Quantifying nucleosome dynamics and protein binding with PIE-FCCS and spFRET

Martens, C.L.G.

CHAPTER 1

INTRODUCTION

1.1 DNA: compaction vs. accessibility

The code of life is written in only four symbols: A, C, G and T, representing the nucleotide bases Adenine, Cytosine, Guanine and Thymine of DNA[1]. After being read during transcription, DNA is translated into the myriad of proteins, peptides and enzymes that build, maintain and regulate all the structures and functions of an organism[2]. The human genomic DNA has a contour length of about two meters, six orders of magnitude larger than the cell's nucleus, which has a diameter of approximately six micrometers[3]. In order to fit inside the nucleus, the genome is dramatically reduced in size through the formation of nucleosomes and chromatin[4][5]. And, although being in this highly condensed state, DNA is involved in fast and frequent processes such as transcription and DNA repair[6]. These processes need direct DNA access, and hence depend on a change in chromatin compaction.

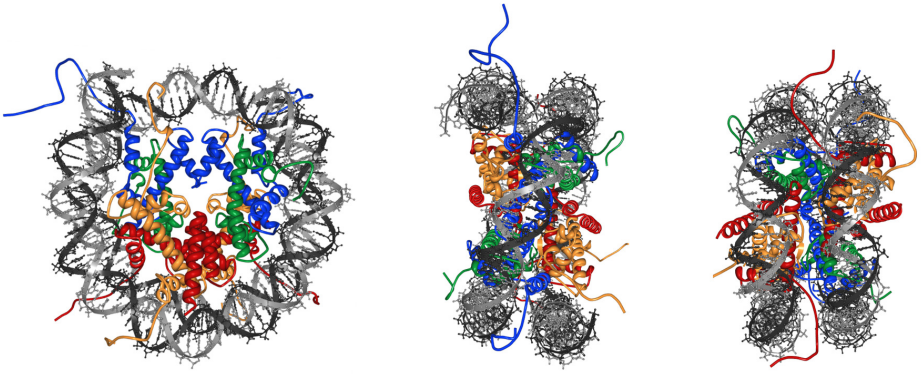

FIGURE 1.1: **The crystal structure of the nucleosome core particle (NCP) shows it near-prefect symmetry.** In the center of the NCP is the histone octamer (HO) consisting of two copies of H2A (yellow), H2B (red), H3 (blue), and H4 (green). Wrapped around the histone core is 147 base pairs of DNA (light and dark grey). DOI:10.2210/pdb1EQZ/pdb NDB:PD0137, rendered with RSCB PDB Protein workshop v.42.0.

In eukaryotic cells, the first order of compaction is the nucleosome (figure **1.1**). This basic unit of chromatin consists of 147 base pairs wrapped 1.6 times around a histone core, an octameric ensemble of compactor proteins[7]. The nucleosome has been extensively characterized and will be discussed in the next paragraph. Larger order structures of compacted

DNA are however less well-elucidated. The 'textbook' explanation has long been that nucleosomes form an almost evenly spaced beads-on-a-string configuration (figure **1.2a**).

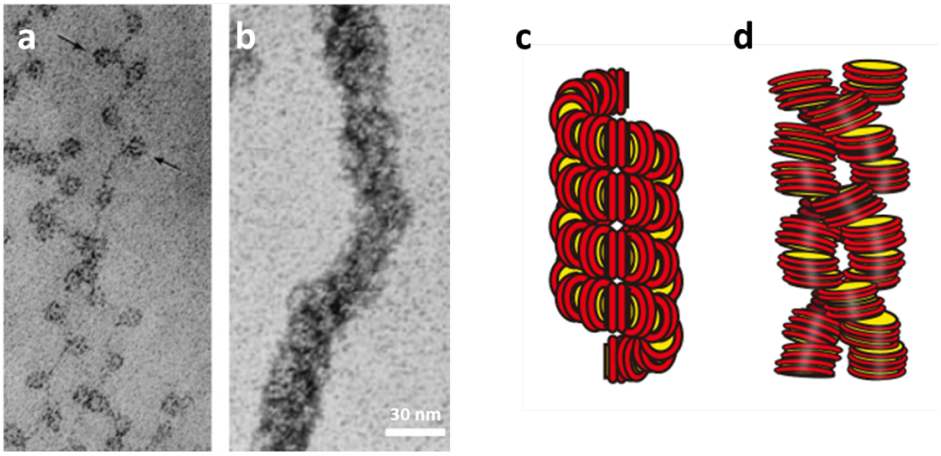

FIGURE 1.2: **Strings of nucleosomes are compacted into fibers under specific conditions**. One of the first observations of the nucleosomes (**a)**) and chromatin fiber (**b)**) was in chicken erythrocytes; picture from Olins et al.[8]. Modelling nucleosome compaction into a 30 nm fiber results into one- (**c)**) or two-start fibers (**d)**), depending on linker DNA length.[9]

