The physics of epigenetics

Ruggero Cortini, Maria Barbi, and Bertrand R. Caré

Sorbonne Universités, UPMC Université Paris 06, UMR 7600 LPTMC F-75005 Paris, France, CNRS, UMR 7600, LPTMC, F-75005 Paris, France, and Nuclear Architecture and Dynamics, CNRS GDR 3536, UPMC Université Paris 6, 75005 Paris, France

Christophe Lavelle

Nuclear Architecture and Dynamics, CNRS GDR 3536, UPMC Université Paris 6, 75005 Paris, France, and Genomes Structure and Instability, Sorbonne Universités, National Museum of Natural History, Inserm U 1154, CNRS UMR 7196, 75005 Paris, France

Annick Lesne

Sorbonne Universités, UPMC Université Paris 06, UMR 7600 LPTMC F-75005 Paris, France, CNRS, UMR 7600, LPTMC, F-75005 Paris, France, and Nuclear Architecture and Dynamics, CNRS GDR 3536, UPMC Université Paris 6, 75005 Paris, France, and Institut de Génétique Moléculaire de Montpellier, CNRS UMR 5535, Montpellier, France

Julien Mozziconacci

Sorbonne Universités, UPMC Université Paris 06, UMR 7600 LPTMC F-75005 Paris, France, CNRS, UMR 7600, LPTMC, F-75005 Paris, France, and Nuclear Architecture and Dynamics, CNRS GDR 3536, UPMC Université Paris 6, 75005 Paris, France

Jean-Marc Victor*

Sorbonne Universités, UPMC Université Paris 06, UMR 7600 LPTMC F-75005 Paris, France, CNRS, UMR 7600, LPTMC, F-75005 Paris, France, and Nuclear Architecture and Dynamics, CNRS GDR 3536, UPMC Université Paris 6, 75005 Paris, France, and Institut de Génétique Moléculaire de Montpellier, CNRS UMR 5535, Montpellier, France

(published 26 April 2016)

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 multiscale physical mechanisms that govern the biological processes behind the initiation, spreading, and inheritance of epigenetic states. These include not only the changes in the molecular properties associated with the chemical modifications of DNA and histone proteins, such as methylation and acetylation, but also less conventional changes, typically in 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 dynamical systems, and the electrostatic and mechanical properties related to chemical modifications of DNA and histones. By putting the complex biological literature in 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 one to rationalize the normal cellular functions, but also helps to understand the emergence of pathological states, in which the epigenetic landscape becomes dysfunctional.

DOI: 10.1103/RevModPhys.88.025002

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

*Corresponding author.
Current address: LPTMC case courrier 121 Université Pierre et Marie Curie, 4 place Jussieu, 75252 Paris cedex 05, France.
victor@lptmc.jussieu.fr

 © 2016 American Physical Society

C. Chromosomes as polymers 6
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 8
3. Histone tail acetylation: Direct effects on chromatin accessibility 8
a. Experiments 8
b. Models 9
4. H4K16 deacetylation is a silencing mark in budding yeast 10
5. Histone tail methylation: Indirect effects on chromatin condensation 10
a. HP1-mediated heterochromatin 11
b. Polycomb-mediated heterochromatin 11
C. How epigenetic marks organize the chromosomes in the cell nucleus: General rules and physical modeling of epigenome wide studies 11
1. Epigenome wide studies 11
2. The physics of TADs: Finite-size effects in the coil-globule transition of copolymers 13
IV. Physical Mechanisms Involved in the Initiation, Spreading, Maintenance, and Heritability of Epigenetic Marks 14
A. Mathematical modeling 14
B. Zero-dimensional models 14
C. Higher-dimensional models 16
1. One-dimensional models 16
2. Three-dimensional models 18
D. Biological relevance of the models 18
1. Waddington's epigenetic landscape revisited 18
2. Hysteresis 19
E. Example: Plant vernalization 19
V. Toward a More Complex Scenario: DNA Methylation, Role of RNAs, and Supercoiling in Epigenetics 20
A. DNA methylation 20
1. Mechanical properties of DNA change upon methylation 21
2. Impact of cytosine methylation on DNA-protein interactions 22
3. Relationship between nucleosome positioning and DNA methylation 22
4. Remarks and perspectives 22
B. Parental imprinting 23
C. Chromosome X inactivation 23
D. Noncoding RNA and microRNA 24
E. Supercoilingomics: Supercoiling as a physical epigenetic mark and its role in the initiation and maintenance of epigenetic marks 25
VI. Conclusion and Perspectives 26
Acknowledgments 26
References 26

I. INTRODUCTION

A. An intricate history

The word "epigenetic" was 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 preexisting concept of epigenesis, namely, the fact that the organism is not fully achieved in the initial cell but experiences complex developmental processes (Gilbert, 2012). 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 "paragenetic." Epigenetic systems, as opposed to the genetic system, were conceived as "signal interpreting devices" (Nanney, 1958), i.e., mediators between signals—environmental or physiological cues—and the genomic response, mainly at the level of transcriptional regulation.

Because of this dual origin of the word epigenetic, the associated concepts 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; and (2) transgenerational inheritance, that is, transmission through meiosis of nongenomic information.

Because of 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 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 is the definition we use in this review. To be even more specific, we note the following:

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*.

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 Sec. II the physical template of epigenetic marking, namely, chromatin.

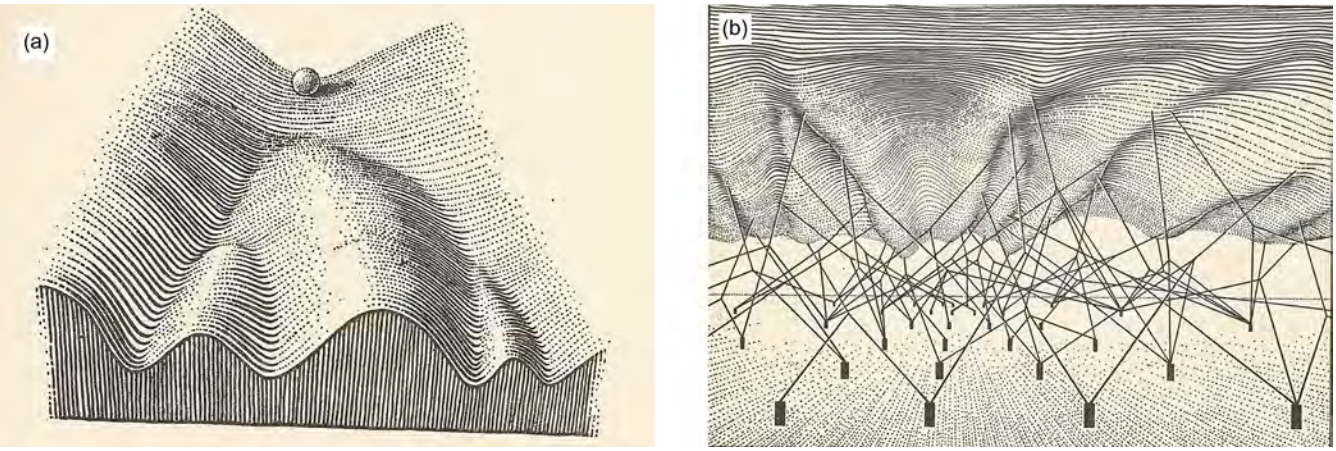

FIG. 1. (Left) The epigenetic landscape described by Waddington (1957) represents the process by which the cell (represented by the ball) faces different possible paths during development (i.e., choose one of the permitted trajectories), leading to different cell fates. (Right) The landscape is dynamically determined by hidden wires that symbolize genes expression and interactions; see Secs. IV.D and VI for a physical reanalysis of this metaphorical picture. Adapted from Waddington, 1957.

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

Because of the fundamentally different logic of transcriptional regulation in prokaryotes and eukaryotes (Struhl, 1999), we put 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 to tackle 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 amino acid 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 small 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's 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 an overview of the molecular structure of chromatin, and we introduce the concept of epigenetic marks. In Sec. II.B we 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: histone 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

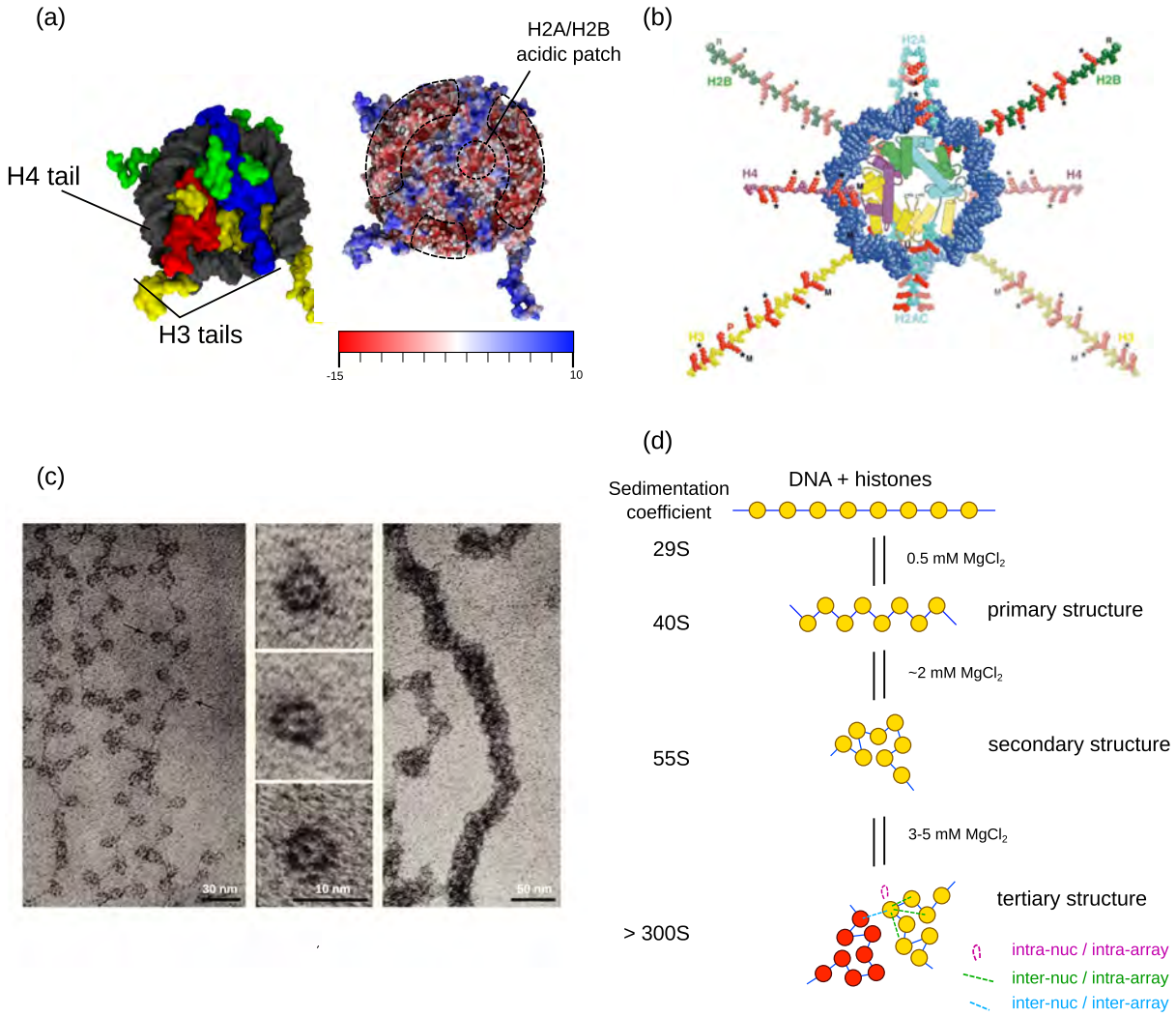

FIG. 2. Detailed structure of the nucleosome core particle (NCP) and chromatin fiber. (a) Left: internal tertiary structure of the NCP147 (PDB: 1KX5). Right: electrostratic potential at the NCP surface [computed using the PDB2PQR/APBS plugin (Baker *et al.*, 2001; Dolinsky *et al.*, 2007) of the Visual Molecular Dynamics (VMD) software package (Humphrey, Dalke, and Schulten, 1996), ionic strength 0.15 M monovalent salt]. (b) The NCP showing histone tails and the globular core. Lysine and arginines residues are marked by asterisks. From Wolffe and Hayes, 1999. (c) Electron microscopy images of nucleosome arrays with low (left) or high (far right) ionic forces, with details of nucleosome (middle). From (Olins and Olins, 2003. (d) Reconstituted chromatin sedimentation assay principles. Adapted from Pepenella, Murphy, and Hayes, 2014.

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 (Luger *et al.*, 1997; Davey *et al.*, 2002). In most species, this assembly, referred to as the nucleosome core particle (NCP) [see Figs. 2(a) and 2(b)], may also integrate a copy of H1 (linker histone) at the DNA entry-exit site, 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 from 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 Figs. 2(c) and 2(d)], 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, Mangenot *et al.*, 2007; Yang and Hayes, 2011). Additionally, the charge distribution in the NCP is not spatially homogeneous [see Fig. 2(a)].

Epigenetic marks are chemical covalent modifications of either DNA (namely, DNA methylation, see Sec. V.A) 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. V.A). Histone PTMs are inherited in a fundamentally different way, which is the principal subject of Sec. IV.

