The physics of epigenetics

Ruggero Cortini,1,2,3 Maria Barbi,1,2,3 Bertrand R. Caré,1,2,3 Christophe Lavelle,3,4 Annick Lesne,1,2,3,5 Julien Mozziconacci,1,2,3 and Jean-Marc Victor1,2,3,5∗
1 Sorbonne Universités UPMC Univ. Paris 06 UMR 7600 LPTMC F-75005 Paris France.
2 CNRS UMR 7600 LPTMC, F-75005 Paris France
3 Nuclear Architecture and Dynamics, CNRS GDR 3536, UPMC Université Paris 6, 75005 Paris France
4 Genomes Structure and Instability, Sorbonne Universités, National Museum of Natural History, Inserm U 1154, CNRS UMR 7196, 75005 Paris France
5 Institut de Génétique Moléculaire de Montpellier CNRS UMR 5535 Montpellier France.

(Dated: September 15, 2015)

In higher organisms, all cells share the same genome, but every cell expresses only a limited and specific set of genes that defines the cell type. During cell division, not only the genome, but also the cell type is inherited by the daughter cells. This intriguing phenomenon is achieved by a variety of processes that have been collectively termed epigenetics: the stable and inheritable changes in gene expression patterns. This article reviews the extremely rich and exquisitely multi-scale physical mechanisms that govern the biological processes behind the initiation, spreading and inheritance of epigenetic states. These include not only the change in the molecular properties associated with the chemical modifications of DNA and histone proteins – such as methylation and acetylation – but also less conventional ones, such as the physics that governs the three-dimensional organization of the genome in cell nuclei. Strikingly, to achieve stability and heritability of epigenetic states, cells take advantage of many different physical principles, such as the universal behavior of polymers and copolymers, the general features of non-equilibrium dynamical systems, and the electrostatic and mechanical properties related to chemical modifications of DNA and histones. By putting the complex biological literature under this new light, the emerging picture is that a limited set of general physical rules play a key role in initiating, shaping and transmitting this crucial "epigenetic landscape". This new perspective not only allows to rationalize the normal cellular functions, but also helps to understand the emergence of pathological states, in which the epigenetic landscape becomes disfunctional.

CONTENTS

I. Introduction 2
A. An intricate history 2
B. Scope of this review 2

II. The physical template of epigenetics: chromatin 3
A. Molecular picture of chromatin and its modifications 3
B. Large-scale picture of chromatin 5
C. Chromosomes as polymers 5

III. From epigenetic marks to regulation of gene expression through the 3D organization of the genome 7
A. General principles of gene silencing. The paradigm of DNA accessibility 7
B. Histone modifications as chromatin structural modulators 7
1. Histone tails and their role in internucleosomal interactions 7
2. Histone tail post-translational modifications (PTMs) 8
3. Histone tail acetylation: direct effects on chromatin accessibility 8
4. H4K16 acetylation is a silencing mark in budding yeast 9
5. Histone tail methylation: indirect effects on chromatin condensation 10
C. How epigenetic marks organize the chromosomes in the cell nucleus. General rules. Physical modeling of epigenome wide studies. 12
1. Epigenome wide studies 12
2. The physics of TADs: finite-size effects in the coil-globule transition of copolymers 12

IV. Physical mechanisms involved in the initiation, spreading, maintenance and heritability of epigenetic marks 14
A. Mathematical modelling 14
B. Zero-dimensional models 14
C. Higher-dimensional models 16
1. One-dimensional models. 16
2. Three-dimensional models. 17
D. Biological relevance of the models 17
1. Waddington’s epigenetic landscape revisited. 17
2. Hysteresis 18
E. Example: plant vernalization 19

V. Toward a more complex scenario: DNA methylation, role of RNAs, supercoiling in epigenetics 20
A. DNA methylation 20
1. Mechanical properties of DNA change upon methylation 22
2. Impact of cytosine methylation on DNA-protein interactions 22
3. Relationship between nucleosome positioning and DNA methylation 23
4. Remarks and perspectives 23
B. Parental imprinting 23

∗ Corresponding author: Jean-Marc Victor LPTMC case courrier 121 Université Pierre et marie Curie 4 place Jussieu 75252, Paris cedex 05 France. Email: victor@lptmc.jussieu.fr
C. Chromosome X inactivation
D. Non-coding RNA and microRNA
E. Supercoilingomics: supercoiling as a physical epigenetic mark, and its role in the initiation and maintenance of epigenetic marks

VI. Conclusion and perspectives
Acknowledgments
References

I. INTRODUCTION

A. An intricate history

The word “epigenetic” has been introduced by Waddington in 1942 in the context of development, to qualify all the processes relating the genotype and the phenotype of an organism (Waddington, 1942). The associated investigations belonged to the domain, novel at that time, of developmental genetics. The word “epigenetic” in this original meaning was imprinted by the pre-existing concept of epigenesis, namely the fact that the organism is not fully achieved in the initial cell but experiences complex developmental processes (Gilbert, 2011).

Epigenetics was at that time the mechanistic study of how genes guide the epigenesis (development) of an organism, what is captured in a metaphoric way by the famous Waddington’s epigenetic landscape: a landscape, shaped by the genes, on which the organism would evolve during its development as a rolling stone on the landscape, following one of the possible epigenetic pathways (see Fig. 1).

In parallel, the adjective has been used with the meaning of “para-genetic”. Epigenetic systems, as opposed to the genetic system, were conceived as “signal interpreting devices” (Nannes, 1958), i.e. mediators between signals – environmental or physiological cues – and the genomic response, mainly at the level of transcriptional regulation.

Due to this dual origin of the word “epigenetic”, the associated concepts have developed in several ways (see (Haig, 2004) for a detailed historical account) and one could find in 1994 the two following complementary definitions of epigenetics (Holliday, 1994): (1) changes in gene expression which occur in organisms with differentiated cells, and the mitotic inheritance of the associated patterns of gene expression; (2) transgenerational inheritance, that is, transmission through meiosis of non-genomic information.

Due to this intricate history, a consensus definition of epigenetics is still lacking today (Dawson and Kouzarides, 2012). Notably the transgenerational inheritance, albeit largely documented in plants, remains a matter of debate in animals and especially in humans.

Recently, some authors recently proposed an operational definition of epigenetics: “An epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence”. This will be the definition we use in this review. To be even more specific, we note that:

Epigenetics is the modification of the function(s) of a gene, that is stable and heritable during mitosis, possibly during meiosis.

Epigenetics is not the reversible regulation of transcription in response to metabolic cues, because this is not stable nor heritable.

According to a scenario proposed by Berger et al. (Berger et al., 2008), there are 3 categories of signals that culminate in the establishment of a stably heritable epigenetic state:

1. Epigenetor: signal (cue) from the environment that triggers an intracellular signaling pathway (e.g. by means of membrane receptors, notably G Protein-Coupled Receptors).

2. Epigenetic initiator: epigenetors activate transcription factors (TFs) that bind to specific DNA targets;

3. Epigenetic maintainer: molecular covalent modifications of DNA or histones (DNA-binding proteins most strongly bound to DNA).

The molecular covalent modifications that eventually result from epigenetors are the so called “epigenetic marks”. The epigenetic marking of the genome is thus a key component of the dialogue between genes and environment in the eukaryotic realm.

B. Scope of this review

In this review, we not only intend to analyze the physics that drives or accompanies epigenetic marking, but we also aim at understanding the rationale behind this marking. And physics is a beautiful, yet underrated guide to reach this goal.

Several epigenetic mechanisms will be distinguished: those occurring at the level of DNA, those involving histone post-translational modification, and less conventional ones involving chromatin topology (supercoiling) and nuclear architecture.

We first introduce in section [I] the physical template of epigenetic marking, namely chromatin.

Section [II] is devoted to the physics behind the family of processes at work in the way epigenetic marks control gene expression in different cell types.

Section [IV] addresses the issue of the initiation, spreading, maintenance, and heritability of the epigenetic marks in the framework of dynamical systems.

In Section [V] we review other epigenetic processes that have a less clear-cut physical interpretation: DNA
methylation, imprinting, chromosome X inactivation, supercoiling marking. In conclusion we finally propose a list of currently significant and challenging issues.

Due to the fundamentally different logic of transcriptional regulation in prokaryotes and eukaryotes (Struhl 1999), we will let aside the realm of bacteria, although epigenetic switches have been observed as well in prokaryotic cells and have been modeled successfully (Lim and Van Oudenaarden 2007; Norregaard et al. 2013).

We hope this review will be a stimulating introduction to epigenetics for physicists as well as an “alternative reading frame” of epigenetics for biologists that will help tackling cutting-edge advances in current topics ranging from nuclear organization and cell differentiation up to cancer progression and chronic diseases.

II. THE PHYSICAL TEMPLATE OF EPIGENETICS: CHROMATIN

In all living organisms, DNA encodes the genetic instructions required to synthesize proteins, the basic bricks ensuring the proper functioning of the cell. The main steps of protein synthesis are DNA transcription into RNA, then RNA translation into an aminoacid chain and chain folding to form a functional protein (Alberts et al. 2013).

The very same genome is found in each cell. It has to be packaged inside its tiny volume, and has to be retrieved at will for physiological purposes. DNA is therefore embedded in an orderly and dynamically retrievable architecture. Two main organizational strategies can be identified. In prokaryotes (bacteria), DNA is located in the same compartment as all other intracellular components. In eukaryotes (from the unicellular yeast up to multicellular organisms, including fungi, animals and plants), DNA is sequestered in the nucleus, a dedicated compartment enclosed within a membrane.

In the cell nucleus, multiple long linear DNA molecules are organized by architectural proteins to form chromosomes. From a physicist point of view, chromosomes are giant polymers. During mitosis, i.e. cell division, chromosomes duplicate and then condense in the well known “X” shape, with each DNA copy forming one of the two rods (the sister chromatids, bound together at the centromere). The rest of the time (i.e. during interphase), chromosomes are less condensed and fill the whole nucleus, more or less homogeneously (Leblond and El-Alfy 1998). To give a quantitative idea of the composition of an interphase nucleus, the dry matter of a yeast nucleus is about \(\sim 70-80\%\) in protein, \(\sim 20-30\%\) in RNA, and only \(\sim 2\%\) in DNA (Rozijn and Tonino 1964).

In this section we introduce the basic concepts that come into play in the study of epigenetics. In Sec. II.A we give a synthetic overview of the molecular structure of chromatin, and we introduce the concept of epigenetic marks. In Sec. II.B we will give an overview of the large-scale organization of chromatin in the cellular nucleus, stressing the importance of this organization in gene expression. Finally, in Sec. II.C we give a synthetic picture of these two aspects, in the framework of polymer physics.

A. Molecular picture of chromatin and its modifications

In eukaryotic organisms, chromosomal DNA is associated with proteins to form chromatin. The principal proteins associated with DNA are called histones. Histones are polypeptidic monomers of five types: His-
tone 1 (H1) class, Histone 2A (H2A) class, Histone 2B (H2B) class, Histone 3 (H3) class and Histone 4 (H4) class. Each histone family has variants whose presence in chromatin depends on the species, the cell type, and the development stage. The classical structure of the histone-DNA assembly consists of 1.7 left-handed turns of double strand DNA (approximately 147 base pairs or bp) wrapped around a histone octamer composed of two copies of each histone monomer H2A, H2B, H3 and H4 (Davey et al. 2002; Luger et al. 1997). In most species, this assembly, referred to as the nucleosome core particle (NCP) (see Fig. 2a,b), may also integrate a copy of H1 (“linker histone”) at the DNA entry/exit point, although H1 does not share the ubiquity of the other histone classes.

In addition, consecutive NCPs are separated by linker DNA whose length ranges from 20 to 60 bp. Indeed, chromosomes are a succession of NCPs and DNA linkers. The basic structural unit (monomer) is made of one NCP and one DNA linker, and is called the nucleosome.
The number of DNA base pairs inside one nucleosome is the Nucleosome Repeat Length (NRL) (see Fig. 2.4d), which is not constant and may vary along the genome and across various tissues.

Electrostatic interactions are important because the NCP has a charge of -150e, to which DNA contributes -294e and histones +144e. The NCP is therefore not electrically neutral, so the folding of nucleosome arrays is highly dependent on the presence of positive counterions (Bertin et al., 2007). Yang and Hayes (2011). Additionally, the charge distribution in the NCP is not spatially homogeneous (see Fig. 2b).