In vivo studies have shown that there indeed seems to be a discrete spacing between nucleosomes, but that this spacing adheres to a *10n+5* rule, with *n* the number of base pairs of linker DNA[5]. The next order of compaction supposedly formed is the 30 nm fiber (figure **1.2b**), named for its width. Based on EM images[10][11] models were constructed, in which the 30nm fiber is formed by a string of nucleosomes compacted via a repetitive structure, dubbed the one-start helix, or solenoid, configuration (figure **1.2c**). When the length of the linker DNA between its nucleosomes is shortened, the fiber takes on a two-start, or zig-zag, configuration (figure **1.2d**). Both structures have been confirmed through *in vitro* experiments[12][13]. However, *in vivo*, no evidence of chromatin compaction through these configurations has been found[9]. The predominant reason for this lies in the set-up of the *in vitro* experiments; DNA containing nucleosome positioning elements alternated with the same length of linked DNA (following the 10n rule) is used. In this way, a high degree of control in synthesis and

probing of the fibers is exerted. Although using synthetic DNA sequences and predetermined linker lengths has its advantages, it provides a limited insight in the mechanisms of accessibility of higher order compacted DNA. The 'problem' *in vivo* is that, even excluding the influence of chromatin remodelers and other interactors, the sequence of the genome is much more varied, influencing the positioning and stability of nucleosomes.

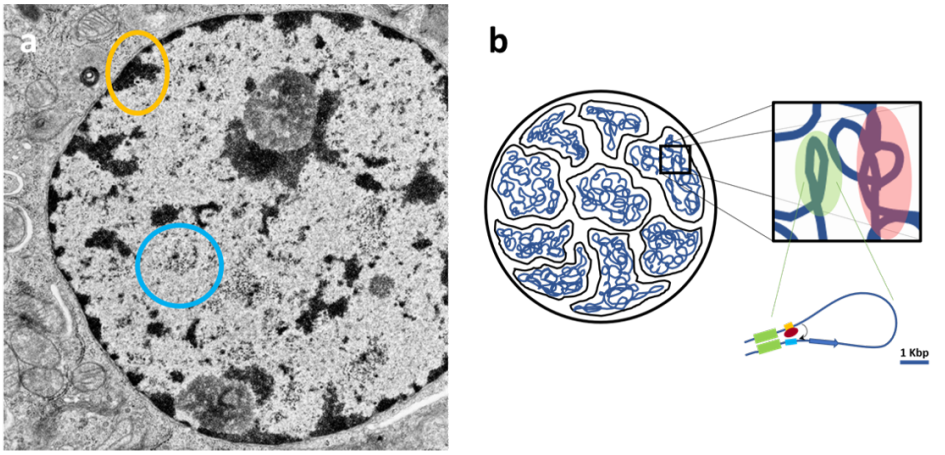

FIGURE 1.3: **In vivo chromatin is compacted into defined regions. a)** Globular structures in chromatin have long been classified as either eu- or heterochromatin (in blue , respectively orange circle) depending on their accessibility[14]. **b)** Recently the classification of topologically associated domains (TADs) has gained notice. TADs are defined as 3D interactions of DNA sequences separated from each other in 2D by 100 to 10.000 base pairs[15].

Over the past years, through the improvement of techniques able to probe chromatin *in vivo*, new insights into the nature of its compaction have arisen. Instead of different highly compacted structures, more global structures of accessible and in- or less accessible chromatin have been observed, categorized as euchromatin (eu- deriving from the Hellenic word for good) or heterochromatin (hetero- because it has not a single definable structure and is generally in a state of inactivity)[16][14] (figure **1.3a**; euchromatin in blue circle, heterochromatin in orange circle). Euchromatin, or a part thereof, is sometimes referred to as hyperaccessible nucleosomes. Mnase seq and other digestion experiments have shown that chromatin, regardless of being in eu- or heterochromatin form, is compacted not in 30 nm sized

structures, but in smaller units[17][18]. One can argue that these structures consist of small numbers of nucleosomes, as well as less defined in linker lengths and nucleosomal stability[19], and hence cannot distribute the DNA bending energy over its local structure as efficiently as repetitive, highly compacted 30 nm fibers.