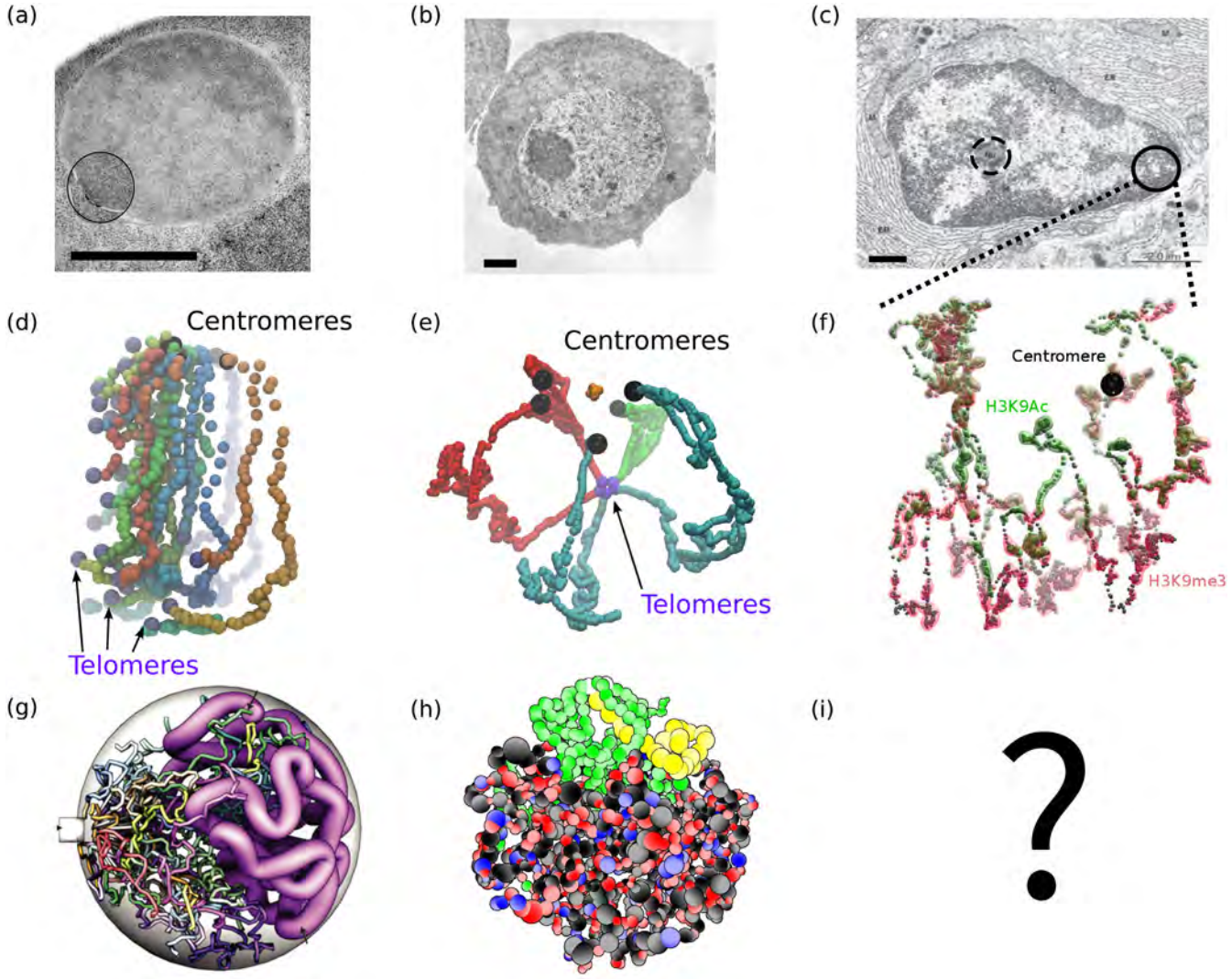

FIG. 3. Nuclear organization in yeast (*S. Cerevisiae*), drosophila, and mammals. (a)–(c) Electron microscopy images of nuclei in (a) yeast, (b) drosophila, and (c) human. Scale bars correspond, respectively, to 1, 2, and 2 μm. (a) The nucleolus is the darker region in the upper part of the nucleus. The spindle pole body (SPB) is shown by a circle. (b) The nucleolus is the dark circular region and heterochromatin can be seen as darker spots. (c) The nucleolus is marked by a dashed circle. (a) http://scienceblogs.com/transcript/2006/08/16/the-centrosome-and-the-spindle/, (b) http://pixgood.com/nuclear-pore-em.html, and (c) 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 (Duan *et al.*, 2010; Dixon *et al.*, 2012; Sexton *et al.*, 2012). (d), (e) All the chromosomes are represented with different colors. (f) Only chromosome 1 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, 40, and 40 kb. (g), (h) Polymer models of the genome in yeast and drosophila: (g) Each chromosome is labeled with a different color. From (Wong *et al.*, 2012. (h) Colors correspond to the colors of chromatin (see Sec. III.C). From Giacomo Cavalli and Pascal Carrivain (IGH, Montpellier). (i) To our knowledge, physical models of the complete human genome have not been developed so far.

B. Large-scale picture of chromatin

Eukaryotic chromosomes are giant polymers, each formed by a large 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, Dechassa, and Tremethick, 2012; Pepenella, Murphy, and Hayes, 2014); see Fig. 2(d).

When observed by electron microscopy [see Figs. 3(a)–3(c)], interphase chromatin appears to fill the entire nucleus volume. As the genome length may vary considerably from organism to organism, the nucleus size varies accordingly: orders of magnitude go from ∼10 Mbp (mega base pairs) for a diameter of the order of 2 μm in yeast [Fig. 3(a)], to ∼100 Mbp and 4 μm in drosophila fly [Fig. 3(b)], and up to ∼1000 Mbp and 10 μm in mammals [Fig. 3(c)]. These differences in size are certainly correlated with the differences

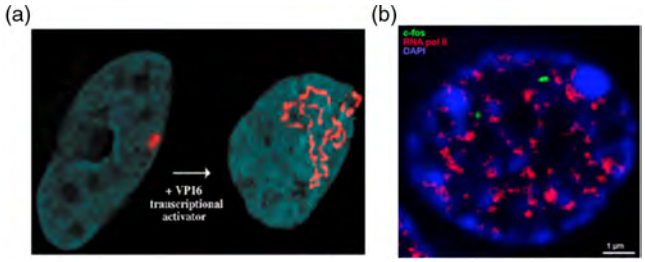

FIG. 4. Fluorescence microscopy of nuclei. (a) A specfic region of the genome is decompacted after the induction of transcription. Adapted from Tumbar, Sudlow, and Belmont, 1999. (b) A human cell nucleus. DNA is labeled in blue, PolII in red. From Crepaldi *et al.*, 2013.

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 Figs. 3(a), 3(d), and 3(g)].

When the nuclei of multicellular organisms are considered [Figs. 3(b) and 3(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 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. 4(a)].

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. 4(b)] (Jackson *et al.*, 1993). These factories are located within the euchromatin domain and each factory has a propensity to gather coregulated 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) enables 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 in Figs. 3(d)–3(f). The results confirmed the tethering of centromeres in yeast and drosophila, but also of telomeres. In humans the reconstruction of chromosome 1 gives a visual illustration of decondensed euchromatin loops emanating from globular heterochromatin globules. These are decorated with two different specific histone marks. We return to 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 whether polymer physics is the main player that drives chromosome organization.

(i) In the simplest case of yeast, where chromosomes are shorter and all anchored at the SPB by their centromeres, it seems to be 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. 3(g).[1] 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 Figs. 3(d) and 3(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 complexify the picture [see Figs. 3(e) and 3(h)]. Roughly speaking, it was 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 *et al.* (2010) identified five principal chromatin states, called chromatin "colors" from the analysis of 53 chromatin protein genome-binding profiles in drosophila cells. Among these states, some essentially correspond to active, transcribing euchromatin, and others to dense, repressed heterochromatin. These chromatin states result from the recruitment of DNA-binding proteins that are specific to the underlying epigenetic marks (see Sec. III.B). Note, however, that the colors are only an arbitrary choice of the authors and do not correspond to actual coloring of the chromatin, due, e.g., to staining.

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

[1]Note that in yeast the whole genome is actively transcribing most of the time, with the only exception of the regions that govern the cell sexual behavior, called "hidden mating-type loci," and of the chromosome extremities, called telomeres, which protect the ends of the chromosome from damaging or from fusion with other chromosomes. Therefore heterochromatin is restricted to telomeres and mating-type loci in this case.

(iii) In mammals heterochromatin is mainly located at the nuclear membrane and euchromatin at the center of the nucleus [see Fig. 3(c)]. 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. 3(f)].

In summary, the conformation adopted by chromatin is affected by its intrinsic structural parameters such as the NRL [see Boulé, Mozziconacci, and Lavelle (2015) for a review] on top of which lies an additional layer of modulation by internucleosomal electrostatic interactions (Hansen, 2002; Pepenella, Murphy, and Hayes, 2014) 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. Section III is 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 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 result in 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 (Kouzarides, 2007; Cantone and Fisher, 2013), 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, and whole nucleus (Zhou *et al.*, 2007; Poirier *et al.*, 2009).

The potentially large number of combinations of epigenetic marks has led to the hypothesis that transcription factors (TFs) might be targeted to nucleosomes endowed with specific combinations of histone-tail post-translational modifications. This was coined as the "histone code" hypothesis at the beginning of the millenium and remained popular in biology for the better part of a decade. While the genetic code encodes the sequence of a protein using the four bases of the DNA, this histone code would encode the regulatory events involved in triggering its expression, and, in particular, the differential expression observed in cells of different types (Jenuwein and Allis, 2001). However, it was then realized that only a few combinations of histone marks are actually observed (Lennartsson and Ekwall, 2009; Rando, 2012). These relevant combinations can even be reduced to five classes in the case of drosophila, the so-called "five colors of chromatin" as identified by Filion *et al.* (2010). Moreover, it was also realized that the histone code is actually not a genuine code insofar as the correspondence between the code words (combinations of epigenetic marks) and their meaning (the TFs they code for) is not gratuitous (Lesne, 2006; Kühn and Hofmeyr, 2014). Instead, the correspondence is a biochemical recognition of the pattern of histone marks by the TF. This is why it was recently proposed that chromatin may have evolved as an allosteric enzyme able to mediate a gratuitous correspondence between epigenetic marks and TF binding to the underlying DNA sequence (Lesne *et al.*, 2015).

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, H2B, H3, and H4 extend from the globular histone core to form the so-called histone tails [see Fig. 2(b)]. 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, and hence adopt a random coil configuration, as suggested by crystallographic studies (Luger *et al.*, 1997; Davey *et al.*, 2002) and proteolytic cleavage assays. Tails contribute differently to intranucleosomal stability and internucleosomal interactions (Allan *et al.*, 1982; Arya and Schlick, 2006, 2009; Zhou *et al.*, 2007; Sinha and Shogren-Knaak, 2010). 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, located on the H2A-H2B interface [see Fig. 2(a)]. A H4 tail on one nucleosome may interact with an H2A/H2B acidic patch on an adjacent nucleosome, acting as a tether connecting the two nucleosomes (Kan, Caterino, and Hayes, 2009; Kalashnikova *et al.*, 2013). 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 internucleosome interactions, although they are required

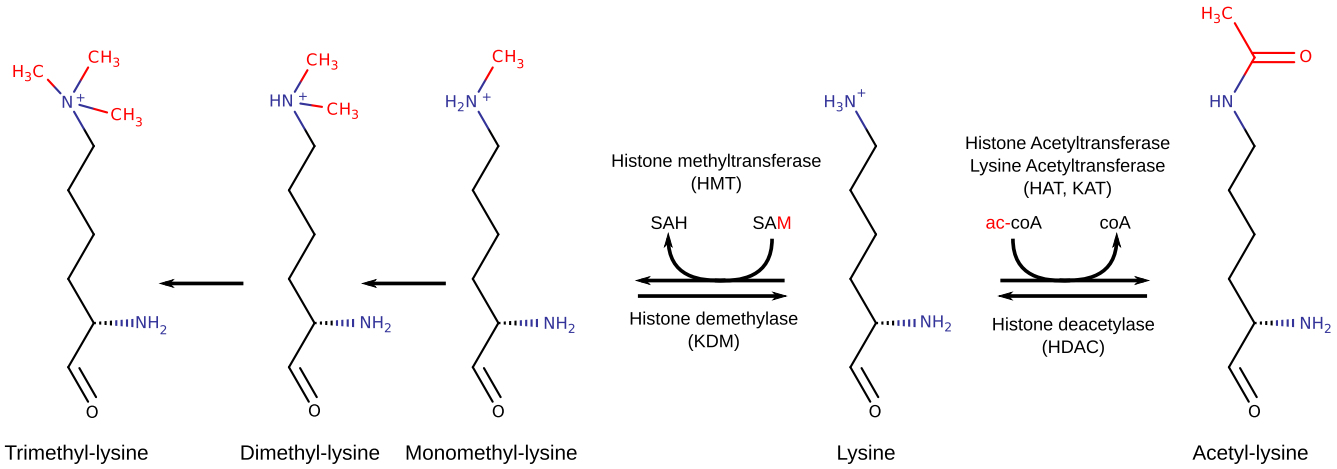

FIG. 5. Chemical formulas of unmodified, acetylated, and monomethylated, dimethylated, trimethylated lysine residues. Methylation (from lysine to the left) is achieved by successive additions of single methyl groups, using the S-adenosylmethionine (SAM) metabolite as a source of methyl groups. Acetylation (from lysine to the right) is achieved using the acetyl-coA cofactor as a source of acetyl groups. Acetylation reduces the charge at biological pH, whereas methylation preserves the charge.

for proper nucleosome reconstitution (Bertin, Durand *et al.*, 2007).

2. Histone-tail post-translational modifications

Histone tails, besides their role in the structuration of nucleosome arrays, are also the support of virtually all PTMs targetting histones, which consist of 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 (Tropberger and Schneider, 2013; Tessarz and Kouzarides, 2014), 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 (Zentner and Henikoff, 2013; Fierz, 2014; Pepenella, Murphy, and Hayes, 2014).

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

(i) Lysine methylation and notably the two main histone PTMs that are involved in gene silencing: trimethylation of the lysine 9 of H3, noted as H3K9me3, which recruits HP1 and trimethylation of the lysine 27 of H3, noted as 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 comprehensive list of the known writers and erasers is given in Table I.

