 MWCO, Ref: UFC501024, Merck Milipore Ltd.)
- TCEP 0.5M pH7.5 (Thermo Scientific Prod# 77720)
- Maleimide-conjugated dyes (Dyomics, ATTO, Alexa, etc.)
Info

• CV = 1 column volume
• Spin down = see the spin concentration device supplier’s notice. The goal is to leave 1/10 or less or the initial volume. For Amicon Ultra 10,000MWCO 0.5mL, use 14,000g, 10mins.
• RT = room temperature

Buffers

1. TE2000: 10 mM Tris Hcl pH7.5-8.0, 1 mM EDTA pH8.0
2. TE2000β: TE2000 + 5 mM β-mercaptoethanol

Protocol

1. Clean the membrane: add 1 CV of TE2000β. Spin down at RT.
2. Break down disulfite bonds in the histone octamers preparation: add TCEP to the histone octamers to a final concentration of 10 mM. Incubate at RT for 10 mins.
3. Add the octamers to the spin concentration device.
4. Remove TCEP from the buffer: add TE2000 to 1 CV. Spin down at 4°C. Repeat 2 times.
5. Dye labeling: Add a 5X excess (or more) of maleimide-conjugated dye to the octamers, in the spin column. Protect from light. Incubate at room temperature for 1h.
6. Remove excess dye: add TE2000β to 1 CV. Spin down at 4°C. Repeat at least 5 times or until flow-through is clear of dye
7. Collect octamers: In a new collection tube: invert the column, 1000xg, 5 mins, 4°C
8. Quantify the histone octamers and degree of labeling
C  Assembly of functionalized oligonucleotides on λ DNA

Goal

To assemble lambda phage DNA molecules with biotin and digoxigenin, or any other small molecule, such as organic dyes.

Background

The lambda phage genomic DNA molecule is linearized for packaging at the cos site, leaving to 12-bp overhangs. These overhangs can be ligated with commercially prepared functionalized oligonucleotides. Because 12 bp is relatively stable, the reaction is very efficient. However, because the cos sites are compatible, the reaction conveyed at too high concentration of λ DNA will result in multimers. In addition, a large amount of the oligonucleotides is lost due to their dimerization, although this is not a problem sufficient enough to be worth fixing, as the reactions remain extremely efficient.

The lambda DNA has a “left” side and a “right” side. The left side is positioned with the cos sequence 5'−GGGCGGCGACCT−3', whereas the right side is positioned with the cos sequence 5'AGTTCGCCGCCC−3'. Consequently, oligonucleotides for the left and right sides, respectively, will have the sequence 5'AGTTCGCCGCCC−3' and 5'−GGGCGGCGACCT−3', respectively. Typical oligonucleotides are BioL and DigR: 5'Phos/AGTTCGCCGCCC/3'Bio and 5'Phos/GGGCGGCGACCT/3'DigN.

Material

• Functionalized oligonucleotides, such as BioL and DigR, 100 μM
• Lambda DNA, typically from NEB (#N3011)
• DNA ligase & buffer
• 65°C bath
Protocol

1. Mix the λ DNA and the oligonucleotides to final concentration 2.5 nM and 250 nM, respectively. Note that it is 250 nM per oligonucleotide. Higher concentrations of λ DNA will lead to concatemers.

2. Heat the mixed reaction at 65°C for 5 minutes, to release the interacting cos sites.

3. Let the reaction cool at room temperature so that the oligonucleotides can assemble on the λ DNA molecules.

4. Add ligase buffer and ligase to the recommended final concentrations of the manufacturer.

5. Incubate at room temperature, overnight.

6. Remove the excess oligonucleotides using the protocol “Purification of excess oligonucleotides and short fragments from λ DNA”.

D Purification of excess oligonucleotides and short fragments from λ DNA

Goal

To separate unligated “handles” and other contaminants <= 100bp from Lambda DNA.

Principle

High concentrations of PEG will precipitate DNA out of solution in a size-dependent manner. The higher the PEG concentration, the shorter the precipitated DNA fragments. At 10% PEG8000, only the fragments larger than 300bp are pulled out of solution. Below 10% PEG, precipitation efficiency drops considerably. Below 0.05 g/L of DNA, the efficiency of precipitation will also drop. At 0.005 g/L, precipitation does not happen. The efficiency of precipitation is increased by cold, incubation time and centrifugation.

273
Buffers

1. PEG8000 30% (m/v) + 10 mM MgCl2
2. TE150: 10 mM Tris HCl pH 8.0, 1mM EDTA pH 8.0, 150 mM NaCl

Protocol

1. Make sure your DNA is in a buffer with at least 10mM Mg2+. If not, bring it to that value. Mg2+ helps with the precipitation of the DNA. 1X NEB T4 ligase buffer contains 10 mM MgCl2. I often bring the [Mg2+] to 20 mM before adding the PEG solution, mostly because I can, maybe because it helps the precipitation (but this is a anecdotal observation).

2. Add ½ volume of PEG8000 30% + 10 mM MgCl2 to your DNA preparation, bringing the concentration of PEG8000 to 10%. Incubate for 30 mins to overnight on ice or at 4°C with rotation. The longer the incubation, the cleaner the separation and the more efficient the precipitation. Rotation does seem to help, but this statement was not carefully tested.

3. Spin down at 14,000 xg, 5 mins. Remove supernatant. The pellet may be invisible (I could observe it at 50mg DNA, though) so be careful. It may be worth it to centrifuge at stronger speeds to make the pellet more stable, though I never tried or needed it. This may also backfire and make the DNA harder to resuspend.

4. Optional: clean with 70% Ethanol to remove excess PEG, Mg2+ and co.

5. Resuspend in TE150 (or your favorite buffer) to get the desired final concentration of DNA (the precipitation efficiency is often close to 100%, so you can predict the mass of DNA you are dissolving).

6. Optional: Test the efficiency of separation/precipitation by loading both supernatant and resuspended pellet on a 0.7% agarose gel. Make sure not to run the gel for too long so that the short handles stay in the gel. They run faster than bromophenol blue. I usually run the gel 10 to 15 mins at 120V.
E Purification of short DNA molecules for nucleosome reconstitution

While the purification of short DNA fragments for nucleosome assembly has been done via plasmid purification and digestion (Luger et al., 1999), I found it easier and faster to achieve by PCR, with the added advantage of easy labeling.

Simply, oligonucleotides are ordered with or without a fluorophore at one end. Several (usually 10) PCR reactions are amplified in parallel. They are subsequently merged and purified on the same commercial PCR purification column. The DNA is eluted with 50-100 μL to achieve a high enough concentration, ideally of several micro molars, since nucleosome reconstitutions are most efficient in the micromolar range. The DNA is eluted directly in buffer with 2M NaCl to minimize the volumes added to the reconstitution reactions.

Note that the DNA has to be labeled with a fluorophore for measuring the amount of shifter products. Indeed, nucleosomes compete with the binding of intercalating dyes such as ethidium bromide or Sybr Safe, and not all reconstitution will have the same amount of aggregation and precipitation, making the comparison of the unshifted fraction between reconstitutions inappropriate.