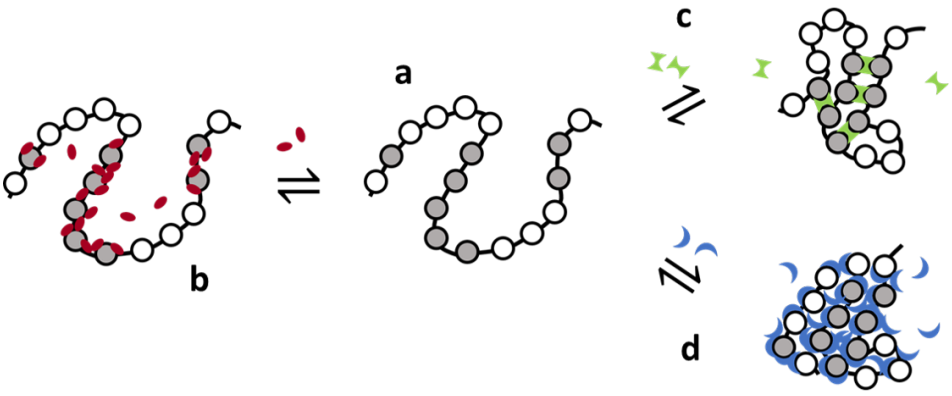

FIGURE 1.4: **Different protein binding mechanisms induce further compaction of nucleosomes on a string through phase separation. a)** nucleosomes on a string are shown without modifiers. **b)** Proteins binding cooperatively to nucleosomes do not necessarily induce compaction for effective functioning. **c)** Bridging proteins compact nucleosomes by cross-linking. **d)** Other molecules induce multivalent interactions with nucleosomes, thereby driving out other molecules and collapsing the chromatin structure to a more condensed state.

Recently a strong case has been made for the existence of topological associated domains (TADs)[20][15], in which sites of non-neighbouring compacted DNA interact with each other (figure **1.3b**).
Different biophysical concepts are currently used to explain the formation of these domains in a self-organizing manner and without (or minimal) energy consumption. In addition to cooperative binding to a preformed chromatin structure, as depicted in figure **1.4b**, two different mechanisms for the formation of phase-separated chromatin sub-compartments have been proposed[21]. One is based on bridging proteins that cross-link polymer segments with particular properties (also seen in the compaction of archeal DNA), inducing a collapse of the chromatin into an ordered globular phase (figure **1.4c**). The other mechanism is based on multivalent interac-

tions among soluble molecules that bind to chromatin (figure **1.4d**). These interactions can induce liquid-liquid phase separation, which drives the assembly of liquid-like nuclear bodies around the respective binding sites in chromatin.

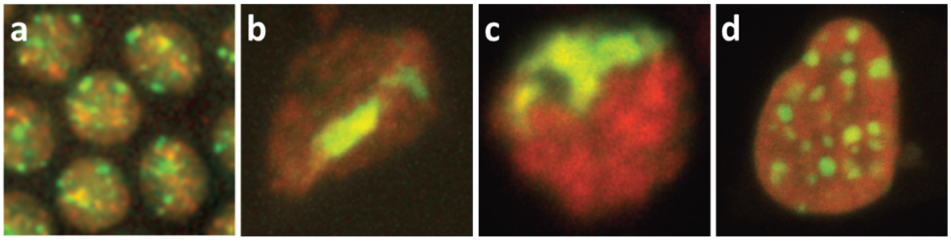

FIGURE 1.5: **In vivo phase separation induced by HP1α.** HP1α labeled with GFP (green), H2A labelled with mCherry (red). Different cell types show the formation of globular structures where HP1α is present, but H2A is not. Cells are **a)** Drosophila embryo and **b)** adult gut, **c)** cultured Kc and **d)** mouse fibroblast NIH3T3. Images adapted from [22]

.

Persuading evidence that certain proteins drive chromatin to different states of compaction *in vivo* has been found by Strom et al.[22], who show that HP1α (heterochromatin protein 1α) undergoes liquid-liquid demixing *in vitro* as well as *in vivo* as compared to histone H2A. Distinct globular domains of heterochromatin containing HP1α but no or little amounts of H2A can be found throughout different cell types, indicating the assembly of these domains depends on the (multivalent) interactions of the protein with the chromatin instead of being a cell types specific effect (figure **1.5**). Results supporting the hypothesis DNA is compartmentalized through LLPS keep being added[23][24][25][26].
What should be concluded from the recent findings *in vivo* presented here is that, regardless what and how higher order structures of chromatin form, or how they interact, they are made up of nucleosomes. And that when one attribute of the nucleosome is changed, for iinstance through binding of a transcription factor or remodeller, there is strong evidence this brings about a change in the higher-order structure as well. Thereby gaining a better understanding in the mechanisms of the nucleosome will result in more insight into the more complex structures or domains it is a part of.