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 acts on chromatin indirectly by recruiting additional architectural proteins (see Sec. III.B.5). For these reasons, acetylation mechanisms are more easily studied by *in vitro* experiments, while methylation effects are more generally studied in the *in vivo* context in the presence of their multiple partners. We now sum up some of the main experimental results and theoretical interpretations concerning both 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 amino acid 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 or without combinations of histone tail acetylation (Shogren-Knaak *et al.*, 2006; Wang and Hayes, 2008; Allahverdi *et al.*,

TABLE I. Principal histone lysine acetylations and methylations of the human epigenome and their writers and erasers. Adapted from Kouzarides (2007), Yang and Seto (2007), Sapountzi and Ct (2011), Zentner and Henikoff (2013), Fierz (2014), and Pepenella, Murphy, and Hayes (2014). Other PTMs not shown include arginine methylation and acetylation, serine/threonine phosphorylation, ubiquitination, and crotonylation [see Sadakierska-Chudy and Filip (2015)].

Type	Family/class	Proteins/complexes	Targets
	GNAT	HAT1	H4(K5, K12)
	GNAT	PCAF	H3(K9, K14, K18)
Acetyltransferases	GNAT	SAGA	H3
(HAT/KAT)	CBP/p300	CBP/P300	H3(K14, K18), H4(K5, K8), H2AK5, H2B(K12, K15)
	MYST	TIP60	H4(K5,K8, K12, K16), H3K14
	MYST	HB01	H4(K5, K8, K12)
	MYST	MOZ, MORF, MOF	H4
	Class I	HDAC1-3,8	H3, H4
	Class IIA	HDAC4,5,7,9	H3, H4
Deacetylases	Class IIB	HDAC6,10	H3, H4
(HDAC)	Class III (sirtuins)	SIRT1-7	H3, H4, H4K16
	Class IV	HDAC11	H3, H4
	None	SIN3-HDAC1,2	H3, H4
	None	N-COR/SMRT-HDAC3	H4
	SET	MLL1-5	H3K4
	SET	SET1A, SET1B	H3K4
	SET	G9A/GLP	H3K9
	SET	SETDB1	H3K9
Methyltransferases	SET	SUV39H1, SUV39H2	H3K9
(HMT/KMT)	SET	EZH2 (PRC2)	H3K27
	SET	NSD1	H3K36
	SET	SET2	H3K46
	SET	SUV420H1, SUV420H2	H4K20
	None	DOT1L	H3K79
	LSD1	BHC110	H3K4
	LSD1	COREST-LSD1	H3K4
Demethylases	JmjC	JHDM1A, JHDM1B	H3K36
(HDM/KDM)	JmjC	JHDM2A, JHDM2B	H3K9
	JmjC	JMJD2A, JHDM3A	H3(K9, K36)
	JmjC	JMJD2B	H3K9

2011; Liu *et al.*, 2011). In addition, small-angle x-ray scattering assays on folded nucleosome arrays give estimations of internucleosome interaction energies (Bertin, Renouard *et al.*, 2007; 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 (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 insight into how the electrostatic nature of histone tail PTMs influences chromatin folding.

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 a 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 [$\sim(5-6)k_BT$] with DNA than the unmodified one ($\sim 2k_BT$), 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, 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 the next H2A/H2B dimers of neighboring nucleosomes (Zhou *et al.*, 2007) [see Figs. 2(a)

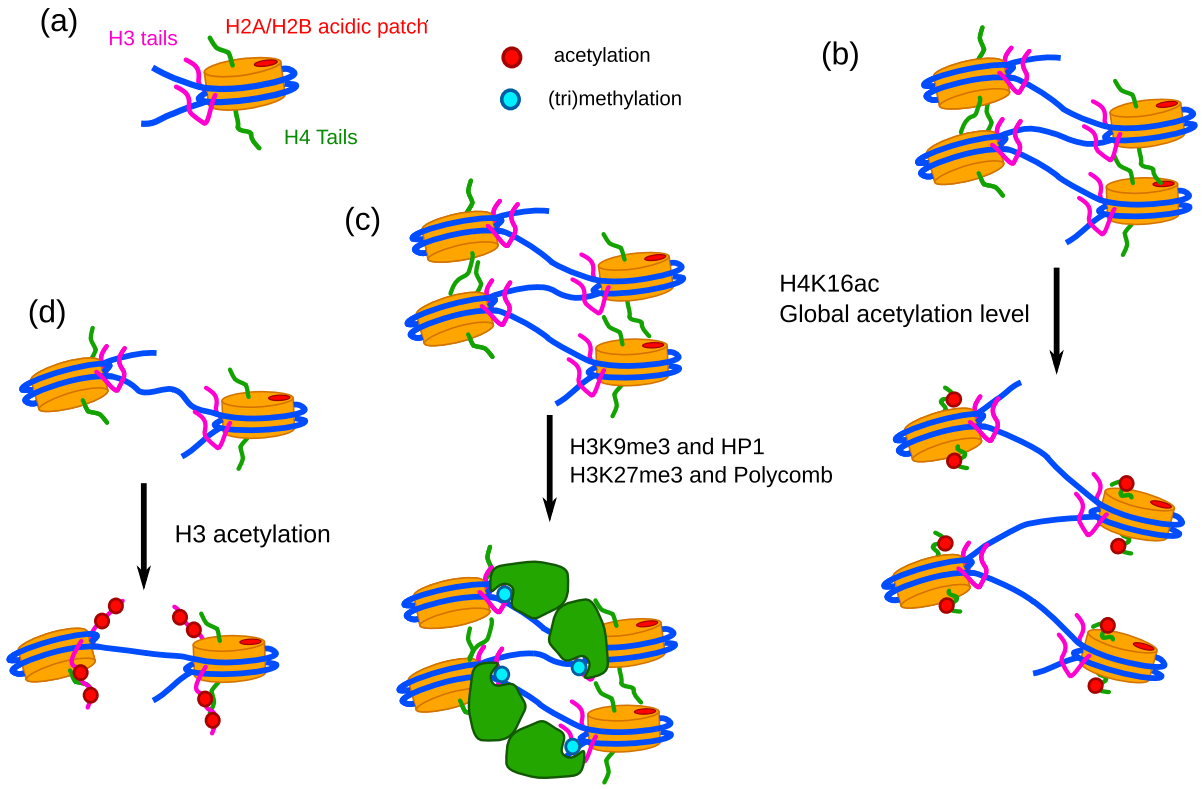

FIG. 6. Nucleosome arrays and histone PTMs. (a) A nucleosome core particle. (b)–(d) Examples of chromatin structural modulation by histone PTMs. Right: H4K16 acetylation induces the collapse of the N-terminal H4 tails on their own nucleosomal DNA, preventing them from binding the H2A/H2B acidic patches of adjacent nucleosomes. 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.

and 6(a). Hence, while modified H4 tails may contribute to the nucleosome-nucleosome attractive interaction by the so-called "tail bridging" effect (Mühlbacher, Holm, and Schiessel, 2006), the acetylation of lysine 16 might oppose this effect, leading to weakened nucleosome-nucleosome interactions (Potoyan and Papoian, 2012) [see Figs. 6(b)–6(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* when acetylated (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 and Schlick (2006, 2009) 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, 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 (Yang *et al.*, 2009; Allahverdi *et al.*, 2011; Liu *et al.*, 2011), showing that the effects of PTMs also largely depend on the valency and the concentration of bulk counterions, consistent with sedimentation assays.

4. H4K16 deacetylation 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 silent information regulatory (SIR) complexes which are recruited by deacetylated nucleosomes. This mechanism is crucially relying on H4K16 (Dayarian and Sengupta, 2013) (see below Sec. IV.C.1 for the detailed mechanism).

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 previously, does not directly induce chromatin fiber compaction [a notable exception was reported by 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 (nonspecific) 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. III.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 (Hathaway *et al.*, 2012; Azzaz *et al.*, 2014) 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 heterochromatinization.

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 (Zentner and Henikoff, 2013; Gosalia *et al.*, 2014; Ong and Corces, 2014; Mulligan, Koslover, and Spakowitz, 2015) or remodeling proteins ((Becker and Hrz, 2002); (ii) there is considerable cross talk among histone tail PTMs (Kouzarides, 2007; Li and Shogren-Knaak, 2008; Bannister and Kouzarides, 2011) 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 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 Journal of Physics: Condensed Matter (Everaers and Schiessel, 2015), devoted to the physics of chromatin, two contrasted models were proposed: the group of Andrew Spakowitz (Mulligan, Koslover, and Spakowitz, 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 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 mononucleosomes 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 heterochromatin and euchromatin (Mulligan, Koslover, and Spakowitz, 2015), whereas Rippe's group evidenced a dependence of the binding stoichiometry on the NRL 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

Polycombs are family proteins that mediate transcriptional silencing (Di Croce and Helin, 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 (Paro, Strutt, and Cavalli, 1998; Beuchle, Struhl, and Muller, 2001). It provides a mechanism for "cellular memory" (Ringrose and Paro, 2004) that has been speculated to be an 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. III.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; Cooper *et al.*, 2014). It was also 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 modeling 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 and 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. 7(a)]. 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 one to use the generation of a list of pairwise contacts between distal parts of the genome in various organisms and cell types [e.g., in drosophila, Fig. 7(b)

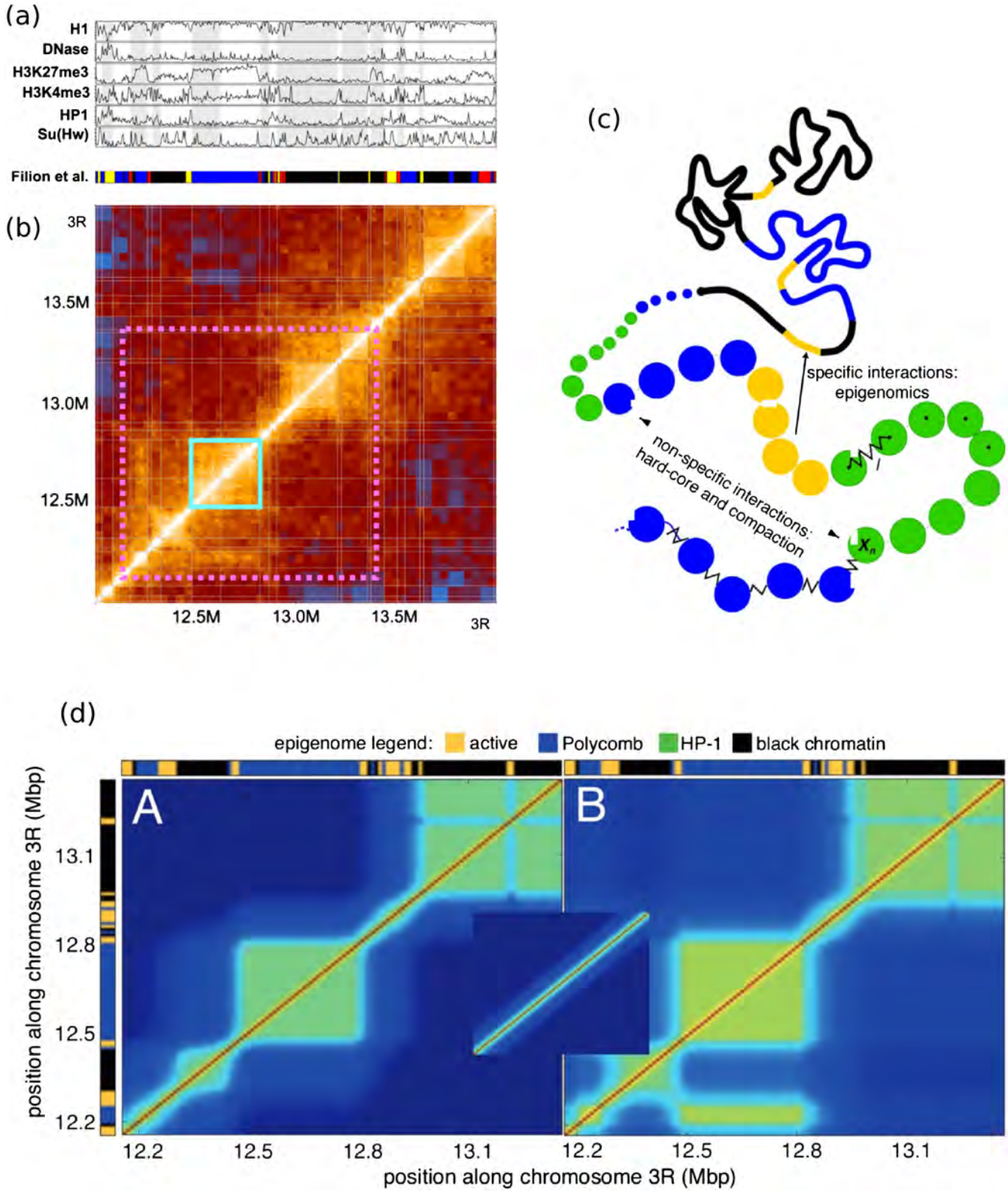

FIG. 7. Modeling of chromosomal contact maps from the epigenomic landscape. (a) Profiles of H1 occupancy, DNA accessibility, H3K27me3, H3K4me3, HP1, and a histone modifier, Su(Hw) along a region of the dorsophila chromosome 3R. At the bottom, the colors corresponding to these profiles are shown (Filion *et al.*, 2010). (b) The corresponding contact map. (a), (b) Adapted from Sexton *et al.*, 2012. (c) Schematics of the copolymer model used in Jost *et al.* (2014). (d) Two predicted contact maps corresponding to the region indicated by the pink dashed square in (b). (c), (d) Adapted from Jost *et al.*, 2014.

