

# Chromatin Computation: Epigenetic Inheritance as a Pattern Reconstruction Problem

Christian Arnold<sup>a,b,d,1</sup>, Peter F. Stadler<sup>b,c,h,e,f,g</sup>, Sonja J. Prohaska<sup>a,b</sup>

<sup>a</sup>Computational EvoDevo Group, Department of Computer Science, Universität Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany

<sup>b</sup>Interdisciplinary Center for Bioinformatics, Universität Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany

<sup>c</sup>Bioinformatics Group, Department of Computer Science, Universität Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany

<sup>d</sup>Harvard University, Department of Human Evolutionary Biology, 11 Divinity Avenue, Cambridge MA 02138, USA

<sup>e</sup>Max-Planck-Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany

<sup>f</sup>Fraunhofer Institut für Zelltherapie und Immunologie Perlickstraße 1, 04103 Leipzig, Germany

<sup>g</sup>Department of Theoretical Chemistry University of Vienna, Währingerstraße 17, 1090 Wien, Austria

<sup>h</sup>Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe NM 87501, USA

---

## Abstract

Eukaryotic histones carry a diverse set of specific chemical modifications that accumulate over the life-time of a cell and have a crucial impact on the cell state in general and the transcriptional program in particular. Replication constitutes a dramatic disruption of the chromatin states that effectively amounts to partial erasure of stored information. To preserve its epigenetic state the cell reconstructs (at least part of) the histone modifications by means of processes that are still very poorly understood. A plausible hypothesis is that the different combinations of reader and writer domains in histone-modifying enzymes implement local rewriting rules that are capable of “recomputing” the desired parental modification patterns on the basis of the partial information contained in that half of the nucleosomes that predate replication.

To test whether such a mechanism is theoretically feasible, we have developed a flexible stochastic simulation system (available at <http://www.bioinf.uni-leipzig.de/Software/StoChDyn>) for studying the dynamics of histone modification states. The implementation is based on Gillespie’s approach, i.e., it models the master equation of a detailed chemical model. It is efficient enough to use an evolutionary algorithm to find patterns across multiple cell divisions with high accuracy.

We found that it is easy to evolve a system of enzymes that can maintain a particular chromatin state roughly stable, even without explicit boundary elements separating differentially modified chromatin domains. However, the success of this task depends on several previously unanticipated factors, such as the length of the initial state, the specific pattern that should be maintained, the time between replications, and chemical parameters such as enzymatic binding and dissociation rates. All these factors also influence the accumulation of errors in the wake of cell divisions.

**Keywords:** chromatin complexity, Gillespie algorithm, evolutionary algorithm, epigenetic inheritance, histone modifications, chromatin computation

---

## 1. Background

Eukaryotic genomes are organized as chromatin in the nucleus of the cell. With the notable exception of dinoflagellates (Herzog and Soyer, 1981), the basic structure of chromatin consists of histones with DNA wrapped around them, together referred to as nucleo-

some. Chromatin controls DNA accessibility and contributes to the recruitment of transcription factors. Eukaryotes have evolved a complex system of chemical modifications of histone proteins, which play an important role in gene regulation and cellular differentiation. Therefore, chromatin can be seen as an advanced signaling module (Turner, 2012) that is a critical responder to external cues such as stress (Smith and Workman, 2012). It also carries a partial “annotation” of genomic features such as promoters and enhancers (Heintzman et al., 2007), the exon/intron struc-

---

*Email addresses:* [achristian@bioinf.uni-leipzig.de](mailto:achristian@bioinf.uni-leipzig.de) (Christian Arnold), [studla@bioinf.uni-leipzig.de](mailto:studla@bioinf.uni-leipzig.de) (Peter F. Stadler), [sonja@bioinf.uni-leipzig.de](mailto:sonja@bioinf.uni-leipzig.de) (Sonja J. Prohaska)

ture (Andersson et al., 2009; Schwartz et al., 2009) or the concurrent deposition of characteristic histone modifications (e.g. H3K4me3 as marker of active transcription (Young et al., 2011; Tippmann et al., 2012)), altogether keeping track of ongoing activities.

Mounting evidence suggests that histone modifications co-occur so that function may be determined by combinatorial patterns of modifications that form recurrent and spatially coherent combinations (Ernst and Kellis, 2010). While some of these chromatin states generically mark functional elements like promoters, others are directly associated with cell-type-specific gene expression programs (Heintzman et al., 2009; Gacek and Strauss, 2012) and other complex processes such as mitosis and meiosis (Xu et al., 2009), DNA repair (Miller and Jackson, 2012), alternative splicing (Luco et al., 2011; Ameyar-Zazoua et al., 2012), and pre-mRNA processing (Brown et al., 2012). The biochemical basis for the complexity of histone modification patterns (Jenuwein and Allis, 2001; Gardner et al., 2011) is provided by the large array of proteins capable of recognizing two or even more histone modifications (e.g., see Wang and Patel, 2011).

It is now debated whether the term “histone code” is appropriate to describe the system of histone modifications (e.g., see Rando, 2012; Turner, 2012). Clearly, there is no simple set of rewriting rules that would translate, akin to the genetic code, a particular modification or a combination of modifications from a single nucleosome into an immutable, well-defined, and context-independent biological function. Instead, the “meaning” of histone modifications appears to be dependent on their context with emphasis on the enzymatic machinery available to access the encoded information in a particular cell type and cell state.

The coupling of reading and writing of histone modifications, which is a common feature of many chromatin-modifying proteins especially in crown-group Eukaryotes, may have converted chromatin into a powerful computational device capable of storing and processing large amounts of information (Prohaska et al., 2010). A recent theoretical study showed that a simple model of chromatin computation, very similar to that proposed in Prohaska et al. (2010), is computationally universal and hence conceptually more powerful than the logic circuits of cis-regulatory networks (Bryant, 2012). Although it is plausible that the computational capacities of chromatin play a role in the integration of external environmental signals and internal status information—and hence in cell-fate decision—these computational aspects have remained largely unexplored so far. It is thus still unclear to what extent the

potential power of chromatin computation is harnessed in real biological systems.

The terms “epigenetic inheritance” and “epigenetic memory” generally designate the capability of cells to transmit alternative gene expression programs and cellular phenotypes across cell divisions without altering the underlying DNA sequence. The carrier of this heritable information is still subject to intense research (Petruk et al., 2012). The most plausible candidates are DNA methylation (for which a faithful copying mechanism is known), histone modifications, histone variants, histone-modifying enzymes, or the general expression state of the cell. For example, it has been argued that Trithorax and Polycomb group proteins play a pivotal role in epigenetic inheritance for certain modifications (e.g., H3K4 or H3K27 methylation) and may in fact be the true epigenetic carriers rather than the histone modifications (Petruk et al., 2012). However, recent work in *Drosophila* indicates the opposite, namely that H3K27 (and its various modifications, particularly methylation) is indeed the crucial substrate for heterochromatin formation via Polycomb proteins (Pengelly et al., 2013).

The inheritance of histone modifications is not a trivial achievement since replication is associated with the partial replacement of histones and the deposition of newly assembled and hence unmodified histones (Anunziato, 2005; Radman-Livaja et al., 2011). In other words, replication and the subsequent re-constitution of chromatin constitutes a dramatic disruption of the chromatin states that amounts to a partial erasure of the information stored in histone modifications.

Here, we propose that the reconstitution of local histone modification patterns is one of the biologically important computational tasks that is naturally solved by the “chromatin computer”. The need to propagate epigenetic information to subsequent generations comes in two variants. The more stringent version concerns stable heritable bistability in which epigenetic information can be transmitted for, in principle, an infinite number of generations, depending on the strength of the bistability of the underlying system. As shown by Dodd et al. (2007), this requires cooperative, positive feedback recruitment reactions as well as non-local interactions. A recent review of formal modeling approaches to the dynamics and propagation of histone modifications has been provided by Rohlf et al. (2012). In contrast, the maintenance of local patterns of histone modifications over a limited number of somatic cell divisions can potentially violate the conditions for long-term stability and tolerate slow accumulation of errors. In this setting it makes sense, therefore, to dispense with the stringent requirements outlined by Dodd et al. (2007). Indeed,

it appears that this less stringent version is the relevant mechanism in multi-cellular organisms, as cells only replicate a limited number of times (the Hayflick limit). Furthermore, it has been reported repeatedly that particular epigenetic modifications can be gradually changed over generations through a number of different settings. In particular, these include the progressive reduction of higher histone methylation levels to lower methylation forms (Katan-Khaykovich and Struhl, 2005), epigenetic reprogramming (Jeong et al., 2007; Katz et al., 2009), epigenetic silencing / heterochromatin formation (Mutskov and Felsenfeld, 2003; Millar and Grunstein, 2006), and transcription-coupled histone modifications (Tippmann et al., 2012). Furthermore, histone modification gradients for a number of different modifications have been observed along a particular genomic region (for a review, see Henikoff and Shilatifard, 2011). This effect may also indirectly result from multiple cell division due to the preferential retention of parental histones at the 5' end of genes (Radman-Livaja et al., 2011). Gradual changes of histone modification levels have been implicated to play a crucial rule in aging (Przybylla et al., 2012; De Vos et al., 2011). DNA methylation changes, finally, are also intimately linked to histone modifications (Mutskov and Felsenfeld, 2003), and therefore may be a direct result of the dilution of one or more histone modifications.

One of the best-studied mechanisms proposed for epigenetic memory is based on positive feedback loops in nucleosome modification (Grunstein, 1998; Turner, 1998), suggesting that a cluster of nucleosomes may be able to stably maintain a particular state of modifications (Dodd et al., 2007; Dodd and Sneppen, 2011). Although this mechanism is not limited to silencing, histone modifications that establish and maintain repressive states (such as H3K9me3) have served as a model for this effect. Mounting evidence indicates that histone-modifying enzymes can have higher binding affinities for a particular location if neighboring nucleosomes carry the same or chemically similar modifications (Jacobson et al., 2000; Owen et al., 2000; Schotta et al., 2002; Margueron et al., 2009; Yun et al., 2011), in particular for H3K9me3 (Nakayama et al., 2001; Lachner et al., 2001; Bannister et al., 2001). This may subsequently facilitate localized spreading *in cis* and the (re-)establishment of chromatin domains. Further contributing to the complexity of how an epigenetic memory, long non-coding RNAs seem to have crucial roles in regulating gene expression by guiding chromatin-modifying enzyme complexes to particular genomic loci (Magistri et al., 2012; Lee, 2012; Mondal et al., 2010; Khalil et al., 2009). An increasing num-

ber of other molecular pathways involving small non-coding RNAs or DNA methylation in concert with histone modifications have been described (Moazed, 2011) as well.

The coupled reading, writing, and erasing of histone modifications is therefore of crucial importance. Given the multiple tasks histone modifications are involved in, we hereafter focus on whether a simple “chromatin computer” is capable of solving the pattern completion problem for a diverse set of chromatin input states, despite the highly disruptive nature of frequent cell divisions. More to the point, we ask whether it is feasible to find combinations of reader/writer enzymes that are capable of propagating, with high accuracy, pre-set chromatin states across several cell divisions. To answer this question we implement a generic stochastic simulation of rule-based chromatin modifications as a model of the “chromatin computer”. We then employ an evolutionary algorithm to evolve “programs” representing mixtures of rewriting rules to solve various pattern reconstruction tasks.

