Understanding how chromatin folding and enzyme competition affect rugged epigenetic landscapes

Daria Stepanova[1*], Meritxell Brunet Guasch[2], Helen M. Byrne[3,4], Tomás Alarcón[5,1,6,7]

[1*]Centre de Recerca Matemàtica, Campus de Bellaterra, Edifici C, Bellaterra, 08193, Barcelona, Spain.
[2]School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh, James Clerk Maxwell Building, Mayfield Rd, Edinburgh, EH9 3FD, Scotland, United Kingdom.
[3]Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe, Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, England, United Kingdom.
[4]Ludwig Institute for Cancer Research, Nuffield Department of Medicine, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford, OX3 7DQ, England, United Kingdom.
[5]Institució Catalana de Recerca i Estudis Avançats, Passeig de Lluís Companys, 23, Barcelona, 08010, Barcelona, Spain.
[6]Departament de Matemàtiques, Universitat Autònoma de Barcelona, Campus de Bellaterra, Edifici C, Bellaterra, 08193, Barcelona, Spain.
[7]Barcelona Collaboratorium for Predictive and Theoretical Biology, Wellington, 30, Barcelona, 08005, Barcelona, Spain.

*Corresponding author. E-mail: dstepanova@crm.cat

Abstract

Epigenetics plays a key role in cellular differentiation and maintaining cell identity, enabling cells to regulate their genetic activity without altering the DNA sequence. Epigenetic regulation occurs within the context of hierarchically folded chromatin, yet the interplay between the dynamics of epigenetic modifications and chromatin architecture remains poorly understood. In addition, it remains unclear what mechanisms drive the formation of rugged epigenetic patterns, characterised by alternating genomic regions enriched in activating and repressive

marks. In this study, we focus on post-translational modifications of histone H3 tails, particularly H3K27me3, H3K4me3, and H3K27ac. We introduce a mesoscopic stochastic model that incorporates chromatin architecture and competition of histone-modifying enzymes into the dynamics of epigenetic modifications in small genomic loci comprising several nucleosomes. Our approach enables us to investigate the mechanisms by which epigenetic patterns form on larger scales of chromatin organisation, such as loops and domains. Through bifurcation analysis and stochastic simulations, we demonstrate that the model can reproduce uniform chromatin states (open, closed, and bivalent) and generate previously unexplored rugged profiles. Our results suggest that enzyme competition and chromatin conformations with high-frequency interactions between distant genomic loci can drive the emergence of rugged epigenetic landscapes. Additionally, we hypothesise that bivalent chromatin can act as an intermediate state, facilitating transitions between uniform and rugged landscapes. This work offers a powerful mathematical framework for understanding the dynamic interactions between chromatin architecture and epigenetic regulation, providing new insights into the formation of complex epigenetic patterns.

Keywords: epigenetic regulation, rugged landscapes, bifurcation analysis, stochastic modelling

MSC Classification: 92-10 , 60J20 , 92D10

1 Introduction

In eukaryotes, packaging genetic information within cells relies on the efficient organisation of DNA. Structured as long polymers, chromosomal DNA must fit within the confines of the cell nucleus while simultaneously ensuring accessibility for transcription of genomic regions containing genes essential for a particular cell phenotype and function. The primary level of packaging involves wrapping the DNA double helix around nucleosomes, protein complexes consisting of eight histones, two of each type – H2A, H2B, H3, H4 (Fig. 1a) (Gross et al., 2015). The resulting structure of DNA and its associated proteins resembles 'beads on a string' and is called chromatin. Chromatin is compacted by coiling and folding to form a condensed 30-nm fibre. Higher-order packaging includes the formation of loops, facilitated by DNA-binding proteins, such as the CCCTC-binding factor (CTCF) and the cohesin complex (Grubert et al., 2020) (Fig. 1a). Loop domains, with sizes ranging from a few kilobases (kb) to hundreds of kb (Rao et al., 2014; Grubert et al., 2020), help to regulate gene expression by enabling enhancer-promoter interactions and insulating regulatory elements from neighbouring regions to prevent abnormal activities (Grubert et al., 2020). At the megabase (Mb) scale, chromatin is organised into Topological Associating Domains (TADs), which are characterised by frequent interactions between genomic loci within the same TAD and limited inter-domain interactions (see Figs. 1a, d) (Szabo et al., 2019). Chromatin may undergo further condensation to form chromatids, and two identical sister chromatids together constitute a chromosome.

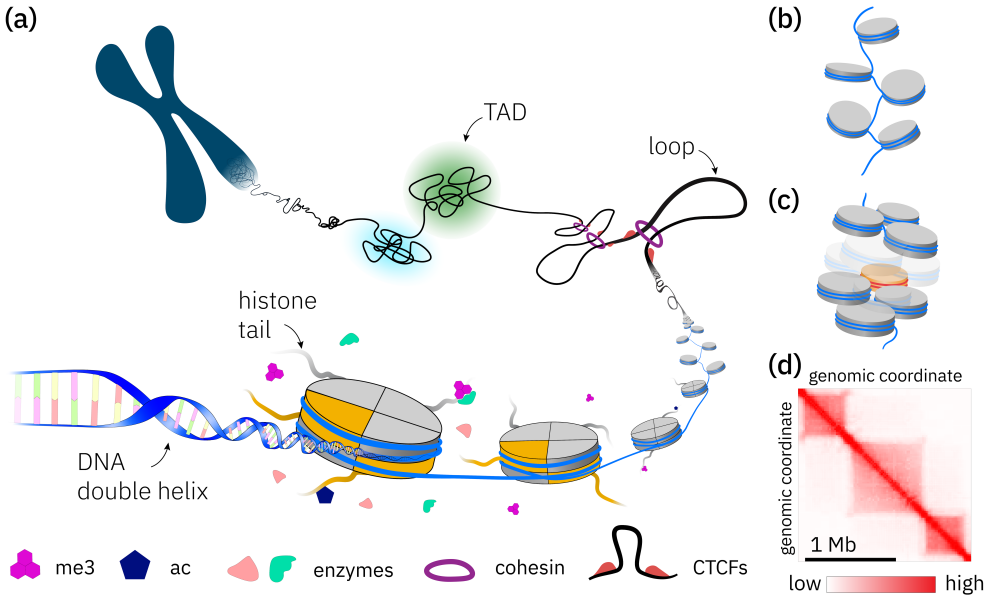

Fig. 1 **a** A cartoon illustrating the hierarchy of DNA packaging. Organisation of the DNA double helix (in blue) begins with it wrapping around nucleosomes, octamers consisting of eight histone proteins (the H3 core histones are highlighted in orange). The DNA fibre in each structural subunit comprises 147 base pairs (bp) wrapped around the nucleosome core and linker DNA, which can vary from a few nucleotides to 80 bp (Gross et al., 2015). The complex of DNA and its associated proteins is called chromatin. Further DNA compaction involves chromatin coiling to form a denser 30-nm fibre (illustrated in black). Higher-order structures, such as loop domains, can form with the help of additional protein complexes, such as CTCF binding sites and cohesin (Gross et al., 2015). CTCF insulators, oriented in a convergent manner, function as barriers for loop extrusion mediated by ring-shaped cohesin complexes that cannot slide past the CTCF sites. Subsequent compartmentalisation of chromatin involves its separation into Topological Associating Domains (TADs), which are characterised by frequent interactions between genomic loci within the TAD, whereas interactions between different TADs are limited (see also **(d)**). This schematic also depicts the addition of epigenetic marks, such as trimethyl (me3) and acetyl (ac), to histone terminal tails mediated by specific enzymes. **b, c** Two schematics showing open and compacted states of the chromatin fibre, respectively. One nucleosome whose DNA cannot be accessed for transcription is coloured in orange (panel **(c)**). **d** A characteristic map showing interaction frequencies between distinct genomic loci. High-frequency diagonal interactions correspond to contacts between neighbouring genomic loci, while non-local square domains of increased interaction frequencies correspond to higher-order chromatin organisational units such as TADs (Szabo et al., 2019)

The packaging of DNA filaments is tightly coupled with the DNA's epigenetic state. Epigenetics regulate which genomic loci are accessible for transcription, without modifying the chromosomal DNA sequence. It involves the addition of specific epigenetic marks to the DNA itself (e.g. DNA methylation) or to the tails of core histones that induce conformational changes in the three-dimensional folding of the chromatin fibre, thus changing its accessibility for binding of transcription factors. One of the most common epigenetic modifications involves covalent modifications of residues of histone tails. Histone tails are long terminal sequences of amino acids which protrude from each core histone and at which post-translational changes can occur (Fig. 1a) (Gross et al., 2015). As the balance between different epigenetic marks changes, the chromatin

architecture also changes to facilitate or inhibit access to genetic information. For example, the addition of acetyl groups to lysine residues at histone tails (acetylation) generally facilitates chromatin accessibility (Strahl and Allis, 2000). The cartoons in Figs. 1b-c show open and compacted genomic loci, respectively. While open chromatin conformations allow access to the DNA (panel b), genetic information encoded in loci that are tightly coiled cannot be accessed (panel c). Chromatin architecture, in turn, also influences the epigenetic profiles via feedback between different epigenetic marks (Sneppen and Ringrose, 2019). In this way, if the configuration of chromatin folding allows distant genomic loci to interact in space, feedback between existing epigenetic modifications at these loci can alter their state (Sneppen and Ringrose, 2019).

Epigenetics plays a key role in the development and maintenance of organism homeostasis. Without changing the underlying DNA sequence, it enables cells to adjust their gene expression profile according to their specialised function and environmental cues. It is directly implicated in cell differentiation, lineage commitment, maintenance of cell identity and adaptation to environmental changes. Specifically, a conceptual model proposed by (Waddington, 1957) illustrates cell differentiation as a trajectory of a ball rolling down a hilly terrain. Hills and valleys in this metaphor define the epigenetic landscape, with each valley representing a distinct cell phenotype characterised by a specific pattern of open and silenced gene loci. This type of 'rugged' epigenetic profile, characterised by alternating patterns of activating and repressive epigenetic modifications, plays a crucial role in determining cellular identity. Dysregulation of epigenetic mechanisms can lead to cancer, neurological disorders, metabolic dysregulation and autoimmune diseases (Zoghbi and Beaudet, 2016; Tollefsbol, 2018; Zhang et al., 2020). Despite extensive research (Felsenfeld, 2014), many aspects of epigenetic regulation remain poorly understood. In particular, data are typically obtained as snapshots at specific time points, and little is known about the interplay between the dynamics of epigenetic modifications and chromatin architecture. Theoretical modelling offers a complementary approach that can be used to gain insight into these interactions and the formation of epigenetic signatures.