(Sexton *et al.*, 2012)]. A high contact probability is characterized by a bright pixel in the Hi-C contact map [see Fig. 7(b), around the diagonal]. The first results confirmed the physical segregation of the genome into heterochromatin and euchromatin regions (Lieberman-Aiden *et al.*, 2009). An attractive model of chromosome folding was proposed in this seminal paper, named "fractal globule," which explains at the same time the existence of chromosome territories, the absence of knots in chromosomes, and the power-law decay (with exponent close to -1) of the contact probability as a function of the genomic distance (Mirny, 2011). This long-lived metastable state was introduced 20 years before on theoretical grounds (Grosberg, Nechaev, and Shakhnovich, 1988) for explaining the kinetics of collapse of homopolymers. At finer scales, Hi-C further led to the identification of domains along the genome in which contacts are numerous, whereas very few contacts are established in between different domains. These regions are termed topologically associating domains (TADs) (Dixon *et al.*, 2012; Nora *et al.*, 2012). TADs can be seen as high intensity blocks along the diagonal of the chromosomal contact maps [Fig. 7(d), cyan squares]. It was then realized that chromosomes are actually block copolymers, each block corresponding to an epigenetic domain [compare Figs. 7(a) and 7(b)]. Combining Hi-C results with the linear epigenomic annotations of the genome (i.e., the biological information of the underlying sequences) is in principle a powerful method to comprehend the functional architecture of the genome.

Several physical models have been developed so far in order to understand the 3D folding properties of block copolymers. The main goal of these studies is to recover the chromosomal contact maps observed from the Hi-C data. Two main classes can be distinguished: simulations that explicitly compute the 3D chromosome conformations (Barbieri *et al.*, 2012; Benedetti, Dorier, and Stasiak, 2014) and implicit models in which average contact maps are directly computed in a self-consistent Gaussian approximation (Jost *et al.*, 2014). The different explicit models can account for the formation of TADs, either by preferential binding of cofactors along specific regions of the genome (Barbieri *et al.*, 2012) or by topological constraints (Benedetti, Dorier, and Stasiak, 2014) but so far, the direct comparison with experimental results has been done using only the implicit approach (Jost *et al.*, 2014). In this study, they used the previously described colors of chromatin drosophila (Filion *et al.*, 2010), which assign to each subregion of the genome an epigenetic color, based on the specific protein binding and histone marks found in this region [Fig. 7(a)]. They then assign specific pair potentials [see also Saberi *et al.* (2015)] between beads of the same or different colors [Fig. 7(c)] and compute the corresponding contact maps using a statistical approach previously described (Timoshenko, Kuznetsov, and Dawson, 1998). With well-chosen parameters, they were able to retrieve the contacts found experimentally [Fig. 7(d)]. An important outcome of their study is that a fixed epigenetic landscape is compatible with several 3D conformations of the chain, a phenomenon which they call multistable folding (see Sec. III.C.2). Note that these copolymer models, all at thermal equilibrium, drastically deviate from the fractal globule hypothesis.

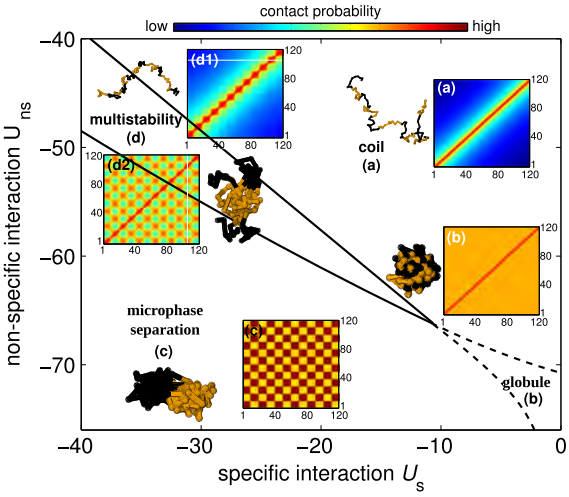

FIG. 8. Phase diagram of a toy model copolymer as a function of specific and nonspecific interactions. From Jost *et al.*, 2014.

2. The physics of TADs: Finite-size effects in the coil-globule transition of copolymers

Jost *et al.* (2014) showed a phase diagram of a toy model copolymer as a function of the intensity of (i) block-specific and (ii) nonspecific interactions that we show in Fig. 8. On top of the coil-globule transition of the whole copolymer, there is also a 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 was proposed by Caré *et al.* (2014).[2]

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, undergoes a coil-globule transition around the 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.

We then consider a copolymer $ABAB\ldots$ made of small blocks A and long blocks B. For example, the A blocks could be HP1-like heterochromatin, and the B blocks could be polycomblike heterochromatin. We consider monomer-monomer attractive interactions represented by an energy of interaction E_{ij} between monomers with epigenetic states i and j of the following kind: $E_{ij} = U_{ns} + \delta_{ij}U_s$, where U_{ns} is a nonspecific term (does not depend on i and j), δ_{ij} is the Kronecker delta, and U_s is a specific interaction term.

[2]See also arXiv:cond-mat/0004273.

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. 7). This corresponds to the "multistate folding" region calculated by Jost *et al.* and depicted in 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 $\Theta(N) - a/\sqrt{\ln(N)}$ all the B globules are permanently collapsed in a macroglobule with the A blocks located at the macroglobule surface (because of interfacial tension). Crucially the A blocks are still coils, hence in the euchromatin phase, so that their genomic sequence is expressed, whereas the B blocks are globular and as such in the heterochromatin phase, hence their underlying sequence is repressed (Caré *et al.*, 2015).

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 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 of 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 review the main modeling efforts that have been made in order to address these questions.

A. Mathematical modeling

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 one 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

$$\sum_{j=1}^{n_S} n_j = N. \tag{1}$$

Many theoretical works use the silenced mating-type locus of the fission yeast *Schizosaccharomyces pombe* [reviewed by 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 Klar, 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 is 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 daugther 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 write a general expression for the time evolution of the variables n_j:

$$\frac{dn_j}{dt} = \sum_{k\neq j}^{n_S} R^+_{jk} n_k - \sum_{k\neq j}^{n_S} R^-_{jk} n_j + \text{noise}. \tag{2}$$

Here R^+_{kj} is the rate of transition of nucleosomes from the state k to the state j, while R^-_{jk} is the rate of transition from state j to k (obviously, $R^+_{kj} = -R^-_{jk}$). 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 Micheelsen *et al.* (2010). They considered the case of $n_S = 2$, that is, they considered only a modified (M) or unmodified (U) state. Using Eq. (1), the system may be described by only one variable n_M, the number of modified nucleosomes. The transition rates are given by

$$\begin{aligned} R^+_{UM} &= \alpha n_M^2 + (1-\alpha), \\ R^-_{MU} &= \alpha(1-n_M)^2 + (1-\alpha). \end{aligned} \tag{3}$$

This model supposes that the creation of a modified state involves a cooperative transition [as evidenced by the quadratic terms in Eq. (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: $n_{M1} = 0$ and $n_{M3} = N$, which are stable, and $n_{M2} = N/2$, 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.* (2009) thoroughly investigated a more general model that still has $n_S = 2$. Their rates read

$$\begin{aligned} R^+_{UM} &= \chi + \alpha n_M^H, \\ R^-_{MU} &= \gamma + \eta(1-n_M)^K. \end{aligned} \tag{4}$$

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 generalization of the model of Micheelsen *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 χ and γ 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 halve the concentration of modified nucleosomes for each daugher 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.* (2007) in a very similar form. They considered 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. 10(a)]. A hypothesis is that only the $U \to M$ and $U \to A$ are allowed, but the $M \to A$ transition is not (i.e., $R^+_{MA} = 0$). They write the following transition rates:

$$\begin{aligned} R^+_{UA} &= \alpha_A n_A + \chi_A, \\ R^+_{UM} &= \alpha_M n_M + \chi_M, \\ R^-_{AU} &= \beta_M n_M + \gamma_A, \\ R^-_{MU} &= \beta_A n_A + \gamma_M. \end{aligned} \tag{5}$$

The study of the system in the case where the basal rates χ_M, γ_M, χ_A, and γ_A vanish already shows the existence of four fixed points: two stable fixed points $\{n_A = 1, n_M = 0\}$ and $\{n_A = 0, n_M = 1\}$, an unstable saddle point $\{n_A = 0, n_M = 0\}$, and an unstable fixed point

$$\begin{aligned} \{n_A &= \alpha_M\beta_A/[\alpha_A\beta_M + (\alpha_M + \beta_M)\beta_A], \\ n_M &= \alpha_A\beta_M/[\alpha_A\beta_M + (\alpha_M + \beta_M)\beta_A]\}. \end{aligned}$$

The two latter points are aligned along the $n_A = n_M$ line and create a barrier between the two basins of attraction (David-Rus *et al.*, 2009). The phase flow diagram of such system is depicted in Fig. 9.

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

$$\begin{aligned} R^+_{UA} &= \epsilon_A n_A + k_0, \\ R^+_{UM} &= \epsilon_M n_M + k_0, \\ R^-_{AU} &= \epsilon_M n_M + k_0, \\ R^-_{MU} &= \epsilon_A n_A + k_0, \end{aligned} \tag{6}$$

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 one to map the system to the zero-dimensional Ising model, with the correspondence $A = +1$, $U = 0$, and $M = -1$. Within this analogy, recruitment corresponds to coupling between spins and random transitions are associated with thermal fluctuations. A new observable, equivalent to the magnetizaton in the Ising model, is introduced here: $\mu = a - m$.

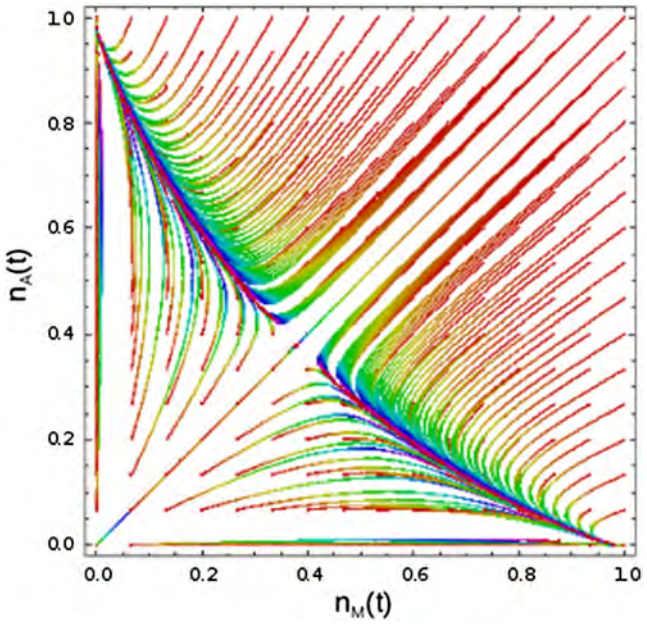

FIG. 9. Phase flow portrait for the three-state model described by David-Rus *et al.* (2009), Eqs. (5). The values of the parameters here were $\alpha_A = \alpha_M = 5$ and $\beta_A = \beta_M = 3$. One can clearly observe the presence of the stable fixed points and the unstable fixed point. From David-Rus *et al.*, 2009.

Some known results can thus be recalled for the symmetric recruitment case $\epsilon_A = \epsilon_M = \epsilon$. Similarly to what was 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 ϵ, $\mu = 0$ becomes unstable and bistability settles down with the appearance of two stable fixed points $\mu_\pm = \pm(k_0/\epsilon)\sqrt{(\epsilon/k_0 + 1)(\epsilon/k_0 - 3)}$.

The nonlocal character of the nucleosome-nucleosome interaction, which is the main hypothesis of the zero-order models, was further justified by Zhang *et al.* (2014). They proposed a two-layer Potts model in which in one layer they described the nucleosomes, and in the other they included explicitly the enzymes that modify the nucleosomes. The interaction between the nucleosomes is 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.* (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 antimodified 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 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 is relevant for the mating-type loci in yeast, it probably fails when multicellular organisms are considered. It is known, for example, that nearly all noncentromeric H3K9me3 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 (2011) realized 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*. In this organism, the MAT (mating type) gene has two variants, MATa and MATα, and switching of the mating type occurs when the expressed MAT variant changes from one type to the other. This mechanism is possible because each of the two variants comes also with a silenced allele: the HML (hidden MAT left) carries a silenced MATα allele, and HMR (hidden MAT right) bears a silenced MATa allele. The HML and HMR are able to spontaneously flip between high and low expression states (Xu, Zawadzki, and Broach, 2006), thereby allowing for switching of mating type. These domains are stable over 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 and Sneppen therefore considered 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. 10(c)]. To limit the long-range interaction between DNA sites one can introduce a distance dependent cooperativity, i.e., by making the reaction rate R_{UM} 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. Because of 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 in gene promoters (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 makes the silencing spread out. This effect can be limited by introducing regularly spaced weak barriers, modeled as antisilencers (enhancers) of the $U \to A$ 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.