## 2. Methods

### 2.1. A Coarse-Grained Chemical Model of Chromatin Computation

We begin by introducing our model of chromatin as a computational device. Computation abstractly consists of a system of states and transitions between them. Our intention is to stay close to a physical model of chromatin. Similar to much of the literature, we define a *chromatin state* as the set of chemical modifications of histone molecules (or their absence) located at specific genomic positions. Although this is a simplified view, the full biological complexity of the components that make up a particular chromatin state cannot be integrated at this time given that the exact underlying mechanisms are incompletely understood. Therefore, we disregard effects such as changes in nucleosome occupancy that can have an impact on gene expression (Tillo et al., 2010; Wang et al., 2011), the presence of histone variants, the effects of the three-dimensional structure of chromatin and nucleosomes (e.g. Andrews and Luger, 2011), chromatin remodeling events that increase the dynamic nature of chromatin (histone turnover, histone tail clipping, histone passback; see Santos-Rosa et al., 2008; Radman-Livaja et al., 2011), and DNA methylation as an epigenetic mechanism that is known to at least partly interact with histone modifications. We suggest that these details of the underlying “hardware” are not required for the investigation of the the capabilities

of the computational paradigm of the “chromatin computer” as a stochastic rewriting system.

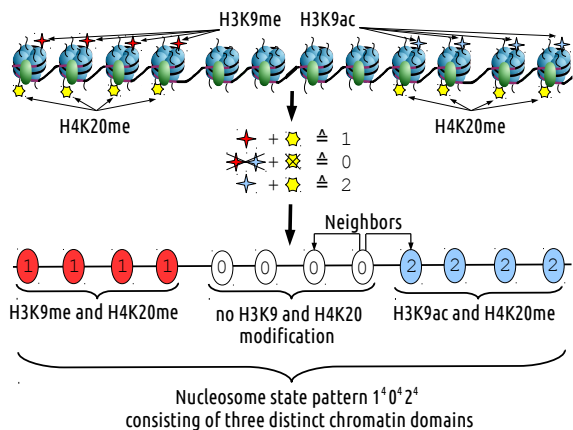


Figure 1: Illustration of nucleosomes, their corresponding states and terminology that will be used throughout this contribution. A genomic region composed of 12 nucleosomes is shown, with H3K9 and H4K20 histone modifications present at particular nucleosomes. For convenience, a translation table can then be used to assign combinations of histone modifications (or the lack thereof) to particular symbols. Each nucleosome can then be assigned one of three distinct states (chromatin states): 0 (white), 1 (red) or 2 (blue). Homogeneously modified regions form a chromatin domain that carries a particular signature (0: unmodified H3K9 and H4K20, 1: H3K9me and H4K20me, 2: H3K9ac and H4K20me). Collectively, these 12 nucleosomes represent the local nucleosome state pattern or, abbreviated, simply pattern  $1^4 0^4 2^4$ . Such patterns therefore usually consist of multiple distinct chromatin domains (see Table 3 for examples). The chromatin string uses a modified version of the Wikimedia Commons file “Nucleosome organization.png” (licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license).

These simplifications allow us to view chromatin as a linear sequence of  $n_n$  nucleosomes, analogous to the well-known “beads-on-a-string” picture of chromatin. Consequently, each nucleosome (except the two boundary nucleosomes) has two immediate adjacent neighboring nucleosomes or abbreviated simply “neighbors” (see also Figure 1). As each nucleosome is completely specified by the collection of histone modifications that it carries, it can be represented by a single character  $a \in \mathcal{A}$  that encodes its modification state (nucleosome or chromatin state). The symbol 0 is reserved for the unmodified state. As a technical simplification to save memory and to expedite the recognition of patterns in practical simulations, we use a single character instead of a string to represent the state of the more than 130 known human histone modifications that have been observed (Tan et al., 2011). We define a chromatin domain as a set of adjacent nucleosomes that are in the

same modification state  $a_i \in \mathcal{A}$ . Each chromatin domain has a particular characteristic signature (e.g. methylated H4K20 and acetylated H3K9 residues) and length  $l$  and therefore can be represented by a sequence of nucleosome states  $a^l$ . We call this a local nucleosome state pattern (or simply pattern) to emphasize that the chromatin state is solely determined by the nucleosome state. Naturally, such patterns may also be composed of multiple adjacent distinct chromatin domains (e.g., see Figure 1 and Table 3).

Transitions between chromatin states are mediated by histone-modifying enzymes. These enzymes catalyze the writing or erasing of histone modifications in a context-dependent manner. Therefore, they are implemented as string rewriting rules acting on the nucleosome state pattern. (Figure 2 A). They recognize parts of the (local) nucleosome state patterns and cause a change in the modification state of one or more nucleosomes. These rewriting rules are described in detail in the following section.

Not all chromatin-modifying enzymes are present in the cell at the same time or in same concentrations, and they may differ substantially in their affinity to their target patterns or in their catalytic efficiency. Furthermore, different enzymes may compete for the same chromatin locations and, *vice versa*, different chromatin positions may compete for low abundance enzymes (Figure 2 C). We therefore model the dynamics of the “rewriting reactions” in a way that follows the mass action kinetics of chemical reaction systems, distinguishing, for each enzyme and each chromatin position, their elementary reactions (Figure 2 B):

1. Binding of an enzyme to a specific locus (i.e., one or more nucleosomes). We assume that enzymes only bind to nucleosomes on the chromatin string that match the pre-condition of the rewriting rule that they embody. Also, enzymes cannot bind if any of the nucleosomes that are decisive for the applicability of the rewriting rule are bound by other enzymes, thereby blocking the accessibility.
2. Dissociation of an enzyme and the application of a rewriting rule (i.e., writing or erasing of one or multiple histone modifications). After dissociating, the enzyme is again available for new reactions.

Contrary to most previous work and simulation systems for chromatin state dynamics, the computational paradigm used here for the enzyme kinetics is a stochastic one. We argue that this level of chemical realism is crucial, since concentrations of regulatory molecules, rather than their mere presence or absence, are very well

known to be of crucial importance in the regulation of gene expression. Indeed, few regulatory events are qualitative – typically changes in expression levels of regulators are gradual and rarely exceeding a few-fold increases or decreases.

Chromatin state dynamics are thus dependent on enzyme abundances, the availability of local patterns on which they can act, the current state of the system, and the rate constants for each chemical reaction (which generally quantify the speed of a chemical reaction and may differ substantially among different enzymes) (Figure 2 C).

Histone modifications have strikingly different life-times and are deposited at different rates. Acetylation events are measured in the order of minutes, while methylation events are stable for days (Barth and Imhof, 2010). These rate differences are determined by the enzymes that catalyze the corresponding reactions (Hathaway et al., 2012). A given mark can be removed either by specific de-modification enzymes or through chromatin remodeling (e.g., histone turnover or histone tail clipping). Since chromatin remodeling phenomena are at present not explicitly included in our model, different life-times can be modelled by neighbor-independent rewriting rules with different rate constants. To our knowledge, spontaneous (i.e., enzyme-independent) decay has not yet been described for histone modifications, although it cannot be excluded that some of the more exotic or yet undescribed modifications may not require an enzyme for de-modification.

For simplicity, we use a single reaction rate parameter for the binding of histone-modifying enzymes or enzyme complexes, although mechanistically, this may require multiple steps (e.g., binding, recruitment of other factors and oligomerization). The propensity for a particular binding reaction is computed as the product of its reaction rate and the number of free (i.e., not bound) molecules for that enzyme, while for any dissociation reaction, the propensity equals its reaction rate and is therefore independent of the number of free molecules.

The time course of the simulation between replications can be subdivided into discrete “phases” (Figure 3) that can have varying durations. For each phase, enzymes can be arbitrarily set to be present or absent (which we will subsequently call enzyme availability) (Figure 3). However, enzyme concentrations either have a constant value as specified by the “chromatin computer’s” programming (present), or a value of 0 (absent). If an enzyme is still bound at the transition from one phase to the next, it dissociates without performing a state change at the nucleosome it binds to. Collectively, these phases abstract the cell’s gene expression

program. We may interpret them for instance as the G1, S, and G2 phase of the cell cycle. Alternatively, phases may be used to model distinct developmental stages.

We also include replication in our model. At regular time intervals, a replication takes place and the parental histones are distributed between the two daughter strands. Several models for histone segregation and deposition have been proposed in the last years, and due to the crucial importance of replication for any chromatin-based inheritance, a few of them will be presented in some detail in the following. The random or conservative model assumes that parental histones are deposited randomly between both strands while the gaps are filled with newly assembled and unmodified nucleosomes (Jackson and Chalkley, 1985; Corpet and Almouzni, 2009; Hansen et al., 2008). Thus, at any position in each of the strands, the probability that parental histones are incorporated is 50% (Figure 2 D). Alternatively, a semi-conservative replication mechanism in which parental nucleosomes are split into halves and distributed equally is theoretically possible (Margueron and Reinberg, 2010). Furthermore, asymmetric models suggest a non-random distribution of parental histones (e.g., to one particular strand only). Lastly, based on findings in *Drosophila*, Petruk et al. (2012) proposed a model in which only newly assembled histones are incorporated, but histone-modifying enzymes remain bound during replication or quickly rebind to restore the parental modification pattern.

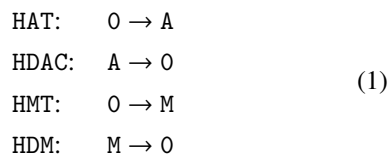
The model with the best experimental support is a random model in which nucleosomes disassemble into one (H3-H4)<sub>2</sub> tetramer and two H2A-H2B dimers that are randomly deposited (Annunziato, 2005; Zhu and Reinberg, 2011; Radman-Livaja et al., 2011). All other models are either purely hypothetical or only weakly supported. For example, a truly semi-conservative replication would require that, within one nucleosome, histones of the same type carry identical histone modifications, a condition that however is not always fulfilled (Voigt et al., 2012; Tran et al., 2012).

For the purpose of the present study, we therefore adopted a random model (Figure 2). We treat nucleosomes as indivisible units, which however is only a simplification if multiple histone modifications from different histones are modeled. We note, however, that the simulation environment can easily be extended to other replication models if the need arises. Again, we argue that this additional level of biological realism is irrelevant for the questions addressed in this contribution. We also assume that parental histones are re-deposited at their pre-replication locus, which is consistent with the finding that most parental histones in budding yeast

are re-incorporated in close vicinity, i.e. within 400 bp, of their original locus (Radman-Livaja et al., 2011). Lastly, analogous to the phase change transitions, enzymes that are still bound at the time of replication dissociate without performing any reaction.

## 2.2. Chromatin Enzymes as Rewriting Rules

In the simplest case, the enzymes evaluate the state of the nucleosome that they modify. The model of Dodd et al. (2007), for example, considers three distinct states: unmodified (O), methylated (M), and acetylated (A). Each state can be interconverted by the catalytic actions of histone acetyltransferases (HATs), histone deacetylases (HDACs), histone methyltransferases (HMTs), and histone demethylases (HDMs). The corresponding set of rewriting rules is

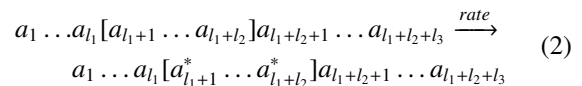


As opposed to these simple, neighbor-independent rewriting rules, more complex ones, such as the ones considered in Sneppen and Dodd (2012) or Bryant (2012), also depend on the neighboring nucleosomes (neighbor-dependent rewriting rules). In our implementation, arbitrarily complex rewriting rules can be specified (see Table 1 and Figure 2 for a few examples). For convenience, we allow wildcards in the definition of the rewriting rule and the state of the neighboring nucleosomes can be incorporated as well. Rules are interpreted as symmetric with respect to their left and right neighbors as it is not possible for the enzyme complexes to determine directionality (e.g. towards the centromere) from the local chromatin structure. In reality, it appears that asymmetries are induced locally and thus are oriented relative to features such as promoters, insulators, or transcription factor binding sites (Kundaje et al., 2012).