Mathematical and computational modelling has been used to understand the mechanisms underlying key features of epigenetic regulation at the scale of small chromatin regions ($10-60$ nucleosomes, or equivalently $2-12$ kb). These locus-specific models focus on either covalent modifications of histone tails (Dodd et al., 2007; Sneppen et al., 2008; David-Rus et al., 2009; Sneppen and Dodd, 2012; Sneppen and Ringrose, 2019; Zhang et al., 2019), DNA methylation (Sneppen and Dodd, 2016; Chen et al., 2021), or both (Thalheim et al., 2017, 2018). A crucial requirement of these models is their ability to exhibit bistability, allowing the formation of epigenetic landscapes corresponding to open and closed chromatin states. These states must remain stable under mitotic perturbations, where approximately half of the nucleosomes are replaced by newly synthesised ones that lack epigenetic modifications. Robustness in transmitting epigenetic information through cell division is referred to as epigenetic memory. One class of theoretical models focuses specifically on the formation of bivalent chromatin (e.g. (Zhao et al., 2021; Sneppen and Ringrose, 2019; Alarcón et al., 2021)), characterised by a mixture of activating and repressive marks. Bivalent chromatin is believed to preserve the developmental potential of stem cells by keeping

key regulatory genes in a poised, undecided state. This poised state allows for rapid and flexible gene expression during development, enabling cells to quickly activate or silence genes in response to signals that determine their specific lineage (Macrae et al., 2023). Sneppen and Ringrose (2019) conducted an extensive literature review to summarise known feedback mechanisms between different epigenetic modifications. They then formulated a computational model to investigate whether these interactions could lead to the formation of bivalent chromatin. Their simulation results suggest that the experimentally observed mixture of activating and repressive marks is due to the rapidly switching bistable dynamics rather than the existence of a chromatin state with intermediate levels of both types of epigenetic marks. Another class of models investigates the interplay between epigenetic regulation and gene expression (Folguera-Blasco et al., 2019; Chen et al., 2021; Alarcón et al., 2021). These models investigate how epigenetic modifications influence gene activity and contribute to the complex regulatory networks governing cellular function and identity.

In recent work, Newar et al. (2022) introduced a stochastic model to study the mechanisms driving the formation of epigenetic patterns of covalent modifications of H3K27 (histone H3 at lysine residue 27) around transcription start sites. By exploring experimental data on enzyme occupancy around target sites, they defined the binding profiles of enzymes that participate in the epigenetic modification of H3K27. The authors also introduced a matrix for contact frequency of interactions between different genomic loci. Nonetheless, for simplicity, they neglected any locus-specific chromatin architecture and assumed that the probability of contact between two genomic coordinates is inversely proportional to the distance along the chromatin polymer that separates them. Their simulation results confirmed that binding profiles of histone-modifying enzymes allow for the emergence of methylation and acetylation patterns in chromatin regions around transcription sites similar to those observed experimentally (Newar et al., 2022).

Another type of computational model focuses exclusively on the three-dimensional genome architecture. These studies employ polymer-based models to elucidate the specific features of the chromatin polymer required to replicate experimental data on chromatin conformation (e.g. (Buckle et al., 2018)). More extensive reviews of existing models of epigenetic regulation can be found in (Cortini et al., 2016; Ringrose and Howard, 2017).

Most existing models (Dodd et al., 2007; Sneppen et al., 2008; David-Rus et al., 2009; Sneppen and Dodd, 2012, 2016; Thalheim et al., 2017, 2018; Sneppen and Ringrose, 2019; Zhang et al., 2019) focus on the stability of a uniform epigenetic state in an isolated locus, whereas the mechanisms underlying the formation of rugged epigenetic landscapes, which consist of patterns of open and closed chromatin loci, remain largely unexplored. Additionally, existing models frequently neglect the influence of chromatin architecture on the formation of epigenetic profiles. Interactions between nucleosomes in these models are simplified to either nearest-neighbour or all-to-all interactions. Several models also account for distance-dependent contact between genomic sites, assuming that the frequency at which these contacts occur decreases with separation distance (Dodd et al., 2007; Sneppen and Dodd, 2016; Newar et al., 2022; Nickels and Sneppen, 2023). In practice, however, chromatin folding is more

complex and can significantly influence the deposition of epigenetic marks through the activity of various 'reader-writer' ('reader-eraser') enzymes (Sneppen and Ringrose, 2019). These enzymes can recognise existing modifications and catalyse the deposition of new marks (or the removal of existing marks) at contact sites. Thus, chromatin folding can cause non-local feedback mechanisms that affect the spatial distribution of epigenetic modifications.

In this work, we propose a novel mesoscopic approach that considers epigenetic regulation between genomic loci of fixed length (on the scale of kb), each corresponding to DNA wrapped around several nucleosomes. We focus on covalent modifications of histone H3, accounting for covalent modifications of H3 tails at lysine residues 4 and 27 (H3K4 and H3K27, respectively). Trimethylation of H3K27 (H3K27me3) is recognised as a repressive mark, whereas acetylation of this residue (H3K27ac) and trimethylation of H3K4 (H3K4me3) represent activating epigenetic modifications. Despite our focus on these three modifications, we generalise their behaviour to represent broader classes of epigenetic marks, which result in closed and open chromatin states, respectively. Following previous work (e.g. (Dodd et al., 2007; David-Rus et al., 2009; Sneppen and Ringrose, 2019)), our model integrates feedback mechanisms governing the addition and removal of these marks, influenced by the availability of catalysing enzymes and also by chromatin folding. We define chromatin architecture, a key input to our model, via a matrix of contact strengths reflecting interactions in 3D space between different genomic sites.

Experimental data for generating interaction frequency heatmaps among genomic loci (Fig. 1d) can be acquired through chromatin conformation capture techniques. Variants of this method offer insights into chromatin organisation at different scales and resolutions (Sati and Cavalli, 2017). For instance, the Hi-C method combines DNA proximity ligation with high-throughput sequencing on a genome-wide scale to capture pairwise interactions between loci (Lieberman-Aiden et al., 2009), typically at the resolution of several kb (in the original work, the resolution was 1 Mb (Lieberman-Aiden et al., 2009)).

The main goal of our work is to understand the mechanisms that drive the formation of rugged (inhomogeneous) epigenetic landscapes, with a particular focus on the roles played by chromatin conformation and enzyme competition. The structure of the paper is as follows. In Sect. 2, we present our modelling framework, formulated in a stochastic setting, to describe the dynamics of epigenetic modifications for a given chromatin conformation. We also derive the mean-field deterministic equations associated with the stochastic model. We use the mean-field equations to perform stability analysis for two interacting genomic sites. This analysis, presented in Sect. 3.1, enables us to identify parameter regimes corresponding to different epigenetic landscapes: open, closed, bivalent, and rugged epigenetic landscapes. Section 3.2 discusses the scalings which must be introduced to extend our model to multiple interacting sites. The multi-site model accounts for chromatin architecture, availability of enzymes, and their competition for binding to target sites. We then perform stability analysis and stochastic simulations for the particular case of two interacting regions spanning several genomic sites (Sect. 3.3). We conclude in Sect. 4 with a summary of our findings and directions for future work.

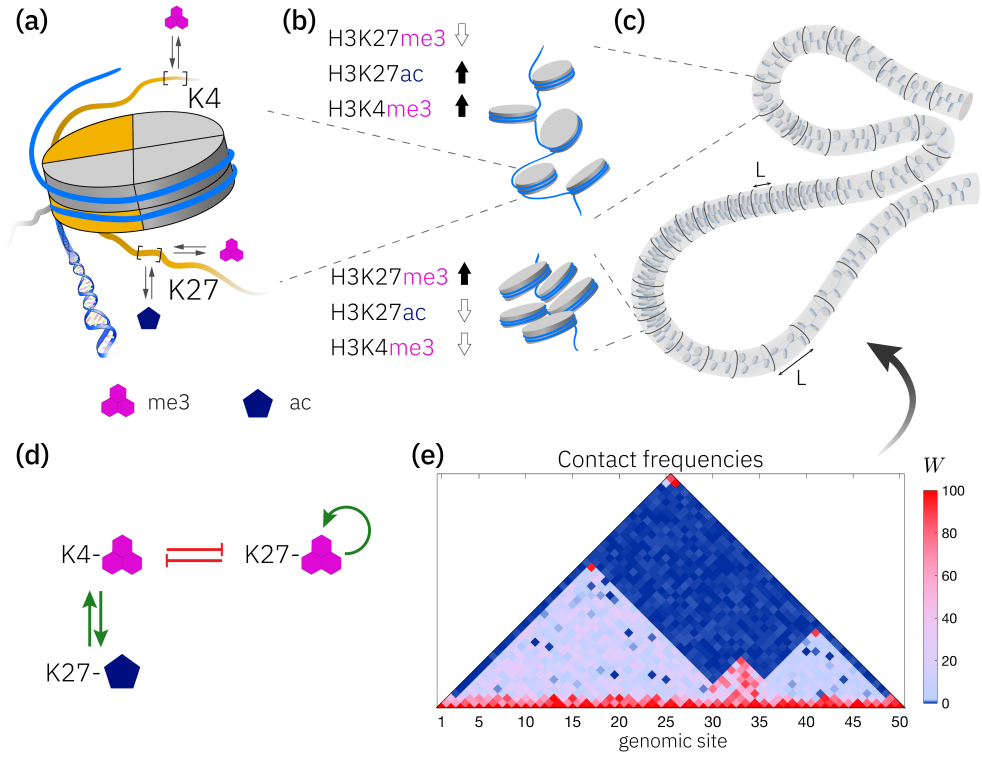

Fig. 2 A graphical summary of the modelling components. **a** A schematic highlighting the epigenetic modifications of lysine residues of the H3 histone (in orange) considered in our model. **b** We employ a mesoscopic approach in which the chromatin polymer is divided into genomic regions of fixed length, L (kb), which may contain multiple nucleosomes. Each genomic region can be characterised by its epigenetic landscape. In particular, compacted chromatin regions are characterised by high levels of H3K27me3 repressive marks, whereas open chromatin contains high levels of activating marks, H3K4me3 and H3K27ac. **c** A cartoon illustrating the division of the chromatin fibre into subregions of constant length, L. **d** Epigenetic marks are known to affect the dynamics of histone modifications via multiple reinforcing and antagonistic feedback mechanisms that are mediated by reader-writer (-eraser) enzymes (Sneppen and Ringrose, 2019). This diagram depicts the crosstalk between different epigenetic marks accounted for in this work. Green arrows indicate reinforcing interactions, while red, flat-end arrows indicate inhibitory interactions. **e** An example of a contact frequency map defining chromatin folding (see **(c)**), which serves as an input parameter in our model. Only the upper-diagonal part of the matrix, W, is shown due to its symmetry. The elements of the matrix are dimensionless