1.2 Nucleosome structure and dynamics

The nucleosome, or more accurately, the nucleosome core particle (NCP), consists of 147 base pairs of DNA wrapped around a histone core consisting of two copies of four histones: H2A, H2B, H3 and H4 (figure **1.6a**). For decades the structure of the nucleosome has been the subject of extensive research[27][28][29][30]. Although the composition of the nucleosome had

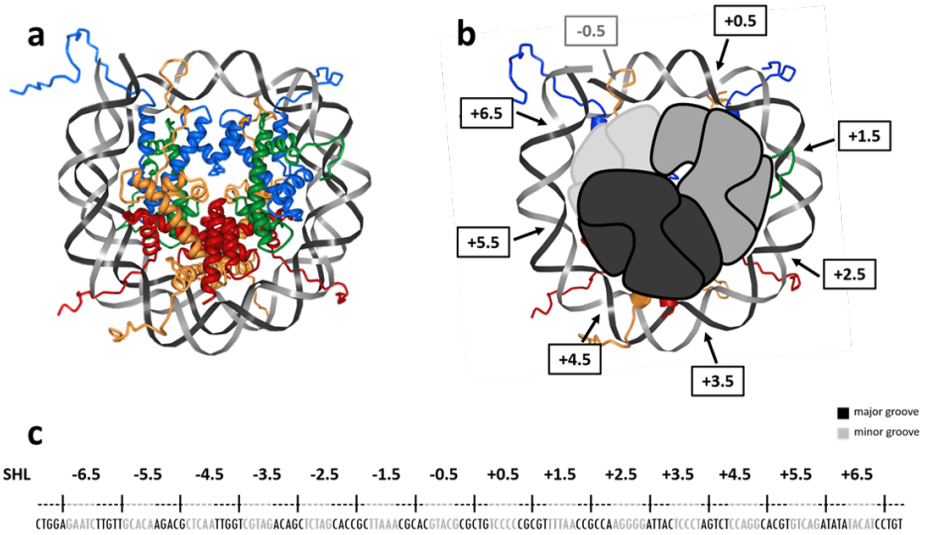

FIGURE 1.6: **Preferred positioning of base pairs in the nucleosome pre-bends the DNA around the histone core. a)** The H3/H4 tetramer (H3 (blue), H4 (green)) forms a diagonal ramp through the dyad, aligning the major and minor grooves of the DNA gyres in neighbouring planes. In this frontal view, the two H2A-H2B dimers perfectly overlap (H2A (yellow), H2B (red)). **b)** Projecting half of the wrapped DNA (75 base pairs) shows clearly how every 10 base pairs the DNA approaches the histone core through the minor groove. **c)** The Widom 601 sequence shows the preferred positioning of the A and T nucleotides in the minor grooves (grey) and the C and G nucleotides in the major grooves (black). (SHL = superhelical locations)

been realized before[31][32][33], it was the crystal structure at 2.8 angstrom resolution resolved by Luger et al. in 1997 that offered the first atomic depiction of the NCP[34]. It showed that the nucleosome has a pseudo-2-fold symmetry axis centered around a single base pair

(the dyad)[35]. What also became apparent from the crystal structure was that the path of nucleosomal DNA is nonuniform: the H2A/H2B dimers bind DNA in two planes perpendicular to DNA superhelical axis, while the central H3/H4 tetramer forms a diagonal ramp through the nucleosomal dyad, connecting these two planes(figure **1.6a**). This way, the DNA gyres in neighbouring planes align their major and minor grooves as they track along the octamer surface. The DNA superhelix contacts the histone octamer at regular intervals (figure **1.6b**) each $\sim$10 base pairs, resulting in 14 superhelical locations (SHLs) denoting the minor grooves (figure **1.6c**). These histone-DNA interfaces are mediated by extensive direct and water-mediated bonds, ionic interactions and nonpolar contacts.[36][37][38]