Such complex rewriting rules indeed seem to be common for eukaryotic systems, as histone-modifying enzymes are often part of large enzyme complexes with multiple protein domains, each of which having a particular function such as DNA sequence recognition or histone binding. Individual histone modifications in the vicinity of the binding may substantially alter the binding affinity of the enzyme complex or its formation. This crosstalk has been repeatedly demonstrated within histones, among histones of the same nucleosome (Wang and Patel, 2011; Lindroth et al., 2004;

Fuchs et al., 2011; Ruthenburg et al., 2011), and among histones of neighboring nucleosomes (for an overview, see Musselman et al., 2012). Even for single domain proteins, the presence of multiple modifications may be required for binding (Morinière et al., 2009).

We represent these complex rules as follows: Each rule is specified by a pair of strings, the left part is the template that must be matched to the local nucleosome state pattern. This precondition must be met so that the rewriting rule can be applied to the matched string. Once a match is found, the left string is replaced by the right string. In our implementation, both strings consist of up to three parts: (i) a string of length  $l_1 \geq 0$  for the required state(s) of the neighboring nucleosome(s) left of the actual binding site of the corresponding enzyme; (ii) a string of length  $l_2 > 0$  for the required state(s) of the nucleosome(s) that the enzyme binds to; (iii) a string of length  $l_3 \geq 0$  for the required state(s) of the neighboring nucleosome(s) right of the binding site. Only the nucleosomes of the binding part may be subject to change in the right part of the rewriting rule (as indicated by  $a^*$  in the equation below). Thus, rules are not reversible per se; instead, reversible reactions must be formulated as separate forward and backward reaction. Furthermore, each rewriting rule is associated with a rate constant. In summary, it is formally specified in the following form:



where each  $a_i \in \mathcal{A}$  denotes a particular nucleosome state. Table 4 compiles the rewriting rules with  $l_1, l_2, l_3 \leq 1$  that are used throughout this contribution.

The nucleosome string can be either linear or circular. For the former, rewriting rules that require the state of both neighboring nucleosomes cannot match the two boundary nucleosomes. This entails that they keep their parental status unless replication replaces them with unmodified nucleosomes, after which their original state is lost permanently (see Figure 5). In the latter case, the two boundary nucleosomes are directly connected, and rewriting rules may match.

## 2.3. Stochastic Simulation using the Gillespie Algorithm

Chromatin state dynamics are modeled using Gillespie's well-established stochastic simulation algorithm (SSA) (Gillespie, 1976, 1977). This approach reproduces the master equation of the underlying stochastic model with a discrete number of objects and continuous time. The algorithm is rejection-less and hence efficient in terms of the simulation, but it requires that possible

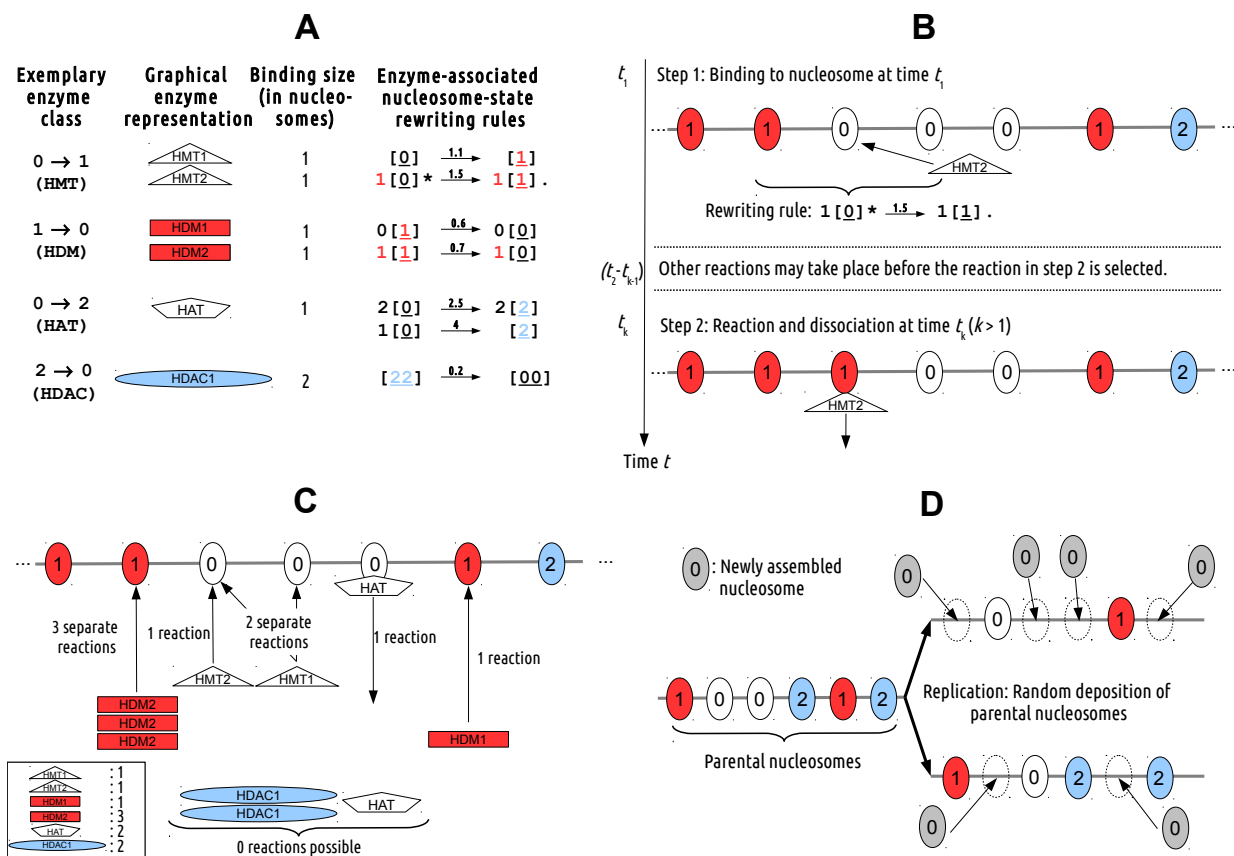


Figure 2: Basic ingredients of the chromatin model. In A-D, different aspects of the model are illustrated. An example of a genomic region is shown where nucleosomes can either be in state 0 (white), 1 (red) or 2 (blue).

(A) Enzymes and rewriting rules. A total of seven enzymes are defined that can be broadly divided into the following four well-known classes (for illustration purposes): histone methyltransferases (HMT), histone demethylases (HDM), histone acetyltransferases (HAT), and histone deacetylases (HDAC). Each enzyme recognizes and binds to a particular, local nucleosome state pattern (embraced by square brackets in the rewriting rules) and thereby blocks accessibility of these nucleosomes for other enzymes. Such recognition patterns may be dependent or independent of neighboring nucleosome states and can be represented as rewriting rules. They may proceed at different rates and change the state of at least one nucleosome.

(B) Enzymes and their reactions. Binding of an enzyme to one or more nucleosomes and the corresponding dissociation at a later time are modeled as separate reaction steps in our model. First, enzymes may bind to one or more nucleosomes as described by their rewriting rules. They then remain bound until the corresponding dissociation reaction is selected in the stochastic simulation. When the enzyme dissociates, the bound nucleosome(s) change their state(s), as specified by the corresponding rewriting rule(s).

(C) Enzyme competition and reaction dynamics. All eight possible reactions that can occur for various enzymes and their corresponding concentrations are shown (see legend), as well as an exemplary genomic region with seven nucleosomes. Each arrow indicates a possible reaction that may take place at this particular time point. If enzymes are bound to particular positions (e.g., HAT), they block the accessibility of other enzymes at bound nucleosomes. Note that particular enzymes can be present multiple times (e.g., HDM2) and that they may be able to bind to the same nucleosome (e.g., HMT1 and HMT2), resulting in competition for nucleosome binding. Other enzymes may not be able to perform any reaction due to the inapplicability of their rewriting rules given the current state of the nucleosomes (e.g., HDAC1 and one of the two HAT molecules).

(D) Replication and nucleosome segregation. The random distribution model of chromatin replication is shown. In this model, for each of the two strands and nucleosomal positions, the probability that the parental nucleosome is retained or re-incorporated, resulting in the preservation of the parental modification pattern, is 50%. The remaining 50% are filled with newly assembled, unmodified nucleosomes (random model).

“reaction channels” (i.e., the transitions between two particular states) are enumerated and finite. Each reaction channel is weighted with its reaction rate, which is crucial for exactness. In essence, given an initial time

$t_0$  and population state vectors (here: enzyme concentrations), the algorithm works as follows: In each iteration, two random numbers  $r_1$  and  $r_2$  are drawn from the uniform distribution in the unit interval, which are

Table 1: Example of different valid nucleosome state rewriting rules for a chromatin-based system with three possible nucleosome states  $\mathcal{A} = \{0, 1, 2\}$ . The symbol “\*” in the left part of the rewriting rule is a wildcard matching any nucleosome state, while “.” on the right part of a rewriting rule means that the nucleosome state is left unchanged. The nucleosomes bound and rewritten by the corresponding enzyme are embraced by square brackets.

Rewriting rule	Matching patterns for the rewriting rule	Neighbor-dependent
$[1] \xrightarrow{r_1} [0]$	$[1] \xrightarrow{r_1} [0]$	no
$[(1 2)] \xrightarrow{r_2} [0]$	$[1] \xrightarrow{r_2} [0], [2] \xrightarrow{r_2} [0]$	no
$1[1] \xrightarrow{r_3} 1[0]$	$1[1] \xrightarrow{r_3} 1[0]$	yes
$1[1] * 0 \xrightarrow{r_4} 1[0].0$	$1[1]00 \xrightarrow{r_4} 1[0]00, 1[1]10 \xrightarrow{r_4} 1[0]10, 1[1]20 \xrightarrow{r_4} 1[0]20$	yes
$[11] \xrightarrow{r_5} [01]$	$[11] \xrightarrow{r_5} [01]$	no
$1[11]0 \xrightarrow{r_6} 1[10]0$	$1[11]0 \xrightarrow{r_6} 1[10]0$	yes

subsequently used to determine the value of the time increment  $\tau$  and to select a reaction channel, respectively. Importantly, time increments are not fixed but are of variable length, depending on the current state of the system. The selected reaction channel is then executed and the state of the system changes accordingly. This process is repeated until a stop criterion has been reached (e.g., a specified simulation time). As this algorithm is well established, we refer to Gillespie (1976, 1977) for more details.

Each replication event occurs periodically after a fixed time interval  $t_r$ . Similarly, each phase  $p_i$  has a defined duration and stops after a particular time  $t_n$ . This requires a correction for the last reaction event in each period—i.e., the one for which  $t + \tau > t_r$  and  $t + \tau > t_n$ , respectively. Here, we draw a random number  $r_3$  and accept the reaction if  $r_3 > (t_r - t)/\tau$  and  $r_3 > (t_n - t)/\tau$ , respectively.