Focusing instead on mammal silenced regions, Hathaway *et al.* (2012) were able to reproduce the sharp peaks observed in the experimental modification patterns by including a "source" term in their model. This is a model in which the

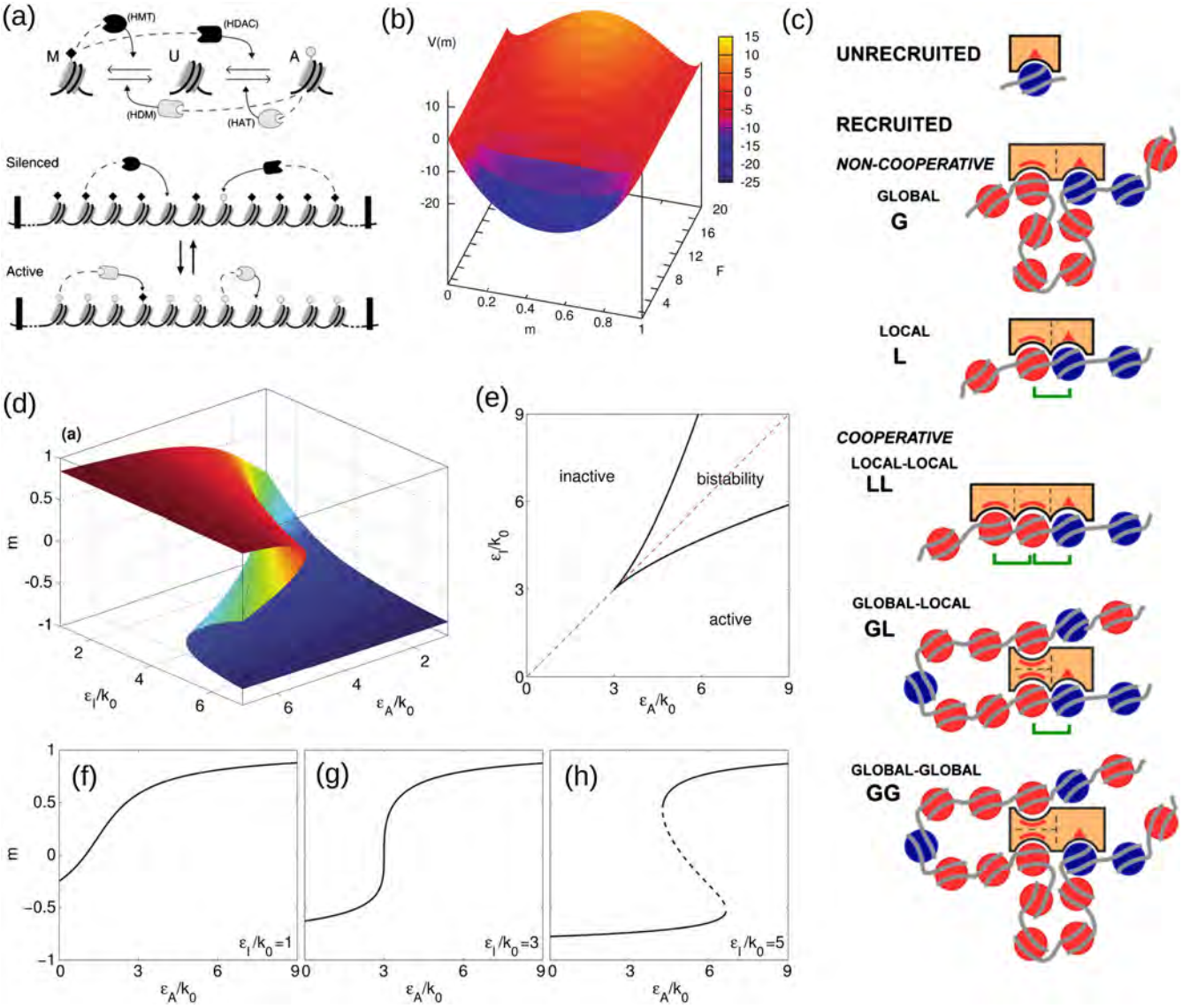

FIG. 10. A modern view on epigenetic landscapes. (a) The basic mechanism of a three-state nucleosome modification model, depicting the modified (M), unmodified (U), and antimodified states (A). The transition between M and U is catalyzed by histone methyltransferases (HMTs) and histone demethylases (HDMs), which depends on an antimodified histone; between U and A the transition is catalyzed by histone acetylases (HATs) and histone deacetylases (HDACs), which depend on a modified histone. 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. From Micheelsen *et al.*, 2010. (c) Different modes of coupling between histone modification states, as described by Dodd and Sneppen (2011). Unrecruited enzymes may modify histones directly. Otherwise, recruited enzymes may operate in noncooperative (global or local) or cooperative (local-local, global-local, or global-local) modes. From Dodd and Sneppen, 2011. (d)–(h) The "epigenetic landscape" as proposed by Jost (2014). As a function of the control parameters ϵ_A and ϵ_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. (d)–(h) From Jost, 2014.

initiation and spreading are explicitly separated, and in turns this allows one to reproduce spatial patterning. They write rates as

$$\begin{aligned} U_0 &\xrightarrow{k_+^{\text{target}}} M_0, \\ \{M_i, U_{(i+1 \text{ or } i-1)}\} &\xrightarrow{k_+} \{M_i, M_{(i+1 \text{ or } i-1)}\}, \\ M_i &\xrightarrow{k_-} U_i. \end{aligned} \tag{7}$$

This description means that at site 0 there is an active modification source with rate k_+^{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 $0.1 - 0.2\ \text{h}^{-1}$ (in agreement with different experimental estimates of k_-). However, as pointed out, this model fails to predict the bistable nature of the system, thus not allowing one to describe this crucial feature.

Hodges and Crabtree (2012) presented a more detailed study of the model. The source term ensures that the resulting

mark distributions are peaked at the nucleation site, as experimentally observed, provided k_+^{target}/k_- is large enough ($\gtrsim 0.2$), with increased amplitude and formation rate for increasing k_+^{target}.

Still referring to the mating-type loci in budding yeast, Dayarian and Sengupta (2013) considered a four-state model with site-dependent rate equations. The fourth state they considered is a double-acetlyated state, which corresponds 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 a 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). Interestingly, indeed, this model also allows for the existence of a silenced and an active domain in stable coexistence, i.e., with an immobile boundary domain, whose position depends on the balance between environmental self-adjusting parameters as the concentration of active proteins, a mechanism that they explored extensively (Sedighi and Sengupta, 2007; Dayarian and Sengupta, 2013; Mukhopadhyay and Sengupta, 2013).

2. Three-dimensional models

Erdel, Müller-Ott, and Rippe (2013) addressed some more specific questions about the establishing of epigenetic domains, such as how are the chromatin-modifying enzymes targeted or excluded from given chromatin regions, how exactly the modification can propagate from one nucleosome to another, or 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 setup of the correct architecture in the initial enzyme binding and in defining the functional chromatin domains remains open.

It also was recently proposed that pericentromeric heterochromatin spreads its silenced state with a "nucleation and looping" mechanism (Müller-Ott *et al.*, 2014). Chromatin-bound SUV39H1/2 complexes act as nucleation sites and propagate a spatially confined heterochromatin domain with elevated H3K9me3 modifications via chromatin loops. It is therefore relevant to include the three-dimensional structure in the theoretical modeling of the spreading of epigenetic marks.

D. Biological relevance of the models

In this section we examined 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 we return to the discussion on the Waddington landscape started in the Introduction. The classical image in the original Waddington representation (Waddington, 1957) of a marble rolling down a hill rather suggests a fixed landscape, leading to erroneous interpretation when one goes beyond the metaphorical level (see Fig. 1).

In the simplest model discussed, the one by Micheelsen *et al.* (2010), showed that the model can be reformulated by a Fokker-Planck equation for the 1D diffusion of a particle in an effective potential $V(n_M)$ [see Fig. 10(b)]. The latter accounts at a time for drift (external forces) and noise events (with a term of the type D/μ, with D the diffusion coefficient and μ the mobility). The Waddington idea of an epigenetic landscape is translated by Micheelsen *et al.* (2010) in more modern terms, by defining a physically consistent energy profile. Note, however, then the mechanism invoked here is not an evolution along the profile of Fig. 10(b) 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 $F = \text{const}$ section of the two-dimensional potential surface of Fig. 10(b). This allows one 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 (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. Using the notation of Jost (2014) [Eqs. (6)], we notice that if the recruitment of enzymes by modified or antimodified marks is different, the stability diagram and the boundaries between the monostable and bistable regions can be traced as a function of the two parameters ϵ_A and ϵ_M. Bistability is observed only for strong recruitment ($\epsilon_{A,M} < \epsilon_c$) and small asymmetry.

The epigenetic landscape may also be viewed as a complex, multidimensional dynamical system in which different cell identities correspond to different dynamical attractors of the system. In one approach, such a landscape is modeled to be shaped by gene regulatory networks. In a recent study, it was

proposed that stem cell differentiation may be viewed as the process of transition between steady-state attractors of genes that induce or repress cell pluripotency (Zhang and Wolynes, 2014). Also, it was suggested that the cellular identities are characterized by a mixture of several states, and external signals may drive the transition from one cell state to another one. By analyzing existing data sets, Lang *et al.* (2014) were able to provide direct evidence for this, demonstrating that epigenetic landscapes are a very powerful tool to understand cellular dynamics.

2. Hysteresis

For even stronger recruitment, a typical hysteretic behavior appears that may have important biological consequences. One can expect that, while for differentiated, stable cells recruitment parameters are almost symmetric, and modifications of the environment might actively induce asymmetric recruitment. The increase of one recruitment parameter can thus bring the system along the metastable branch, and then make it abruptly switch to the alternative state, which will remain stable even when the recruitment parameters come back to their initial values, thanks to the hysteretic shape of the bifurcation curve. In Fig. 10(d), starting for instance from the low m state and symmetric recruitment, one can increase ϵ_I/k_0 and switch to the upper, high m branch, then come back to $\epsilon_A = \epsilon_I$ without switching back [see also Figs. 10(e)–10(h)].

Close to $\epsilon_A \sim \epsilon_M \sim \epsilon_c$, the system becomes ultrasensitive to perturbations and highly unstable. This regime may be associated with 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 three-state model proposed by Dodd *et al.* (2007) was 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. 11) (Song, Irwin, and Dean, 2013). This latter feature is characteristic of annual plants, while in FLC perennial plants is repressed only transiently.

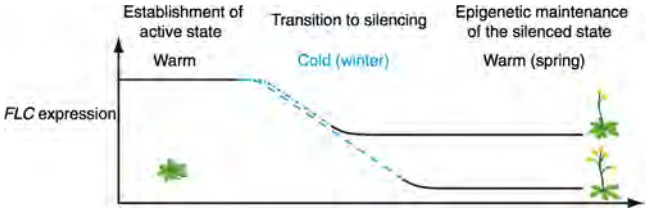

FIG. 11. Mechanism of vernalization. The expression of the floral repressor gene, FLC, is repressed when plants are exposed to cold and remains stably repressed on the return to warm temperatures. Since this repression increases with the duration of the exposure to cold, flowering is more abundant for longer cold duration. From Song, Irwin, and Dean, 2013.

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

$$R_{U\to A} = \beta n_A + \chi, \tag{8}$$

$$R_{U\to M} = u(T)(\beta n_M + \chi), \tag{9}$$

$$R_{A\to U} = v(T)(\beta n_M + \chi), \tag{10}$$

$$R_{M\to U} = \epsilon(\beta n_A + \chi), \tag{11}$$

where $u(T)$ and $v(T)$ account for the temperature tuning and take 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. They proved 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 division rate and rapidity of the deposition of epigenetic marks after vernalization. Changes in these parameters may lead to a much shorter memory extent (from more than 1 yr to a few days), potentially explaining the different behavior observed in annual and perennial plants.

While the previous work addressed the question of bistability behavior in vernalization, the question of the establishment of epigenetic marks induced by cold was discussed by Angel *et al.* (2011), both theoretically and by experiments in *Arabidopsis thaliana*. They focused on the fact that, when subjected to cold, repression (H3K27me3, *M* state) concerns only a small (1 kb) nucleation region inside the FLC (8 kb), close to the first exon (coding region) after the promoter. While during the cold only the nucleation region is marked, after warm restoring the profile changed rather little in the nucleation region but rose quantitatively across the rest of the FLC locus according to the length of the cold period. They asked whether the small size of the nucleation region would be sufficient to cause a quantitative switch in the epigenetic state of the whole FLC locus after returning to the warm. Experimental results are shown to be compatible with a three-state zero-dimensional model, provided that two supplementary ingredients are included: (i) a site-specific nucleation of the silencing modification during cold, described as an increased probability to switch to the *M* state for a subensemble of the nucleosomes, and (ii) a permanent bias in the histone dynamics toward the *M* modification on return to the warm. Within these assumptions simulated population-averaged levels of the *M* modification are found to be approximately stable up to 30 days after the cold, with a modification intensity which depends on the duration of cold, in good agreement with experiments.

Together these two studies show that relative simple models displaying a strong bistability can be usefully employed to model epigenetic mechanisms involved in real systems as, in the case discussed here, in plants, even if a real system typically includes a few additional features needed to specifically respond to the particular functional task they are designed for.

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

Up to now we 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 have been known for a long time, such as DNA methylation (see Sec. V.A), while others have been evidenced quite recently, as chromosome coating with (long) noncoding RNAs as in *X* inactivation (see Sec. V.C), messenger (i.e., protein coding) RNA silencing by interaction with micro RNAs (see Sec. V.D), or the coupling between epigenetics and supercoiling (see Sec. V.E). An exhaustive description of the overall picture represents 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.

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. I). In fact, in early days of research on DNA methylation, it was found that methylation states are propagated through mitosis (Wigler, Levy, and Perucho, 1981).

DNA methylation is the substitution of a methyl ($-CH_3$) group to the carbon atom in position 5 at the cytosine base (5 mC). Importantly, DNA methylation is coupled to metabolism through SAM [see Fig. 12(a)].