Epigenetic marks are chemical covalent modifications of either DNA (namely DNA methylation, see Sec. III.B), or histones (so-called post-translational modifications, PTMs, see Sec. III.B). The DNA methylation state and the histone PTMs are transmitted through cell division both because they are covalent and thanks to specific mechanisms. DNA methylation is accurately transmitted by a specific molecular mechanism (see Sec. III.B). Histone PTMs are inherited in a fundamentally different way, which will be the principal subject of Sec. IV.

**B. Large-scale picture of chromatin**

Eukaryotic chromosomes are giant polymers, each formed by a huge string of nucleosomes. The conformation of this string at different length scales is generally described using an analogy with proteins: the string of nucleosomes itself can be viewed as the primary structure of chromatin; the conformation adopted by an array of a few dozen successive nucleosomes forms the secondary structure of chromatin. The 3D structural arrangement of several arrays can finally be viewed as the chromatin tertiary structure (Luger et al., 2012; Pepenella et al., 2013, see Fig. 2b). When observed by electron microscopy (see Fig. 3a,b,c), interphase chromatin appears to fill the entire nucleus volume. As genome length may vary considerably from organism to organism, the nucleus size varies accordingly: orders of magnitude go from ~10 Mb (Mega base pairs) for a diameter of the order of 2 µm in yeast (Fig. 3b), to ~100 Mb and 4 µm in drosophila fly (Fig. 3a), and up to ~1000 Mb and 10 µm in mammals (Fig. 3c). These differences in size are certainly correlated with the differences in chromatin organization that can be directly deduced by simple inspection of electron microscopy images.

Yeast nuclei are the most homogeneously filled ones, with a large, denser region called the nucleolus, which is known to be the site of very intense ribosomal RNA synthesis. A smaller, dark linear body can also be seen in the inset, connected with a star-shaped structure, the spindle pole body (SPB) from which tubular proteic assemblies, microtubules, stem and “hold” chromosomes at their centromeres. In contrast with multicellular organisms, in yeast this microtubule bundle is preserved all along the cell cycle. It is a crucial organization center for the assembly of chromosomes in interphase and for chromosome segregation during mitosis (see Fig. 3d,g).

When the nuclei of multicellular organisms are considered (Fig. 3b,c), their most striking feature is the coexistence of distinct denser and less compact regions. These regions are persistent and are not simply the result of temporal fluctuations of chromatin density. These features have been shown to strongly correlate with the transcription activity of genes. Active genes tend indeed to gather at the center of the nucleus, in a region where chromatin is less dense and more accessible, which is called euchromatin. Inactive genes are found instead in denser regions, called heterochromatin, and tend to associate with the nuclear periphery. As a stunning example of chromatin compaction and localization changes induced by transcription, the activation of a genomic locus results in a dramatic change of its topology (Fig. 4a).

With the improvement of imaging and labeling techniques, gene transcription by the RNA polymerase PolII in multicellular organisms has been shown to occur in well-defined loci, called factories (Fig. 4b) (Jackson et al., 1993). These factories are located within the euchromatin domain and each factory has a propensity to gather co-regulated genes (Jackson et al., 1998). In this picture, it appears that the functional differences between cell types are related to the way the genome is folded in the nucleus of these cells.

In the last two decades, impressive advances in experimental techniques in measurements of 3D chromosomal contacts have been made, starting from the “Chromosome conformation capture” approach (Dekker et al., 2002). Its genome wide derivative (Hi-C) enable the generation of contact maps at the genome scale (Lieberman-Aiden et al., 2009). From these maps, it is possible to reconstruct the underlying 3D structure of the genome and such structures are represented on Fig. 3i,e,f. The results confirmed the the tethering of centromeres in yeast and drosophila, but also of telomeres. In humans the reconstruction of the first chromosome gives a visual illustration of decondensed euchromatin loops emanating from globular heterochromatin globules. These are decorated with two different specific histone marks. We will come back on the results of these investigations in Sec. III.C.1.

**C. Chromosomes as polymers**

Most of the modeling efforts addressing the question of the nuclear organization have been so far oriented by polymer physics. The question then arises as to understand whether polymer physics is the main player that drives chromosome organization.

(i) In the simplest case of yeast, where chromosomes...
FIG. 3  Nuclear organization in yeast (S. Cerevisiae), drosophila and mammals. (a-c) Electron microscopy images of nuclei in yeast (a) in drosophila (b) and human (c). Scale bars correspond respectively to 1 µm, 2 µm and 2 µm. In (a), the nucleolus is the darker region in the upper part of the nucleus. The SPB is shown by a circle. In (b), the nucleolus is the dark circular region and heterochromatin can be seen as darker spots. In (c), the nucleolus is marked by a dashed circle. (Figure (a) from [http://scienceblogs.com/transcript/2008/08/16/the-centrosome-and-the-spindle/](http://scienceblogs.com/transcript/2008/08/16/the-centrosome-and-the-spindle/), (b) from [http://pixgood.com/nuclear-pore-em.html](http://pixgood.com/nuclear-pore-em.html), (c) from [http://tinyurl.com/m6phpf8](http://tinyurl.com/m6phpf8). (d-f) 3D models reconstruction (Lesne et al., 2014) from contact maps obtained using the Hi-C protocol in these three organisms (Dixon et al., 2012; Duan et al., 2010; Sexton et al., 2012). In (d) and (e) all the chromosomes are represented with different colors. In (f), only the first chromosome is shown and the colors correspond to regions harboring different epigenetic marks: H3K9Ac and H3K9me3 (see Sec. III.B). On each reconstruction, centromeres are shown as black beads and telomeres (chromosome ends) as purple beads. Each bead represents respectively 12 kb, 40 kb and 40 kb. (g-h) Polymer models of the genome in yeast and drosophila: In (h) each chromosome is labeled with a different color (Wong et al., 2012). In (g) colors correspond to the colors of chromatin (see Sec. III.C, courtesy of Giacomo Cavalli and Pascal Carrivain (IGM, Montpellier)). (i) To our knowledge, physical models of the human genome have not been developed so far.
are shorter and all anchored at the SPB by their centromeres, it seems to be indeed the case. Indeed, several polymer simulations have been able to reproduce the structure of interphase yeast nuclei (Tjong et al. 2012; Wong et al. 2012), see Fig. 3k, l. Moreover, fluorescent microscopy has been used to check the dynamical behavior in vivo of given chromosomal loci, (Albert et al. 2013; Hajjoul et al. 2013). Single particle tracking has revealed a quite uniform response within the genome, characteristic of polymers in confined spaces. Except for telomeres and for the highly transcribed DNA in the nucleolus, yeast chromosomes behave as a polymer brush, and are essentially organized by simple physical principles (Huet et al. 2014) (see Fig. 3l,g).

(ii) In the well-studied, intermediate-size case of the drosophila, recent investigations tend to indicate that this polymer behavior is partially conserved, but with some significant changes that go in the sense of greater complexity (see Fig. 3h,i). Roughly speaking, it has been proposed that euchromatin and heterochromatin have intrinsically different biochemical and physical properties, due to a deeply different protein “dressing” of the DNA molecules. More precisely, Filion and co-workers have identified five principal chromatin states, called chromatin “colors” from the analysis of 53 chromatin protein genome-binding profiles in drosophila cells (Filion et al. 2010). Among these states, some essentially correspond to active, transcribing euchromatin, other to dense, repressed heterochromatin. These chromatin states result from the recruitment of DNA-binding proteins that are specific of the underlying epigenetic marks (see Sec. III.B).

As a consequence, drosophila chromosomes are more properly described as co-polymers, i.e. polymers containing more than one type of monomer. A model of the resulting copolymer brush is depicted on Fig. 3h.

(iii) In mammals heterochromatin is mainly located at the nuclear membrane and euchromatin at the center of the nucleus (see Fig. 3j). The reconstituted 3D structure of chromosome 1 (the longest human chromosome) shows an alternance of long loops of euchromatin and dense parts of heterochromatin tethered to the nuclear membrane (see Fig. 3j).

In summary, the conformation adopted by chromatin is affected by its intrinsic structural parameters such as the NRL (the reader may find an extensive review in (Boulé et al. 2015)), on top of which lies an additional layer of modulation by internucleosomal electrostatic interactions (Hansen 2002; Pepenella et al. 2013) and binding of architectural proteins. This conformation is essential for gene regulation. The epigenetic marks present on DNA and histones, by mediating specific interactions between portions of chromatin, alter its conformation and hence its function. The next section will be devoted to understanding the complex relationship between epigenetic marking and genome structure and function.

III. FROM EPIGENETIC MARKS TO REGULATION OF GENE EXPRESSION THROUGH THE 3D ORGANIZATION OF THE GENOME

A. General principles of gene silencing. The paradigm of DNA accessibility

During development, the determination of the cell type (cell fate) involves progressive restrictions in its developmental potency and results from differential gene expression. DNA methylation is a key control parameter of this process: genes that are specific for the desired tissue are kept unmethylated, whereas the others are methylated. Moreover, patterns of DNA methylation are faithfully propagated throughout successive cell divisions (see Sec. V.A). However the physics of DNA methylation is still elusive and we therefore postpone further developments on DNA methylation to the last part of this review (see Sec. V.A).

Epigenetic regulation of gene expression involves silencing, i.e. a permanent and heritable inhibition of gene transcription (transcriptional gene silencing) or translation (post-transcriptional gene silencing). The current paradigm is that gene silencing is achieved through chromatin condensation, in a so-called heterochromatinization process (Grewal and Moazed 2003). Can we characterize the physical properties of heterochromatin and euchromatin? What are the physical consequences of heterochromatinization in terms of structure, dynamics and how do these physical consequences turn out into functional consequences?
Histones simultaneously play a crucial role in determining the structure of chromatin; they are the substrate of a vast catalog of epigenetic markings (Cantone and Fisher, 2013; Kouzarides, 2007), which is not a coincidence. This supports the hypothesis that epigenetic histone marks modulate gene expression through chromatin structural rearrangements at each level of the nuclear organization: nucleosome, chromatin fiber, chromatin loops, chromosome territories, whole nucleus (Poirier et al., 2009; Zhou et al., 2007).

### B. Histone modifications as chromatin structural modulators

Most epigenetic marking occurs on the histones that coat DNA. What are the physical consequences of this marking and what is its effect on chromatin organization?

#### 1. Histone tails and their role in internucleosomal interactions

As already mentioned, nucleosomes are formed by wrapping DNA around an octameric protein assembly formed by histone proteins. The N-terminal sequences of H2A, H3 and H4 extend from the globular histone core to form the so-called histone tails (see Fig. 2b). The H3 and H4 tails consist respectively of 35 and 20 residues, of which respectively 13 and 9 are positively charged (lysines, K and arginines, R). These tails are intrinsically disordered protein domains, hence adopt a random coil configuration, as suggested by crystallographic studies (Davey et al., 2002; Luger et al., 1997) and proteolytic cleavage assays. Tails contribute differently to intranucleosomal stability and internucleosomal interactions (Aljan et al., 1982; Arya and Schlick, 2006, 2009; Sinha and Shogren-Knaak, 2010; Zhou et al., 2007). The two H3 tails exit from the histone core close to the DNA entry-exit site of the nucleosome, and associate preferentially with DNA to “lock” its wrapping around the histone core. The H4 tails are known to associate with a set of seven residues referred to as the H2A/H2B acidic patch on adjacent nucleosome, acting as a tether connecting the two nucleosomes (Kalashnikova et al., 2013; Kan et al., 2009). The H2A and H2B tails, much shorter than their H3 and H4 counterparts and the subject of a much smaller literature, do not seem to significantly contribute to internucleosomal interactions, although they are required for proper nucleosome reconstitution (Bertin et al., 2007a).

#### 2. Histone tail post-translational modifications (PTMs)

Histone tails, besides their role in the structuration of nucleosome arrays, are also the support of virtually all PTMs targeting histones, which consist in replacing groups of atoms on one residue by another, chemically different one (see Fig. 5). For an historical account of their discovery, see (Morange, 2013). The globular histone core and the lateral surface of the nucleosome may also undergo post-translational modifications, which modulate the nucleosome stability, DNA wrapping (Tesarz and Kouzarides, 2014; Tropper and Schneider, 2013), hence chromatin architecture. The repertory of histone-tail PTMs is vast both in terms of types of modifications and in terms of where the modification can take place (Fierz, 2014; Pepenella et al., 2013; Zentner and Henikoff, 2013).