2 Mathematical model

In this section, we describe our model for epigenetic regulation (see Fig. 2). As mentioned above, we focus on the epigenetic regulation of two lysine residues of histone H3, namely, H3K4 and H3K27. The dynamics of epigenetic marks at these two sites attract particular attention due to their role in regulating gene expression (Miller and Grant, 2012). Figure 2a illustrates those post-translational modifications of H3 residues that are included in the model – trimethylation of H3K4 and H3K27 (H3K4me3 and H3K27me3, respectively) and acetylation of H3K27 (H3K27ac). Although we consider

only three epigenetic modifications, we view them as representative of generic classes of marks with similar behaviour. For simplicity, we do not account for the intermediate stages of adding and removing individual methyl groups.

We also take into account the spatial folding of the chromatin fibre. Specifically, we decompose the chromatin polymer into fixed-length genomic regions that can include multiple nucleosomes (see Figs. 2b-c). In this study, we assume that each genomic site spans a DNA chain of length L kb (Fig. 2c). This mesoscopic approach enables us to account for non-local interactions between distinct epigenetic marks. The pattern of these non-local interactions is given by the shape of chromatin folding, which can allow two genomic regions to be in close physical proximity without being neighbours in the DNA chain. Figure 2c depicts a simple domain with loops which bring into close 3D proximity genomic loci that are separated by many nucleotides along the genome. In general, interactions between different genomic sites are specified by contact frequency maps similar to those obtained through chromatin conformation capture techniques. Figure 2e shows an example of such a map. Since the contact frequency matrix is symmetric, it suffices to plot only its upper triangular part, which is commonly rotated by 45° to align the main diagonal horizontally. This approach is widely used in experimental studies and will be employed in our work (Fig. 2e). Higher frequency contacts at the diagonal entries (along the x-axis) indicate more frequent interactions between neighbouring loci, while loops and TADs are characterised by high contact frequencies off the main diagonal. In our work, the contact frequency map, W, is assumed to be fixed. While epigenetic marks are known to affect chromatin folding, here, we consider only the forward problem of determining the epigenetic landscape for a given (fixed) chromatin architecture.

Each genomic site is characterised by its pattern of epigenetic marks. In particular, closed chromatin regions are known to carry H3K27me3 repressive marks (Fig. 2b, bottom panel), while open chromatin is typically characterised by acetylation of H3K27 and trimethylation of H3K4 (see Fig. 2b, upper panel). Regions with excess of H3K27ac and H3K4me3 marks are associated with activation of the gene expression within those regions. We will also distinguish bivalent chromatin, which is characterised by a mixture of repressive and activating marks.

In what follows, $i = 1, \ldots, N$ indicates the index of a particular genomic region, where N is the total number of regions considered. Table 1 lists the epigenetic modifications considered in our model along with the enzymes responsible for adding and removing these marks (Sneppen and Ringrose, 2019). Specifically, the addition of trimethyl groups to H3K4 is catalysed by the Mixed-Lineage Leukemia (MLL) complex of the Trithorax group, while removal of this mark is mediated by the histone lysine demethylase (KDM) family of enzymes. For the lysine residue H3K27, we include its trimethylation by the methyltransferase EZH2 (Enhancer of Zeste 2), which is an enzymatically-active component of the Polycomb Repressor Complex 2 (PRC2). Removal of this mark is mediated by the demethylase UTX (Ubiquitously Transcribed Tetratricopeptide Repeat on chromosome X). The acetyltransferase CBP (CREB-binding protein) catalyses addition of the acetyl group to H3K27, whereas deacetylation of this residue is mediated by the NuRD (Nucleosome Remodeling and Deacetylase) complex and the sirtuin family (SIRT) of proteins (Park and Kim, 2020).

A comprehensive review of histone-modifying enzymes and the feedback mechanisms between them for Drosophila and vertebrates is provided in (Sneppen and Ringrose, 2019).

Table 1 Histone modifications and histone-modifying enzymes considered in the model. The corresponding model variables are indicated in blue. We detail enzymes known to add and remove epigenetic marks in vertebrates (Sneppen and Ringrose, 2019)

Unmodified residues	**Modifications**	**Writer enzyme (free)**	**Eraser enzyme (free)**
H3K4, S_4	H3K4me3, P_A	MLL, E_M	KDM, E_K
H3K27, S_{27}	H3K27me3, P_I	EZH2, E_Z	UTX, E_U
	H3K27ac, Q	CBP, E_C	NuRD and SIRT, E_S

Free enzyme refers to the enzyme molecules that are not bound to its substrate (lysine residues in the case of our model).

We model (de-)methylation and (de-)acetylation of lysine residues as enzymatic reactions catalysed by (eraser) writer enzymes (see Table 1). Each reaction leads to the formation of an intermediate lysine residue-enzyme complex, C_{J_i}, which later yields the product reactant and the enzyme that initiated that reaction. If we introduce the subscript, $J \in \{M, Z, C, K, U, S\}$ (see Table 1) to indicate the enzyme that catalyses a particular reaction, then the following reactions represent the addition of epigenetic marks

$$S_{27_i} + E_Z \underset{\bar{k}_{2i}}{\overset{\bar{k}_{1i}}{\rightleftarrows}} C_{Z_i} \xrightarrow{\bar{k}_3} P_{I_i} + E_Z, \tag{1}$$

$$S_{4_i} + E_M \underset{\bar{k}_{5i}}{\overset{\bar{k}_{4i}}{\rightleftarrows}} C_{M_i} \xrightarrow{\bar{k}_6} P_{A_i} + E_M, \tag{2}$$

$$S_{27_i} + E_C \underset{\bar{k}_8}{\overset{\bar{k}_{7i}}{\rightleftarrows}} C_{C_i} \xrightarrow{\bar{k}_9} Q_i + E_C, \tag{3}$$

and the following reactions their removal:

$$P_{I_i} + E_U \underset{\bar{k}_{11}}{\overset{\bar{k}_{10}}{\rightleftarrows}} C_{U_i} \xrightarrow{\bar{k}_{12}} S_{27_i} + E_U, \tag{4}$$

$$P_{A_i} + E_K \underset{\bar{k}_{14}}{\overset{\bar{k}_{13}}{\rightleftarrows}} C_{K_i} \xrightarrow{\bar{k}_{15}} S_{4_i} + E_K, \tag{5}$$

$$Q_i + E_S \underset{\bar{k}_{17}}{\overset{\bar{k}_{16}}{\rightleftarrows}} C_{S_i} \xrightarrow{\bar{k}_{18}} S_{27_i} + E_S. \tag{6}$$

Here, the rate parameters $\bar{k}_r$ for $r = \{3, 6, 8, 9, \ldots, 18\}$, are assumed to be constant, while $\bar{k}_r$ for $r = \{1i, 2i, 4i, 5i, 7i\}$ account for reinforcing and antagonistic interactions of unmodified lysine residues with existing histone modifications (Sneppen and Ringrose, 2019). Since Eqs. (1)-(6) define a counting process, the molecular species participating in these equations are dimensionless quantities, and all the rate constants $\bar{k}_r$ have units of $\left[t^{-1}\right]$. Figure 2d illustrates the feedback mechanisms incorporated in our model. The crosstalk between different histone modifications is mediated by specific reader-writer and reader-eraser enzymes. These enzymes typically possess both reader and writer domains. The reader domain is capable of recognising specific epigenetic marks, which can induce allosteric changes within the enzyme and, in turn, modulate the activity of the writer domain. Depending on the nature of the feedback, these allosteric changes can either enhance or inhibit the catalytic activity of the writer domain, thereby influencing the addition or removal of epigenetic marks at nearby genomic loci (Lee et al., 2018; Kim et al., 2013; Schmitges et al., 2011). For reinforcing feedback, the recognition of a mark by the reader domain boosts the enzyme's ability to deposit the same or similar marks (e.g. (Lee et al., 2018)). Conversely, for antagonistic feedback, the presence of an opposing mark can disrupt the enzyme's catalytic function, either by directly inhibiting its activity or destabilising the enzyme-chromatin interaction (Kim et al., 2013; Schmitges et al., 2011; Finogenova et al., 2020). Chromatin folding can enhance these interactions by bringing distant genomic sites into close spatial proximity (see Fig. 2e).