1.2.1 Sequence-dependent properties

Proof that the DNA sequence has an effect on nucleosome stability has been shown by Jonathan Widom through his SELEX (Systematic Evolution of Ligands by EXponential enrichment) experiments[39][40][41]. In a SELEX experiment, DNA sequences with high affinities for the histone octamer are selected by repeating the following steps: first, random DNA sequences are synthesized and reconstituted into nucleosomes. Their affinities for the histone octamer, as well as their ability to position nucleosomes are measured with several techniques (salt titration, etc). The best sequences are selected and amplified by PCR. This selection is repeated several times. Widom found that his selected sequences had higher affinity than natural nucleosome positioning sequences previously found. His most famous sequence is number 601, known by biologists, chemists and biophysicists as 'Widom 601' (601 has lesser-known siblings, among them 603, which is also often used for similar nucleosome experiments[40][42]).
Examination of the 601 sequence (figure **1.6c**) and other high scoring sequences revealed common features: nucleotides arginine (A) and thymine (T) prefer positions in minor grooves, closer to the histone octamer, while cytosine (C) and guanine (G) prefer the major grooves[43]. These preferences stem from the difference in the number of hydrogen bonds between A-T (two bonds) and C-G (three bonds) making the latter pair more rigid[44][45]. Preferred positioning of base pairs also results in pre-bending DNA in such a way that less energy is needed to further bend the DNA around the histone core[43][46].

Recent micromanipulation experiments performed by Ngo et al. have shown nucleosomes respond highly asymmetrical to external forces[47]. The experiment combined optical tweezers capable of exerting piconewton forces with fluorescence microscopy on single nucleosomes, each nucleosome containing a pair of fluorescent labels at positions making FRET possible. Combining these techniques showed the nucleosome predominantly unwraps from the same end. It was proposed that this unwrapping preference is caused by the fact that the Widom 601 sequence making up the wrapped DNA portion is non-palindromic. This asymmetry makes one-half of the wrapped portion more rigid than the other and the researcher propose the stiffer half unwraps first because it cannot distribute the exerted stress as well as his flexible counterpart. This hypotheses was tested by flipping the inner two quarters of the 601 sequence, resulting in a nucleosome that unwraps from the other end. These experiments have been validated by the development of a theoretical framework by the Schiessel group[48][49]. Their model describes how the possible asymmetric metastable conformations of the nucleosome depend on the distribution of the kinetic energy over the bonds between base pairs.

1.2.2 Nucleosome dynamics

More work from the Widom group[50][51] has shown that nucleosomal DNA is not statically wrapped around the histone core but is in equilibrium between a wrapped and unwrapped state (figure **1.7**). Through stopped-flow and time-resolved fluorescence experiments it was shown DNA at the exit of a nucleosome unwraps for 10 to 50 ms and rewraps for 250 ms[52]. This means that DNA is fully accessible for some period of time without being actively dissociated from the nucleosome. The stochastic switching between conformations has important functional implications for DNA-binding proteins that interact with chromatin. Widom and others have implied that nucleosome breathing is the rate-limiting step in some passive binding steps of transcriptional processes[50][53].
Nucleosomes are also able to reposition themselves by sliding DNA through the histone gyres without disrupting the histone octamer. This process is not stochastic but mediated by ATP-dependent chromatin remodelling enzymes[54][55].

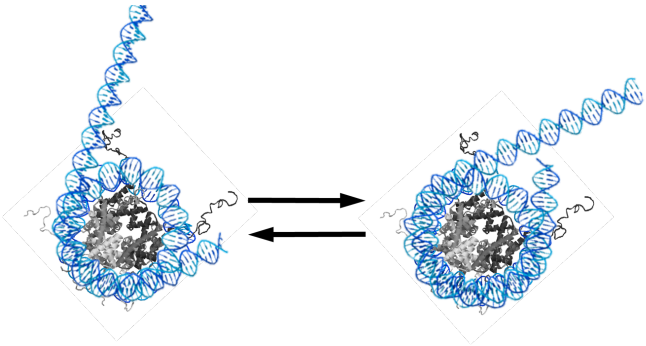

FIGURE 1.7: **Depiction of unwrapping and (re-)wrapping of the DNA exits or tails of the nucleosome, also known as nucleosomal breathing.**

1.2.3 Histone tail modifications

A way cells regulate transcription is through posttranslational modifications (PTMs) of histone tails. Modifications of the tails such acetylation, methylation and phosphorylation change the structure and as a consequence the function of the tail and can act as markers for enzymes and proteins[56][57]. Acetylation of H4-tail lysines results in reduced compactness of chromatin, allowing access to DNA in acetylated chromatin regions to transcription factors[58][59]. Some modifications have been associated with increased chances of cancer cell formation, like methylation of histone tail H3K27[60]. Recent studies suggested that during mitosis, chromosome formation is triggered by a combination of modifications of the H3 and H4 tails[61][62][63]. Figure **1.8** provides an overview of most known histone tail modifications. The modification occurring most often is the acetylation of lysine and induces gene expression associated with metabolism[64][65]. Another common modification is methylation, which occurs as mono-, di- or trimethylation. These different degrees of methylation may be the reason why this modification seems more precise than acetylation. Acetyltransferases tend to act on several adjacent lysines to perhaps have a similar effect on chromatin structure as methylation. Histone methylation is associated with both gene upregulation and repression[66], and specific modifications are involved in DNA repair[67] and in some species are even linked to an increased lifespan[68][69].
One of the methylations associated with DNA repair is H3K36me3[70][71][72]. Several proteins involved in DNA mismatch