In order to reach a comprehensive physical picture we oversimplify the daunting complexity of epigenetic histone PTMs (Kouzarides, 2007) to focus here on:

(i) lysine methylation and notably the two main histone PTMs that are involved in gene silencing: tri-methylation of the lysine 9 of H3, noted H3K9me3, which recruits HP1 and tri-methylation of the lysine 27 of H3, noted H3K27me3, which recruits the Polycomb architectural complex;

(ii) lysine acetylation and specifically the acetylation of lysine 16 of H4, H4K16ac, which is a hallmark of active chromatin (actively expressed genes).

Epigenetic marks are deposited on or removed from histone tails by dedicated enzymes, so-called “writers” and “erasers” (Fierz, 2014). Writers devoted to acetylation are histone acetyltransferases (HAT), notably lysine acetyltransferases (KAT), and writers devoted to methylation are histone methyltransferases (HMT), notably lysine methyltransferases (KMT). Erasers are histone deacetyltransferases (HDAC) and histone demethyltransferases (HDM), notably lysine demethyltransferases (KDM), see Fig. 5.

A wealth of data exists regarding the presence of histone tail modifications in different species, development stages and cell types – the so-called epigenome – but efforts for characterizing the effect of histone PTMs are currently limited by the difficulty of examining in vivo chromatin structure. Interestingly, the two main modifications discussed here – lysine acetylation and lysine methylation – seem to act on the chromatin architecture and state of activity through rather different mechanisms. In the case of acetylation, a direct effect on nucleosome-nucleosome interactions is at play, with a certain but subtle relationship with the associated loss of a positive charge (see Sec. III.B.3). In contrast, methylation preserves electric charges, while introducing significant steric hindrance and potentially hydrophobic interactions, and mainly act on chromatin indirectly by recruiting additional architectural proteins
(see Sec. II.B.5). For this reasons, acetylation mechanisms are more easily studied by in vitro experiments, while methylation effects are more generally studied in the in vivo context in presence of their multiple partners. We will now sum up some of the main experimental results and theoretical interpretations concerning both these PTMs.

3. Histone tail acetylation: direct effects on chromatin accessibility

a. Experiments Experimental studies of the role of histone tail acetylation in the architecture of nucleosomal arrays are conducted using reconstituted, in vitro chromatin. In this approach, nucleosomes are reconstituted by incorporating recombinant histones with tailored aminoacid sequences on tandem repeats of a DNA sequence with very high histone affinity (the so-called "601 sequence"). The sedimentation coefficient of such arrays is then measured as a proxy for their folding propensity, comparing the sedimentation coefficient of arrays with of without combinations of histone tail acetylation [Allahverdi et al., 2011; Liu et al., 2011; Shogren-Knaak et al., 2006; Wang and Hayes, 2008]. In addition, small-angle X-ray scattering assays on folded nucleosome arrays give estimations of internucleosome interaction energies [Bertin et al., 2007c; Howell et al., 2013]. Taken together, these studies show that H4 tail acetylation decreases internucleosomal intra-array associations [Hizume et al., 2010].

Acetylation of lysine 16 of histone H4 (H4K16ac) has the strongest effect in this regard, and may lead to massive disruption of dense chromatin fibers in vitro [Shogren-Knaak et al., 2006]. Structural effects of H4K16 acetylation on chromatin compaction are also confirmed by the observation of a weakening of chromatin packing in vivo [Shahbazian and Grunstein, 2007], and are in general associated with actively transcribed genes (e. g. , [Taylor et al., 2013]). Surprisingly enough, histone H3 acetylation, which also reduces the charge of the tails, does not seem to modify the folding propensity of nucleosome arrays [Wang and Hayes, 2008] pointing to a specific mechanism of H4K16 acetylation.

b. Models Experimental studies are often combined with computational models to provide deeper insights on how the electrostatic nature of histone tail PTMs influence chromatin folding.

Potoyan & Papoian [Potoyan and Papoian, 2012] addressed the question of the decompaction induced by H4K16 acetylation, and carried out all-atom simulations in explicit solvent to compare the conformation of H4 tail with and without this modification. For the isolated histone tails, H4K16ac leads to slightly more compact and significantly more structured globular H4 tails. At this level, compaction is not surprising since the net charge reduction weakens self repulsion between the tail residues. When DNA is present, i.e. when the entire nucleosome is considered, tails have a similar behavior: acetylated tails are more compact, less fluctuating, and are more frequently bound to their own nucleosomal DNA. However the less charged acetylated tail interacts much more strongly (∼ 5 − 6kB T) with DNA than the unmodified one (∼ 2kB T), in contrast to what is expected from electrostatic reasons. This counterintuitive effect is achieved thanks to an important tail reorganization that brings other lysines closer to DNA. While the overall electrostatic attraction is basically unchanged,
FIG. 6 Nucleosome arrays and histone PTMs. (a) Cartoon of a nucleosome core particle (histone core in yellow, DNA in blue) (b-d) Examples of chromatin structural modulation by histone PTMs. Right: H4K16ac decreases nucleosome stacking by preventing the H4 tail from binding the H2A/H2B acidic patch. Global acetylation removes positive charges on H4 tails and decreases the electrostatic screening of NCP electrostatic repulsion. Middle: H3 methylation recruits chromatin associated proteins to form heterochromatin (e.g. H3K9me3 and HP1, H3K27me3 and the Polycomb family complexes). Left: acetylation of H3 tails decreases their affinity for nucleosomal or linker DNA and reduces the electrostatic screening of DNA negative charges, leading to changes in the mechanical properties of the linker and accessibility of nucleosomal DNA for other proteins.

the collapse of the tail is favored by hydrophobic interaction and entropic gain. In contrast, unmodified H4 tails are more extended and flexible. They showed a preferential interaction with linker DNA (Angelov et al., 2001) and with an acidic patch exposed on the surface of next H2A/H2B dimers of neighboring nucleosomes (Zhou et al., 2007) (see Figs. 2a and 6a). Hence, while modified H4 tails may contribute to the nucleosome-nucleosome attractive interaction by the so-called “tail bridging” effect (Mühlbacher et al., 2006), the acetylation of lysine 16 might oppose this effect, leading to weakened nucleosome-nucleosome interactions (Potokan and Papoian, 2012) (see Fig. 6a,c,d). Of note this is qualitatively consistent with experiments on the disordered C-terminal tail of the p53 protein where a significant increase of its site-specific DNA binding is observed both in vitro and in vivo (Luo et al., 2004).

Other computational models generally rely on coarse-grained approximations of the nucleosome core particle and linker DNA which integrate the mechanical dynamics of nucleosome as well as its distribution of charges. Arya & Schlick used their Discrete Surface Charge Optimization framework to provide estimations of the contribution of tails to electrostatic interaction energies, showing that H3 tails principally screen the negative charge of linker DNA, while H4 tails mediate internucleosomal interactions (Arya and Schlick, 2006, 2009), in agreement with previous experimental findings. However, these studies do not compare interaction energies with or without histone PTMs. Several other coarse-grained models have been used so far to specifically investigate histone tail acetylation (Aliabavi et al., 2011; Liu et al., 2011; Yang et al., 2009), showing that the effect of PTMs also largely depend on the valency and the concentration of bulk counterions, consistent with sedimentation assays.
4. H4K16 acetylation is a silencing mark in budding yeast

In budding yeast, and this is specific to budding yeast, silencing is not achieved by histone methylation. Instead heterochromatin is induced by SIR (Silent Information Regulatory) complexes which are recruited by deacetylated nucleosomes, crucially relying on H4K16 (Dayarian and Sengupta, 2013) (see below Sec. [V.C.1]).

5. Histone tail methylation: indirect effects on chromatin condensation

In animals, notably in drosophila and mammals, silencing is mainly achieved through histone tail methylation which, as mentioned above, does not directly induce chromatin fiber compaction (a notable exception was reported in (North et al., 2014)) but leads to the recruitment of additional architectural proteins, typically heterochromatin proteins.

Importantly, such architectural proteins are included in the set of proteins that have been used to define the chromatin colors in drosophila (Filion et al., 2010). Precisely, chromatin colors are specific combinations of epigenetic marks and associated proteins belonging to the following set: histone-modifying enzymes, proteins that bind specific histone modifications, general transcription machinery components, nucleosome remodelers, insulator proteins, heterochromatin proteins, structural components of chromatin, and a selection of DNA binding factors (Filion et al., 2010). Histone tail methylation seems therefore to act as a (region specific) substrate to recruit (non specific) proteins. In turn, these proteins induce different chromatin-chromatin interactions in different regions, and eventually different chromatin folding leading in particular to different compaction degrees (see Sec. [II.C.1]).

There are various kinds of heterochromatin in animals (e.g. black, blue and green chromatin in drosophila; even more “colors” in mammals). We focus here on the physical mechanisms that drive the two main silencing processes in animals, namely the recruitment and spreading of HP1 (Heterochromatin Protein 1) by the H3K9me3 mark (Azzaz et al., 2014; Hathaway et al., 2012) and the recruitment of the Polycomb architectural complex (PcG) by the H3K27me3 mark (Tie et al., 2009). We moreover discuss the role of these architectural proteins in the physical process of heterochromatization.

Unlike acetylation, results obtained in vitro using reconstituted chromatin arrays are not directly transferrable to in vivo contexts for at least two reasons: (i) lysine methylation has no direct physical effect (recall that, unlike lysine acetylation, lysine methylation does not change electric charges), instead, lysine methylation is recognized as a biochemical tag by dedicated chromatin proteins, either architectural (Gosalia et al., 2014; Muligan et al., 2015; Ong and Corces, 2014; Zentner and Henikoff, 2013) or remodeling proteins ((Becker and Hrz, 2002); (ii) there is considerable cross-talk among histone tail PTMs (Bannister and Kouzarides, 2011; Kouzarides, 2007; Li and Shogren-Knaak, 2008) which can then form networks comparable to signaling pathways, eventually resulting in a structural effect. An example of such a pathway is given by (Wilkins et al., 2014) in the context of budding yeast cell division where phosphorylation of Serine 10 of the H3 tail induces H4K16 deacetylation, which eventually leads to chromatin compaction.

a. HP1-mediated heterochromatin The family of Heterochromatin Protein 1 (HP1) are fundamental components of heterochromatin. They are abundant at the centromeres and telomeres (which correspond roughly, as we have seen, to central and ending regions of the chromosomes, respectively) in nearly all eukaryotes.

They display high binding affinity for the H3K9me3 mark and are therefore specifically targeted to nucleosomes harboring this mark. However the spreading of HP1 along an H3K9me3 epigenetic domain is still a matter of debate. Thus in the latest special issue of JPCM (Everaers and Schiessel, 2015), devoted to the physics of chromatin, two contrasted models have been proposed: the group of Andrew Spakowitz (Mulligan et al., 2015) claims that bridging interaction between HP1 dimers is critical for HP1 spreading, at odds with the group of Karsten Rippe (Teif et al., 2015) who claims that the binding of one HP1 dimer can stabilize a stacked nucleosome conformation and facilitate the binding of a second dimer via an allosteric change of the nucleosome substrate, with no need for a direct interaction between neighboring HP1 dimers. It is to be noted that both groups could reproduce the in vitro binding curves of the yeast analog of HP1 (Swi6) on mono- and dinucleosomes as well as on arrays of nucleosomes. Moreover Spakowitz’s group claims that HP1 bridging interaction between different chromatin fibers explains the phase separation of hetero- and euchromatin (Mulligan et al., 2015), whereas Rippe’s group evidenced a dependence of the binding stoichiometry on the NRL (nucleosome repeat length) due to allosteric cooperativity of binding for nucleosome arrays with long but not with short DNA linkers, pointing to a facilitated spreading of HP1 on long NRL substrates.

b. Polycomb-mediated heterochromatin Polycomb are a family proteins that mediate transcriptional silencing (Di Croce and Belin, 2013; Simon and Kingston, 2013). In drosophila, it was found that two distinct regulatory complexes (PRC1 and PRC2) are able to silence the Hox genes in a stable and inheritable way (Beuchle et al., 2001; Paro et al., 1998). It provides a mechanism for
“cellular memory” (Ringrose and Paro, 2004), that has been speculated to be alternative to DNA methylation (Bird, 2002).