Since the set of available reaction channels and their weights are, in general, dependent on the current state of the system, the “book-keeping” of all reaction channels and their status is an important issue for the practical implementation of the Gillespie algorithm. This is particularly relevant for models with large numbers of different molecules and reactions. We next describe the peculiarities and design decisions that are specific to our chromatin-based model. To do so, we make use of abbreviations for relevant parameters that are summarized in Table 2. Consider a system with  $n_e$  enzymes, each of which has a particular number of rewriting rules. The sum of all rewriting rules is then  $n_{\text{rules}} = \sum_{k=1}^{n_e} |e_k|$ , where  $|e_k|$  denotes the number of rewriting rules that are defined for that particular enzyme. In our system, each rewriting rule adds a total of  $2n_n$  reaction channels (one corresponding to the binding and one to the dissociation for each of the  $n_n$  nucleosomes). Thus, the total number

of reaction channels is  $2n_n \times n_{\text{rules}}$ . Equivalently, each of the  $n_n$  nucleosomes has  $2n_{\text{rules}}$  different reaction channels).

Internally, each rewriting rule is associated with a state change vector, which describes how the concentrations of the available  $n_l$  enzymes are affected upon execution of the reaction that specifies the rewriting rule. For example, any binding reaction decreases the concentration of the corresponding enzyme by 1, whereas any dissociation reaction frees the enzyme and thereby increases its concentration by 1.

To save computing time, only the propensities of the reactions in the vicinity of the nucleosome(s) subject to the last reaction are re-computed. A recomputation of the reaction propensities is also necessary after each replication and phase transition.

#### 2.4. Evolutionary Optimization of the Rewriting Rule Sets

Our “chromatin computer” may operate with rather complex instructions that correspond to a particular gene expression pattern. More formally, an instruction consists of a list  $\mathcal{L}$  of rewriting rules and associated enzyme concentrations (see Figure 3). This is similar to the computational model of Bryant (2012), except that we have incorporated a concentration associated with each rewriting rule that modulates the probability with which it is applied. If the time course of the simulation is divided into  $n_p$  phases, more complex programs can be implemented as sequences  $(\mathcal{L}_i, \tau_i), i = 1, \dots, n_p$  of instructions that are valid for a prescribed time period  $\tau_i$  (i.e., a “phase”) before being supplanted by the next instruction. Thus, if a particular rewriting rule  $r_k \in \mathcal{L}_i$ , the enzyme performing this reaction is available in phase  $i$ .

Following previous studies that used evolutionary optimization, we define mutation and recombination oper-



Start time $t_k$	Duration	Phase	Enzyme availability for each phase (presence / absence)				Instruction $\mathcal{L}_i$
			1 HMTT	2 HDM1	1 HAT	3 HDAC1	
	$\tau_r = 0$	Replication ( $\mathcal{L}_r, \tau_r$ )					$\mathcal{L}_r = \{ \}$
	$\tau_1 = 2$	Phase 1 ( $\mathcal{L}_1, \tau_1$ )	✓	✓		✓	$\mathcal{L}_1 = \{ \begin{array}{l} 1 \text{ HMTT} \\ 2 \text{ HDM1} \\ 3 \text{ HDAC1} \end{array} \}$
	$\tau_2 = 4$	Phase 2 ( $\mathcal{L}_2, \tau_2$ )	✓		✓		$\mathcal{L}_2 = \{ \begin{array}{l} 1 \text{ HMTT} \\ 1 \text{ HAT} \end{array} \}$
	$\tau_3 = 1$	Phase 3 ( $\mathcal{L}_3, \tau_3$ )			✓		$\mathcal{L}_3 = \{ 1 \text{ HAT} \}$
	$\tau_r = 0$	Replication ( $\mathcal{L}_r, \tau_r$ )					$\mathcal{L}_r = \{ \}$

Constant time  $t_r = 7$  between each two replications

Time  $t_k + \tau_1 + \tau_2 + \tau_3$  Phases repeat

Figure 3: Illustration of the phases concept. An example with four different enzymes (each of which has a particular concentration) and three phases is represented. After each replication, these phases periodically follow each other in sequential order. Formally, each phase  $i$  is described by a tuple  $(\mathcal{L}_i, \tau_i)$ . Each phase  $i$  can be described by (i) a particular combination  $\mathcal{L}_i$  of enzymes that are available in that particular phase (along with their individual fixed concentrations), and (ii) its duration  $\tau_i > 0$ . Although replication is not strictly considered as a phase, it can similarly be represented by  $(\mathcal{L}_r, \tau_r)$  with  $\mathcal{L}_r = \emptyset$  and  $\tau_r = 0$ , as each replication is modeled as an instantaneous event where enzymes cannot perform any of their reactions.

ators for the individual instructions as follows:

- The concentration of a random enzyme can be decreased by up to two molecules. The concentration of a different random enzyme is correspondingly increased to keep the total number of enzymes constant.
- We can replace up to two enzymes and their corresponding rewriting rules by randomly picked different enzymes. The concentration variables and rate constants remain unchanged.
- In principle, the rate constants can be changed. However, this option has been disabled in the simulations, as discussed below.
- The recombination (cross-over) operator builds a convex combination of two instructions  $\mathcal{L}_1$  and  $\mathcal{L}_2$  using the formula  $\xi\mathcal{L}_1 + (1 - \xi)\mathcal{L}_2$  with a randomly drawn weight  $\xi \in [0, 1]$ .
- A second cross-over operator may similarly construct a combination of two already present enzymes and their associated rules, therefore adding a new enzyme in a non-random fashion. Due to the relative simplicity of the rules used in simulations, however, such operators have not been used here.
- If multiple instructions are used, their individual durations  $\tau_i$  can be changed. For this, two randomly picked phases increase and decrease their individual durations by 10%, respectively (with respect to the time  $t_r$  between two replications). In addition, independently for each phase, enzyme availabilities (presence or absence, see Figure 3) of up to two random enzymes can be inverted.

The fitness of an instruction (or a schedule of instructions) is evaluated by comparing the patterns immediately before each of the cell divisions to the start pattern using the normalized Hamming distance. The initial patterns used here are compiled in Table 3. This yields  $r + 1$  distance values  $d(i)$ . For each start pattern and  $n_i$  independent starting point, we run the Gillespie simulation  $n_g$  times with different random number seeds. Then, we average the distance values over the Gillespie realizations, obtaining  $1 - \langle d(i) \rangle$  as the autocorrelation function of the pattern. The fitness value is next computed as the sum of this autocorrelation function over the  $n_r$  cell divisions. It turns out that a simple hill-climbing approach is sufficient to obtain good solutions. Hence, a proposed mutation of the instruction (or schedule of instructions) is accepted if the estimated fitness increases. We stop the search if the best solution among all runs does not improve for  $n_i$  iterations.

## 2.5. Simulations

For all patterns, we generated  $n_s$  initial instructions by randomly selecting  $n_l$  out of a total of 28 rewriting rules (i.e., the enzymes that implement them) as listed in Table 4. The individual concentrations of these  $n_l$  enzymes were also assigned randomly so that the total number of molecules equaled  $n_m$ . All other parameters were kept constant between the independent runs. Table 2 summarizes the parameters that were used for the evolutionary algorithm and the individual stochastic simulations required for the fitness evaluation.

We simulated local chromatin state dynamics for a genomic region of 12 kb to 30 kb (depending on the pattern), a range that was also used in previous approaches (Dodd et al., 2007) and that provides a good balance

Table 2: Summary of the most relevant parameters that we used for the evolutionary algorithm. See text for details.

Par.	Value	Description
<i>Specific to the evolutionary algorithm</i>		
$n_i$	1000	Stop criterion
$n_s$	10	No. of independent starting points / runs
$n_p$	1-4	No. of phases
<i>Specific to the biology and the stochastic simulation</i>		
$n_r$	50	No. of replications
$n_n$	30-150	Total no. of nucleosomes
$c$	circular	Nucleosome organisation
$t_r$	20	Time between two replications
$n_g$	20	No. of independent Gillespie realizations
$n_a$	5 or 10	Maximal no. of distinct active enzymes (pattern-dependent) (see Table 4)
$n_b$	1-5	No. of distinct chromatin domains per pattern (see Table 3 and Figure 1)
$k$	30	Chromatin domain length (in nucleosomes)
$n_m$	10	No. of enzymes in the cell (per chromatin domain)
$n_e$	$n_b \times n_m$	Total no. of enzymes in the cell (all chromatin domains)
$r_b$	1	Binding reaction rate constant
$r_d$	5	Dissociation reaction rate constant

between computational speed and biological verisimilitude. We used a value of 50 for  $n_r$ , as this reflects the Hayflick limit for how many times a cell can divide. Also, we used a circular nucleosome organization throughout the simulations to avoid artifacts for the boundary nucleosomes due to the rewriting rules and their neighbor-dependence. For a suitable number of enzymes available for the modeled chromatin region, we chose a value of 10 per chromatin domain (parameter  $n_m$ ). Dependence on the number of chromatin domains is necessary to ensure comparability among patterns with varying length due to the concentration dependence of the enzyme reactions in the Gillespie al-

Table 3: Summary of the start patterns used for the fitness evaluations. State 0 designates an unmodified nucleosome, whereas states 1 and 2 designate two distinct modified states of a particular histone modification. The parameter  $k$  (here set to 30) is the length of each individual chromatin domain in each pattern, see Table 2 and Figure 1 for details.

Pattern summary	Pattern length ( $n_n$ )	Pattern abbreviation
<i>Elementary patterns</i>		
$1^k$	30	1
$0^k 1^k$	60	01
$1^k 2^k$	60	12
<i>Composite patterns</i>		
$1^k 2^k 1^k$	90	121
$1^k 2^k 2^k$	90	122
$1^k 2^k 1^k 2^k$	120	1212
$1^k 2^k 2^k 2^k$	120	1222
$0^k 1^k 0^k$	90	010
$0^k 1^k 1^k$	90	011
$1^k 0^k 1^k$	90	101
$0^k 1^k 0^k 1^k$	120	0101
$1^k 0^k 1^k 1^k$	120	1011
$1^k 0^k 2^k$	90	102
$1^k 2^k 0^k$	90	120
$1^k 2^k 0^k 0^k$	120	1200
$1^k 2^k 0^k 2^k 1^k$	150	12021

gorithm. For the elementary patterns, up to five different enzymes may be selected, whereas we increased  $n_a$  to 10 for composite patterns. We found these values to be sufficient to obtain good solutions in our simulations. The values for  $t_r$ ,  $r_b$ , and  $r_d$  were chosen so that (i) enough reactions can take place between two subsequent replications, and (ii) the system tends to have free molecules available rather than a condition where all molecules are bound.

The unmodified state behaves differently from modifications since it is not distinguishable from erased information after replication. It may be helpful to allow more than one phase to obtain good solutions. Coordinated, phase-dependent enzyme availabilities may make it easier to systematically recompute the parental modification state. To identify the optimal number of phases  $n_p$ , we first ran the evolutionary algorithm for the pattern 01 for solutions with one, two, three, and four

Table 4: Summary of the 28 rewriting rules with patterns on  $\mathcal{A} = \{0, 1, 2\}$  used in this contribution. For rewriting rules that are not intrinsically symmetric we also list their mirror image. The position in brackets is bound and modified, the two flanking positions remain invariant. The second column gives a rewriting rule abbreviation that will be used hereafter.