repair have a high affinity for H3K36me3, and it has been shown that because of elevated activity of mismatch repair, regions of the human genome containing high levels of H3K36me3 are less prone to somatic mutations[73]. H3K36me3 is also associated with heterochromatin[74]. Another protein recruited by H3K36me3 is LEDGF/p75. This protein is a co-activator in transcription, but has gained more attention due to its role in HIV DNA integration into host DNA[75][76][77]. LEDGF/p75 is hijacked by a pre-integration complex containing the HIV DNA and used as a bridge to the host DNA. It is not known whether the LEDGF/p75 binds first to the pre-integration complex and then to the host DNA, as LEDGF/p75 contains both an integrase-binding domain, through which it interacts with integrase proteins, as well as a N-terminal domain with a high affinity for chromatin[75].

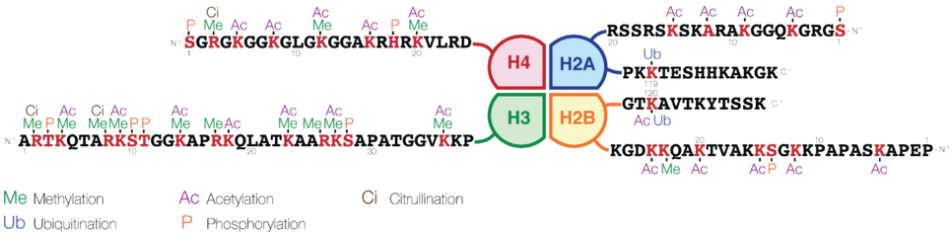

FIGURE 1.8: **Overview of most know posttranslational histone tail modifications.** Most PTMs are found in the tail of H3, presumably due to it having the highest lysine (K) content[61].

1.2.4 Transcription factors

Transcription factors (TFs) are proteins regulating the activation and repression of genes by preparing chromatin for the recruitment of the transcription pre-initiation complex (PIC)[78][79], an ensemble of transcription factors and RNA polymerase. TFs have a DNA binding domain (DBD) through which they interact with a response element (RE) close to the gene that they regulate.

The Glucocorticoid Receptor (GR) is a transcription factor involved in activating as well as repressing gene expression. The GR protein itself is activated by the hormone cortisol[80]. In the absence of hormone, GR is part of a complex including heat shock proteins hsp70 and hsp90 and resides in the cytosol[81][82]. Upon activation GR is released from the complex and translocates via active transport to the nucleus[83]. Depending on the cell

type and presence of other TFs, the GR either binds as a homodimer directly to DNA and transactivates gene transcription, or complexes with other TFs to repress or downregulate the genes these TFs would transactivate[84][85].

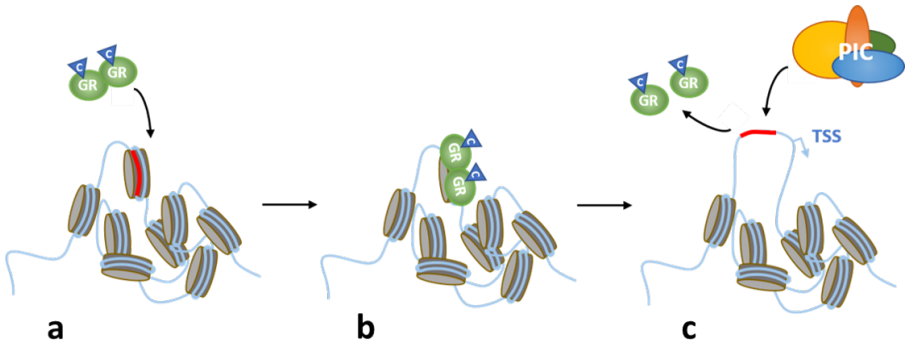

FIGURE 1.9: **Pioneering factors like the glucocorticoid receptor unwrap compacted DNA to initiate transcription. a)** GR prefers to bind to a GRE (red) positioned in nucleosomes. **b)** As a homodimer, GR is able to bind its response element without other TFs. **c)** After unwrapping the nucleosome by competing with the histone core, GR detaches and the pre-initiation complex (PIC) moves in to start transcription at a nearby transcription start site (TSS, blue arrow).