The precise mechanism underlying the heritability of the repressed state of genes silenced by the Polycomb complexes is still debated. It is known that the repressive histone mark H3K27me3 (see Sec. II.B.5) is recruited by the PRC2 complex. In turn, H3K27me3 recruits PRC1, which then induces histone H2AK119 ubiquitination. However, recent studies showed that this relationship may also work in the opposite sense (Blackledge et al., 2014). It has also been suggested that in X chromosome inactivation (see Sec. V.C), histone ubiquitination, and Polycomb proteins are mechanistically related to propagate the silenced state (de Napoles et al., 2004).

A physical modelling of the cross-talk between histone marks and the Polycomb complexes would be useful and is, to the best of our knowledge, still missing.

C. How epigenetic marks organize the chromosomes in the cell nucleus. General rules. Physical modeling of epigenome wide studies.

1. Epigenome wide studies

One of the current paradigms in the field is that the epigenetic landscape is driving the 3D genome folding and by extension the functional state of the cell. In order to tackle this issue at the genome scale level, epigenomic techniques based on Next Generation Sequencing (NGS) are increasingly used (Rivera and Ren, 2013). These techniques are commonly used to map accessibility, protein binding sites, and biochemical modification of histones or DNA along the linear genome (e.g. in drosophila Fig. 7a). A new technique, genome-wide Chromosomal Conformation Capture (Hi-C) has been developed in order to address the issue of genome 3D folding using NGS. This technique allows to the generation of a list of pairwise contacts between distal parts of the genome in various organisms or cell types (e.g. in drosophila, Fig. 7a). A physical modelling of the cross-talk between histone marks and the Polycomb complexes would be useful and is, to the best of our knowledge, still missing.

Jost et al. (Jost et al., 2014) show a phase diagram of a toy model copolymer as a function of the intensity of (i) block-specific and (ii) non-specific interactions, that we show in Fig. 8. On top of the coil-globule transition of the whole copolymer, there is also coil-globule transition restricted to each separate block. Importantly, both coil and globule phases coexist in a region of the phase diagram, the size of which depends on the (average) size of the blocks. This is consistent with the finite-size scaling analysis of the coil-globule transition which has been proposed in (Caré et al., 2014) (see also arXiv: arXiv:cond-mat/0004273).

Let us show that both transitions, namely the coil-globule transition inside a given block and the segregation of different blocks of the same color into separated microphases, overlap in the phase diagram because of finite-size effects.

We first remember that a polymer of N monomers, with monomer-monomer attractive interactions, under-
FIG. 7 Modeling of chromosomal contact maps from the epigenomic landscape. (a) Profiles of H1 occupancy, DNA accessibility, H3K27me3, H3K4me3, HP1 and a histone modifier, Su(Hw) along a region of the Drosophila chromosome 3R. At the bottom, the colors corresponding to these profiles are shown. Yellow and red correspond to active chromatin, blue to polycomb bound regions. (Filion et al. 2010) (b) The corresponding contact map (Sexton et al. 2012) (c) Schematics of the co-polymer model used in (Jost et al. 2014) (d) Two predicted contact maps corresponding to the region indicated by the pink dashed square in (b).

FIG. 8 Phase diagram of a toy model copolymer, as a function of specific and non-specific interactions. Figure taken from (Jost et al. 2014).

We then consider a copolymer ABAB... made of small blocks A and long blocks B, with monomer-monomer attractive interactions represented by an energy of interaction $E_{ij}$ between monomers with epigenetic states $i$ and $j$ the following kind: $E_{ij} = U_{ns} + \delta_{ij}U_s$, where $U_{ns}$ is a non-specific term (does not depend on $i$ and $j$), $\delta_{ij}$ is the Kronecker delta, and $U_s$ is a specific interaction term. According to the preceding results on the coil-globule transition of finite-size polymers, long blocks B go into globules when small blocks A are still coils. When lowering the temperature (or equivalently increasing the interactions), blocks B start to transiently bind together into a macroglobule: this is now the coil-globule transition of the whole copolymer which is equivalent to a chain of B globules separated by A linkers; and while this chain collapses (folds) the A linkers start to go into globules, so that both transitions overlap.

Importantly the macroglobule fluctuates between coil and globule conformations (as well as any B globule) so that it transiently dissociates thus permitting the small A blocks, even in remote locations on the genome, to come transiently into contact (see fig. 5). This corresponds to the multistate folding region calculated by Jost et al. and depicted on Fig. 8. Note that the width of this multistate folding region varies as $1/\ln(n)$ where $n$ is the typical size of the small(est) blocks.

Below the lower critical temperature $\Theta(N) = \Theta(1 - b\sqrt{\ln(N)}/N$ where $b$ is a dimensionless prefactor of order unity. More precisely, there is an equilibrium between coil and globule conformations over a temperature range between $\Theta(N) - a/\sqrt{\ln(N)}$ and $\Theta(N) + a/\sqrt{\ln(N)}$ where $a$ is a dimensionless prefactor of order unity. At $T = \Theta(N)$ both coil and globule conformations are in equal proportions. Therefore, at a given temperature $T$, longer polymers are more globular than small polymers of the same kind.
B blocks are globular and as such in the heterochromatin phase, hence their underlying sequence is repressed.

IV. PHYSICAL MECHANISMS INVOLVED IN THE INITIATION, SPREADING, MAINTENANCE AND HERITABILITY OF EPIGENETIC MARKS

Stem cells are capable of differentiating to the desired fate depending on the tissue. Dramatic changes in gene expression occur during development. These changes are then stabilized and become heritable. Epigenetic modifications take part in both initiating, stabilizing and propagating the patterns of gene expression. Gene regulation by epigenetic modifications is indeed stably propagated through cell divisions (and, in some cases, across generations). At each cell division, the whole DNA is replicated. Chromosomes then consist of two sister chromatids which both have identical genetic information, joined together at their centromere. Then, during mitosis, the two chromatids are separated and segregated into the two nuclei of the daughter cells.

Eukaryotic replication involves both DNA synthesis and chromatin assembly. As the two double helices are synthesized from the two single strands of the mother-cell DNA, nucleosomes on the mother-cell DNA strand should also be distributed to both daughter double helices, and completed by de novo nucleosome assembly. In order to ensure the transmission of epigenetic marks to daughter cells, mother-cell nucleosomes should be shared by both newly formed chromosomes, even if the detailed mechanisms of this distribution are still debated (MacAlpine and Almouzni, 2013).

While it is clear that histone modifications are involved in gene silencing, hence gene regulation, the questions how epigenetic marking is initiated, how it may spread over specific chromosome regions (and not beyond), and how it can be stably maintained along the cell cycle and through the cell division are still under investigation. In this section, we will review the main modeling efforts that have been made in order to address these questions.

A. Mathematical modelling

Many recent theoretical works addressed the question of how epigenetic marks are initiated, spread, and maintained. The main objective of these models is to reproduce a few essential features observed in vivo: (a) the multistability of the epigenetic marks; (b) their spatial patterns and (c) their heritability.

By multistability, it is generally meant that the epigenetic marks act as switches between different functional states. In the simplest case, different patterns of epigenetic marks allow to switch between two states that have a well-defined functional characterization (bistability). Such functional states are then inherited by the daughter cells through mitosis, which is what we call heritability. As observed in genome-wide studies, the epigenetic patterns correspond to distinct epigenomic domains that are separated by boundaries (see Sec. III.C).

We consider a system of N nucleosomes that can be in $n_S$ different states. In the simplest case, $n_S = 2$ and one refers to “modified” or “unmodified” states, which can be related to active or inactive genes.

The state of the system is described by the variables $\{s_1, s_2, \ldots, s_N\}$, where $s_i$ is the state of nucleosome $i$. If we define $n_j$ as the number of nucleosomes in the state $j$, then one can write the conservation of the number of nucleosomes as

$$\sum_{j=1}^{n_S} n_j = N \quad (1)$$

Many theoretical works use the silenced mating-type locus of the fission yeast *Schizosaccharomyces pombe* (reviewed in (Grewal and Elgin, 2002)) as a model system. In this system, the region containing the two mating-type regions is normally “silenced”, i.e. not expressed. The expression of the mating-type genes may become bistable in mutants, flipping between a silenced state and an active state (Grewal and Kl 1996; Thon and Friis, 1997. Each state is stable and heritable; transition between them occurs apparently stochastically. The *S. pombe* HMT, HDAC and other proteins are necessary for silencing, and all may bound H3K9me directly or indirectly.

In the following, we review the models of this behavior proposed so far.

B. Zero-dimensional models

In zero-dimensional models, neither the spatial organization of the N nucleosomes, nor the notion of distance are introduced. In general, the model concerns rate equations on how the variables $n_j$ vary as a function of time, and the objective of the models is to show how bistable or multistable states can appear. In this class of models, the initiation of the epigenetic mark is implicitly defined as the initial state of the dynamical system, and the spreading is described as the time evolution of the initial state. Mitosis can be modeled as an instantaneous process in which the concentrations of all species (modified and unmodified nucleosomes) are diluted and the system restarts. The dilution is due to sharing of mother-cell nucleosomes between both daughter cells. Nucleosomes are not necessarily shared into equal parts between daughter chromosomes, but this may be assumed without loss of generality as is done for convenience in most models.

We can write a general expression for the time evolu-
tion of the variables $n_j$:

$$\frac{dn_j}{dt} = \sum_{k \neq j} R_{kj}^+[n_k - n_j] + \text{noise} \quad (2)$$

Here, $R_{kj}^+$ is the rate of transition of nucleosomes from the state $k$ to the state $j$, while $R_{kj}^-$ is the rate of transition from state $j$ to $k$ (obviously, $R_{kj}^- = -R_{kj}^+$). In general, these coefficients are not constant, but depend on the other dynamical variables. The “noise” may be included to describe the effect of stochastic processes involved in the system.

The simplest possible model of this kind was proposed by Michaelsen et al. (Michaelsen et al. 2010). The authors consider the case of $n_S = 2$, that is, they consider only a modified (M) or unmodified (U) state. Using equation (1), the system may be described by only one variable $n_M$, the number of modified nucleosomes. The transition rates are given by

$$R_{UM}^+ = \alpha n_M^2 + (1 - \alpha)$$
$$R_{MU}^- = \alpha (1 - n_M)^2 + (1 - \alpha). \quad (3)$$

This model supposes that the creation of a modified state involves a cooperative transition (as evidenced by the quadratic terms in equation (3)) or a spontaneous conversion to the unmodified state (which is described by the $(1 - \alpha)$ term). Despite its simplicity, the model can account for the emergence of bistability. The parameter $F = \alpha/(1 - \alpha)$ (feedback to noise ratio) governs the behavior of the system. For $F > 4$, three fixed points emerge in the system: $n_M = 0$ and $n_M = N$, which are stable, and $n_M = N/2$, which is unstable. The $F$ parameter is possibly under active control by the cell, which then can regulate its function (notably by HDAC inhibitors (Dayarian and Sengupta 2013)). Heritability can be partially accounted for by this model, since one can speculate that cell division brings the system close to the unstable point, which then returns to its stable attractor.

David-Rus et al. thoroughly investigated a more general model that still has $n_S = 2$ (David-Rus et al. 2009). Their rates read:

$$R_{UM}^+ = \chi + \alpha n_M^H$$
$$R_{MU}^- = \gamma + \eta (1 - n_M)^K \quad (4)$$

The first interesting result they obtained is that this model can reproduce bistability only for $H, K > 1$. The simple quadratic case $H = K = 2$ is a generalisation of the model of Michaelsen et al. (Michaelsen et al. 2010), where the cooperative transition probability (rate) from U to M is independent from that to M to U. If the basal rates $\chi$ and $\gamma$ are small, one again obtains three fixed points, with the intermediate unstable point being $n_{M2} \approx \eta/(\alpha + \eta)$. Assuming now that cell divisions exactly halves the concentration of modified nucleosomes for each daugther cell, if $n_{M2} > n_{M3}/2$, then the system will always fall in the basin of attraction of $n_{M1}$ after a cell division, hence the only stable point is the unmodified state $n_{M1} \approx 0$. Conversely, for $n_{M2} < n_{M3}/2$ (hence $\eta < \alpha$), the system will converge to the modified state fixed point $n_{M3} \approx N$ for initial conditions larger than $2n_{M2}$, and bistability becomes effective.

This scenario is however modified by the presence of noise in the system. In fact, if the probability of transition from U to M is larger than the probability of transition from M to U (that is, $\eta < \alpha$), then the $n_{M1}$ fixed point is no longer stable. Noise drives the system out of the $n_M = 0$ state, and brings it to the fully modified $n_M \approx N$ state. This consideration highlights the importance of asymmetric recruitment rates.