<b>Rewriting rules realizing <math>0 \rightarrow 1</math> reactions</b>	$\alpha$
$[0] \rightarrow [1]$	$\alpha_1$
$0[0]0 \rightarrow 0[1]0$	$\alpha_2$
$1[0]1 \rightarrow 1[1]1$	$\alpha_3$
$2[0]2 \rightarrow 2[1]2$	$\alpha_4$
$0[0]1 \rightarrow 0[1]1$ or $1[0]0 \rightarrow 1[1]0$	$\alpha_5$
$0[0]2 \rightarrow 0[1]2$ or $2[0]0 \rightarrow 2[1]0$	$\alpha_6$
$1[0]2 \rightarrow 1[1]2$ or $2[0]1 \rightarrow 2[1]1$	$\alpha_7$
<b>Rewriting rules realizing <math>0 \rightarrow 2</math> reactions</b>	$\beta$
$[0] \rightarrow [2]$	$\beta_1$
$0[0]0 \rightarrow 0[2]0$	$\beta_2$
$1[0]1 \rightarrow 1[2]1$	$\beta_3$
$2[0]2 \rightarrow 2[2]2$	$\beta_4$
$0[0]1 \rightarrow 0[2]1$ or $1[0]0 \rightarrow 1[2]0$	$\beta_5$
$0[0]2 \rightarrow 0[2]2$ or $2[0]0 \rightarrow 2[2]0$	$\beta_6$
$2[0]1 \rightarrow 2[2]1$ or $1[0]2 \rightarrow 1[2]2$	$\beta_7$
<b>Rewriting rules realizing <math>1 \rightarrow 0</math> reactions</b>	$\gamma$
$[1] \rightarrow [0]$	$\gamma_1$
$0[1]0 \rightarrow 0[0]0$	$\gamma_2$
$1[1]1 \rightarrow 1[0]1$	$\gamma_3$
$2[1]2 \rightarrow 2[0]2$	$\gamma_4$
$0[1]1 \rightarrow 0[0]1$ or $1[1]0 \rightarrow 1[0]0$	$\gamma_5$
$0[1]2 \rightarrow 0[0]2$ or $2[1]0 \rightarrow 2[0]0$	$\gamma_6$
$2[1]1 \rightarrow 2[0]1$ or $1[1]2 \rightarrow 1[0]2$	$\gamma_7$
<b>Rewriting rules realizing <math>2 \rightarrow 0</math> reactions</b>	$\delta$
$[2] \rightarrow [0]$	$\delta_1$
$0[2]0 \rightarrow 0[0]0$	$\delta_2$
$2[2]2 \rightarrow 2[0]2$	$\delta_3$
$1[2]1 \rightarrow 1[0]1$	$\delta_4$
$0[2]2 \rightarrow 0[0]2$ or $2[2]0 \rightarrow 2[0]0$	$\delta_5$
$0[2]1 \rightarrow 0[0]1$ or $1[2]0 \rightarrow 1[0]0$	$\delta_6$
$1[2]2 \rightarrow 1[0]2$ or $2[2]1 \rightarrow 2[0]1$	$\delta_7$

phases and identified the solution with the highest score. We then ran all composite patterns that contain the un-

modified state with only  $n_p$  phases (instead of variants with one, two, three, or four phases, respectively).

### 3. Results

Using the flexible software system that we developed (see end of results for details) to study the dynamics of histone modification states, the difficulty of pattern reconstruction problem depends on the structure of the start pattern. Therefore, we initially summarize the observations for simple, elementary start patterns (see Tables 5, 6 and Figure 4). Stable solutions for constant patterns and patterns that consist of only modified nucleosomes were achieved with relative ease. Neighbor-independent rewriting rules were only rarely selected (or only in low concentrations), as they easily introduce noise to the system. Similar to the results of Dodd and Sneppen (2011); Hodges and Crabtree (2012), we found that chromatin domains can transiently multifurcate to form multiple smaller domains that remain stable for a particular amount of time, which was particularly pronounced for patterns that contain patches of unmodified nucleosomes (Figure 5). Noteworthy, we sometimes observed a gradual accumulation of errors during the lifetime of a cell (e.g., see the least stable solution for pattern 12 in Figure 4).

**Constant patterns.** As expected, it is trivial to find optimal solutions for the constant pattern 1 as the only rewriting rules required to recompute the parental pattern are either  $\alpha_1$  ( $[0] \rightarrow [1]$ ) or  $\alpha_5$  ( $0[0]1 \rightarrow 0[1]1$  or  $1[0]0 \rightarrow 1[1]0$ ). These must be present in higher concentrations than rewriting rules that change 1 to 0 (class  $\gamma$ ). Consequently, a large number of simulations achieved the optimal score. Also, contrary to other patterns, the inclusion of neighbor-independent rewriting rules, such as  $\alpha_1$ , pose no disadvantage to the system.

**Pattern 12.** We also found that it is relatively easy to evolve a system that can stably maintain patterns when the parental nucleosome state consists of several chromatin domains of modified nucleosomes, as exemplified by patterns 12 and 121. Four rewriting rules with approximately equal concentrations were sufficient for stable inheritance over 50 generations:  $\alpha_3$  ( $1[0]1 \rightarrow 1[1]1$ ),  $\alpha_5$  ( $0[0]1 \rightarrow 0[1]1$  or  $1[0]0 \rightarrow 1[1]0$ ),  $\beta_4$  ( $2[0]2 \rightarrow 2[2]2$ ), and  $\beta_6$  ( $0[0]2 \rightarrow 0[2]2$  or  $2[0]0 \rightarrow 2[2]0$ ). Notably, the boundary between differentially modified regions was subject to stochastic fluctuation given that it was controlled solely by the available rewriting rules (Figure 4).

**Pattern 01.** It is substantially more difficult to find good solutions for patterns that contain a mixture of modified nucleosomes and unmodified nucleosomes

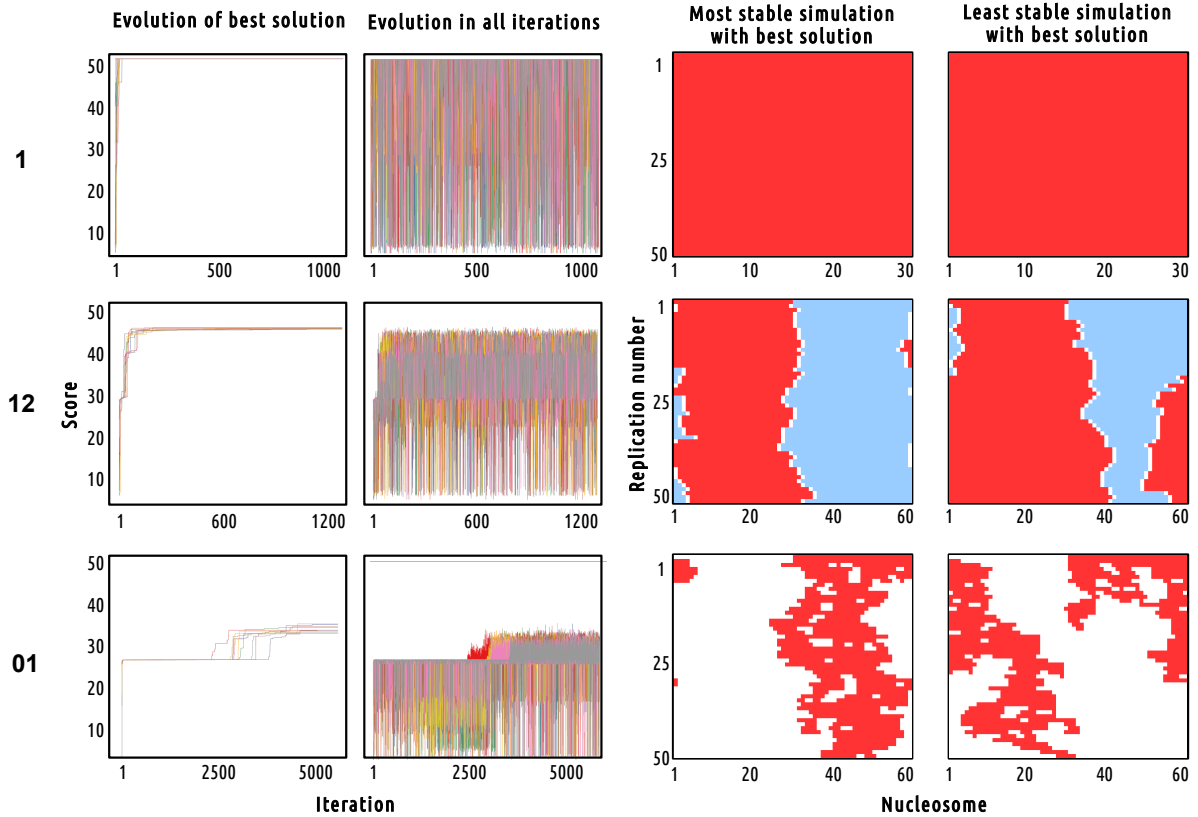


Figure 4: Results from the evolutionary algorithm for the three elementary patterns (top: 1, middle: 12, bottom: 01, 2 phases). The leftmost figure in each row shows the evolution of the best score (separately for each of the  $n_s$  independent starting points). The second leftmost figure shows the evolution of the score of the solution that was proposed in each particular iteration. The two rightmost figures in each row display the stochasticity of the best solution among all  $n_s$  independent stochastic simulations (left: most stable (highest score), right: least stable (lowest score)). For each visualization, the state of the system is shown directly before each replication (with the initial state at the top, and the state after the last replication at the bottom). The coloring of the different nucleosome states is analogous to the previous Figures (0: white, 1: red, 2: blue). For pattern 1, no variation was detectable between the best and the worst performing run when states were only compared directly before each replication event.

such as 01, 101, or 102. The intuitive reason is that the 0-state in the target pattern is indistinguishable from the information lost during the replication event. Thus, the chromatin computer's programs lack means to determine where states need to be regenerated and where the target state has already been reached. No good solutions seem to exist for this problem when only a single phase is allowed between replication events. To illustrate this, consider the pattern  $0^k 1^k$ . Rewriting rules are needed to re-establish the right part of the pattern that contains the nucleosomes in state 1 after each cell division. These can principally be constructed in two ways: (i) The rewriting rule  $\alpha_3$  ( $1[0]1 \rightarrow 1[1]1$ ) ensures that it can only be applied in the right part. However, having only this rewriting rule is too strict, as patterns such as 01001 cannot be repaired otherwise. This would be

possible with (ii)  $\alpha_5$  ( $0[0]1 \rightarrow 0[1]1$  or  $1[0]0 \rightarrow 1[1]0$ ), however this is also applicable at the boundary of the 0 and 1 region and can therefore slowly spread into the 0-region, leading to gradual loss of the ancestral signal. Rewriting rules that do not incorporate the state of neighboring nucleosomes (e.g.,  $\alpha_1$  ( $[0] \rightarrow [1]$ )) are also not helpful in this regard, since they introduce additional noise to the system.

Admitting more phases, i.e., variation of the gene expression program through the cell cycle, can mitigate this difficulty. Different phases can then serve a particular purpose and, collectively, they aim at restoring the parental pattern. In practice, we found that solutions with more than one phase become only marginally better with two or more phases, possibly due to the greatly expanded parameter space of the solution (Table 5). It

may therefore be necessary to run the evolutionary algorithm for a substantially longer time to obtain good solutions that consist of more than one phase. Nevertheless, the stability of the best solutions we found (independent of the number of phases) is not comparable to solutions for patterns 1 and 12. This is indicated by both the attainable scores and by the time course visualization of the patterns in Figure 4.