Specifically, we assume that trimethylation of H3K27 is self-reinforcing (Sneppen and Ringrose, 2019; Lee et al., 2018). Consequently, rate constant, $\bar{k}_{1i}$, in Eq. (1) is upregulated if the same mark, H3K27me3, is present in the vicinity of genomic site i (see Eq. (7)). The proximity in 3D space of two genomic loci is defined by their contact frequency (see Fig. 2e). Modifications H3K27me3 and H3K4me3 are known to inhibit each other (Sneppen and Ringrose, 2019; Kim et al., 2013; Schmitges et al., 2011; Finogenova et al., 2020) (Fig. 2d). Therefore, we assume that the dissociation rates of the lysine residue-enzyme complex, $\bar{k}_{2i}$ and $\bar{k}_{5i}$, (from Eqs. (1)-(2)) are increasing functions of their antagonising epigenetic mark (H3K4me3 and H3K27me3, respectively; see Eq. (7)). Additionally, the activating marks, H3K4me3 and H3K27ac, have been reported to reinforce each other's production (Sneppen and Ringrose, 2019) (Fig. 2d). Therefore, we assume that the rate constants, $\bar{k}_{4i}$ and $\bar{k}_{7i}$, for the formation of complexes to catalyse the deposition of trimethyl in H3K4 and acetyl in H3K27, respectively, (Eqs. (2)-(3)) are functions of the corresponding reinforcing mark. While these assumptions provide one possible way of incorporating feedback mechanisms into the model, we expect alternative functional forms will lead to qualitatively similar model behaviours. The above feedback mechanisms are incorporated into our model as follows:

$$\bar{k}_{1i} = \bar{k}_1 \left(\bar{k}_{1_0} + \sum_{l=1}^{N} \overline{w}_{il} P_{I_l} \right), \qquad \bar{k}_{2i} = \bar{k}_2 \left(\bar{k}_{2_0} + \sum_{l=1}^{N} \overline{w}_{il} P_{A_l} \right),$$

$$\bar{k}_{4i} = \bar{k}_4 \left(\bar{k}_{4_0} + \sum_{l=1}^{N} \overline{w}_{il} Q_l \right), \qquad \bar{k}_{5i} = \bar{k}_5 \left(\bar{k}_{5_0} + \sum_{l=1}^{N} \overline{w}_{il} P_{I_l} \right), \tag{7}$$

$$\bar{k}_{7i} = \bar{k}_7 \left(\bar{k}_{7_0} + \sum_{l=1}^{N} \overline{w}_{il} P_{A_l} \right).$$

The functional form of the rate constants appearing in Eq. (7) has been motivated by the existing models (Dodd et al., 2007; Sneppen and Ringrose, 2019; Zhao et al., 2021) and extended to account for the influence of chromatin folding. Specifically, we assume that $\bar{k}_r$ with units of $\left[t^{-1}\right]$ sets the general rate for the enzyme (un)binding while the dimensionless term in the parenthesis can be viewed as enzyme activity at the given genomic coordinate. The first term, given by $\bar{k}_{r_0}$, sets the *baseline* for enzyme activity, and the second term corresponds to the *recruited* activity due to existing epigenetic marks at neighbouring site l that can recruit reader-writer enzymes and enhance or decrease the stability of the enzyme complexes at site i (Dodd et al., 2007; Sneppen and Ringrose, 2019; Zhao et al., 2021). The matrix $\overline{W} = \{\overline{w}_{il}\}_{il}$ accounts for the spatial embedding of the chromatin fibre in 3D space and describes the pattern of interactions among genomic regions, both proximal and distant along the fibre. Its dimensionless elements, $\overline{w}_{il}$, quantify the strength of feedback from site l, based on the number of relevant histone modifications present at this neighbouring genomic locus.

In this work, we will consider two specific choices for $\overline{W}$. In the first scenario, we assume that the interaction strength between genomic sites i and l is determined by their contact frequency, w_{il}, and the level of H3K27me3 at those sites. The entries in contact frequency matrix, $W = \{w_{il}\}_{il}$ (dimensionless), illustrated in Fig. 2e, are assumed to be fixed. These data can be obtained experimentally using chromatin conformation capture techniques. The formation of loops and non-local interactions between chromatin loci has been reported to be mediated by Polycomb Repressor Complex 1 (PRC1) recruited by trimethylation of H3K27 (Loubiere et al., 2019). This motivated us to assume that the contact strength between sites i and l is proportional to the level of this histone modification, i.e. $\overline{w}_{il} = w_{il} P_{I_i} P_{I_l}$. For the second scenario, we assume that the interaction strength depends solely on the contact frequency between the relevant genomic loci, i.e. $\overline{w}_{il} = w_{il}$. Nonetheless, for most of our results, we use the H3K27me3-dependent interaction between genomic sites, unless otherwise stated. We also note that while the functional form of the rate constants in Eq. (7) is unbounded with respect to neighbouring epigenetic marks, the model operates under the assumption of competition for enzyme binding within the simulated genomic region, which imposes a natural saturation limit. Additionally, contact frequencies between genomic loci decrease exponentially with their linear distance along the chromatin polymer. As a result, extending the genomic domain beyond the scale of TADs would not lead to unbounded interaction rates between sites.

The reactions, given by Eqs. (1)-(6) are complemented by the conservation laws for the total number of lysine residues (unmodified and modified) at each genomic region i,

$$S_{27_0} = S_{27_i} + P_{I_i} + Q_i + C_{Z_i} + C_{U_i} + C_{C_i} + C_{S_i}, \tag{8}$$
$$S_{4_0} = S_{4_i} + P_{A_i} + C_{M_i} + C_{K_i}, \tag{9}$$

and the total numbers of the writer (eraser) enzymes,

$$E_J + \sum_{l=1}^{N} C_{J_l} = E_{J_0}, \qquad J \in \{M, Z, C, K, U, S\}, \tag{10}$$

where the summation is taken over all the genomic regions, $l = 1, \ldots, N$. Here, E_{J_0}, $J \in \{M, Z, C, K, U, S\}$, represents the total amount of enzyme J in the system. We note that our approach explicitly accounts for the competition of enzyme binding to the available lysine residues.

To simplify the model and reduce its dimensionality, we follow the Briggs-Haldane approach and assume the following scalings for the model variables (Briggs and Haldane, 1925; Keener and Sneyd, 2009). Lysine residues (modified and unmodified) scale with Ω_S representing the characteristic number of lysine residues in a genomic site, i.e. $(S_{27_i}, S_{4_i}, P_{I_i}, P_{A_i}, Q_i) = \Omega_S\,(s_{27_i}, s_{4_i}, p_{I_i}, p_{A_i}, q_i)$ for all i. By contrast, enzymes and the complexes they form scale with the characteristic enzyme level, Ω_E, which leads to $(E_J, E_{J_0}, C_{J_i}) = \Omega_E\,(e_J, e_{J_0}, c_{J_i})$ for all enzyme types, $J \in \{M, Z, C, K, U, S\}$, and all genomic sites, i. Based on the available experimental data and modelling results (Stafford et al., 2018; Owen et al., 2023), we assume that histone-modifying enzymes are less abundant compared to the numbers of lysine residues, i.e. $\Omega_E \ll \Omega_S$. In particular, Owen et al. (2023) showed that this is a key assumption to ensure that the general epigenetic pattern remains preserved under cell division (epigenetic memory). Under this scaling, the dynamics of intermediate enzymatic complexes are fast compared to those associated with product formation. As such, we can assume that the enzyme distributions are at a quasi-equilibrium on the slow timescale of epigenetic modifications. In the stochastic version of the quasi-steady state (QSS) approximation (Ball et al., 2006; Folguera-Blasco et al., 2019; Alarcón et al., 2021), the fast variables quickly settle on to their QSS distribution, which is conditioned to the (instantaneous) values of the slow variables. The evolution of the slow variables takes place on a much slower timescale than that required for the fast variables to reach the quasi-equilibrium. Their evolution takes a simplified form where the values of the fast variables are sampled from their QSS distribution. It can be shown (see Sect. I of Supplementary Material) that this results in multinomial QSS distributions for the probabilities of finding an enzyme J in a complex at one of the genomic sites or in a free state. The corresponding QSS probability density function (PDF), ϕ_J, reads:

$$\phi_J(E_J, C_{J_1}, \ldots, C_{J_N} \mid \mathbf{p_I}, \mathbf{p_A}, \mathbf{q}) = \frac{(E_{J_0})!}{(E_J)! \prod_{l=1}^{N} (C_{J_l})!} \; (p_{E_J})^{E_J} \; \prod_{l=1}^{N} \left(p_{C_{J_l}}\right)^{C_{J_l}}, \tag{11}$$

where E_{J_0} is the total amount of enzyme J (free and bound in complexes) and p_{E_J} and $p_{C_{J_l}}$ are the probabilities that an enzyme of this type is in a free or bound state at genomic site l, $l = 1, \ldots, N$. In particular, we have that $p_{E_J} + \sum_{l=1}^{N} p_{C_{J_l}} = 1$. Furthermore, in Eq. (11), we note the dependency of the QSS PDFs on the levels of existing histone modifications, $(\mathbf{p_I}, \mathbf{p_A}, \mathbf{q})$, where $\mathbf{p_I} = \{p_{I_1}, \ldots, p_{I_N}\}$ denotes the H3K27me3 landscape, $\mathbf{p_A} = \{p_{A_1}, \ldots, p_{A_N}\}$ denotes the levels of H3K4me3 marks and $\mathbf{q} = \{q_1, \ldots, q_N\}$ denotes the H3K27ac profile. A detailed derivation of the QSS approximation is included in Sect. I of Supplementary Material. For instance, the QSS probabilities for free and bound EZH2 methyltransferase, p_{E_Z} and $p_{C_{Z_i}}$, are given by the following expressions:

$$p_{E_Z}(\mathbf{p_I}, \mathbf{p_A}, \mathbf{q}) = \frac{1}{1 + \sum_{j=1}^{N} \frac{\kappa_{1j}}{\kappa_{2j} + \kappa_3} (s_{27_0} - p_{I_j} - q_j)}, \tag{12}$$

$$p_{C_{Z_i}}(\mathbf{p_I}, \mathbf{p_A}, \mathbf{q}) = \frac{\frac{\kappa_{1i}}{\kappa_{2i} + \kappa_3} (s_{27_0} - p_{I_i} - q_i)}{1 + \sum_{j=1}^{N} \frac{\kappa_{1j}}{\kappa_{2j} + \kappa_3} (s_{27_0} - p_{I_j} - q_j)}, \tag{13}$$

where κ_r, $r = \{1i, 2i, 3\}$, are appropriately rescaled rate constants. Here, the number of unmodified H3K27 residues available for enzyme binding at genomic region i is approximated by $(s_{27_0} - p_{I_i} - q_i)$.