The same authors also considered the case of $n_S = 3$, which was already considered by Dodd et al. (Dodd et al. 2007) in a very similar form. They consider the case of an “antimodified” state (A), that is, possibly an acetylated state (active chromatin mark) that is opposed to the M state which is possibly a methylated (repressive) state (see Fig. 3). A hypothesis is that only the $U \rightarrow M$ and $U \rightarrow A$ are allowed, but the $M \rightarrow A$ transition is not (i.e., $R_{MA}^+ = 0$). They write the following transition rates:

$$R_{UA}^+ = \alpha A n_A + \chi A$$
$$R_{UM}^+ = \alpha M n_M + \chi M$$
$$R_{AU}^- = \beta M n_M + \gamma A$$
$$R_{MU}^- = \beta A n_A + \gamma M. \quad (5)$$

The study of the system in the case where the basal rates $\chi_M$, $\gamma_M$, $\chi_A$ and $\gamma_A$ vanish already shows the existence of four fixed points: two stable fixed points, $\{a = 1, m = 0\}$ and $\{a = 0, m = 1\}$, an unstable saddle point $\{a = 0, m = 0\}$ and an unstable fixed point $\{a = \alpha M / (\alpha A + \beta M), m = \alpha A / (\alpha A + \beta M)\}$. The two latter points are aligned along the $a = m$ line and create a barrier between the two basins of attraction (David-Rus et al. 2009).

The last model of this class that we consider is the one proposed by Jost (Jost 2014). The author considers a special case of the three-state model outlined above:

$$R_{UA}^+ = \epsilon_A n_A + k_0$$
$$R_{UM}^+ = \epsilon_M n_M + k_0$$
$$R_{AU}^- = \epsilon_M n_M + k_0$$
$$R_{MU}^- = \epsilon_A n_A + k_0. \quad (6)$$

that is, it is the same model with $\alpha_{A,M} = \beta_{A,M} = \epsilon_{A,M}$ and $\chi_{A,M} = \gamma_{A,M} = k_0$. Interestingly, this particular choice allows to map the system to the zero-dimensional
Ising model, with, e.g., the correspondence $A = +1$, $U = 0$, $M = -1$. Within this analogy, recruitment corresponds to coupling between spins and random transitions are associated with thermal fluctuations. A new observable, equivalent to the magnetization in the Ising model, is introduced here: $\mu = a - m$.

Some known results can thus be recalled for the symmetric recruitment case $\epsilon_A = \epsilon_M = \epsilon$. Similarly to what previously discussed, three fixed points exist. The first one, $\mu = 0$, is stable for weak recruitment, i.e. for $\epsilon < \epsilon_c = 3k_0$. Above this critical value of $\epsilon$, $\mu = 0$ becomes unstable and bistability settles down with the appearance of two stable fixed points, $\mu = \pm (k_0/\epsilon)\sqrt{(\epsilon/k_0 + 1)(\epsilon/k_0 - 3)}$.

The non-local character of the nucleosome-nucleosome interaction, which is the main hypothesis of the zero-order models, has been further justified by a recent work (Zhang et al. 2014). The authors proposed a two-layer Potts model in which in one layer they describe the nucleosomes, and in the other they include explicitly the enzymes that modify the nucleosomes. The interaction between the nucleosomes is the effectively mediated by the modifying enzymes. Interestingly, by integrating out the effect of the modifiers, it is possible to prove the exact equivalence to the model proposed by Dodd et al. (Dodd et al. 2007).

To conclude this section, we stress the main results of this comparative analysis. Bistability is obtained by this class of models in two ways: in two-state models only when including nonlinear rates, and in three-state models even having linear rates. The reason for this is that in three-state models the transitions from a modified to an unmodified state can proceed only in a two-step process, effectively requiring cooperativity, hence producing bistable states (Dodd et al. 2007).

C. Higher-dimensional models

An inherent limit of the models discussed above is that they cannot reproduce spatial patterning of the epigenetic marks. Hence, limitation of the mark spreading should be included by limiting the extent of the concerned domain, i.e. the total number of nucleosomes. If this assumption may be relevant, e.g., for the mating-type loci in yeast, it probably fails when multicellular organisms are considered. It is known, for example, that nearly all noncentromeric H3K4me3 domains in mouse embryonic stem cells have a peaked shape, with continuously decaying mark densities on both sides (Hathaway et al. 2012).

1. One-dimensional models.

Even when mark spreading is surrounded by boundaries, the question arises how to model their presence and effects. Dodd and Sneppen realize in their 2011 work (Dodd and Sneppen 2011) that positive feedback can lead to spreading of the modifications to genome regions other than the target. They refer in particular to the silent mating-type loci in budding yeast 
Saccharomyces cerevisiae, HML and HMR, that are able to spontaneously flip between high and low expression states (Xu et al. 2006). These domains are stable up to 80 cell generations, and are surrounded by boundary elements that prevent silencing to spread out of the domains. These “barriers” are specific sequences, and may simply be target sites for certain DNA-binding proteins, strong gene promoters, or nucleosome-excluding structures. Dodd an Sneppen therefore consider a model in which all nucleosomes are explicitly treated, and the long-distance interaction between nucleosomes is modeled in a “local-local”, “local-global”, or “global-global” scheme (see Fig. 8). To limit the long-range interaction between DNA sites one can introduce a distance dependent cooperativity, i.e. by making the reaction rate $R_{U,M}$ dependent on the nucleosome distance. A power law dependence, typical of the three-dimensional probability of contact, can be assumed.

Then, the confinement of silenced regions can be obtained by introducing local barriers, modeled as single nucleosomes fixed in the active (A) state. Due to the local character of the modification step, a single silencing-resistant nucleosome (e.g. H3K4me3 (Venkatasubrahmanyam et al. 2007)) or a nucleosome-depleted region (notably promoters of TEF genes (Bi et al. 2004)) is enough to stop the silencing spreading, provided that the flanking regions are entirely in the active state. However, an occasional inactivation of the barrier make the silencing spread out. This effect can be limited by introducing regularly spaced weak barriers, modeled as anti-silencers (enhancers) of the $U \rightarrow A$ reaction, or by implementing in the model a Michaelis-Menten saturation effect when the number of U state nucleosomes increases. The combination of both effects results in robust prevention of silencing spreading.

Still referring to the mating-type loci in budding yeast, Dayarian and Sengupta consider a four-state model with site-dependent rate equations (Dayarian and Sengupta 2013). The fourth state they consider is the diacetylated state, which would correspond to acetylation of two H4K16 sites. Importantly, in this model, the modified state M is supposed to be a state where nucleosomes are bound to silencing (Sir) proteins, and depends therefore on their availability (concentration). In its most general form, this is a one-dimensional model that explicitly describes cooperative transitions that involve any nucleosome pair. However, it can be simplified into a zero-
dimensional model when considering uniform solutions, which again show bistability and a characteristic bifurcation diagram. Moreover, such concentration dependent model allows for additional interesting effects, involving a fine balance between the silencing of mating-type loci, which have a definite extent, and of the telomeres, whose extent may vary depending on the protein availability (Dayarian and Sengupta, 2013).

Focusing instead on mammal silenced regions, Hathaway et al. were able to reproduce the sharp peaks observed in the experimental modification patterns by including a “source” term in their model (Hathaway et al., 2012). This is a model in which the initiation and spreading are explicitly separated, and in turns this allows to reproduce spatial patterning. They write rates as

\[ U_0 \xrightarrow{k_{\text{target}}} M_0 \]
\[ \{M_i, U_{i+1} \text{ or } i-1\} \xrightarrow{k_+} \{M_i, M_{i+1} \text{ or } i-1\} \]
\[ M_i \xrightarrow{k_-} U_i. \] (7)

This description means that at site 0 there is an active modification source with rate \( k_{\text{target}} \), which then spreads to the neighboring nucleosomes with rate \( k_+ \). Fitting to experimental results leads to \( k_+ \) and \( k_- \) rates both of the order of 0.1–0.2 h\(^{-1}\) (in agreement with different experimental estimates of \( k_+ \)).

In Ref. (Hodges and Crabtree, 2012) a more detailed study of the model is presented. The source term ensures that the resulting mark distribution are peaked at the nucleation site, as experimentally observed, provided \( k_{\text{target}}/k_- \) is large enough (\( \gtrsim 0.2 \)), with increased amplitude and formation rate for increasing \( k_{\text{target}} \).

2. Three-dimensional models.

Erdel et al. (Erdel et al., 2013) addressed some more specific questions about the establishment of epigenetic domains, as how are the chromatin modifying enzymes targeted or excluded from given chromatin regions, and how exactly the modification can propagate from one nucleosome to another, how is this state reestablished or maintained during replication. The proposed model focuses on the permanent binding of enzymes to a scaffold, either on chromatin itself or on the nuclear membrane, this leading to the definition of a limited chromatin region allowed to interact with the enzyme by short-range diffusion. The spatial distribution of the enzyme hence may result in a spatially limited enzymatic activity, and results in the definition of epigenetic domains. This first attempt to take into account the chromatin architecture in a three-dimensional model is noteworthy, despite the difficulty in estimating many of the geometrical and physical parameters involved in the model, as the linear base-pair density along the chromatin fiber, the fiber stiffness, or the nucleosome local density. Moreover, the question of how the set up of the correct architecture in the initial enzyme binding and in defining the functional chromatin domains remains open.

D. Biological relevance of the models

In this section we intend to examine the biological relevance of a few key points that emerged in the discussion of models of initiation and spreading of epigenetic modifications.

1. Waddington’s epigenetic landscape revisited.

First, let us return to the discussion on the Waddington landscape that we started in the Introduction. The classical image in the original Waddington representation (Waddington, 1957) of a marble rolling down a hill does rather suggest a fixed landscape, leading to erroneous interpretation when one goes beyond the metaphorical level (see Fig. 1).

In the simplest model we discussed, the one by Micheelsen et al. (Micheelsen et al., 2010), the authors show that the model can be reformulated by a Fokker-Planck equation for the 1D diffusion of a particle in an effective potential \( V(M) \) (see Fig. 9). The latter accounts at a time for drift (external forces) and noise events (with a term of the type \( D/\mu \), with \( D \) the diffusion coefficient and \( \mu \) the mobility). The Waddington idea of an epigenetic landscape is translated in (Micheelsen et al., 2010) in more modern terms, by defining a physically consistent energy profile. Note however than the mechanism invoked here is not an evolution along the profile of Fig. 9, toward the minimum energy states, since different values of the \( F \) parameter correspond to different system parameters, hence different external constraints. In other words, the equivalent of an epigenetic landscape corresponds here to a given \( F = \text{constant} \) section of the two dimensional potential surface of Fig. 9. This allows in turn to suppose that external constraints may be included in the parameter \( F \), which may vary as a function of metabolism (level of activity) or drug delivering of “writers” or “erasers” (see Sec. III.B), notably HDAC inhibitors (Dayarian and Sengupta, 2013), thus typically making the system switch from bistable to monostable conditions. As discussed by Jost (Jost, 2014), this may also represent a strategy to gain in system sensitivity hence plasticity during development. Note that the switching mechanism between bistable and monostable conditions can be interpreted as the result of an active process bringing the system out of bistability and favoring its switching to a different state.

We then stress that it is important to consider the asymmetry of the modification rates. Taking the no-
FIG. 9 A modern view on epigenetic landscapes. (a) The basic mechanism of a three-state nucleosome modification model, depicting the modified (M), unmodified (U) and antimodified states (A). The transition between M and U is catalyzed by histone methyltransferases (HMTs) and histone demethylases (HDMs), which depends on an antimodified histone; between U and A the transition is catalyzed by histone acetylases (HATs) and histone deacetylases (HDACs), which depends on a modified histone. Figure taken from (Dodd et al., 2007). (b) The coarse-grained potential $V(m)$ as defined in the main text, as a function of the mean fractional number of modified nucleosomes $m$, and the feedback-to-noise ratio $F$ (taken from (Micheelsen et al., 2010)). (c) Different modes of coupling between histone modification states, as described in (Dodd and Sneppen, 2011). Unrecruited enzymes may modify histones directly. Otherwise, recruited enzymes may operate in non-cooperative (global or local) or cooperative (local-local, global-local, or global-local) modes. Figure taken from (Dodd and Sneppen, 2011). (d-h) Illustration of the “epigenetic landscape” as proposed by Jost (Jost, 2014). As a function of the control parameters $\epsilon_A$ and $\epsilon_I$, the system may undergo a transition between an inactive, active or bistable state. As shown in (h), the system may also develop a hysteretic behavior.