**Composite patterns.** For the composite patterns, the results were as follows: Patterns that are combinations of all three elementary patterns (i.e., with 0-1, 0-2, and 1-2 chromatin domain transitions) produced the lowest scores, despite allowing the evolutionary algorithm to increase the maximal number of active enzymes (see below). The best scores were produced by variations of the pattern 12. For the pattern 12021 and 1011, the 0 patch in the middle of the patterns was consistently lost, even in the best solutions.

We also found that solutions for the three elementary patterns can be used with comparable quality for repetitions of elementary patterns. A good example is the pattern 12, where all of the variations tested produced scores that differed by less than 10% from the original score. This occurred even though the patterns were either more complex (121, 1212) or contained chromatin domains of unequal length (122, 1222). For the pattern 01, we observed a similar result with the patterns 010, 011, 101, 0101, and 1011, although the scores were up to 20% smaller (data not shown). In all cases, the score may be improved by adjusting enzyme concentrations and phase durations, particularly for patterns with a different length than the original pattern.

However, we found that individual solutions from elementary patterns cannot merely be combined for a more complex pattern, particularly with multiple phases. For example, the combination of the individual solutions for the patterns 01, 02, and 12 does not produce a good result for patterns that contain all these three types of transitions such as 102. The combination of different solutions—i.e. the simultaneous presence of more enzymes—apparently interferes with the “strategy” of the partial solutions: The additional enzymes act at the newly produced modifications and obliterate the nascent pattern.

We also investigated whether the performance could be improved by optimizing the rate constants for binding and/or dissociation rates, along with the rewriting rules themselves. Although the scores were comparable, we obtained solutions that are much more tailored

Table 5: Summary of the results from the evolutionary algorithm. For each pattern, the number of iterations  $n_{it}$  after which the evolutionary algorithm finished is given, as well as the score of the best solution, the number of phases  $n_p$  that solution consisted of, and the number of distinct enzymes that were selected ( $n_a$ ). For more details on the composition of the solutions, see Table 6.

Pattern	$n_p$	$n_{it}$	$n_a$	Best score
<i>Elementary patterns</i>				
1	1	1001	4	51
12	1	1180	5	45.6
	1	4083	5	35.1
01	2	5360	3	36.6
	3	2259	5	35.5
	4	4823	5	35.6
<i>Composite patterns</i>				
121	1	2717	4	47.8
122	1	1346	6	44.7
1212	1	1367	4	45.0
1222	1	3595	4	48.4
010	2	5229	7	43.2
011	2	6574	6	36.6
101	2	5834	5	35.0
0101	2	5106	6	35.9
1011	2	1002	5	38.4
102	2	5116	5	33.4
120	2	3840	6	33.2
1200	2	5231	6	34.4
12021	2	3429	7	39.1

towards the reconstruction of a particular initial pattern length (data not shown). This is because the reaction rate constants also control how many reactions may take place during a particular time in the stochastic simulation. By allowing them to vary, the enzymes and the corresponding reaction rates become tailored to the specific pattern length. Neighbor-independent enzymes were then also more frequently selected, because their frequency of selection can be controlled by the reaction rates.

Finally, we examined to what extent the parameter  $n_a$ , which limits the maximal number of active enzymes, has an influence on the quality of the solutions. We found that selection sometimes tends to increase the number of enzymes by including rewriting rules that are rarely applicable. This was particularly true for the pattern 01, where the score for the best solution with one

phase was increased to that of solutions with multiple phases using the original value of  $n_a$  (data not shown). This appears to be a means of adjusting reaction rates to decrease the number of reactions that take place between replications.

Table 6: Summary of the best simulation for each of the elementary patterns. The second column presents the number of phases  $n_p$  that compose the solution. The third column summarizes the best solution and lists the enzymes active in a particular phase, together with their abundance (in brackets). For the patterns 1 and 01 with a single phase, many distinct optimal solutions have been found (see text), and only one representative is included here. For the pattern 01, multiple independent evolutionary optimizations with a different number of phases have been performed (see text). If multiple phases were set, the individual phase durations are also presented (in percent).

Pat-tern	$n_p$	Best solution ( $\tau_i$ ) : $\mathcal{L}_i$
1	1	$\mathcal{L}_1:\alpha_3(3), \alpha_5(4), \alpha_7(1), \delta_2(2)$
12	1	$\mathcal{L}_1:\alpha_3(3), \alpha_5(5), \beta_4(3), \beta_6(5), \gamma_7(4)$
	1	$\mathcal{L}_1:\alpha_3(2), \alpha_5(8), \alpha_7(2), \delta_4(4)$
	2	$\tau_1(17.8\%), \mathcal{L}_1:\alpha_3(6), \alpha_5(1), \delta_5(13)$ $\tau_2(82.2\%), \mathcal{L}_2:\alpha_3(6)$
		$\tau_1(45.5\%), \mathcal{L}_1$ :see phase 3, + $\alpha_5(7)$
01	3	$\tau_2(31.8\%), \mathcal{L}_2$ :see phase 1 $\tau_3(22.8\%), \mathcal{L}_3:\alpha_3(4), \alpha_6(4), \alpha_7(3), \delta_4(2)$
		$\tau_1(35\%), \mathcal{L}_1:\alpha_5(7), \alpha_7(4), \beta_5(4)$
	4	$\tau_2(35\%), \mathcal{L}_2:\alpha_5(7), \alpha_7(4), \delta_4(3)$ $\tau_3(15\%), \mathcal{L}_3:\alpha_5(7), \alpha_7(4), \beta_5(4), \delta_4(3)$ $\tau_4(15\%), \mathcal{L}_4:\alpha_3(2), \alpha_7(4), \delta_4(3)$

To verify that the best solutions for the different patterns are not specific to the parameterization of the model and the specific pattern length, we also tested the sensitivity to parameter variations (Table 7). Specifically, we tested the effect of a linear nucleosome string rather than a circular one, the time between two replications, the number of nucleosomes and replications, and the dissociation rate of all enzymes. In summary, we found that the solutions produce very similar scores in most of the parameter space for the patterns 1 and 12, while for the pattern 01 and its compositions, the best solutions strongly depend on the kinetic parameters. We also found that the number of nucleosomes must not be too low (a value around 40 was sufficient for robustness). Otherwise, stochastic effects may irreversibly destroy the parental signal. Similarly, the time  $t_r$  between

two replications must be long enough to allow for re-computation of the parental pattern. Some noteworthy effects that we observed while varying the parameters are summarized in Figure 5.

The source code of a C implementation of our software system can be obtained under the GNU Public License from <http://www.bioinf.uni-leipzig.de/Software/StoChDyn> and consists of two separate programs: the stochastic simulation of histone modification dynamics using Gillespie’s approach (*StoChDyn*) and the evolutionary algorithm (*Evo-ES*) that uses *StoChDyn* to evaluate its solutions.

#### 4. Discussion

In this work, we have queried whether the propagation of histone modification patterns across cell divisions can be seen as a computational problem and if so, whether chromatin is organized in a way that is amenable to the solution of this problem. Our answer is twice affirmative. We demonstrated that the faithful propagation of patterns of histone modification can be interpreted as a computational problem that is achievable through a small collection of rewriting rules. These rewriting rules are abstractions of a well-described class of enzymes and enzyme complexes combining reader, writer, and eraser domains for specific histone modifications. For the best solutions, the evolutionary algorithm selected almost exclusively enzymes that are dependent on the states of neighboring nucleosomes (except for the trivial pattern 01). This highlights that context-dependency is crucial for such inheritance systems, as context-independent enzymes easily introduce too much noise to the system that further complicate the recomputation of parental state patterns after cell division. Indeed, for many histone-modifying enzymes, it is well-known that their binding affinities are highly influenced by the presence or absence of particular histone modifications or other signals nearby.

This modification process is intrinsically stochastic and crucially depends on concentrations of the available enzymes and histone modifications. Hence, we propose a cellular automata-like 1-D string as the computational paradigm for a chromatin computer, on which sets of local rewriting rules are applied asynchronously with time-dependent probabilities. As a practical implementation, we employ a detailed stochastic simulation of chromatin state dynamics to approximate the physico-chemical constraints of our approach. Here, the probability for applying rewriting rules is modeled explicitly in reaction rates for binding and dissociation following the laws of mass action. The “software” for this type of

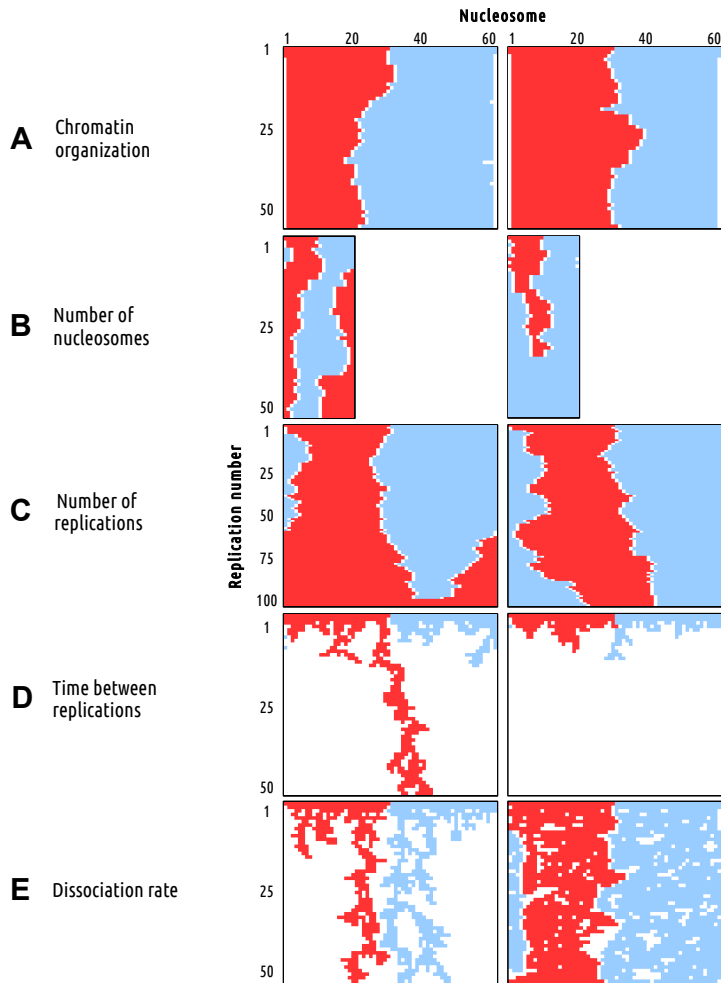


Figure 5: Visualizations from selected simulations for pattern 12 of the robustness analysis. One distinct parameter is fluctuated in each row (A-E, as indicated on the left) as compared to the best solution for this pattern, and two exemplary runs of the same simulation (except for E) are shown that highlight some noteworthy effects that were observed. The coloring is analogous to Figure 4.

(A) Variations in the chromatin organization at the boundaries (here: linear instead of circular chromatin string).

(B) Variations in the number of nucleosomes (here: 20 instead of 60), illustrating that reducing the number of nucleosomes increases the likelihood that the signal is lost due to stochasticity.

(C) Variations in the number of replications (here: 100 instead of 50), showing that chromatin domains can gradually disappear (left) or change their exact location due to stochasticity (right).

(D) Variations in the time between replications (here: 0.5 instead of 20), highlighting that the original signal may gradually get lost (either fully or partly) if the cell has not enough time to recompute the parental pattern.