The reactions for addition and removal of epigenetic modifications of the reduced model under the assumption of QSS distribution of enzyme complexes are then simplified as follows:

$$S_{27_i} \underset{\kappa_{12} c_{U_i}}{\overset{\kappa_3 c_{Z_i}}{\rightleftarrows}} P_{I_i}, \qquad S_{4_i} \underset{\kappa_{15} c_{K_i}}{\overset{\kappa_6 c_{M_i}}{\rightleftarrows}} P_{A_i}, \qquad S_{27_i} \underset{\kappa_{18} c_{S_i}}{\overset{\kappa_9 c_{C_i}}{\rightleftarrows}} Q_i. \tag{14}$$

Here, κ_r, $r = \{3, 6, 9, 12, 15, 18\}$, are appropriately rescaled rate constants. These reactions can be simulated using any algorithm for stochastic simulations (e.g. Gillespie algorithm (Gillespie, 1976)). At each time step of these simulations, the levels of enzymatic complexes, c_{J_i}, of type J are sampled from the QSS multinomial distribution (see Supplementary Material). Thus, our stochastic model consists of reactions for the dynamics of epigenetic marks given by Eq. (14) complemented by appropriate initial conditions for each genomic site i, $P_{I_i}(t = 0)$, $P_{A_i}(t = 0)$ and $Q_i(t = 0)$. Eq. (14) is

used for the stochastic simulations presented in this work. While we do not directly simulate the microscopic dynamics of the full model described in Eqs. (1)-(6), we note that deriving the model from the enzyme dynamics is essential to obtain the correct QSS probabilities needed to accurately simulate the reduced model from Eq. (14).

We also derive the associated mean-field equations for the dynamics of epigenetic modifications which we use to perform the stability analysis of our system. The resulting system of coupled ordinary differential equations is given by:

$$\begin{aligned}
\frac{\mathrm{d}p_{I_i}}{\mathrm{d}\tau} &= \kappa_3 e_{Z_0} p_{C_{Z_i}}(\mathbf{p_I}, \mathbf{p_A}, \mathbf{q}) - \kappa_{12} e_{U_0} p_{C_{U_i}}(\mathbf{p_I}), \\
\frac{\mathrm{d}p_{A_i}}{\mathrm{d}\tau} &= \kappa_6 e_{M_0} p_{C_{M_i}}(\mathbf{p_I}, \mathbf{p_A}, \mathbf{q}) - \kappa_{15} e_{K_0} p_{C_{K_i}}(\mathbf{p_A}), \\
\frac{\mathrm{d}q_i}{\mathrm{d}\tau} &= \kappa_9 e_{C_0} p_{C_{C_i}}(\mathbf{p_I}, \mathbf{p_A}, \mathbf{q}) - \kappa_{18} e_{S_0} p_{C_{S_i}}(\mathbf{q}).
\end{aligned} \tag{15}$$

Here, $i = 1, \ldots, N$ and we emphasise the dependency of the QSS probabilities, $p_{C_{J_i}}$, on the levels of epigenetic marks along chromatin, $(\mathbf{p_I}, \mathbf{p_A}, \mathbf{q})$, leading to nonlocal coupling between genomic sites. These equations are supplemented with initial conditions $\mathbf{p_I}(t = 0)$, $\mathbf{p_A}(t = 0)$ and $\mathbf{q}(t = 0)$. In our stochastic simulations, unless otherwise stated, we start with random initial conditions, where the initial levels of epigenetic marks are uniformly distributed on $[0, 1]$.

3 Results

3.1 Emergence of rugged epigenetic landscapes: Bifurcation analysis of the two-site model

As can be seen from Sect. 2, our model contains a large number of parameters. Although estimates exist for the (de-) methylation and (de-) acetylation rates of individual lysine residues (e.g. see (Newar et al., 2022) and references therein) for several cell types, little is known about the feedback between epigenetic modifications and the effects of chromatin folding. Epigenetic landscapes and chromatin organisation can vary significantly between organisms and cell differentiation stages. For these reasons, we focus on understanding the qualitative behaviour of our model without attempting to calibrate it against experimental data. To ensure that our predictions are robust, we performed an extensive parameter sweep to discern different modes of behaviour as enzyme levels, e_{J_0}, and contact frequencies, w_{il}, vary. We note that all Supplementary Figs. and Tables can be found in Sect. V of Supplementary Material.

We start by performing a model analysis for the simplified case when the entire chromatin domain under consideration is split into two interacting genomic loci, as shown in Fig. 3. Specifically, we used the mean-field, deterministic equations associated with our stochastic model (see Sect. II of Supplementary Material). Within this simplified scenario, we can perform an exhaustive bifurcation analysis of the system regarding the stable epigenetic landscapes that our model can produce. We are particularly interested in proving that our model produces both uniform and rugged

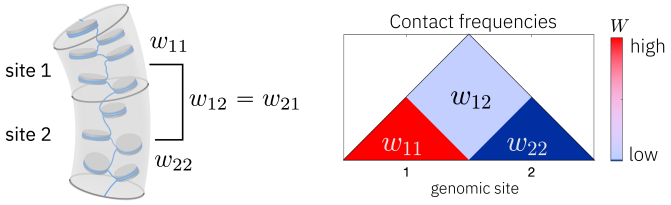

Fig. 3 A cartoon illustrating the two-site system with the frequency of interactions within site 1 (site 2), w_{11} (w_{22}), and the cross-interaction between the sites given by w_{12}. The cross-interactions are symmetric, i.e. $w_{21} = w_{12}$

epigenetic landscapes and in determining the parameter regimes that lead to either of them. Rugged epigenetic profiles are characterised by an alternating pattern of domains enriched in activating marks and regions exhibiting higher levels of repressive marks. We introduce a ruggedness measure, $\mathcal{R}$, which quantifies the heterogeneity of the epigenetic profiles for the two genomic sites:

$$\mathcal{R} = \max_{s.s.s.} \left(|p_{I_1} - p_{I_2}| \right). \tag{16}$$

Here, the maximum is taken over all stable steady states (s.s.s.) of the system for a given parameter set. Although this measure quantifies differences in H3K27me3 levels between sites, similar metrics can be considered for activating epigenetic marks, H3K4me3 and H3K27ac, due to their inverse relationship with H3K27me3 (see Fig. 2b).

We begin by imposing heterogeneous behaviour between the two sites, assuming that one site has a high frequency of interactions while the other is weakly connected. Thus, we fix $w_{11} = 200$ and $w_{22} = 0.01$ and vary the strength of the cross-interaction, w_{12}, and enzyme levels. In Supplementary Table 1, we list the baseline values of parameters used in our model, except where we explicitly note the variations in certain parameters. Figure 4 illustrates the characteristic behaviour of our system as we vary levels of EZH2 enzyme, e_{Z_0}, and the interaction strength between the sites, w_{12}. The phase diagram shown in Fig. 4a$_1$ delineates different stability regimes as these parameters change.

We determine that low levels of the methyltransferase, e_{Z_0}, are insufficient to induce trimethylation of H3K27, resulting instead in mostly hypomethylated chromatin configurations. At higher values of e_{Z_0}, the system switches to a stable steady state corresponding to hypermethylated chromatin. A rugged pattern, where one site is open, and the other closed, appears when $e_{Z_0} \gtrsim 0.3$ and cross-interactions, w_{12}, are low (see Fig. 4a$_2$). Additionally, given the values of w_{11} and w_{22}, site 1 is more connected and tends to be closed (high H3K27me3), while site 2 remains open (low H3K27me3). This pattern is shown in Fig. 4a$_3$, which reports the steady-state levels of this mark at both sites, p_{I_1} and p_{I_2}. The colouring in this plot indicates the maximum value of the real part of the eigenvalues of the linearised system for the corresponding steady state, which describes its local stability.