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 was detected in any sequence context in both 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), illustrated in Fig. 13. The precise mechanism behind the establishment of the initial methylation pattern during development is largely unknown. It was 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 deamination [see Fig. 12(a)], leading to a common DNA mutation. The hydrophobic methyl group in the DNA major groove gives a structural similarity between thymine and 5 mC [see Fig. 12(b)]. It is important to note that this allows for the possibility of a base readout in the major groove as proposed by Machado *et al.* (2014). We 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 (Mohandas, Sparkes, and Shapiro, 1981; Graves, 1982; Venolia *et al.*, 1982), and gene imprinting (Li, Beard, and Jaenisch, 1993; Razin and Cedar, 1994). It was later established that when CpG-island promoters are methylated, then the gene is irreversibly silenced (Jones, 2012). With the advent of technologies that enable for genome-wide screening of the methylation state of DNA, it has become clear that gene silencing is not the only role of DNA methylation (Lister *et al.*, 2009; Jones, 2012), 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

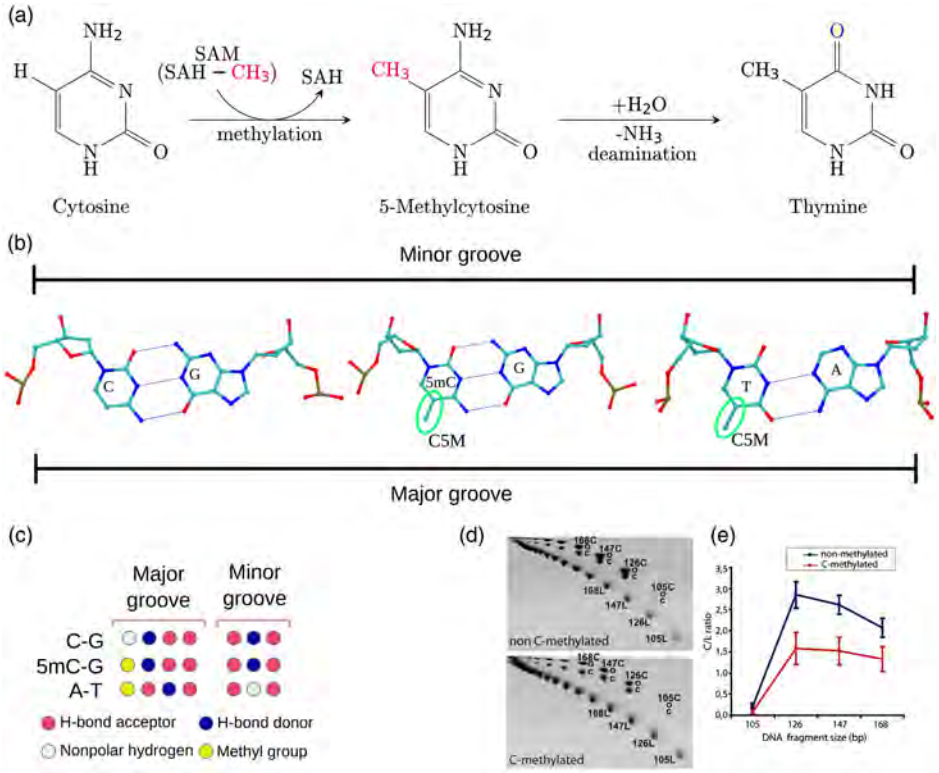

FIG. 12. Physical aspects of DNA methylation. (a) Conversion of cytosine to 5-methylcytosine occurs using SAM (S-adenosylmethionine) as a methyl group donor. Spontaneous deamination may convert 5 mC to thymine, leaving the methyl group in the major groove. (b) Structural similarities between 5 mC and thymine in the DNA major groove. (c) Specific patterning of H-bond donors, acceptors, methyl groups, and nonpolar hydrogen atoms allows for "base readout" of the DNA sequence without strand opening (see main text). (d), (e) Experimental cyclization experiments on methylated and unmethylated DNA show distinct elastic properties of the two species. From Pérez *et al.*, 2012. (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.

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; Spruijt and Vermeulen, 2014; Marchal and Miotto, 2015), suggesting that the role of DNA methylation is not at all limited to repression of transcription.

DNA methylation was also 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 the prediction of a precancerous cell state (Feinberg and Tycko, 2004; Timp and Feinberg, 2013).

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

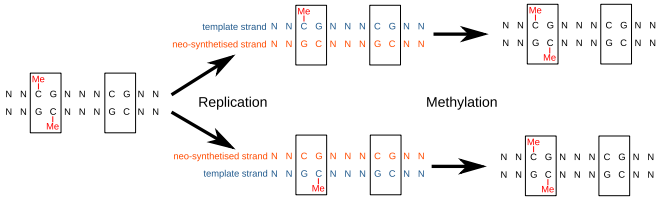

FIG. 13. Methylation transmission. After replication, only hemimethylated bases are converted into fully methylated bases.

1. Mechanical properties of DNA change upon methylation

It was 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 one to prove that oligomers stiffen significantly upon methylation [see Figs. 12(d) and 12(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 toward the DNA major groove and a stiffening of the sequence. However, when a polydinucleotide of type $\mathrm{d(CpG)}_n$ 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, Li, and Olson, 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." Moreover, by combining single-molecule cyclization experiments with all-atom molecular dynamics simulations, Ngo *et al.* (2016) showed that 5 mC increases the DNA local rigidity and also reached the conclusion that this destabilizes nucleosomes.

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 (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, however leaving 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, Severin *et al.* showed 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 a methylated A-form DNA oligomer (Mayer-Jung, Moras, and Timsit, 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. 12(c), 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 readout in the major groove that does not require strand separation (Sasai, Nakao, and Defossez, 2010; Zou *et al.*, 2012).

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.* (2013) found that roll and minor groove width were excellent predictors of sequence specificity for DNaseI endonuclease. 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 fluorescence resonance energy transfer (FRET) to measure relative displacement between two sites on nucleosomal DNA, Lee and Lee (2012) showed that methylation leads to tightening of DNA around the nucleosome. Along the same lines, Choy *et al.* (2010) suggested 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.* (2010) showed that there is a small increase (about 1.2%, from 75 to 76.2%) in preference of nucleosome positioning for methylated DNA. These data were contested by Felle *et al.* (2011) that instead showed that the nucleosome occupancy of methylated DNA was twofold lower compared to unmethylated sequences.

Other studies showed that the increased rigidity of methylated DNA disfavors nucleosome positioning. We already discussed the works of Yusufaly, Li, and Olson (2013) and Ngo *et al.* (2016), but many other studies reached the same conclusions (Felle *et al.*, 2011; Pérez *et al.*, 2012; Portella, Battistini, and Orozco, 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 was 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 a careful investigation of the differences between their physical properties is needed, which only very recently has begun (Ngo *et al.*, 2016).

Traditional sequencing methods for detection of cytosine methylation are not able to distinguish among 5mC, 5hmC, 5fC, and 5caC. Accurate methods to identify the genome-wide map of cytosine methylation states have only 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 (Spruijt and Vermeulen, 2014; Schübeler, 2015). We suggest that physics may help advancing our understanding of the biological aspects of this important epigenetic mark.

A very recent study has also shown by all-atom molecular dynamics and single-molecule fluorescence that DNA methylation is able to promote attraction between DNA fragments mediated by polyvalent cations (Yoo *et al.*, 2016). This study suggests yet another fascinating role for the role of physics in determining the large-scale organization of DNA in cell nuclei.

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 by Cedar and Bergman (2009). However, to our knowledge there has not been 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 (Lesne *et al.*, 2015). 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 transcriptional repressor CTCF protein that prevents the interaction between regulatory sequences (enhancers) located downstream of H19 and the Igf2 gene located farther 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 100 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, the X chromosome is much longer and contains thousands of genes. Females thus carry twice as many X-linked genes as males, 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 body (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 the paternal or maternal X chromosome (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 center (Xic). It was proven that the starting mechanism in X inactivation is mediated by the noncoding (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 noncoding RNA gene that is antisense to the Xist gene (it is transcribed from the complementary DNA strand). Because of this compementarity, 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, and 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 was proposed (Giorgetti *et al.*, 2014) that relies on the polymer physics properties of chromatin, on its organization in 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 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.* (2014) and previously discussed. The distribution of conformations observed thanks to carbon copy chromosome conformational capture (5C) experiments, illustrated by 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 (Imbert, Lesne, and Victor, 1997; Caré *et al.*, 2014).

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 confirmations indeed changes 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 ensure that Xist is transcribed only 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, and chromatin condensation, i.e., the same general mechanisms that we previously introduced.

D. Noncoding 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, noncoding RNAs have more recently emerged as one of the main epigenetic mechanisms. Two main classes of epigenetically active noncoding 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 noncoding RNA (lncRNA), it has been shown that Xist RNA directly recruits chromatin-modifying factors such 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 involved (Xist Tsix and others) (Pontier and Gribnau, 2011). According to a proposed model (Lee, 2012), long noncoding 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 work 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 lncRNAs and chromatin modifiers in epigenetics]. Recently, a single-molecule, single cell RNA fluorescence *in situ* hybridization (FISH) has allowed one to quantify and categorize the subcellular localization patterns of a representative set of 61 lncRNAs in three different cell types, giving further insight into linking their putative role to their position in the cell (Cabili *et al.*, 2015). lncRNAs exhibit a diversity of localization patterns, including a dispersed distribution in the nucleus or in the cytoplasm, and condensed in nuclear foci, a configuration which may be consistent with a role for these lncRNAs in chromatin regulation, as shown for Xist and other lncRNAs involved in imprinting. Most lncRNAs present, however, stronger nuclear localization than most messenger RNAs. Interestingly, lncRNAs do not persist at nuclear foci during mitosis, suggesting that retention at specific regulatory regions through mitosis is likely not a mechanism of mitotic inheritance.

A whole new realm of small noncoding RNAs was discovered in the late 1990s. 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 slightly longer RNAs, playing 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. Again, we note here that unnecessary and potentially deleterious genes are often under epigenetic control (Cowley and Oakey, 2010).

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[3] that targets repetitive DNA sequences (Grewal and Jia, 2007). These are DNA sequences of up to several million base pairs and consist of a

[3]The RNA interference machinery includes different factors (such as Dicer, Argonaute, and RNA-dependent RNA polymerases) that produce the small RNAs or bind them to form functional complexes.

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 the production of 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 play together in the setting up and maintenance of different functional chromatin states. DNA methylation and histone modifications often act together to regulate a 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 probably need 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.* (2014). A typical basic regulatory network involving miRNAs and epigenetic regulators is the double negative feedback 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 was discussed in Sec. IV, it is possible to describe the system. More precisely, the model variables are the number of miRNAs, the number of messenger RNAs that miRNAs repress, and the number of proteins (epigenetic regulator) resulting from the messenger RNA translation (and repressing, in turn, the miRNAs). The interesting input from the miRNA regulation step is its role in keeping fluctuations of gene expression under control, either by suppressing translation or by promoting RNA degradation. Both effects helps indeed in reducing the burstiness in protein production (Friedman, Cai, and Xie, 2006; Osella *et al.*, 2014). The resulting set of rate equations includes therefore, on the one hand, the highly nonlinear and bistable character of epigenetic regulation and, on the other hand, the stabilizing effect of miRNAs regulation, leading to an increased stability of the system and to an increased range of bistability of the switch.

This result suggests 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 Vinograd *et al.* (1965) and it took some more years for Wang (1971) to discover the first enzyme able to relax these topological constraints *in vivo*. The 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 alter only 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 (Kouzine *et al.*, 2004, 2008; Liu *et al.*, 2006; Belotserkovskii, Mirkin, and Hanawalt, 2013). Therefore supercoiling would act as a transient mechanotransducer as well as a physical epigenetic mark. Moreover, nucleosome conformational changes might help to smooth 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 (Sheinin *et al.*, 2013). It remains to be seen how much the structural differences provided by histone variants (Shaytan, Landsman, and Panchenko, 2015) would help to build "elongation friendly" regions that could be transmitted through cell division.

Supercoiling of DNA was 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 were generated (Bermudez *et al.*, 2010; Joshi, Piña, and Roca, 2010; Kouzine, Gupta *et al.*, 2013; Kouzine *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, Zhao, and Li, 2009; Maizels and Gray, 2013). The emerging picture is that supercoiling is associated with 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 are induced by supercoiling (Lesne *et al.*, 2015).

But how can 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, Martinez-Perez, and Merkenschlager, 2013; Hirano, 2014).

The NRL (see Sec. II.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 (Padinhateeri and Marko, 2011), and (ii) topoisomerase activity. Importantly both these adenosine triphosphate (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 transmits 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, multiscale approaches articulating different models at different levels of organization are to be developed.

Mainly, epigenetics display an intrication 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. Importantly, theoretical models are now amenable to experimental verification, thanks to the many technologies that are available to that end. It is worth noting that the field of synthetic chromatin biology (Keung *et al.*, 2015), which allows one to modify and manipulate epigenetic states, is a promising new avenue for putting theories to the test.

Significant and challenging issues remain such as the following:

(i) Coupling nuclear architecture and epigenetic marking: Understanding the interplay between 3D nucleus architecture and 1D epigenetic marking (including barrier positions).

(ii) Physical signals that stimulate epigenetic marking: 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, 2015). Further understanding of these transduction mechanisms is a challenging perspective.

(iii) Equilibrium versus nonequilibrium physics: An implicit, yet overlooked, assumption of the models of bistability introduced by Sneppen and co-workers in the context of epigenetics (Dodd *et al.*, 2007; Micheelsen *et al.*, 2010) is that the system is open and far from equilibrium. The nonequilibrium nature of their models lies in the asymmetry of the transitions: the recruitment of an unmarked nucleosome by a marked nucleosome does not affect the latter, in strong contrast with an equilibrium transition between two species. 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, and ionic pumps at the cell membrane. In our opinion this is a very challenging and timely topic for biology-oriented physicists.

ACKNOWLEDGMENTS

We thank all the members of the CNRS GDR 3536 for the stimulating discussions that inspired this work. Special thanks to Aurélien Bancaud, Philippe Bertrand, Pascal Carrivain, Giacomo Cavalli, Thierry Forné, Daniel Jost, and Cedric Vaillant for their help in the preparation of the manuscript. This work was funded by the French Institut National du Cancer, Grant No. INCa_5960 and by the French Agence Nationale de la Recherche, Grant No. ANR-13-BSV5-0010-03.

REFERENCES

Albert, B. *et al.*, 2013, J. Cell Biol. **202**, 201.

Alberts, B., D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, 2013, *Essential Cell Biology* (Garland Science, New York).

Allahverdi, A., R. Yang, N. Korolev, Y. Fan, C. A. Davey, C.-F. Liu, and L. Nordenskiöld, 2011, Nucleic Acids Res. **39**, 1680.

Allan, J., N. Harborne, D. C. Rau, and H. Gould, 1982, J. Cell Biol. **93**, 285.

Angel, A., J. Song, C. Dean, and M. Howard, 2011, Nature (London) **476**, 105.