For even stronger recruitment, a typical hysteretic behavior appears that may have important biological consequences. One can expect indeed that, while for differentiated, stable cells recruitment parameters are almost symmetric, modifications of the environment might actively...
induce asymmetric recruitment. The increase of one recruitment parameter can thus bring the system along the metastable branch, then make it abruptly switch to the alternative state, which will then remain stable even when the recruitment parameters come back to their initial values, thanks to the hysteretic shape of the bifurcation curve. In Fig. 9, starting for instance from the low $m$ state and symmetric recruitment, one can increase $\epsilon_1/k_0$ and switch to the upper, high $m$ branch, then come back to $\epsilon_A = \epsilon_1$ without switching back (see also Fig. 9-c-h).

Close to $\epsilon_A \sim \epsilon_M \sim \epsilon_e$, the system becomes ultra-sensitive to perturbations, and highly unstable. This regime may be associated to diseases. A pathological increase in the frequency of replication, for instance, may result in an increase of the random transition rate $k_0$, which in turn may bring the system close to the critical point and induce epigenetic instability and misregulation.

However, the existence of a critical region may also represent an advantage. During development, the ability to switch between two coherent states when applying a weak asymmetric signal (the developmental signal) may facilitate developmental transitions. Since the random transition rate $k_0$ may be increased by reducing the cell cycle, the system can be brought closer to the critical region and the switch induced by the application of a weak asymmetric signal during a finite period of time (Jost, 2014).

E. Example: plant vernalization

The 3-state model proposed by Dodd et al. (Dodd et al., 2007) has been successfully adapted to the description on vernalization, the mechanism allowing plants to flower after a prolonged cold period.

Plants have the ability to measure the duration of a cold season and to remember this prior cold exposure in the spring. In Arabidopsis thaliana, an annual plant, a prolonged cold exposure progressively triggers the H3K27me3-mediated epigenetic silencing of Flowering Locus C (FLC), a locus encoding for proteins that in turn act as flowering repressors. The accumulation of histone epigenetic marks in the FLC locus keeps increasing during the cold. This slow dynamics of vernalization, taking place over weeks in the cold, generate a level of stable silencing of FLC in the subsequent warm that depends quantitatively on the length of the prior cold. Then, once the FLC is switched off, the silencing persists at the return of the warm season, and is mitotically stable through the rest of the development (often for many months) (see Fig. 10 (Song et al., 2013). This latter feature is characteristic of annual plants, while in FLC perennial plants is repressed only transiently.

Satake and Iwasa (Satake and Iwasa, 2012) show that this behavior can be accounted for by means of the Dodd 3-state model (Dodd et al., 2007), provided that an explicit dependence on temperature of the model parameters is included. Explicitly, the transition rates are written in this case as

$$R_{U \rightarrow A} = \beta n_A + \chi$$  \hspace{1cm} (8)

$$R_{U \rightarrow M} = u(T)(\beta n_M + \chi)$$  \hspace{1cm} (9)

$$R_{A \rightarrow U} = v(T)(\beta n_M + \chi)$$  \hspace{1cm} (10)

$$R_{M \rightarrow U} = \epsilon(\beta n_A + \chi),$$  \hspace{1cm} (11)

where $u(T)$ and $v(T)$ account for the temperature tuning and takes different values in warm conditions before vernalization, in cold conditions during vernalization, and in warm conditions after vernalization. Transition rates are in fact under the control of a series of proteins (and in particular Vernalization Insensitive 3, VIN3) whose expression is temperature dependent. Authors prove that a strong feedback, hence bistability, is necessary to reproduce the experimental observations. Interestingly, when the system evolution is simulated, the $M \leftrightarrow A$ transition is observed at a random time during the cold, for a given system containing $N$ nucleosomes (i.e. a given cell). Different cells switch therefore to the repressed state after different delays after the change from warm to cold. However, the average over a cell population leads to a typical behavior that can be reproduced, if the cell population is large enough (Satake and Iwasa, 2012). The duration of winter memory is also tuned by model parameters, and in particular by those accounting for to call division rate and rapidity of deposition of epigenetic marks after vernalization. Changes in these parameters may lead to a much short memory extent (from more than one year to a few days), this potentially explaining the different behavior observed in annuals and perennials plants.

While the previous work addressed the question of bistability behavior in vernalization, the question of the establishment of epigenetic marks induced by cold is discussed by Angel et al. (Angel et al., 2011), both theoretically and by experiments in Arabidopsis thaliana. These authors focus on the fact that, when subjected to cold, repression (H3K27me3, M state) only concerns a small (1...
A. DNA methylation

Historically, DNA methylation has been the first epigenetic mark to be recognized as a “stable, inheritable chemical modification that alters gene expression and does not modify the sequence” (see Sec. [1]). In fact, in early days of research on DNA methylation, it was found that methylation states are propagated through mitosis (Wigler et al. 1981).

DNA methylation is the substitution of a methyl (−CH₃) group to the carbon atom in position 5 at the cytosine base (5mC). Importantly, DNA methylation is coupled to metabolism through SAM (see Fig. [11b]).

The prevalence of DNA methylation in the genome changes significantly among different organisms: it is very high in vertebrates (where one refers to a “global” methylation), very low in Drosophila, and absent in the nematode worm C. Elegans. In somatic cells, cytosine methylation occurs predominantly at CpG dinucleotides, although it has been detected in any sequence context both in plants (Cokus et al. 2008) and humans (Lister et al. 2009), where 70–80% of CpG dinucleotides are methylated.

The patterns of DNA methylation in the genome are established in early development, and then faithfully propagated throughout successive cell divisions. Crucially, tissue-specific genes are kept unmethylated, whereas the others are heavily methylated. These processes are catalyzed by DNA methyltransferases (DNMTs). It is generally thought that the two methyltransferases DNMT3A and DNMT3B are responsible for establishing the methylation pattern during development (de novo methylation), and DNMT1 propagates the methylation pattern to daughter cells (maintenance methylation) (Bird 2002), which we illustrate in Fig. [12].

The precise mechanism behind the establishment of the initial methylation pattern during development is largely unknown. It has been proposed (Khraiwesh et al. 2010) that initiation of epigenetic silencing by DNA methylation depends on the ratio of the miRNAs and their target messenger RNAs, in a so-called “RNA-directed DNA methylation”.

5-methylcytosine can convert to thymine by spontaneous de-amination (see Sec. [1]), leading to a common DNA mutation. The hydrophobic methyl group in the DNA major groove gives a structural similarity between thymine and 5mC (see Fig. [11b]). It is important to notice that this allows for the possibility of a “base readout” in the major groove, as proposed by some authors (Machado et al. 2014). We shall discuss the implications of this later.

From the biological point of view, the role of DNA methylation is not clearly understood. Early studies on the role of DNA methylation highlighted its importance in gene silencing (McGhee and Ginder 1979; Razin and Cedar 1991), in X-chromosome inactivation (Graves 1982; Mohandas et al. 1981; Venolia et al. 1982) and gene imprinting (Li et al. 1993; Razin and Cedar 1994). It was later established that when CpG-island promoters are methylated, then the gene will be irreversibly silenced.

V. TOWARD A MORE COMPLEX SCENARIO: DNA METHYLATION, ROLE OF RNAs, SUPERCOILING IN EPIGENETICS

Up to now we have focused on histone PTMs and presented them as a crucial issue in the transmission of epigenetic information. However, the global picture is more complex. Among the additional epigenetic mechanisms, some are known since a long time, as DNA methylation (see Sec. [VA]), while others have been evidenced quite recently, as chromosome coating with (long) non-coding RNAs as in X inactivation (see Sec. [VC]), messenger RNA silencing by interaction with micro RNAs (see Sec. [VD]), or the coupling between epigenetics and supercoiling (see Sec. [VE]). An exhaustive description of the overall picture would represent a titanic task, well beyond the aim of this introductory review. Therefore we focus here on the main physical aspects of these biologically relevant mechanisms, drawing on a few concrete examples.
FIG. 11 Physical aspects of DNA methylation. (a) Conversion of cytosine to 5-methylcytosine occurs using SAM (S-adenosylmethionine) as a methyl group donor. Spontaneous deamination may convert 5mC to thymine, leaving the methyl group in the major groove. (b) Structural similarities between 5mC and thymine in the DNA major groove. (c) Specific patterning of H-bond donors, acceptors, methyl groups and non-polar hydrogens allows for “base readout” of the DNA sequence without strand opening (see main text). (d) and (e): experimental cyclization experiments on methylated and unmethylated DNA show distinct elastic properties of the two species (figures taken from [Pérez et al., 2012]). In (d), 2D electrophoresis shows the different migrations of linear (L) and circular (C) DNA species (either covalently closed (c) or nicked open (o)) for nonmethylated and methylated oligomers of 21 bp, respectively. (e) Ratio of circular and linear species as a function of fragment size.

With the advent of technologies that enable genome-wide screening of the methylation state of DNA, it has become clear that gene silencing is not the only role of DNA methylation [Jones, 2012; Lister et al., 2009], and its biological role is highly dependent on the sequence context in which it may be found. For example, DNA methylation has been associated with active gene bodies [Chodavarapu et al., 2010], quite the opposite of its established role of gene silencing. Detailed analyses of the differences in the methylation patterns in different cell types revealed an even richer phenomenology [Lister et al., 2009; Marchal and Miotto, 2015; Spruijt and Vermeulen, 2014], suggesting that the role of DNA methylation is not at all limited to repression of transcription.

DNA methylation has also been implicated in a variety of human diseases (see [Machado et al., 2014] for an extensive review), in particular in cancer, and has therefore received enormous attention. A challenging issue is
the relation of DNA methylation to cancer progression and prediction of pre-cancerous cell state (Feinberg and Tycko, 2004; Timp and Feinberg, 2013).

Can the change in physical properties of DNA upon cytosine methylation help understanding the variety of its roles? To address this question, we shall discuss several aspects of the physics involved in cytosine methylation. In the following, we will review the available knowledge on the following aspects: (a) the change in the elastic/mechanical properties of DNA upon cytosine methylation; (b) the role of the hydrophobic methyl group in determining DNA-protein interactions; (c) the relationship between DNA methylation and chromatin structure in vivo.

1. Mechanical properties of DNA change upon methylation

It has been recently established that the mechanical properties of DNA change when cytosine is methylated. The extent of this change is still unclear though.

A combination of all-atom molecular dynamics simulations and cyclization experiments revealed that a single cytosine methylation at a CpG dinucleotide step has a significant impact on the mechanical properties of DNA (Pérez et al., 2012). Cyclization experiments allowed to prove that oligomers stiffen significantly upon methylation (see Fig. 11a-e). Moreover, the value of the base pair roll was found to increase, whereas the twist and the width of the minor groove decreased. This should lead to a bending of the base pair towards the DNA major groove, and a stiffening of the sequence. However, when a poly-dinucleotide of type d(CpG)n was methylated, no significant difference was found compared to its unmethylated counterpart. This is due to the fact that the GpC dinucleotide has a compensating effect on the change of mechanical properties of the CpG step.

The stiffening of DNA upon methylation was also predicted by another theoretical study that employed van der Waals density functional theory (Yusufaly et al., 2013). There, it was shown that “Methylation of CG-rich stretches of DNA enhances the formation of the A-DNA polymorph, a helical form that is more resistant to bending deformations than B-DNA, and which also bends DNA in the opposite sense. Consequently, interactions with the histones are inhibited, and nucleosome formation is suppressed.”

In mixed sequence DNA, no significant effect of DNA methylation was observed by cyclization of 158–180 bp fragments (Hodges-Garcia and Hagerman, 1995). However, when combining detailed Monte Carlo simulations with cyclization experiments of an EcoRI restriction site, Nathan and Crothers (Nathan and Crothers, 2002) found that methylated sequences change the flexibility and the twist rate of DNA. The emerging picture is that cytosine methylation changes local structural parameters of base pair sequences, leaving however unperturbed the global elastic and mechanical properties of DNA.

Another consequence of cytosine methylation is altered resistance to strand separation (Severin et al., 2011). By combining single molecule force experiments with all-atom molecular dynamics simulations, the authors show that strand separation is strongly affected by cytosine methylation. It inhibits or facilitates strand separation, depending on the sequence context. Again, the sequence context plays an essential role in determining the direction and extent of the impact of DNA methylation.