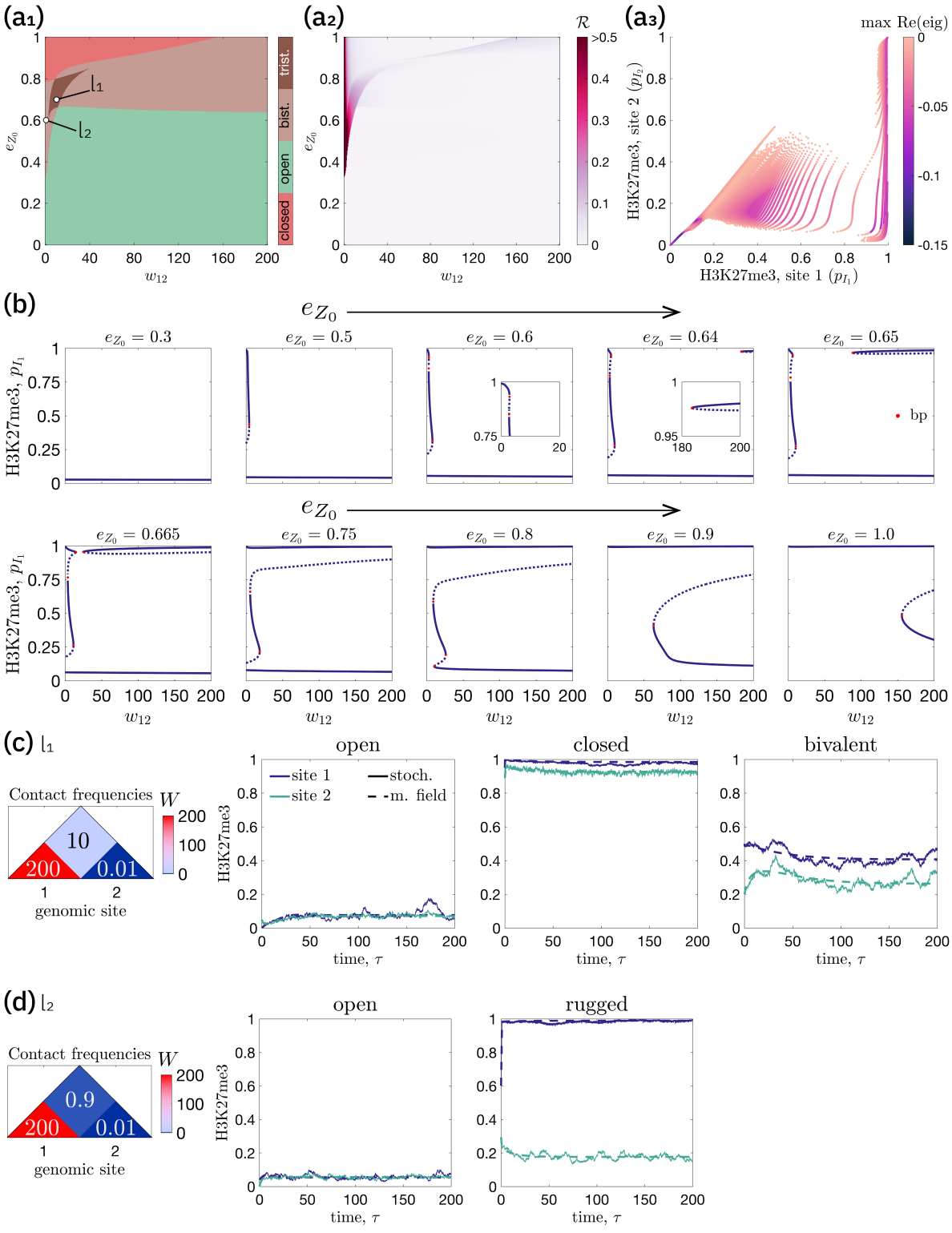

Fig. 4 *(Caption on the next page)*

Fig. 4 **a** Stability analysis of the two-site system in response to changes in levels of the enzyme EZH2, e_{Z_0}, and the interaction strength between sites 1 and 2, w_{12}. $\mathbf{a_1}$ A phase diagram where the monostable region corresponding to open (closed) chromatin is shown in green (red). A state is classified as open or closed depending on the equilibrium values of the epigenetic marks at site 1; an open (closed) state corresponds to $p_{I_1} < 0.3$ ($p_{I_1} > 0.7$). Bistable (tristable) parameter regions are indicated in (dark) brown. $\mathbf{a_2}$ A heatmap of the ruggedness metric, $\mathcal{R}$, defined by Eq. (16). $\mathbf{a_3}$ A scatter plot for the values of H3K27me3 at sites 1 and 2 for all stable equilibria obtained by varying e_{Z_0} and w_{12}. The colour bar indicates the maximum value of the real part of the eigenvalues of the linearised system. **b** A series of bifurcation plots showing how, for fixed values of e_{Z_0}, levels of H3K27me3 at site 1 change as w_{12} varies. Saddle-node bifurcation points are indicated by red dots (bp). **c**, **d** Temporal evolution of the two-site model for two choices of parameter values: $e_{Z_0} = 0.75$, $w_{12} = 10$ (labelled l_1 in $\mathbf{(a_1)}$) and $e_{Z_0} = 0.6$, $w_{12} = 0.9$ (l_2 label in $\mathbf{(a_1)}$). In each case, the stable equilibrium to which the system converges depends on the initial distribution of marks. We include trajectories obtained from stochastic simulations (solid lines) and the associated mean-field equations (dashed lines) for site 1 (deep blue) and site 2 (green). Except for e_{Z_0} and w_{il}, all parameters are fixed at the baseline levels listed in Supplementary Table 1

The transition to bi- and tristability can be better observed in the bifurcation plots shown in Fig. 4b. Here, we demonstrate the bifurcation curve for H3K27me3 levels at site 1, p_{I_1}, as a function of the interaction strength between the sites, w_{12}. The level of EZH2 in each panel is fixed to the value indicated in its title. The system exhibits several saddle-node bifurcations, which facilitate the appearance of the closed state for site 1 for higher values of EZH2 methyltransferase, e_{Z_0}.

It is noteworthy that our model also reproduces the emergence of the so-called bivalent state, which is characterised by intermediate levels of negative (H3K27me3) and positive marks (H3K4me3 and H3K27ac). The bivalent state appears less robust than the open and closed configurations (see Fig. 4a$_3$). We further confirmed this by plotting the distributions of the maximum value of the real part of the eigenvalues of the linearised system in a breakdown for open, closed, and bivalent chromatin. Supplementary Fig. 2a shows these distributions for the data presented in Fig. 4a. It can be seen that for all stable equilibria obtained by varying e_{Z_0} and w_{12}, intermediate bivalent states are characterised by the highest values of their maximum real part of the eigenvalues as compared to open and closed chromatin. Similar results are obtained for other parameter values (see Supplementary Figs. 2b-f, 3c-d and 4).

We illustrate the (stochastic) temporal evolution of the system for two particular combinations of parameters, one within the tristable region (Fig. 4c) and another one within the bistable region (Fig. 4d). These parameter combinations are referred to as l_1 and l_2, respectively (see Fig. 4a$_1$). For l_1, the system possesses three stable steady states corresponding to open, closed, and bivalent chromatin. The convergence of its time evolution to a specific equilibrium depends on the initial conditions. We observe similar tristable regions in the parameter space for variations of all other enzymes incorporated in the model and other kinetic parameters (Supplementary Figs. 1 and 3a-b). For all these cases, the stable equilibria in tristable regions correspond to open, closed, and bivalent chromatin. No rugged pattern is observed for the tristable regime (see Figs. 4a$_1$-a$_2$ and Supplementary Figs. 1 and 3a-b). For asymmetric self-interactions of $w_{11} = 200$ and $w_{22} = 0.01$, a rugged pattern of epigenetic profiles at sites 1 and 2 is observed in the mono- and bistable regimes. Figure 4d illustrates this behaviour for the bistable case (parameter combination, l_2).

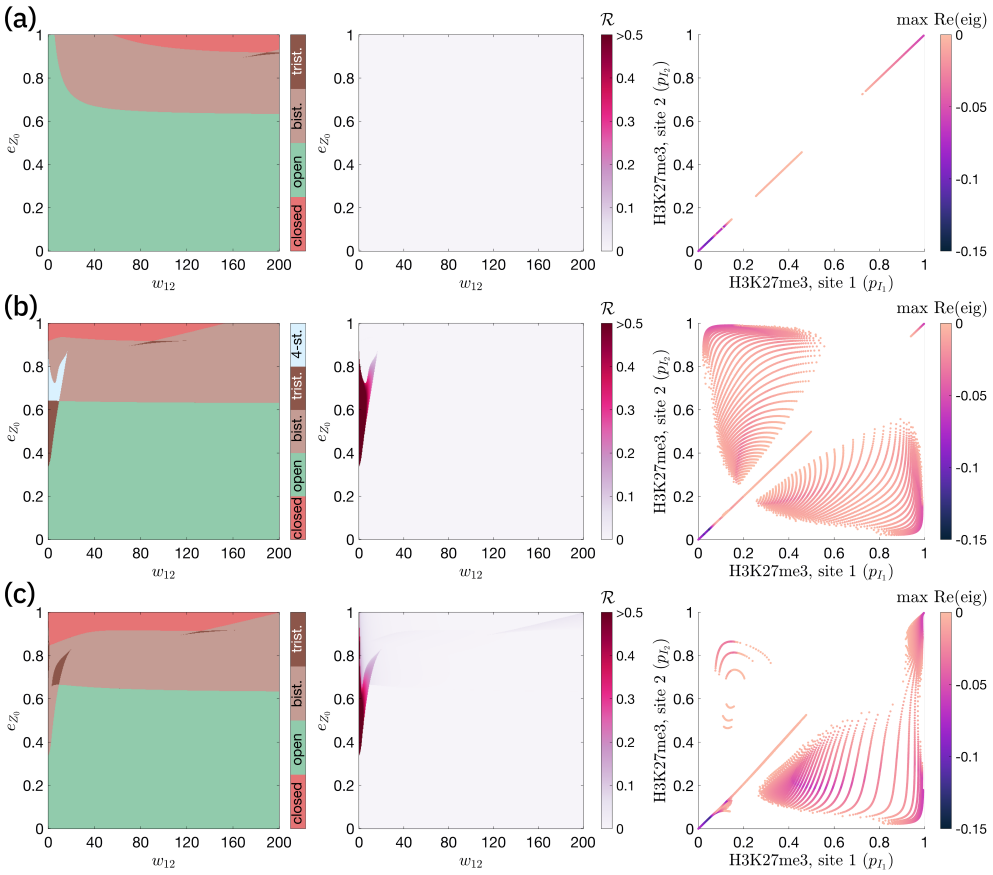

Fig. 5 Stability analysis of the two-site system for varying cross-interaction, w_{12}, and the EZH2 levels, e_{Z_0} for **(a)** $w_{11} = w_{22} = 1$, **(b)** $w_{11} = w_{22} = 100$, and **(c)** $w_{11} = 100$, $w_{22} = 10$. The case of $w_{11} = 200$, $w_{22} = 0.01$ is shown in Fig. 4a. The leftmost plots present phase diagrams with monostable regions corresponding to the open (closed) state are shown in green (red). The classification of the state as open or closed is determined by the equilibrium value of epigenetic marks at site 1; an open (closed) state corresponds to $p_{I_1} < 0.3$ ($p_{I_1} > 0.7$). The bistable and tristable parameter regions are indicated in brown and dark brown, respectively. In panel **(b)**, we observe the emergence of regions where there exist four stable equilibria (uniform open, uniform closed and two opposite rugged states). These regions are coloured in light blue. The middle plots show heatmaps for the value of the ruggedness metric, $\mathcal{R}$, defined by Eq. (16). The rightmost plots present scatter plots of the H3K27me3 levels at sites 1 and 2 for all stable equilibria included in the analysis of the corresponding panel. The colour bar indicates the maximum value of the real part of the eigenvalues of the linearised system. The remaining parameters are set to their baseline levels listed in Supplementary Table 1

We now revise the assumption of asymmetric strengths in connectivity within sites 1 and 2 and investigate the behaviour of the system when interaction frequencies in both sites have the same value, i.e. $w_{11} = w_{22}$. This scenario represents the worst-case setting for observing rugged patterns in epigenetic profiles. Figures 5a-c show these results for increasing frequency of self-interactions. We observe that low frequencies of self-interaction, i.e. $w_{ii} = 1$, fail to generate heterogeneous patterns of epigenetic marks

in sites 1 and 2 (see Fig. 5a). Here, low levels of EZH2 methyltransferase result in a uniform open configuration at both sites, while high levels lead to a uniformly closed configuration. Between these extremes lies a bistable region where both (homogeneous) stable equilibria coexist. The bivalent state can also be observed in a small region of tristability. We note the symmetry in the distributions of H3K27me3 marks for the two sites (all equilibria lie on the diagonal in Fig. 5a, right plot).