Angelov, D., J. M. Vitolo, V. Mutskov, S. Dimitrov, and J. J. Hayes, 2001, Proc. Natl. Acad. Sci. U.S.A. **98**, 6599.

Aragon, L., E. Martinez-Perez, and M. Merkenschlager, 2013, Curr. Opin. Genet. Dev. **23**, 204.

Arya, G., and T. Schlick, 2006, Proc. Natl. Acad. Sci. U.S.A. **103**, 16236.

Arya, G., and T. Schlick, 2009, J. Phys. Chem. A **113**, 4045.

Azzaz, A. M., *et al.*, 2014, J. Biol. Chem. **289**, 6850.

Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, 2001, Proc. Natl. Acad. Sci. U.S.A. **98**, 10037.

Bancaud, A., *et al.*, 2007, Mol. Cell **27**, 135.

Bannister, A. J., and T. Kouzarides, 2011, Cell Res. **21**, 381.

Barbieri, M., M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi, 2012, Proc. Natl. Acad. Sci. U.S.A. **109**, 16173.

Barr, M., and E. Bertram, 1949, Nature (London) **163**, 676.

Becker, P. B., and W. Hrz, 2002, Annu. Rev. Biochem. **71**, 247.

Bellas, E., and C. S. Chen, 2014, Curr. Opin. Cell Biol. **31**, 92.

Belotserkovskii, B. P., S. M. Mirkin, and P. C. Hanawalt, 2013, Chem. Rev. **113**, 8620.

Benedetti, F., J. Dorier, and A. Stasiak, 2014, Nucleic Acids Res. **42**, 10425.

Bermudez, I., J. Garcia-Martinez, J. Perez-Ortin, and J. Roca, 2010, Nucleic Acids Res. **38**, e182.

Bertin, A., D. Durand, M. Renouard, F. Livolant, and S. Mangenot, 2007, Eur. Biophys. J. **36**, 1083.

Bertin, A., S. Mangenot, M. Renouard, D. Durand, and F. Livolant, 2007, Biophys. J. **93**, 3652.

Bertin, A., M. Renouard, J. S. Pedersen, F. Livolant, and D. Durand, 2007, Biophys. J. **92**, 2633.

Beuchle, D., G. Struhl, and J. Muller, 2001, Development **128**, 993.

Bi, X., Q. Yu, J. J. Sandmeier, and Y. Zou, 2004, Mol. Cell. Biol. **24**, 2118.

Bird, A., 2002, Genes Dev. **16**, 6.

Blackledge, N. P., *et al.*, 2014, Cell **157**, 1445.

Booth, M. J., G. Marsico, M. Bachman, D. Beraldi, and S. Balasubramanian, 2014, Nat. Chem. **6**, 435.

Boulé, J.-B., J. Mozziconacci, and C. Lavelle, 2015, J. Phys. Condens. Matter **27**, 033101.

Brockdorff, N., 2013, RNA **19**, 429.

Buck-Koehntop, B. A., and P.-A. Defossez, 2013, Epigenetics **8**, 131.

Buck-Koehntop, B. A., R. L. Stanfield, D. C. Ekiert, M. A. Martinez-Yamout, H. J. Dyson, I. A. Wilson, and P. E. Wright, 2012, Proc. Natl. Acad. Sci. U.S.A. **109**, 15229.

Bulut-Karslioglu, A., *et al.*, 2012, Nat. Struct. Mol. Biol. **19**, 1023.

Cabili, M. N., M. C. Dunagin, P. D. McClanahan, A. Biaesch, O. Padovan-Merhar, A. Regev, J. L. Rinn, and A. Raj, 2015, Genome Biol. **16**, 20.

Cantone, I., and A. G. Fisher, 2013, Nat. Struct. Mol. Biol. **20**, 282.

Caré, B. R., P. Carrivain, T. Forné, J.-M. Victor, and A. Lesne, 2014, Commun. Theor. Phys. **62**, 607.

Caré, B. R., P.-E. Emeriau, R. Cortini, and J.-M. Victor, 2015, Biophysics **2**, 517.

Cedar, H., and Y. Bergman, 2009, Nat. Rev. Genet. **10**, 295.

Chodavarapu, R. K., *et al.*, 2010, Nature (London) **466**, 388.

Choy, J. S., S. Wei, J. Y. Lee, S. Tan, S. Chu, and T.-H. Lee, 2010, J. Am. Chem. Soc. **132**, 1782.

Cokus, S. J., S. Feng, X. Zhang, Z. Chen, B. Merriman, C. D. Haudenschild, S. Pradhan, S. F. Nelson, M. Pellegrini, and S. E. Jacobsen, 2008, Nature (London) **452**, 215.

Cooper, S., *et al.*, 2014, Cell Rep. **7**, 1456.

Court, F., J. Miro, C. Braem, M. Lelay-Taha, A. Brisebarre, F. Atger, T. Gostan, M. Weber, G. Cathala, and T. Forné, 2011, Genome Biol. **12**, R42.

Cowley, M., and R. J. Oakey, 2010, Brief. Funct. Genom. **9**, 340.

Crepaldi, L., C. Policarpi, A. Coatti, W. T. Sherlock, B. C. Jongbloets, T. A. Down, and A. Riccio, 2013, PLoS Genet. **9**, e1003699.

Davey, C. A., D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond, 2002, J. Mol. Biol. **319**, 1097.

David-Rus, D., S. Mukhopadhyay, J. L. Lebowitz, and A. M. Sengupta, 2009, J. Theor. Biol. **258**, 112.

Dawson, M., and T. Kouzarides, 2012, Cell **150**, 12.

Dayarian, A., and A. M. Sengupta, 2013, Phys. Biol. **10**, 036005.

Dekker, J., K. Rippe, M. Dekker, and N. Kleckner, 2002, Science **295**, 1306.

de Napoles, M., *et al.*, 2004, Dev. Cell **7**, 663.

Di Croce, L., and K. Helin, 2013, Nat. Struct. Mol. Biol. **20**, 1147.

Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, 2012, Nature (London) **485**, 376.

Dodd, I. B., M. A. Micheelsen, K. Sneppen, and G. Thon, 2007, Cell **129**, 813.

Dodd, I. B., and K. Sneppen, 2011, J. Mol. Biol. **414**, 624.

Dolinsky, T. J., P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker, 2007, Nucleic Acids Res. **35**, W522.

Du, Z., Y. Zhao, and N. Li, 2009, Nucleic Acids Res. **37**, 6784.

Duan, Z., M. Andronescu, K. Schutz, S. McIlwain, Y. J. Kim, C. Lee, J. Shendure, S. Fields, C. A. Blau, and W. S. Noble, 2010, Nature (London) **465**, 363.

Engler, A. J., S. Sen, H. L. Sweeney, and D. E. Discher, 2006, Cell **126**, 677.

Erdel, F., K. Müller-Ott, and K. Rippe, 2013, Ann. N.Y. Acad. Sci. **1305**, 29.

Everaers, R., and H. Schiessel, 2015, J. Phys. Condens. Matter **27**, 060301.

Feinberg, A., and B. Tycko, 2004, Nat. Rev. Cancer **4**, 143.

Felle, M., H. Hoffmeister, J. Rothammer, A. Fuchs, J. H. Exler, and G. Längst, 2011, Nucleic Acids Res. **39**, 6956.

Fierz, B., 2014, ChemMedChem **9**, 495.

Filion, G. J., *et al.*, 2010, Cell **143**, 212.

Friedman, N., L. Cai, and X. S. Xie, 2006, Phys. Rev. Lett. **97**, 168302.

Gilbert, S. F., 2012, Int. J. Epidemiology **41**, 20.

Giorgetti, L., R. Galupa, E. P. Nora, T. Piolot, F. Lam, J. Dekker, G. Tiana, and E. Heard, 2014, Cell **157**, 950.

Gosalia, N., D. Neems, J. L. Kerschner, S. T. Kosak, and A. Harris, 2014, Nucleic Acids Res. **42**, 9612.

Graves, J. A. M., 1982, Exp. Cell Res. **141**, 99.

Grewal, S. I., and S. C. Elgin, 2002, Curr. Opin. Genet. Dev. **12**, 178.

Grewal, S. I., and A. J. Klar, 1996, Cell **86**, 95.

Grewal, S. I., and D. Moazed, 2003, Science **301**, 798.

Grewal, S. I., and S. Jia, 2007, Nat. Rev. Genet. **8**, 35.

Grosberg, A. Y., S. K. Nechaev, and E. I. Shakhnovich, 1988, J. Phys. (Orsay, Fr.) **49**, 2095.

Haig, D., 2004, in *Cold Spring Harbor Symposia in Quantitative Biology*, Vol. 69 (Cold Spring Harbor Laboratory Press, New York), pp. 67–70.

Hajjoul, H., *et al.*, 2013, Genome Res. **23**, 1829.

Hansen, J. C., 2002, Annu. Rev. Biophys. Biomol. Struct. **31**, 361.

Hathaway, N. A., O. Bell, C. Hodges, E. L. Miller, D. S. Neel, and G. R. Crabtree, 2012, Cell **149**, 1447.

Hirano, T., 2014, Trends Cell Biol. **24**, 727.
Hizume, K., S. Araki, K. Hata, E. Prieto, T. K. Kundu, K. Yoshikawa, and K. Takeyasu, 2010, Archives of Histology and Cytology **73**, 149.
Hodges, C., and G. R. Crabtree, 2012, Proc. Natl. Acad. Sci. U.S.A. **109**, 13296.
Hodges-Garcia, Y., and P. J. Hagerman, 1995, J. Biol. Chem. **270**, 197.
Holliday, R., 1994, Dev. Genet. **15**, 453.
Howell, S. C., K. Andresen, I. Jimenez-Useche, C. Yuan, and X. Qiu, 2013, Biophys. J. **105**, 194.
Huet, S., C. Lavelle, H. Ranchon, P. Carrivain, J.-M. Victor, and A. Bancaud , 2014, Int. Rev. Cell Mol. Biol. **307**, 443.
Humphrey, W., A. Dalke, and K. Schulten, 1996, J. Mol. Graphics **14**, 33.
Imbert, J., A. Lesne, and J. Victor, 1997, Phys. Rev. E **56**, 5630.
Jackson, D. A., A. B. Hassan, R. J. Errington, and P. R. Cook, 1993, EMBO J. **12**, 1059.
Jackson, D. A., F. J. Iborra, E. M. Manders, and P. R. Cook, 1998, Mol. Biol. Cell **9**, 1523.
Jenuwein, T., and C. D. Allis, 2001, Science **293**, 1074.
Jones, P. A., 2012, Nat. Rev. Genet. **13**, 484.
Joshi, R. S., B. Piña, and J. Roca, 2010, EMBO J. **29**, 740.
Jost, D., 2014, Phys. Rev. E **89**, 010701.
Jost, D., P. Carrivain, G. Cavalli, and C. Vaillant, 2014, Nucleic Acids Res. **42**, 9553.
Kalashnikova, A. A., M. E. Porter-Goff, U. M. Muthurajan, K. Luger, and J. C. Hansen, 2013, J. R. Soc. Interface **10**, 20121022.
Kan, P.-Y., T. L. Caterino, and J. J. Hayes, 2009, Mol. Cell. Biol. **29**, 538.
Kaur, P., B. Plochberger, P. Costa, S. M. Cope, S. M. Vaiana, and S. Lindsay, 2012, Phys. Biol. **9**, 065001.
Keung, A. J., J. K. Joung, A. S. Khalil, and J. J. Collins, 2015, Nat. Rev. Genet. **16**, 159.
Khraiwesh, B., M. A. Arif, G. I. Seumel, S. Ossowski, D. Weigel, R. Reski, and W. Frank, 2010, Cell **140**, 111.
Kouzarides, T., 2007, Cell **128**, 693.
Kouzine, F., A. Gupta, L. Baranello, D. Wojtowicz, K. Ben-Aissa, J. Liu, T. Przytycka, and D. Levens, 2013, Nat. Struct. Mol. Biol. **20**, 396.
Kouzine, F., J. Liu, S. Sanford, H.-J. Chung, and D. Levens, 2004, Nat. Struct. Mol. Biol. **11**, 1092.
Kouzine, F., S. Sanford, Z. Elisha-Feil, and D. Levens, 2008, Nat. Struct. Mol. Biol. **15**, 146.
Kouzine, F., *et al.*, 2013, Cell **153**, 988.
Kriaucionis, S., and N. Heintz, 2009, Science **324**, 929.
Kühn, S., and J.-H. S. Hofmeyr, 2014, Biosemiotics **7**, 203.
Lang, A. H., H. Li, J. J. Collins, and P. Mehta, 2014, PLoS Comput. Biol. **10**, e1003734.
Langecker, M., A. Ivankin, S. Carson, S. Kinney, F. Simmel, and M. Wanunu, 2015, Nano Lett. **15**, 783.
Lavelle, C., 2014, Curr. Opin. Genet. Dev. **25**, 74.
Lazarovici, A., *et al.*, 2013, Proc. Natl. Acad. Sci. U.S.A. **110**, 6376.
Leblond, C. P., and M. El-Alfy, 1998, Anat. Rec. **252**, 426.
Lee, J. T., 2012, Science **338**, 1435.
Lee, J. Y., and T.-H. Lee, 2012, J. Am. Chem. Soc. **134**, 173.
Lennartsson, A., and K. Ekwall, 2009, Biochim. Biophys. Acta, Mol. Basis Dis. **1790**, 863.
Lesne, A., 2006, Eur. Phys. J. E **19**, 375.
Lesne, A., N. Foray, G. Cathala, T. Forné, H. Wong, and J.-M. Victor, 2015, J. Phys. Condens. Matter **27**, 064114.
Lesne, A., J. Riposo, P. Roger, A. Cournac, and J. Mozziconacci, 2014, Nat. Methods **11**, 1141.
Li, E., C. Beard, and R. Jaenisch, 1993, Nature (London) **366**, 362.
Li, S., and M. A. Shogren-Knaak, 2008, Proc. Natl. Acad. Sci. U.S.A. **105**, 18243.
Lieberman-Aiden, E., *et al.*, 2009, Science **326**, 289.
Lim, H. N., and A. Van Oudenaarden, 2007, Nat. Genet. **39**, 269.
Lister, R., *et al.*, 2009, Nature (London) **462**, 315.
Liu, F., E. Tøstesen, J. K. Sundet, T.-K. Jenssen, C. Bock, G. I. Jerstad, W. G. Thilly, and E. Hovig, 2007, PLoS Comput. Biol. **3**, e93.
Liu, J., F. Kouzine, Z. Nie, H.-J. Chung, Z. Elisha-Feil, A. Weber, K. Zhao, and D. Levens, 2006, EMBO J. **25**, 2119.
Liu, Y., C. Lu, Y. Yang, Y. Fan, R. Yang, C.-F. Liu, N. Korolev, and L. Nordenskiöld, 2011, J. Mol. Biol. **414**, 749.
Liu, Y., Y. O. Olanrewaju, Y. Zheng, H. Hashimoto, R. M. Blumenthal, X. Zhang, and X. Cheng, 2014, Nucleic Acids Res. **42**, 4859.
Liu, Y., X. Zhang, R. M. Blumenthal, and X. Cheng, 2013, Trends Biochem. Sci. **38**, 177.
Luger, K., M. L. Dechassa, and D. J. Tremethick, 2012, Nat. Rev. Mol. Cell Biol. **13**, 436.
Luger, K., A. W. Mder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, 1997, Nature (London) **389**, 251.
Luo, J., M. Li, Y. Tang, M. Laszkowska, R. G. Roeder, and W. Gu, 2004, Proc. Natl. Acad. Sci. U.S.A. **101**, 2259.
MacAlpine, D. M., and G. Almouzni, 2013, Cold Spring Harbor Persp. Biol. **5**, a010207.
Machado, A. C. D., T. Zhou, S. Rao, P. Goel, C. Rastogi, A. Lazarovici, H. J. Bussemaker, and R. Rohs, 2015, Brief. Funct. Genom. **14**, 61.
Maizels, N., and L. T. Gray, 2013, PLoS Genet. **9**, e1003468.
Marchal, C., and B. Miotto, 2015, J. Cell. Physiol. **230**, 743.
Marchese, F. P., and M. Huarte, 2014, Epigenetics **9**, 21.
Mayer-Jung, C., D. Moras, and Y. Timsit, 1998, EMBO J. **17**, 2709.
McGhee, J. D., and G. D. Ginder, 1979, Nature (London) **280**, 419.
Micheelsen, A., N. Mitarai, K. Sneppen, and I. Dodd, 2010, Phys. Biol. **7**, 026010.
Mirny, L. A., 2011, Chromosome Research **19**, 37.
Mohandas, T., R. Sparkes, and L. Shapiro, 1981, Science **211**, 393.
Morange, M., 2013, J. Biosci. **38**, 451.
Mühlbacher, F., C. Holm, and H. Schiessel, 2006, Europhys. Lett. **73**, 135.
Mukhopadhyay, S., and A. M. Sengupta, 2013, PLoS Comput. Biol. **9**, e1003121.
Müller-Ott, K., *et al.*, 2014, Mol. Syst. Biol. **10**, 746.
Mulligan, P. J., E. F. Koslover, and A. J. Spakowitz, 2015, J. Phys. Condens. Matter **27**, 064109.
Nanney, D., 1958, Proc. Natl. Acad. Sci. U.S.A. **44**, 712.
Nathan, D., and D. M. Crothers, 2002, J. Mol. Biol. **316**, 7.
Naughton, C., N. Avlonitis, S. Corless, J. Prendergast, I. Mati, P. Eijk, S. Cockroft, M. Bradley, B. Ylstra, and N. Gilbert, 2013, Nat. Struct. Mol. Biol. **20**, 387.
Ngo, T. T. M., J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev, and T. Ha, 2016, Nat. Commun. **7**, 10813.
Nora, E. P., *et al.*, 2012, Nature (London) **485**, 381.
Norregaard, K., M. Andersson, K. Sneppen, P. E. Nielsen, S. Brown, and L. B. Oddershede, 2013, Proc. Natl. Acad. Sci. U.S.A. **110**, 17386.
North, J. A., M. Šimon, M. B. Ferdinand, M. A. Shoffner, J. W. Picking, C. J. Howard, A. M. Mooney, J. van Noort, M. G. Poirier, and J. J. Ottesen, 2014, Nucleic Acids Res. **42**, 4922.
Okamoto, I., A. P. Otte, C. D. Allis, D. Reinberg, and E. Heard, 2004, Science **303**, 644.
Olins, D. E., and A. L. Olins, 2003, Nat. Rev. Mol. Cell Biol. **4**, 809.
Ong, C.-T., and V. G. Corces, 2014, Nat. Rev. Genet. **15**, 234.
Osella, M., A. Riba, A. Testori, D. Corà, and M. Caselle, 2014, Front. Genet. **5**.