Although these two interaction modes of GR are described most often in literature, there is also evidence GR is able to interact with DNA in its monomeric form[86], as a heterodimer with other corticoid receptors[87][88], or even as a tri- or tetramer[89][90]. *In vitro* experiments by Wrange have shown that GR's affinity for its response element (GRE) is higher when positioned in a nucleosome compared to bare DNA[91][92]. Other *in vitro*, *in vivo* and *in silico* studies have shown the GRE is usually found near the exits of nucleosomes[93][94] and the GR prefers GREs in the minor grooves of wrapped DNA[95]. These findings support the role of GR as a pioneering factor, i.e. a transcription factor pioneering compacted DNA to find its GRE. The pioneering mechanism is depicted in figure **1.9**; a dimer of two GR proteins recognizes the GRE in a nucleosome (figure **1.9a**) and binds to it (figure **1.9b**). By presumably out-competing the histones, the GR dimer unwraps the nucleosome, making the DNA available for the PIC. The GR detaches and transcription will start from the transcription start site (TSS) close to the GRE (figure **1.9c**).

1.3 Single-molecule microscopy and Förster Resonance Energy Transfer (FRET) for nucleosome research

Methods such as crystallography, gel shift mobility assays, stopped-flow and bulk fluorescence have been used for years to investigate nucleosome structure, stability, conformation and dynamics[35][96][97][98][99][100]. Newer, single-molecule techniques have been developed more recently and provide a extra layer of information by visualizing characteristics of single molecules. These techniques are often a combination of confocal microscopy with low concentrations of the sample of interest. Fluorescent labels are excellent for following single molecules, as labelling and tracking of the signal are non-invasive techniques offering a high degree of accuracy[101][102]. An additional fluorescence method often used in biology is Förster Resonance Energy Transfer (FRET), which provide conformational information at nanometer scale and has proven itself in recent years in nucleosome dynamics research[98][99][103]. Labelling a nucleosome as depicted in figure **1.10a** allows for tracking nucleosome breathing in burst experiments combined with alternating laser excitation (ALEX) where fluorophores are excited by alternating light pulses at microsecond timescale (figure **1.10a**) to generate timetraces such as those in figure **1.10b**. In experiments performed by Koopmans et al.[53] the concentration of nucleosomes was lowered to be able to distinguish single-molecule burst events. These bursts were characterized by their label stoichiometry S and FRET efficiency E[refs] and plotted in an E, S-histogram (figure **1.10c**) to assess the equilibrium between closed ($E > 0.25$) and open ($E < 0.25$) nucleosomes ($S > 0.2$ and < 0.8). Timetraces from burst experiments were also used to determine the difference in diffusion time for closed and open nucleosomes; correlation curves shown in figure **1.10d** were obtained by correlating photon arrival times assigned to closed or open nucleosomes. As depicted, closed nucleosomes diffuse faster (curve shifts to smaller tau) than (partially) open nucleosomes. Burst experiments done by Buning et al.[104] have shown this method is so accurate one is able to distinguish the difference in compaction as measured by FRET efficiency through elongating one of the nucleosomes' DNA exits (figure **1.10e-f**).