For larger values of $w_{11} = w_{22}$, a rugged landscape emerges when the cross-interaction, w_{12}, is small and levels of EZH2, e_{Z_0}, are high (Fig. 5b). By symmetry, two opposite rugged states emerge simultaneously, corresponding to equilibria falling outside the main diagonal in Fig. 5b (right plot). One rugged state corresponds to site 1 being in an open and site 2 in a closed configuration. The other rugged state is characterised by site 1 being closed and site 2 open. In this case, the rugged states lie within a tristable regime (coexisting with a uniform open state) and in regions with four stable equilibria (uniform open, uniform closed, and two rugged states). The two rugged states are also observed when the contact patterns are asymmetric, provided the self-interaction frequencies are sufficiently high. For example, in Fig. 5c, we fix $w_{11} = 100$ and $w_{22} = 10$, so interactions within site 2 are less frequent than for site 1. Nonetheless, Fig. 5c (right plot) confirms the existence of the second rugged state (high H3K27me3 for site 2 and low H3K27me3 for site 1). For completeness, we report the distributions of the maximum value of the real part of the eigenvalues for the data presented in Fig. 5 (see Supplementary Fig. 4).

We observe similar trends when all kinetic parameters and enzyme levels are fixed, and the contact frequencies, w_{il}, vary. Specifically, Supplementary Fig. 5 presents a stability analysis of the two-site system as w_{12} and w_{22} are varied for increasing interaction within site 1, w_{11}. It can be seen that larger values of cross-interaction, w_{12}, induce the formation of uniform epigenetic landscapes. In contrast, rugged patterns emerge for low w_{12} and when self-interaction in at least one site is sufficiently strong. Higher self-interaction values at both genomic sites support the formation of rugged landscapes for a larger range of w_{12} values.

Summarising, our model exhibits bistability where uniform open and closed states coexist for a wide range of parameter values. These regions can be seen in our stability maps in brown colour (Fig. 4a$_1$, leftmost plots in Fig. 5 and Supplementary Figs. 1 and 3a-b) whenever the corresponding values of the ruggedness metric (middle panels) are low. Rugged landscapes can be identified by higher values (greater than 0.5) of the ruggedness metric, $\mathcal{R}$. Typically, rugged equilibria appear in regions of multistability, accompanied by a uniform open and/or closed state(s), for relatively low values of cross-interaction between genomic sites. When the contact frequency between sites 1 and 2 increases, the system tends to have uniform epigenetic profiles (i.e. no ruggedness). Bivalent chromatin emerges in regions of bi- and tristability and is more sensitive to external perturbations. This feature is consistent with the function bivalent chromatin is normally assumed to perform, namely, facilitating rapid transitions to either uniform open or uniform closed equilibria that coexist with it. We did not observe coexistence of bivalent and rugged states. Instead, bivalent states appear to act as intermediaries, facilitating transitions between uniform and rugged landscapes (and vice versa). This behaviour can be observed, for instance, in the bifurcation plot

shown in Fig. 4b for $e_{Z_0} = 0.6$. As we trace the bifurcation curve, starting at high values of the cross-interaction parameter w_{12}, the system initially exhibits monostability, characterised by an open chromatin state. As w_{12} decreases, a second stable steady state, representing a bivalent state (intermediate values of p_{I_1}), emerges via a saddle-node bifurcation. With further reduction in w_{12}, site 1 in this bivalent state shifts towards a closed state (indicated by higher values of p_{I_1}). Finally, when w_{12} reaches lower values (e.g., $w_{12} = 0.9$ as in Fig. 4d), the system transitions to a regime in which open and rugged states coexist.

3.2 Emergence of rugged epigenetic landscapes in multi-site systems: A scaling approach

Under general conditions, an exhaustive bifurcation and parameter sensitivity analysis like the one carried out in Sect. 3.1 for a two-site system is unfeasible when we move on to the more realistic multi-site cases where $N \gg 1$. Rather than attempting to tackle this situation directly, our strategy is to leverage our understanding of the system's behaviour in response to parameter variations obtained from the stability analysis of the two-site system. To do so, we apply a scaling argument. This scaling must ensure the invariance of the system's behaviour as we change the number of genomic sites onto which the entire chromatin domain is decomposed. In particular, the rescaling of the system should enable us to increase the resolution of our system by subdividing the total chromatin domain into smaller bins. Thus, we now assume that the number of genomic sites is greater than two, i.e. $N > 2$.

The scaling argument goes as follows. In our model, since the total concentration of enzymes (free and chromatin-bound) is conserved, the number of genomic sites influences the competition for enzyme binding, and this is reflected in the QSS enzyme distributions of the type given by Eqs. (12)-(13) (see Supplementary Material for the complete list of QSS probabilities). They can be rewritten in the following general way:

$$p_{E_J} = \frac{1}{1 + \sum_{j=1}^{N} \widetilde{\kappa}_j \, s_{_j}}, \qquad p_{C_{J_i}} = \frac{\widetilde{\kappa}_i \, s_{_i}}{1 + \sum_{j=1}^{N} \widetilde{\kappa}_j \, s_{_j}}. \tag{17}$$

Here, $\widetilde{\kappa}_i$ denotes the inverse of a generalised Michaelis constant (Ingalls, 2013) for the enzymatic reaction catalysed by enzyme J and $s_{_i} \in [0, 1]$ is the normalised substrate level at binding position i. Inspection of the expression for p_{E_J} reveals that adding more binding sites will reduce the probability of finding enzyme in an unbound state to zero, i.e. $p_{E_J} \to 0$ as $N \to \infty$.

To avoid this dilution effect, we employ a scaling for which p_{E_J} remains fixed as the resolution is successively refined by dividing the chromatin domain into larger numbers of genomic sites. If then, for simplicity, we assume that all binding sites have the same affinity (i.e. $\widetilde{\kappa}_i = \widetilde{\kappa}$), this is equivalent to the following:

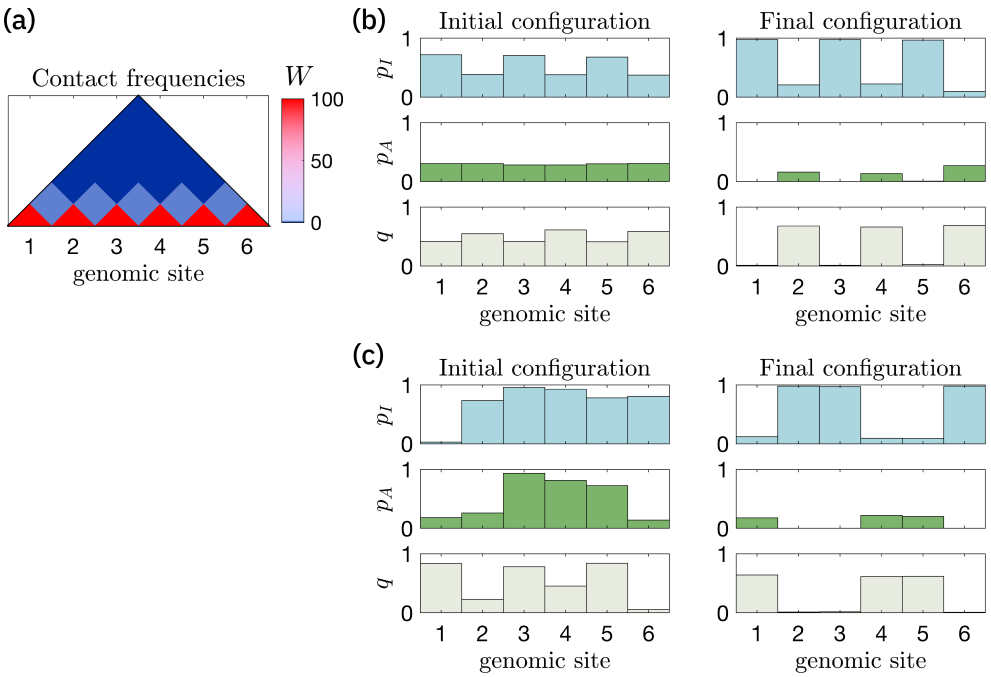

Fig. 6 Two stochastic simulations of the multi-site system for $N = 6$. The upper-diagonal part of the connectivity matrix, W, is shown in **(a)**. Here, $w_{ii} = 100$, $w_{il} = 0.9$ for nearest-neighbour interaction and $w_{il} = 0$ otherwise. **b** Initial and final configurations of a simulation initialised with alternating expressions of repressive and activating marks. **c** Initial and final configurations of a simulation with a random initial condition. Here, p_I denotes levels of H3K27me3, p_A – H3K4me3, and q – H3K27ac. In these simulations, we used $e_{Z_0} = 0.6$ (EZH2 methyltransferase) and the final simulation time, $\tau_{final} = 2000$ (dimensionless). The remaining parameters are set to their baseline levels listed in Supplementary Table 1

$$\lim_{N\to\infty} \frac{1}{N/2} \sum_{j=1}^{N} s_j = \text{const.} \tag{18}$$

Here, the factor $N/2$ is due to the fact that the initial analysis was performed for two sites.

The above scaling is used to rescale the amount of available substrate for all QSS probabilities and leads to the following transformation of Eq. (17):

$$p_{E_J} = \frac{1}{1 + \frac{2}{N}\sum_{j=1}^{N} \widetilde{\kappa}_j s_j}, \qquad p_{C_{J_i}} = \frac{\widetilde{\kappa}_i \frac{2}{N} s_i}{1 + \frac{2}{N}\sum_{j=1}^{N} \widetilde{\kappa}_j s_j}. \tag{19}$$

A full list of the rescaled QSS probabilities is included in Sect. III of Supplementary Material.