Padinhateeri, R., and J. F. Marko, 2011, Proc. Natl. Acad. Sci. U.S.A. **108**, 7799.
Paro, R., H. Strutt, and G. Cavalli, 1998, in *Novartis Foundation Symposium 214-Epigenetics* (Wiley Online Library), pp. 51–66.
Pepenella, S., K. J. Murphy, and J. J. Hayes, 2014, Chromosoma **123**, 3.
Pérez, A., *et al.*, 2012, Biophys. J. **102**, 2140.
Poirier, M. G., E. Oh, H. S. Tims, and J. Widom, 2009, Nat. Struct. Mol. Biol. **16**, 938.
Pontier, D. B., and J. Gribnau, 2011, Hum. Genet. **130**, 223.
Portella, G., F. Battistini, and M. Orozco, 2013, PLoS Comput. Biol. **9**, e1003354.
Potoyan, D. A., and G. A. Papoian, 2012, Proc. Natl. Acad. Sci. U.S.A. **109**, 17857.
Ramdas, N. M., and G. Shivashankar, 2015, J. Mol. Biol. **427**, 695.
Rando, O. J., 2012, Curr. Opin. Genet. Dev. **22**, 148.
Razin, A., and H. Cedar, 1991, Microbiol. Rev. **55**, 451.
Razin, A., and H. Cedar, 1994, Cell **77**, 473.
Recouvreux, P., C. Lavelle, M. Barbi, N. C. e Silva, E. Le Cam, J.-M. Victor, and J.-L. Viovy, 2011, Biophys. J. **100**, 2726.
Ringrose, L., and R. Paro, 2004, Annu. Rev. Genet. **38**, 413.
Rivera, C. M., and B. Ren, 2013, Cell **155**, 39.
Rozijn, T. H., and G. Tonino, 1964, Biochim. Biophys. Acta **91**, 105.
Saberi, S., P. Farré, O. Cuvier, and E. Emberly, 2015, BMC Bioinf. **16**, 171.
Sadakierska-Chudy, A., and M. Filip, 2015, Neurotoxic. Res. **27**, 172.
Sapountzi, V., and J. Ct, 2011, Cell Mol. Life Sci. **68**, 1147.
Sasai, N., M. Nakao, and P.-A. Defossez, 2010, Nucleic Acids Res. **38**, 5015.
Satake, A., and Y. Iwasa, 2012, J. Theor. Biol. **302**, 6.
Schübeler, D., 2015, Nature (London) **517**, 321.
Sedighi, M., and A. M. Sengupta, 2007, Phys. Biol. **4**, 246.
Severin, P. M., X. Zou, H. E. Gaub, and K. Schulten, 2011, Nucleic Acids Res. **39**, 8740.
Sexton, T., E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli, 2012, Cell **148**, 458.
Shahbazian, M. D., and M. Grunstein, 2007, Annu. Rev. Biochem. **76**, 75.
Shaytan, A. K., D. Landsman, and A. R. Panchenko, 2015, Curr. Opin. Struct. Biol. **32**, 48.
Sheinin, M. Y., M. Li, M. Soltani, K. Luger, and M. D. Wang, 2013, Nat. Commun. **4**, 2579.
Shivashankar, G., 2011, Annu. Rev. Biophys. **40**, 361.
Shogren-Knaak, M., H. Ishii, J.-M. Sun, M. J. Pazin, J. R. Davie, and C. L. Peterson, 2006, Science **311**, 844.
Simon, J. A., and R. E. Kingston, 2013, Mol. Cell **49**, 808.
Sinha, D., and M. A. Shogren-Knaak, 2010, J. Biol. Chem. **285**, 16572.
Song, C.-X., *et al.*, 2013, Cell **153**, 678.
Song, J., J. Irwin, and C. Dean, 2013, Curr. Biol. **23**, R807.
Spruijt, C. G., and M. Vermeulen, 2014, Nat. Struct. Mol. Biol. **21**, 949.
Struhl, K., 1999, Cell **98**, 1.
Taylor, G. C., R. Eskeland, B. Hekimoglu-Balkan, M. M. Pradeepa, and W. A. Bickmore, 2013, Genome Res. **23**, 2053.
Teif, V. B., N. Kepper, K. Yserentant, G. Wedemann, and K. Rippe, 2015, J. Phys. Condens. Matter **27**, 064110.
Terweij, M., and F. van Leeuwen, 2013, Front. Life Sci. **7**, 63.
Tessarz, P., and T. Kouzarides, 2014, Nat. Rev. Mol. Cell Biol. **15**, 703.
Teves, S. S., and S. Henikoff, 2014, Nat. Struct. Mol. Biol. **21**, 88.
Thon, G., and T. Friis, 1997, Genetics **145**, 685.
Tie, F., R. Banerjee, C. A. Stratton, J. Prasad-Sinha, V. Stepanik, A. Zlobin, M. O. Diaz, P. C. Scacheri, and P. J. Harte, 2009, Development (Cambridge, U.K.) **136**, 3131.
Timoshenko, E., Y. A. Kuznetsov, and K. Dawson, 1998, Phys. Rev. E **57**, 6801.
Timp, W., and A. Feinberg, 2013, Nat. Rev. Cancer **13**, 497.
Tjong, H., K. Gong, L. Chen, and F. Alber, 2012, Genome Res. **22**, 1295.
Travers, A., and G. Muskhelishvili, 2007, EMBO Rep. **8**, 147.
Tropberger, P., and R. Schneider, 2013, Nat. Struct. Mol. Biol. **20**, 657.
Tumbar, T., G. Sudlow, and A. S. Belmont, 1999, J. Cell Biol. **145**, 1341.
Venkatasubrahmanyam, S., W. W. Hwang, M. D. Meneghini, A. H. Y. Tong, and H. D. Madhani, 2007, Proc. Natl. Acad. Sci. U.S.A. **104**, 16609.
Venolia, L., S. Gartler, E. Wassman, P. Yen, T. Mohandas, and L. Shapiro, 1982, Proc. Natl. Acad. Sci. U.S.A. **79**, 2352.
Vinograd, J., J. Lebowitz, R. Radloff, R. Watson, and P. Laipis, 1965, Proc. Natl. Acad. Sci. U.S.A. **53**, 1104.
Vlijm, R., M. Lee, J. Lipfert, A. Lusser, C. Dekker, and N. H. Dekker, 2015, Cell Rep. **10**, 216.
Waddington, C. H., 1942, Endeavor **1**, 18 [reprinted in the Intl. J. Epidemiology **41**, 10 (2012)].
Waddington, C. H., 1957, *The Strategy of the Genes* (Allen, London).
Wang, J. C., 1971, J. Mol. Biol. **55**, 523.
Wang, J. C., 2002, Nat. Rev. Mol. Cell Biol. **3**, 430.
Wang, X., and J. J. Hayes, 2008, Mol. Cell. Biol. **28**, 227.
Wang, Z., H. Yao, S. Lin, X. Zhu, Z. Shen, G. Lu, W. S. Poon, D. Xie, M. C. mi Lin, and H. fu Kung, 2013, Cancer Letters **331**, 1.
Wigler, M., D. Levy, and M. Perucho, 1981, Cell **24**, 33.
Wilkins, B. J., N. A. Rall, Y. Ostwal, T. Kruitwagen, K. Hiragami-Hamada, M. Winkler, Y. Barral, W. Fischle, and H. Neumann, 2014, Science **343**, 77.
Wolffe, A. P., and J. J. Hayes, 1999, Nucleic Acids Res. **27**, 711.
Wong, H., H. Marie-Nelly, S. Herbert, P. Carrivain, H. Blanc, R. Koszul, E. Fabre, and C. Zimmer, 2012, Curr. Biol. **22**, 1881.
Xu, E. Y., K. A. Zawadzki, and J. R. Broach, 2006, Mol. Cell **23**, 219.
Yang, X.-J., and E. Seto, 2007, Oncogene **26**, 5310.
Yang, Y., A. P. Lyubartsev, N. Korolev, and L. Nordenskiöld, 2009, Biophys. J. **96**, 2082.
Yang, Z., and J. J. Hayes, 2011, Biochemistry **50**, 9973.
Yoo, J., H. Kim, A. Aksimentiev, and T. Ha, 2016, Nat. Commun. **7**, 11045.
Yusufaly, T. I., Y. Li, and W. K. Olson, 2013, J. Phys. Chem. B **117**, 16436.
Zentner, G. E., and S. Henikoff, 2013, Nat. Struct. Mol. Biol. **20**, 259.
Zhang, B., and P. G. Wolynes, 2014, Proc. Natl. Acad. Sci. U.S.A. **111**, 10185.
Zhang, H., X.-J. Tian, A. Mukhopadhyay, K. Kim, and J. Xing, 2014, Phys. Rev. Lett. **112**, 068101.
Zhou, J., J. Y. Fan, D. Rangasamy, and D. J. Tremethick, 2007, Nat. Struct. Mol. Biol. **14**, 1070.
Zou, X., W. Ma, I. A. Solov'yov, C. Chipot, and K. Schulten, 2012, Nucleic Acids Res. **40**, 2747.