(E) Variations in the dissociation rate of the enzymes (here: 0.05 (left) and 0.15 (right), respectively, instead of 1), illustrating that the enzyme kinetics may also have a large effect on pattern stability. The binding rate constants were unmodified; however, the dissociation rate constants have been reduced, which means that enzymes need more time to perform their designated reaction.

Table 7: Summary of the robustness analyses for the best solution for each elementary pattern. The table lists the parameter value intervals when the score of the modified solution (with one changed parameter) achieved more than 90% of the score of the original best solution. The variable  $s$  denotes the step size. When the number of replication  $n_r$  was varied, we calculated the ratio of the score and the corresponding maximal possible score and compared the two ratios using the 90% threshold for better comparability, as  $n_r$  has a direct influence on the maximal possible score.

Parameter	Original value	New, varied values	Pattern		
			1	12	01
Nucleosome organisation ( $c$ )	circular	linear	all	all	all
Number of nucleosomes ( $n_n$ )	60	10-200 ( $s = 10$ )	all	> 40	60-90,140
Number of replications ( $n_r$ )	50	5-100 ( $s = 5$ )	all	all	< 55
Time between replications ( $t_r$ )	20	0.5 - 40 ( $s$ variable)	> 5	> 1	19-20
Dissociation rate ( $r_d$ )	5	0.05-100 ( $s$ variable)	> 0.5	> 0.15	4

chromatin computer is thus a sequence of sets of rewriting rules and concentrations that can be directly interpreted as part of the cell's gene expression program.

We have shown that stable propagation of complex patterns without the need for explicit boundary elements is possible in our model, although not for all

types of patterns. This is consistent with the findings of Hodges and Crabtree (2012), who recently proposed that explicit boundary elements may not be required for H3K9me3 domains, as their size and propagation may be naturally limited by chromatin remodeling processes. Nevertheless, the maintenance of (approximate) boundaries between differentially modified chromatin domains is a main challenge of any epigenetic inheritance mechanism. In practice, additional security measures that prevent spreading may furthermore be required for faithful long-term inheritance. Indeed, several mechanisms that prevent spreading may trivially solve the problem of restoring the parental modification pattern when exact boundaries are important. Examples include nucleosome-depleted regions (see, for example, Jiang and Pugh, 2009; Mavrich et al., 2008; Yadon et al., 2010), bound proteins (e.g., CTCF) or histone variants (e.g., H2A.Z) in the vicinity of positions that border the different chromatin domains, or the marking with a particular histone modification directly after replication due to other signalling factors.

In our model, long-range interactions are not necessarily required for stability, which contrasts to what is argued by Dodd et al. (2007). However, the presence or absence of long-range interactions has no impact on the computational power of the chromatin computer or our conclusions with respect to epigenetic inheritance. Generally, epigenetic inheritance should be seen as an ensemble of different strategies that collectively aim to transmit a particular chromatin state throughout cell division. Histone modification patterns fit this bill. The propagation of one out of several alternative modified states provides the required memory across divisions. Nevertheless, only a few primary histone modifications that typically form large and homogeneous chromatin domains (e.g., H3K9me3) may be copied in a self-propagating manner as described here. The transmission or recomputation of other, more secondary histone modifications, however, likely depends on other factors (Sarkies and Sale, 2011). On the other hand, the inheritance of promoter-specific modifications (which typically cover only a few nucleosomes) is likely implemented in a different way than the propagation of large homogeneous chromatin domains, because short domains are much more difficult to inherit due to pure stochasticity (e.g., see Figure 5B).

Pattern stability is influenced by a number of factors, such as dynamic chromatin remodeling events, the up- and downregulation of genes that code for or regulate the corresponding histone-modifying enzymes, pattern complexity (Sneppen and Dodd, 2012), or the length of the state to be maintained. In summary, histone modifi-

cation patterns are often an ongoing enzymatic competition between their placement and removal. Altering this steady-state balance pushes either towards the accumulation of the mark or its erasure (Hodges and Crabtree, 2012). This has been described most clearly in embryonic reprogramming (Katz et al., 2009; VerMilyea et al., 2009).

Our finding that patterns containing patches of unmodified nucleosomes are more difficult to inherit than modified ones (irrespective of the number of phases) due to the ambiguity of the unmodified state raises the question of biological relevance. Due to the sheer complexity of histone modifications, the vast majority of nucleosomes may carry at least one modification, which could facilitate recomputation of the parental patterns and resolve the difficulty of stably inheriting such domains. Additionally, other chemical signals within the vicinity of a nucleosome such as DNA methylation, the presence of histone variants or spatial contacts with genomic loci or protein complexes that are themselves retained through cell division may be specifically used to backup the information of nucleosome left unmodified intentionally.

We emphasize that the focus of this contribution is on the computational task of re-constructing a complex histone modification pattern typical for somatic cells. We do not claim that epigenetic inheritance across the germ line follows the same paradigm.

Information inherited through the germline for an effectively infinite number of generations is subject to Eigen's error threshold (Eigen, 1971), which links the amount of stably inheritable information to the accuracy of information propagation. While effective proof-reading mechanisms limit replication errors to a single mutation per round of replication for genomic DNA, no mechanism is conceivable that would achieve a similar accuracy for histone modifications. As a consequence, the amount of stably inheritable epigenetic information is severely limited. Consistent with this theory, most, if not all, of the extraneous epigenetic information is erased during spermatogenesis and oogenesis. The resulting totipotent state (Hackett et al., 2012) is characterized by global erasure of DNA methylation, chromatin reorganisation, differential regulation of histone-modifying enzymes (e.g., the tendency for the upregulation of histone de-modifying and downregulation of histone-modifying enzymes). The initial stages of embryogenesis are governed by a gene regulatory network dominated by transcription factors (e.g., reviewed in Adachi and Schöler, 2012), partial ejection of nucleosomes (Watanabe et al., 2013) and therefore a reduction in the availability of a major epigenetic information car-



rier. Indeed, it seems that only few epigenetic modifications are part of the epigenomic basal state (e.g., strong heterochromatin formation of genes linked to differentiation (Watanabe et al., 2013) or imprinting and poised promoters). In contrast, the error threshold does not preclude inheritance of complex patterns of histone marks in somatic cell lines because the number of generations is limited, and usually small. Here, the degradation of the epigenetic information is acceptable for a while, but inevitably leads to daughter cells whose epigenetic patterns are damaged beyond repair. This effect may thus constitute an epigenetic version of aging. Indeed, major changes to the epigenetic information can result in “chromatin diseases” such as cancer (Timp and Feinberg, 2013), for example.

Whether histone modifications and the presence of histone variants are a cause or consequence of the transcriptional status (Henikoff and Shilatifard, 2011; Rando, 2012) is still hotly debated. Here, we are completely agnostic about this issue since it has no impact on our conclusions. The programs that run on the chromatin computer (i.e., the schedules and concentrations of re-writing enzymes), are externally specified in our model. In particular, we make no statement in regards to whether the gene expression program is a direct consequence of, or at least dominated by, the chromatin state, or whether it is entirely determined by classical transcription factor networks that are largely or even completely independent of the chromatin state. In computer science terms, we employ a model of computation that strictly distinguishes between (gene expression) programs and (histone modification) data.

It appears natural, in a next step, to remove this distinction and to ask if chromatin itself can “learn” to re-program itself, by making the gene expression programs an intrinsic function of the histone modification data. While this may be an extreme model that implicitly views transcription factors as being enslaved by histone modification states at their gene loci, it is an important limiting case given that gene expression is clearly not independent of chromatin state.

#### Acknowledgements

We thank two anonymous reviewers for helpful comments and suggestions. Additionally, we thank Charles L. Nunn and Martin A. Smith for proofreading and Lydia Steiner for stimulating discussions. This work was supported in part by the Templeton Foundation, grant # 24332 “Origins and Evolution of Regulation in Biological Systems” and the Sächsische Aufbaubank (SAB), grant # 14494.

#### References

- Adachi, K., Schöler, H. R., 2012. Directing reprogramming to pluripotency by transcription factors. *Current Opinion in Genetics & Development*.
- Ameyar-Zazoua, M., Rachez, C., Souidi, M., Robin, P., Fritsch, L., Young, R., Morozova, N., Fenouil, R., Descostes, N., Andrau, J. C., Mathieu, J., Hamiche, A., Ait-Si-Ali, S., Muchardt, C., Batsché, E., Harel-Bellan, A., 2012. Argonaute proteins couple chromatin silencing to alternative splicing. *Nature Structural & Molecular Biology*.
- Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., Komorowski, J., 2009. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* 19, 1732–1741.
- Andrews, A., Luger, K., 2011. Nucleosome structure (s) and stability: variations on a theme. *Annual review of biophysics* 40, 99–117.
- Annunziato, A. T., 2005. Split decision: what happens to nucleosomes during DNA replication? *J Biol Chem* 280, 12065–12068.
- Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., Kouzarides, T., 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410 (6824), 120–124.
- Barth, T. K., Imhof, A., 2010. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem Sci.* 35, 618–626.
- Brown, S. J., Stoilov, P., Xing, Y., 2012. Chromatin and epigenetic regulation of pre-mRNA processing. *Hum Mol Genet.* 21, R90–R96.
- Bryant, B., 2012. Chromatin computation. *PLoS ONE* 7, e35703.
- Corpet, A., Almouzni, G., 2009. Making copies of chromatin: the challenge of nucleosomal organization and epigenetic information. *Trends in cell biology* 19 (1), 29–41.
- De Vos, D., Frederiks, F., Terweij, M., van Welsem, T., Verzijlbergen, K. F., Iachina, E., de Graaf, E. L., Altelaar, A. M., Oudgenoeg, G., Heck, A. J., et al., 2011. Progressive methylation of ageing histones by dot1 functions as a timer. *EMBO reports* 12 (9), 956–962.
- Dodd, I. B., Micheelsen, M. A., Sneppen, K., Thon, G., 2007. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* 129, 813–822.
- Dodd, I. B., Sneppen, K., 2011. Barriers and silencers: A theoretical toolkit for control and containment of nucleosome-based epigenetic states. *Journal of Molecular Biology* 414, 624–637.
- Eigen, M., 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58 (10), 465–523.
- Ernst, J., Kellis, M., 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol.* 28, 817–825.
- Fuchs, S. M., Krajewski, K., Baker, R. W., Miller, V. L., Strahl, B. D., 2011. Influence of combinatorial histone modifications on antibody and effector protein recognition. *Current Biology* 21 (1), 53–58.
- Gacek, A., Strauss, J., 2012. The chromatin code of fungal secondary metabolite gene clusters. *Appl Microbiol Biotechnol* 95, 1389–1404.
- Gardner, K. E., Allis, C. D., Strahl, B. D., 2011. Operating on chromatin, a colorful language where context matters. *J Mol Biol* 409, 36–46.
- Gillespie, D. T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.* 22, 403–434.
- Gillespie, D. T., 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361.