2. Impact of cytosine methylation on DNA-protein interactions

The addition of a methyl group in the major groove of cytosine bases alters the hydrophobicity of DNA at the base pair step. In crystallographic studies, it was found that of a methylated A-form DNA oligomer (Mayer-Jung et al., 1998) is well hydrated, thereby allowing for the possibility of specific recognition of methylated DNA sequences, through the interaction with the tightly bound water at the methyl group.

Three classes of proteins that specifically bind to methylated DNA are known: MBD (methyl binding domain) proteins, SRA (SET and RING associated-) domain proteins, and zinc-finger proteins (Buck-Koehntop and Defossez, 2013). Some crystal structures of proteins that specifically recognize methylated DNA have recently become available (Buck-Koehntop et al., 2012; Liu et al., 2014). The detailed analysis of the binding modes revealed that although similarities between them exist (Liu et al., 2013), it is yet unclear why three distinct families of methylated DNA-binding proteins were needed in the course of evolution. The sequence flanking the methylated CpG step was shown to be important in determining the specificity of the interaction.

The structural differences between methylated and unmethylated DNA may in part explain the specificity of interactions between proteins and DNA. As shown in Fig. 11, the DNA base pair text determines a specific pattern of chemical groups (hydrophobic, hydrogen bond donor or acceptor) in the major groove, but not in the minor groove. As a consequence, one proposed mechanism of recognition of methylated states involves a “base read-
out” in the major groove, that does not require strand separation (Sasai et al., 2010; Zou et al., 2011).

Yet another consequence of DNA methylation is the change in base pair structural parameters such as twist, roll and minor groove width. Lazarovici et al. found that roll and minor groove width were excellent predictors of sequence specificity for DNase I endonuclease (Lazarovici et al., 2013). This mechanism of recognition of methylated sequences is termed “shape readout”. It is likely that both shape and base readout play a role in most cases (Machado et al., 2014).

3. Relationship between nucleosome positioning and DNA methylation

As discussed earlier, it appears that CpG methylation locally stiffens the DNA, and it was speculated that as a consequence, nucleosome positioning would be disfavored. However, contradictory results exist, which also suggest that methylation actually favors nucleosome formation, or stabilizes DNA wrapped around a nucleosome. We discuss here the available data.

Several lines of evidence suggest that methylation affects the structure of DNA wrapped around a histone core. Using FRET, Lee et al. showed that methylation leads to tightening of DNA around the nucleosome (Lee and Lee, 2011). Along the same lines, Choy et al. suggest that the enhanced rigidity of methylated DNA leads to more compact and closed nucleosomes. Methylated DNA was also shown to be more difficult to remove from a nucleosome (Kaur et al., 2012). By combining data from genome-wide nucleosome positioning with available DNA methylation maps, Chodavarapu et al. (Chodavarapu et al., 2010) have shown that there is a small increase (about 1.2%, from 75 to 76.2%) in preference of nucleosome positioning for methylated DNA. This data has been contested by Felle et al. (Felle et al., 2011), that instead showed that the nucleosome occupancy of methylated DNA was 2-fold lower compared to unmethylated sequences.

Other studies showed that the increased rigidity of methylated DNA disfavors nucleosome positioning. We already discussed the work of Yusufaly et al. (Yusufaly et al., 2013), but many other studies reached the same conclusions (Felle et al., 2011; Pérez et al., 2012; Portella et al., 2013).

Even more recent studies have shown that methylation has a negligible influence on nucleosome stability (Langecker et al., 2015).

4. Remarks and perspectives

It has been discovered that 5-methylcytosine is not the only cytosine variant. In 2009, 5-hydroxymethylcytosine was discovered in mouse brains (Kriaucionis and Heintz, 2009), and since then, two other forms of cytosine methylation were discovered: 5-formylcytosine (5fC), and 5-carboxymethylcytosine (5caC) (see Spruijt and Vermeulen, 2014 for a review). Each of these other forms has been detected in mouse embryonic stem cells in significant amounts (Song et al., 2013), so that a careful investigation of the differences between their physical properties is needed.

Traditional sequencing methods for detection of cytosine methylation are not able to distinguish between 5mC, 5hmC, 5fC and 5caC. Accurate methods to identify the genome-wide map of cytosine methylation states has only very recently become available (Booth et al., 2014). Much study is needed to understand the roles of these epigenetic marks.

We speculate that distinctive physical properties of these alternative forms are in part responsible for determining the variety of roles attributed to DNA methylation (Schüebele, 2015; Spruijt and Vermeulen, 2014). We suggest that physics may help advancing our understanding of the biological aspects of this important epigenetic mark.

Finally, there is already a vast literature on the coupling between DNA methylation and histone post-translational modifications (see Sec. III.B), which was reviewed in Cedar and Bergman (2009). However, to our knowledge there has been not yet an attempt to model this coupling from a physical point of view.

B. Parental imprinting

In animals, it has been observed that certain genes are expressed in a parent-of-origin-specific manner. These so-called “imprinted” genes are DNA methylated on specific sequences named Imprinting-Control Regions (ICRs). The ICR methylation of any imprinted gene occurs either on the paternally herited or on the maternally herited chromosome this gene belongs to. Remember that there is one copy of each gene on the paternal and on the maternal chromosome (except for sexual chromosomes). Both copies are generally different and are called paternal and maternal alleles. ICR methylation starts in germ cells, keeps the memory of the parental origin of the allele and drives monoallelic expression. For example, this mechanism happens on the Igf2/H19 gene locus of mouse chromosome 7 (see Ref. Lesne et al., 2013). The ICR located upstream of the H19 gene is methylated on the paternal allele but it remains unmethylated on the maternal allele. The maternal unmethylated allele is bound by the CTCF protein that prevents the interaction between regulatory sequences (enhancers) located downstream of H19 and the Igf2 gene located further upstream of the ICR. Therefore, the Igf2 gene is not activated on the maternal chromosome 7. Instead, on the
paternal chromosome, the DNA methylation prevents the binding of the CTCF protein and the Igf2 gene can be activated by the regulatory sequences. This differential folding has been evidenced by Chromosome Conformation Capture (3C) experiments during mouse development (Court et al. 2011).

Genomic imprinting was selected during evolution at the transition between placental (e.g. mouse) and marsupial (e.g. kangaroo) mammals. Six genes in marsupials and about a hundred in placental mammals undergo parental genomic imprinting. The reason for the selection of this unusual, epigenetic mechanism of gene regulation during mammalian evolution remains poorly understood at present.

C. Chromosome X inactivation

Another important and historically relevant example of epigenetic silencing is the inactivation of the X chromosome in mammal females.

The pair of sex chromosomes (XY in males, XX in females) is responsible for sex determination in mammals. While the Y is small and carries only a few genes, X chromosome is much longer and contains thousands of genes. Females thus carry twice as many X-linked genes as males, this leading to a potentially lethal dosage problem.

During early embryonic development, one of the two X chromosomes of females is thus inactivated, and condensed to form the so-called Barr (Barr and Bertram, 1949).

In mice, the inactivation of the X chromosome occurs in three phases. First, the paternal X chromosome is inactivated during the preimplantation period, from the stage “two cells”. Then, it remains inactive in the peripheral cells, which will form the placenta, but it is reactivated in the cells which will form the future embryo. Finally, a second inactivation takes place, and this time it concerns, randomly in each cell, either paternal or maternal X (Okamoto et al. 2004).

Intriguing questions then arise: how does the cell manage to silence only one of the two X chromosomes? How is the silenced X chosen?

The X inactivation mechanism seems to be controlled by a complex genetic locus called the X-chromosome-inactivation centre (Xic). It has been proven that starting mechanism in X inactivation is mediated by the non-coding (not translated into protein) transcript of the Xist (X-inactive specific) gene, present within Xic. Once transcribed, many copies of Xist RNA accumulate along the X chromosome (RNA coating), then induce its heterochromatinization.

This process is however under the control of a few other genes included in the Xic region. One of the crucial elements is Tsix, a non-coding RNA gene that is antisense to the Xist gene (it is transcribed from the complementary DNA strand). Due to this complementarity, the Tsix RNA-transcript duplexes with the complementary Xist RNA-transcript into a double stranded RNA which is further degraded. This mechanism prevents the accumulation of Xist, hence inhibits the X inactivation (Okamoto et al. 2004).

Random selection of the inactivated X chromosome may therefore emerge from a detailed balance in the synthesis of Tsix and Xist. Recently, a model to explain this complex regulation path has been proposed (Giorgetti et al. 2014), that relies on the polymer physics properties of chromatin, on its organization in topologically-associating domains (TADs, see Sec. III.C.2), and on a detailed coupling between gene expression and 3D organization at the level of the Xic center.

The Xic is composed of two topologically-associating domains (TADs, Nora et al. 2012), called the Tsix TAD (320kb) and the Xist TAD (600kb). The 3D structure of the Tsix TAD is highly variable among cells and this variability is most probably due to fluctuations of chromatin conformation at time scales shorter than a cell cycle, as in the model proposed by Jost et al. previously discussed (Jost et al. 2014). The distribution of conformations observed thanks to 5C experiments (Carbon Copy Chromosome Conformational Capture), illustrated by the Fig. 5A in (Giorgetti et al. 2014), shows indeed an equilibrium between coil and globule conformations, typical of a coil-globule transition of a polymer with finite-size effects (Caré et al. 2014; Imbert et al. 1997).

The model also assumes that the level of Tsix depends on the activity of two putative regulatory elements (Linx and Chic1, Nora et al. 2012), placed inside the same Tsix TAD.

Inactivation can then be explained as a result of the chromatin conformation of the Tsix TAD, which determines the regulation of specific interactions between all these elements (regulatory elements, Tsix and Xist). The switch between globule and coil conformations changes indeed the spatial proximity between these genes, hence their interactions. As a result, globule conformations induce higher Tsix transcription levels, while coil conformations correspond to lower Tsix levels.

Statistical fluctuations in chromatin conformation within the Tsix TAD may, then, contribute to ensuring asymmetric expression from the Xic at the onset of X chromosome inactivation, as shown by simulations (Giorgetti et al. 2014). If, in a cell, Tsix TAD is similarly compacted on the two alleles, Tsix transcript levels from the two alleles are similar, with little or no heterochromatinization effect. As a fluctuation induces the coil-globule transition for the Tsix TAD on one allele, then the two transcripts tend to be differentially expressed. This mechanism may help ensuring that Xist is only transcribed from the allele with lower Tsix transcription (Fig. 6B in Giorgetti et al. 2014).
Once established, the X inactivation is stably transmitted through mitosis along the following development. Further (and later) features of the inactive X include hypermethylation of DNA, histone deacetylation, chromatin condensation, i.e. the same general mechanisms that we have previously introduced.

D. Non-coding RNA and microRNA

The X inactivation is a clear example of the crucial and early role of RNA in epigenetic silencing. Together with histone and DNA modifications, non coding RNAs have (more recently) emerged as one of the main epigenetic mechanisms. Two main classes of epigenetically active non coding RNAs can be identified: small (<30 nucleotides) and long (>200 nucleotides). Both classes play a role in heterochromatin formation, histone modification, DNA methylation targeting, and gene silencing.

Long RNAs can complex with chromatin-modifying proteins and recruit their catalytic activity to specific sites in the genome, thereby modifying chromatin states and influencing gene expression. In the case of the X inactivation, the first described epigenetic mechanism involving a long non coding RNA, it has been shown that Xist RNA directly recruits chromatin-modifying factors as Polycomb repressive complex PRC2 that mediates histone H3 lysine 27 methylation, but the direct character of such interaction remains to be confirmed (Brockdorff, 2013), as well as the overall mechanism, including the interplay between the different RNA involves (Xist Tsix and others) (Pontier and Gribman, 2011). According to a proposed model (Lee, 2012), long non coding RNAs may function by sequestering chromatin-modifying enzymes away from other interacting partners or by guiding chromatin modifiers to the correct locations in the genome. In other cases, the long RNA seems to works by binding and bringing together different types of proteins that can cooperate in establishing the repressive chromatin state (see Marchese and Huarte, 2014 for a review on the interplay between long non coding RNAs and chromatin modifiers in epigenetics).