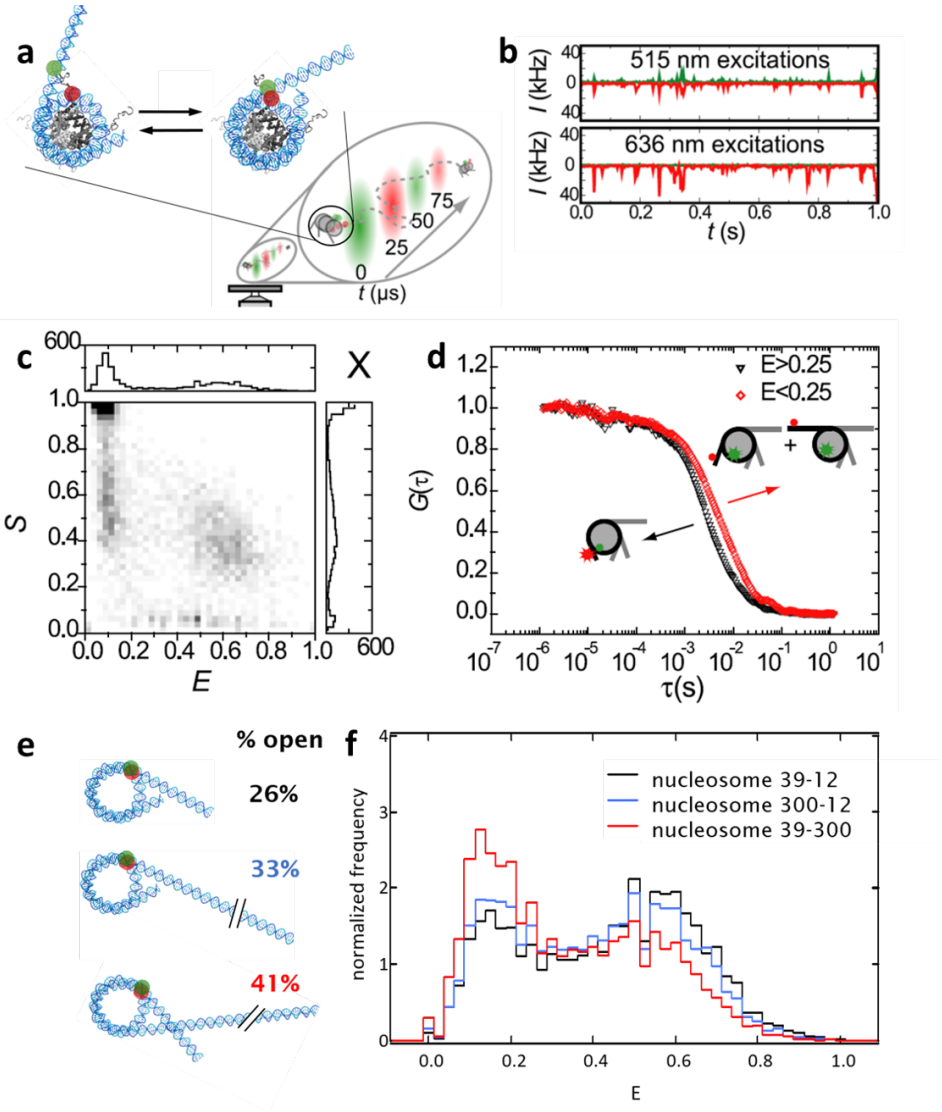

FIGURE 1.10: **Alternating laser excitation (ALEX) combined with FRET and nanomolar concentrations in a confocal microscope gives conformational information of the nucleosome on a single-molecule level. a)** Nucleosomes were labeled with FRET pair Cy3B-Atto647N. At nanomolar concentrations single nucleosomes diffuse through the confocal spot. **b)** The diffusion of single molecules generates timetraces of burst events. When a nucleosome comes through the focus in a closed state, both green and FRET emissions are detected (upper timetrace). *caption continues on next page*

1.4 Scope of this thesis

The next chapters revolve around the nucleosome, our unit of interest. We have investigated and quantified when possible the effects of different attributes of the nucleosome itself, as well as the effects of interactors with the nucleosome, focusing on the dynamics, accessibility and stability of the nucleosome. In chapters 2 and 3 we elaborate on the optical, analytical and biological tools we have used and developed to successfully perform experiments sensitive enough to visualize these effects. Both chapters conclude with validations of the methods from experiments.
In chapters 4, 5 and 6 we combined several fluorescent techniques such as fluorescence correlation spectroscopy, spFRET and burst analysis to resolve nucleosome dynamics and binding affinity of different proteins to DNA and nucleosomes. By combining these techniques we have quantified the interactions in and with the nucleosome. In chapter 4 we show the effects of changes in DNA sequence, linker DNA and buffer composition on nucleosomes. We have elucidated how changing a few base pairs near the DNA exit or changing the number of base pairs of linker DNA of a nucleosome alters nucleosome stability and dynamics. In chapter 5 we quantified the effect of histone posttranslational modification H3K36me3 on nucleosome stability and dynamics. We also showed how this PTM affects the affinity of the LEDGF/p75 protein for nucleosomes. Chapter 6 shows how an ensemble of proteins from an extracted nucleus containing an activated and labeled transcription factor GR interact with bare DNA and nucleosomes.

FIGURE 1.10: *Caption figure 1.10 continued:* **c)** Bursts were characterized by FRET efficiency (E) and label stoichiometry (S) and plotted in histogram format, showing two populations (E < 0.25 and E > 0.25) for 0.2 < S < 0.8, representing open and closed nucleosome, resp. **d)** Arrival times of photons assigned to bursts of open and closed nucleosomes were correlated to generate (auto)correlation curves, showing closed nucleosomes diffuse faster (black curve) than (partially) open nucleosomes (red curve). **e)** and **f)** Addition of free linker DNA drives nucleosomes to the open state. Noticeable is the difference in FRET distribution for 39-300 compared to 300-12; not only is the closed state less populated (59% vs. 67%) it is also less closed (lower $<E>$). Figures adapted from Koopmans et al. 2009 and Buning et al. 2015.