- Grunstein, M., 1998. Yeast heterochromatin: regulation of its assembly and inheritance by histones. *Cell* 93 (3), 325–328.
- Hackett, J. A., Zyllicz, J. J., Surani, M. A., 2012. Parallel mechanisms of epigenetic reprogramming in the germline. *Trends in Genetics* 28 (4), 164–174.
- Hansen, K. H., Bracken, A. P., Pasini, D., Dietrich, N., Gehani, S. S., Monrad, A., Rappilber, J., Lerdrup, M., Helin, K., 2008. A model for transmission of the H3K27me3 epigenetic mark. *Nature cell biology* 10 (11), 1291–1300.
- Hathaway, N. A., Bell, O., Hodges, C., Miller, E. L., Neel, D. S., Crabtree, G. R., 2012. Dynamics and memory of heterochromatin in living cells. *Cell*.
- Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., Kellis, M., Ren, B., 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112.
- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., Ren, B., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet.* 39, 311–318.
- Henikoff, S., Shilatifard, A., 2011. Histone modification: cause or cog? *Trends Genet.* 27, 389–396.
- Herzog, M., Soyer, M. O., 1981. Distinctive features of dinoflagellate chromatin. Absence of nucleosomes in a primitive species *Prorocentrum micans* E. *Eur J Cell Biol.* 23, 295–302.
- Hodges, C., Crabtree, G., 2012. Dynamics of inherently bounded histone modification domains. *Proceedings of the National Academy of Sciences* 109 (33), 13296–13301.
- Jackson, V., Chalkley, R., 1985. Histone segregation of replicating chromatin. *Biochemistry* 24, 6930–6938.
- Jacobson, R. H., Ladurner, A. G., King, D. S., Tjian, R., 2000. Structure and function of a human TAFII250 double bromodomain module. *Science* 288 (5470), 1422–1425.
- Jenuwein, T., Allis, C. D., 2001. Translating the histone code. *Science* 293, 1074–1080.
- Jeong, Y. S., Yeo, S., Park, J. S., Lee, K. K., Kang, Y. K., 2007. Gradual development of a genome-wide H3-K9 trimethylation pattern in paternally derived pig pronucleus. *Developmental Dynamics* 236 (6), 1509–1516.
- Jiang, C., Pugh, B. F., 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nature Reviews Genetics* 10 (3), 161–172.
- Katan-Khaykovich, Y., Struhl, K., 2005. Heterochromatin formation involves changes in histone modifications over multiple cell generations. *The EMBO journal* 24 (12), 2138–2149.
- Katz, D. J., Edwards, T. M., Reinke, V., Kelly, W. G., 2009. A *C. elegans* LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* 137 (2), 308–320.
- Khalil, A., Guttman, M., Huarte, M., Garber, M., Raj, A., Morales, D., Thomas, K., Presser, A., Bernstein, B., Van Oudenaarden, A., et al., 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences* 106 (28), 11667–11672.
- Kundaje, A., Kyriazopoulou-Panagiotopoulou, S., Libbrecht, M., Smith, C. L., Raha, D., Winters, E. E., Johnson, S. M., Snyder, M., Batzoglou, S., Sidow, A., 2012. Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. *Genome Res.* 22, 1735–1747.
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., Jenuwein, T., 2001. Methylation of histone H3 lysine 9 creates a binding site for HPI proteins. *Nature* 410 (6824), 116–120.
- Lee, J., 2012. Epigenetic regulation by long noncoding RNAs. *Science* 338 (6113), 1435–1439.
- Lindroth, A., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., Jenuwein, T., Khorasanizadeh, S., Jacobsen, S. E., 2004. Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *The EMBO Journal* 23 (21), 4146–4155.
- Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., Misteli, T., 2011. Epigenetics in alternative pre-mRNA splicing. *Cell* 144 (1), 16–26.
- Magistri, M., Faghihi, M. A., St Laurent, G., Wahlestedt, C., 2012. Regulation of chromatin structure by long noncoding rnas: focus on natural antisense transcripts. *Trends in Genetics*.
- Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury Iii, W. J., Voigt, P., Martin, S. R., Taylor, W. R., De Marco, V., Pirrotta, V., Reinberg, D., Gamblin, S. J., 2009. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 461 (7265), 762–767.
- Margueron, R., Reinberg, D., 2010. Chromatin structure and the inheritance of epigenetic information. *Nature Reviews Genetics* 11 (4), 285–296.
- Mavrich, T. N., Jiang, C., Ioshikhes, I. P., Li, X., Venters, B. J., Zanton, S. J., Tomsho, L. P., Qi, J., Glaser, R. L., Schuster, S. C., Gilmour, D. S., Albert, I., Pugh, B. F., 2008. Nucleosome organization in the *Drosophila* genome. *Nature* 453 (7193), 358–362.
- Millar, C. B., Grunstein, M., 2006. Genome-wide patterns of histone modifications in yeast. *Nature Reviews Molecular Cell Biology* 7 (9), 657–666.
- Miller, K. M., Jackson, S. P., 2012. Histone marks: repairing DNA breaks within the context of chromatin. *Biochem Soc Trans* 40, 370–376.
- Moazed, D., 2011. Mechanisms for the inheritance of chromatin states. *Cell* 146, 510–518.
- Mondal, T., Rasmussen, M., Pandey, G., Isaksson, A., Kanduri, C., 2010. Characterization of the RNA content of chromatin. *Genome research* 20 (7), 899–907.
- Morinière, J., Rousseaux, S., Steuerwald, U., Soler-López, M., Curtet, S., Vitte, A. L., Govin, J., Gaucher, J., Sadoul, K., Hart, D. J., Krijgsvelde, J., Khochbin, S., Müller, C. W., Petosa, C., 2009. Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* 461 (7264), 664–668.
- Musselman, C., Lalonde, M., Côté, J., Kutateladze, T., 2012. Perceiving the epigenetic landscape through histone readers. *Nature Structural & Molecular Biology* 19 (12), 1218–1227.
- Mutskov, V., Felsenfeld, G., 2003. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *The EMBO journal* 23 (1), 138–149.
- Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., Grewal, S. I. S., 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science Signalling* 292 (5514), 110.
- Owen, D. J., Ornaghi, P., Yang, J. C., Lowe, N., Evans, P. R., Ballario, P., Neuhaus, D., Filetici, P., Travers, A. A., 2000. The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. *The EMBO journal* 19 (22), 6141–6149.
- Pengelly, A. R., Copur, Ö., Jäckle, H., Herzig, A., Müller, J., 2013. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science* 339 (6120), 698–699.
- Petrkov, S., Sedkov, Y., Johnston, D. M., Hodgson, J. W., Black, K. L., Kovermann, S. K., Beck, S., Canaani, E., Brock, H. W., Mazo, A., 2012. TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. *Cell* 150 (5), 922–933.
- Prohaska, S. J., Stadler, P. F., Krakauer, D. C., 2010. Innovation in

- gene regulation: The case of chromatin computation. *J. Theor. Biol.* 265, 27–44.
- Przybilla, J., Galle, J., Rohlf, T., 2012. Is adult stem cell aging driven by conflicting modes of chromatin remodeling? *BioEssays*.
- Radman-Livaja, M., Verzijlbergen, K. F., Weiner, A. W., van Welsem, T., Friedman, N., Rando, O. J., van Leeuwen, F., 2011. Patterns and mechanisms of ancestral histone protein inheritance in budding yeast. *PLoS Biol.* 9, e1001075.
- Rando, O. J., 2012. Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Current opinion in genetics & development*.
- Rohlf, T., Steiner, L., Przybilla, J., Prohaska, S. J., Binder, H., Galle, J., 2012. Modeling the dynamic epigenome: from histone modifications towards self-organizing chromatin. *Epigenomics* 4, 205–219.
- Ruthenburg, A. J., Li, H., Milne, T. A., Dewell, S., McGinty, R. K., Yuen, M., Ueberheide, B., Dou, Y., Muir, T. W., Patel, D. J., Allis, C. D., 2011. Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell* 145 (5), 692–706.
- Santos-Rosa, H., Kirmizis, A., Nelson, C., Bartke, T., Saksouk, N., Cote, J., Kouzarides, T., 2008. Histone H3 tail clipping regulates gene expression. *Nature structural & molecular biology* 16 (1), 17–22.
- Sarkies, P., Sale, J. E., 2011. Propagation of histone marks and epigenetic memory during normal and interrupted DNA replication. *Cellular and Molecular Life Sciences*, 1–20.
- Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., Reuter, G., 2002. Central role of drosophila SU (VAR)3–9 in histone H3-K9 methylation and heterochromatic gene silencing. *The EMBO journal* 21 (5), 1121–1131.
- Schwartz, S., Meshorer, E., Ast, G., 2009. Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* 16, 990–995.
- Smith, K. T., Workman, J. L., 2012. Chromatin proteins: key responders to stress. *PLoS Biol* 10, e1001371.
- Sneppen, K., Dodd, I. B., 2012. A simple histone code opens many paths to epigenetics. *PLoS Comput Biol* 8, e1002643.
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B., Zhao, Y., 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146 (6), 1016–1028.
- Tillo, D., Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Field, Y., Lieb, J. D., Widom, J., Segal, E., Hughes, T. R., 2010. High nucleosome occupancy is encoded at human regulatory sequences. *PLoS One* 5 (2), e9129.
- Timp, W., Feinberg, A. P., 2013. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nature Reviews Cancer*.
- Tippmann, S. C., Ivanek, R., Gaidatzis, D., Schöler, A., Hoerner, L., van Nimwegen, E., Stadler, P. F., Stadler, M. B., Schübeler, D., 2012. Relative contributions of different regulatory layers to steady-state mRNA levels. *Mol. Systems Biol.* 8, 593.
- Tran, V., Lim, C., Xie, J., Chen, X., 2012. Asymmetric division of drosophila male germline stem cell shows asymmetric histone distribution. *Science* 338 (6107), 679–682.
- Turner, B. M., 1998. Histone acetylation as an epigenetic determinant of long-term transcriptional competence. *Cellular and molecular life sciences* 54 (1), 21–31.
- Turner, B. M., 2012. The adjustable nucleosome: an epigenetic signaling module. *Trends in Genetics* 28 (9), 436–444.
- VerMilyea, M. D., O'Neill, L., Turner, B. M., 2009. Transcription-independent heritability of induced histone modifications in the mouse preimplantation embryo. *PLoS One* 4 (6), e6086.
- Voigt, P., LeRoy, G., Drury III, W., Zee, B., Son, J., Beck, D., Young, N., Garcia, B., Reinberg, D., 2012. Asymmetrically modified nucleosomes. *Cell* 151 (1), 181–193.
- Wang, X., Bryant, G., Floer, M., Spagna, D., Ptashne, M., 2011. An effect of DNA sequence on nucleosome occupancy and removal. *Nature structural & molecular biology* 18 (4), 507–509.
- Wang, Z., Patel, D. J., 2011. Combinatorial readout of dual histone modifications by paired chromatin-associated modules. *J Biol Chem* 286, 18363–18368.
- Watanabe, A., Yamada, Y., Yamanaka, S., 2013. Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1609).
- Xu, D., Bai, J., Duan, Q., Costa, M., Dai, W., 2009. Covalent modifications of histones during mitosis and meiosis. *Cell Cycle* 8, 3688–3694.
- Yadon, A. N., Van de Mark, D., Basom, R., Delrow, J., Whitehouse, I., Tsukiyama, T., 2010. Chromatin remodeling around nucleosome-free regions leads to repression of noncoding RNA transcription. *Molecular and cellular biology* 30 (21), 5110–5122.
- Young, M. D., Willson, T. A., Wakefield, M. J., Trounson, E., Hilton, D. J., Blewitt, M. E., Oshlack, A., Majewski, I. J., 2011. ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucl. Acids Res.* 39, 7415–7427.
- Yun, M., Wu, J., Workman, J., Li, B., 2011. Readers of histone modifications. *Cell research* 21 (4), 564–578.
- Zhu, B., Reinberg, D., 2011. Epigenetic inheritance: uncontested? *Cell research* 21 (3), 435–441.