A whole new realm of small non-coding RNAs was discovered in the late 90s. It includes two classes of small RNAs: micro RNAs (miRNAs) and small interfering RNAs (siRNAs), which perform many functions, and in particular are involved in the so called “RNA interference”, a regulation pathway of gene expression at the transcriptional and post-transcriptional level. In other words, RNA interference may act either by inhibition of the target RNA transcription or by degradation of the transcript RNA. Piwi-interacting RNAs (piRNAs) represent a third, large class of lightly longer RNAs, palying a role in epigenetic and post-transcriptional gene silencing of retrotransposons and other genetic elements in germ line cells through their association with piwi proteins.

At the post-translational level, the essential repression mechanism for both miRNAs and siRNAs is through pairing to a complementary sequence of the messenger RNA transcribed from a target gene, which results in the degradation of the RNA and thus the repression of the gene. To recognize its target messenger RNA, miRNAs and siRNAs must be associated with a protein to form the RNA-induced silencing complex.

Interestingly, heterochromatinization may also be initiated by the RNA interference machinery (see (Marchese and Huarte, 2014) for a review on the interaction between the different RNA sequences of up to several million base pairs and consist of a large number of repetitions of a much smaller sequence, and are found, in particular, at centromeres, telomeres, and other regions that remain condensed throughout the cell cycle, referred to as constitutive heterochromatin. At the same time, heterochromatin mediates the spreading of RNA interference machinery to surrounding sequences, hence to produce siRNAs, which in turn are essential for the stable maintenance of heterochromatic structures.

A complex scenario thus emerges in which DNA sequences, RNAs, epigenetic factors and chromatin remodelers plays together in the setting up and maintenance of different functional chromatin states. DNA-methylation and histone-modifications often act together to regulate miRNA expression, while, conversely, some miRNAs can regulate the expression of epigenetic machinery, with important dysregulation effects in cancer (Wang et al., 2013).

The models previously described for the spreading of epigenetic marks needs probably to include these additional features in order to reproduce the epigenetic mechanisms to a larger scale. A first step in this direction may be the study of the interplay between miRNAs and epigenetic regulators, and the particular role of post-translational regulation, as discussed by Osella et al. (Osella et al., 2014). A typical basic regulatory network involving miRNAs and epigenetic regulators is the double negative feed-back loop, in which a miRNA represses an epigenetic regulator, which in turn represses the expression of the same miRNA. Starting from an approach similar to what discussed in Sec. IV is it possible to describe the system. More precisely, the model describes separately the number of miRNAs, of messenger RNAs that miRNAs repress, and of proteins (epigenetic regulator) resulting from the messenger RNA translation (and repressing, in turn, the mi-RNAs). The interesting input from the miRNA regulation step is its role in keeping fluctuations of gene expression under control, either

2 The RNA interference machinery includes different factors (as Dicer, Argonaute and RNA-dependent RNA polymerases) to produce the small RNAs or bind them to form functional complexes.
by suppressing translation and by promoting RNA degradation. Both effects helps indeed in reducing the burstiness in protein production [Friedman et al., 2006; Osella et al., 2014]. The resulting set of rate equations includes therefore, on one hand, the highly non-linear and bistable character of epigenetic regulation and, on the other, the stabilizing effect of miRNAs regulation, this leading to an increased stability of the system, and to an increased range of bistability of the switch.

This result suggest possible reasons for the existence of regulatory pathways combining epigenetic regulators and miRNAs, although both experimental investigations and modeling of such complicated circuits are still at the embryonic stage.

E. Supercoilingomics: supercoiling as a physical epigenetic mark, and its role in the initiation and maintenance of epigenetic marks

DNA supercoiling was first properly described by Jerome Vinograd and colleagues (Vinograd et al., 1965) and it took just some more years to James Wang to discover the first enzyme able to relax these topological constraints in vivo (Wang, 1971). DNA topological state is given by its linking number $(Lk)$, defined as the number of times that a strand of DNA winds in the right-handed direction around the helix axis when the axis is straight (or constrained to lie in a plane for a circular molecule). This integer is the sum of two geometrical parameters: twist (or twisting number $Tw$, a measure of the helical winding of the DNA strands around each other, hence representing a “1D” deformation along the axis) and writhe (or writhing number $Wr$, a measure of the 3D coiling of the axis of the double helix). The partitioning between $Tw$ and $Wr$ for a given $Lk$ is determined by the free energy of DNA (itself dependent on ionic conditions) and by DNA/protein interactions that locally impose some particular DNA torsion and/or writhe. While structural proteins can only alter the $Tw/Wr$ ratio, enzymes such as topoisomerases or gyrases can alter $Lk$ by catalyzing the cleavage of one or both strands of DNA, followed by the passage of a segment of DNA through this break and the resealing of the DNA break [Wang, 2002].

In most living organisms, DNA is negatively supercoiled, which prepares DNA for processes requiring separation of the DNA strands, such as replication or transcription. In eukaryotes, this negative supercoiling is constrained within the nucleosome, so that its removal will simultaneously favor the access and melting of previously occulted DNA, therefore facilitating transcription initiation. The distribution of nucleosomes, and notably the NRL, appears then as an important feature to propagate through mitosis, partly for topological reasons. Regarding the elongation step, as DNA is screwing through the polymerase during transcription, the negative supercoiling induced in the back of the enzyme can propagate through the chromatin fiber and trigger local DNA alterations that have been proposed to serve as a regulatory signal for molecular partners (Belotserkovskii et al., 2013; Konzine et al., 2004, 2008; Liu et al., 2006). Therefore supercoiling would act as a transient mechanotransducer as well as a physical epigenetic mark. Moreover, nucleosome conformational changes might help to smoothen the elongation process by buffering some topological constraint (Bancaud et al., 2007; Recouvreux et al., 2011; Vlijm et al., 2015) and facilitate H2A/H2B dimer loss in front of the polymerase (Sleim et al., 2013). It remains to be seen how much the structural differences provided by histone variants (Shaytan et al., 2015) would help to build “elongation friendly” regions that could be transmitted through cell division.

Supercoiling of DNA has been recently proposed to be considered as a true physical epigenetic mark, entering the family of “omics” data one should consider to get a comprehensive genome-wide epigenetic landscape of a cell at a given state of its development (Lavelle, 2014). Indeed, genome-wide maps of DNA supercoiling states have been generated (Bermudez et al., 2010; Joshi et al., 2010; Konzine et al., 2013; Naughton et al., 2013; Teves and Henikoff, 2014) which add to existing predicted maps of DNA melting (Liu et al., 2007) or G-quartet motifs (Du et al., 2009; Maizels and Gray, 2013). The emerging picture is that supercoiling is associated to the structuration of chromatin topological domains, which largely overlap with TADs (Naughton et al., 2013).

DNA supercoiling is a physical epigenetic mark because it may change the affinity of the underlying DNA sequence to specific transcription factors (Travers and Muskhelishvili, 2007). Supercoiling may also silence a whole topological domain when recruiting TFs which in turn may recruit silencing enzymes, e.g. Suv39h which eventually deposits H3K9me3 epigenetic marks (Bulut-Karslioglu et al., 2012). Note that in this case, transcription factors are used to repressing instead of activating gene expression. This mechanism relies on DNA allostery, i.e. the change of DNA affinity to some transcription factors that is induced by supercoiling (Lesne et al., 2015).

But how may supercoiling be initiated and herited? The distribution of topoisomerases and structural proteins such as condensins should help in transmitting some domain structuration and topological states through cell division (Aragon et al., 2013; Hirano, 2014). The nucleosome repeat length (NRL, see Sec. I.A) might also be a key control parameter. We first note that the twist rate of the DNA double helix in a given topological domain is a function of (i) remodeling activity, notably through active nucleosome removal (Fadinhaterei and Marko, 2011), and (ii) topoisomerase activity. Importantly both these ATP-consuming mechanisms are
under active control of the cell metabolism. Moreover both remodeling and topoisomerase activities regulate the value of the average NRL of a given topological domain. Therefore the average NRL over some genomic domain appears to be a physical epigenetic mark of this domain. And the transmission of this average NRL through mitosis would transmit the twist rate of the domain. Interestingly the recently observed spreading mode of histone PTMs over transcription cycles (Terweij and van Leeuwen 2013) might explain the spreading and maintenance of the NRL on epigenetic domains. In support of this hypothesis active remodeling processes achieved by ATP-consuming remodeling factors - and crucially through active nucleosome removal - have been shown to be essential for driving biologically relevant nucleosome positioning (Padinhateeri and Marko 2011), thus fixing the average NRL. Challenging genome-wide studies are needed to further correlate supercoiling maps to cell differentiation states.

VI. CONCLUSION AND PERSPECTIVES

By putting the rich and diverse biological literature under the new light of a physical approach, the emerging picture is that a limited set of general physical rules play a key role in the epigenetic regulation of gene expression. Processes at work are diffusion-limited and involve a small number of molecules, which precludes simple approaches in terms of concentrations. Instead, multi-scale approaches articulating different models at different levels of organization are to be developed.

Mainly, epigenetics display an intricacy of physical mechanisms and specific biological entities, devised in the course of evolution to achieve an exquisitely coordinated and adaptable regulation of transcriptional activity. Our review demonstrates the need to take into account both aspects, within a dialogue between physics and biology, theory and living-cell experiments.

Significant and challenging issues remain:

(i) coupling nuclear architecture and epigenetic marking: understanding the interplay between 3D nuclear architecture and 1D epigenetic marking (including barrier positions);

(ii) physical epigenetors: coupling nucleus mechanical deformations to epigenetic marking. Cells dramatically change their shape and mechanics during development by integrating physicochemical signals from the local microenvironment (morphogens gradients, cell-cell contact, adhesion to extracellular matrix) to generate lineage-specific gene expression (Engler et al. 2006). Recent studies have begun to uncover the mechanisms by which these signals are integrated into the 3D spatiotemporal organization and epigenetic state of the nucleus and impact cell fate decision (Shivashankar 2011, Bellas and Chen 2014, Ramdas and Shivashankar 2014). Further understanding of these transduction mechanisms is a challenging perspective;

(iii) equilibrium vs nonequilibrium physics: An implicit - yet overlooked - assumption of the models of bistability introduced by Sneppen and coworkers in the context of epigenetics (Dodd et al. 2007, Micheelsen et al. 2010) is that the system is open and far from equilibrium. Indeed epigenetic marks undergo permanent recycling and biochemical transformations, so that epigenetic marks turn out to be steady states and not equilibrium states. Therefore, it is of primary importance to identify, characterize and model the various active physical mechanisms that are at work in the initiation and maintenance of epigenetic marks. In particular, it is crucial to evidence active (ATP-dependent) mechanisms that maintain epigenetic marks, for instance: metabolism, transcription, replication, ionic pumps at the cell membrane. In our opinion this is a very challenging and timely topic for biology-oriented physicists.

ACKNOWLEDGMENTS

The authors wish to thank all the members of the CNRS GDR 3536 for the stimulating discussions that inspired this work. Special thanks to Aurélien Bancard, Philippe Bertrand, Pascal Carrivain, Giacomo Cavalli, Thierry Forné, Daniel Jost, and Cedric Vaillant for their help in the preparation of the manuscript. This work has been funded by the French Institut National du Cancer, grant INCa 5960 and by the French Agence Nationale de la Recherche, grant ANR-13-BSV5-0010-03.

REFERENCES


Aragon, L., E. Martínez-Perez, and M. Merkenschlager (2013), Current opinion in genetics & development 23 (2), 204.


http://www.pnas.org/content/103/44/16236.full.pdf+html.
Graves, J. A. M. (1982), Experimental cell research 141 (1), 99.
Graves, J. A. M. (1982), Experimental cell research 141 (1), 99.
Joshi, R. S., B. Piña, and J. Rocca (2010), The EMBO journal 29 (4), 740.
Kan, P.-Y., T. L. Caterino, and J. J. Hayes (2009), Molecular and Cellular Biology 29 (2), 538.
Kriaucionis, S., and N. Heintz (2009), Science 324 (5929), 929.
Lee, J. T. (2012), Science 335 (6133), 1435.
Li, S., and M. A. Shogren-Knaak (2008), Proceedings of the National Academy of Sciences of the United States of America 105 (47), 18243.
Tjong, H., K. Gong, L. Chen, and F. Alber (2012), Genome research 22 (7), 1295.
Yang, Y., A. P. Lyubartsev, N. Korolev, and L. Nordenskiöld (2009), Biophysical Journal 96 (6), 2082.
Yang, Z., and J. J. Hayes (2011), Biochemistry 50 (46), 9973.