We illustrate the behaviour of the rescaled system by performing stochastic simulations of our model for $N = 6$, using the parameter set corresponding to tristability in Fig. 5b (with $e_{Z_0} = 0.6$ and $w_{12} = 0.9$). We recall that this parameter combination is characterised by the coexistence of a uniformly open and two rugged states. Figure 6 shows stochastic simulations for two different initial conditions. For a suitable choice of the initial conditions, we obtain a regular rugged pattern of alternating open and closed regions (Fig. 6b). For random initial conditions, the rugged pattern persists, but the distribution of open and closed loci varies (Fig. 6c). Nonetheless, a general trend establishes so that half of the loci are open (and half closed), forming a rugged pattern as predicted by the two-site analysis.

3.3 Stability-analysis for two-region system

The scaling argument introduced in Sect. 3.2 enables us to extend our model to an arbitrary number of genomic sites. However, instead of conducting an extensive exploration of characteristic chromatin conformations, we focus on a specific chromatin architecture involving two interacting regions, which allows for simulation via our stochastic model and stability analysis using the associated mean-field equations. We consider a chromatin domain comprising N genomic loci divided into two distinct regions. In region 1, genomic sites interact only within the same site, so that $w_{ii} = w_{diag}$, for $i = 1, \dots R$, where R is the size of region 1, and $w_{il} = 0$ for $i = 1, \dots R$, $l = 1, \dots R$ and $i \neq l$. Region 2 behaves like a TAD-like domain, with all-to-all interactions between genomic sites at a uniform frequency, w_{TAD}, i.e. $w_{il} = w_{TAD}$ for $i = R+1, \dots N$ and $l = R+1, \dots N$. We assume further that regions 1 and 2 do not interact. A schematic of this chromatin architecture is provided in Sect. IV of Supplementary Material. We note that, even though not all genomic sites interact in this setup, their dynamics are coupled due to competition for enzyme binding.

The mean-field equations associated with the stochastic model of this two-region scenario are provided in Sect. IV of Supplementary Material. Specifically, the uniform connectivity of all genomic sites within each region means that their evolution can be described by identical equations. Further, if we assume uniform initial conditions within each region, then the system's dynamics can be reduced to a system of two ordinary differential equations, each representing a distinct region. Consequently, we can perform stability analysis as in Sect. 3.1.

Figure 7a shows the results of the stability analysis as we vary w_{diag} and w_{TAD}, the components of the contact frequency matrix in regions 1 and 2, respectively. Given that TAD-like regions are typically characterised by a high frequency of contacts among genomic loci, we focus on the dynamics for intermediate-to-high values of w_{TAD}. Figure 7a shows that, for any fixed value of $w_{TAD} > 2$, the system undergoes a transition from one to two coexisting rugged states when $w_{diag} \approx 10$. Specifically, for low values of w_{diag}, the system exhibits bistability between uniformly open chromatin and rugged pattern 1. For rugged pattern 1, region 1, characterised by diagonal connections, is enriched with activating marks (H3K4me3 and H3K27ac), while the TAD-like region 2 shows high levels of repressive modifications (H3K27me3). As w_{diag}

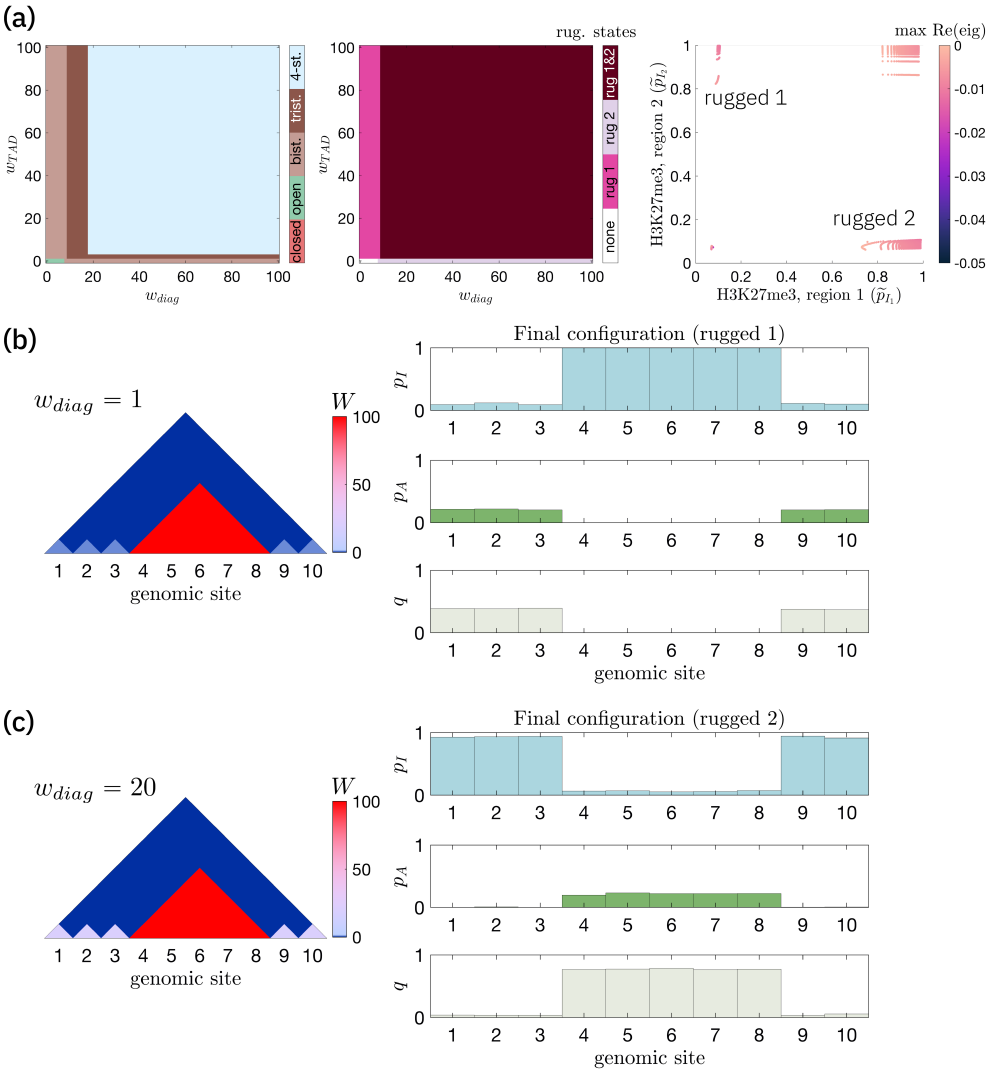

Fig. 7 Analysis and simulations of the two-region system, with region 1 exhibiting only connections within the same site with frequency, w_{diag}, while TAD-like region 2 is characterised by all-to-all interactions with the uniform strength of w_{TAD}. **a** Stability analysis of the system in response to variations of w_{diag} and w_{TAD}. The full modelling equations are presented in Supplementary Material. We note that this analysis is valid only for a particular case when the initial conditions within each region are the same. The colour schemes of the leftmost and rightmost panels are as in Fig. 5b. The middle panel shows a heatmap describing which rugged equilibrium patterns exist for the corresponding values of parameters. Here, rugged pattern 1 (2) shown in magenta (pastel purple) is characterised by region 1 enriched with activating (repressive) marks, whereas TAD-like region 2 displays high levels of repressive (activating) modifications. The parameter space where both rugged patterns coexist is shown in dark red. **b**, **c** Stochastic simulations of the multi-site system for the two-region scenario with a fixed value of $w_{TAD} = 100$. The value of $w_{diag} = 1$ in **(b)** and $w_{diag} = 20$ in **(c)**. Final configurations are marked as rugged pattern 1 and rugged pattern 2 as described for panel **(a)**. Here, we used the total number of sites, $N = 10$, the size of the region 1, $R = 5$, and the final simulation time, $\tau_{final} = 5000$ (dimensionless). All other parameters are set to their baseline values listed in Supplementary Table 1

increases, the system transitions to rugged pattern 2 (high repressive marks in region 1, high activating marks in region 2), which coexists with rugged pattern 1.

We performed stochastic simulations for this chromatin geometry, with a fixed value of w_{TAD} and two values of w_{diag}. The simulation results are presented in Figs. 7b-c. For the first scenario with the low value of $w_{diag} = 1$, only rugged pattern 1 exists. In the second scenario, we employed a higher value of $w_{diag} = 20$, representing the parameter regime where two opposite rugged patterns coexist. We note that for the chromatin geometries considered in these figures, region 1 consists of genomic sites $\{1, 2, 3, 9, 10\}$, while region 2 comprises genomic loci 4 to 8. For low values of w_{diag} (Fig. 7b), we consistently observed an epigenetic landscape corresponding to rugged pattern 1, with the TAD region 2 being closed (high H3K27me3 levels). However, for large values of w_{diag} (e.g. $w_{diag} = 20$ in Fig. 7c), the only equilibrium pattern observed in our stochastic simulations is rugged pattern 2 (TAD region open with acetylated H3K27). The emergence of this epigenetic landscape was robust regardless of the choice of initial conditions used in the simulations. This outcome arises because the stability analysis of the two-region system was conducted for a simplified case (ODE system of 6 equations), which assumed identical initial conditions within each region–homogeneous initial conditions for all sites in the TAD-like region and identical initial conditions for loci in the region with diagonal self-interactions. The complete set of equations associated with the stochastic simulations shown in Figs. 7b-c consists of 30 coupled ODEs, representing the evolution of 3 epigenetic marks across 10 genomic sites. Interactions among 10 genomic sites and competition for enzyme binding in the full system increase the size of the basin of attraction of rugged pattern 2.

Similar results are obtained when the interaction matrix is constant, $\overline{W} = W$ (see Sect. 2). Supplementary Figure 6 presents the stability analysis and simulations that give rise to the two rugged patterns in this scenario. It is important to note that the results shown in Fig. 7 were obtained using an H3K27me3-dependent interaction matrix, $\overline{w}_{il} = w_{il} p_{I_l} p_{I_l}$. This non-linearity necessitates larger values of $W = \{w_{il}\}_{i,l}$ for transitions between distinct behaviours, compared to the constant $\overline{W} = W$ matrix used in Supplementary Fig. 6.

In conclusion, the results presented in Fig. 7 and Supplementary Fig. 6 confirm that, regardless of the functional form of the interaction matrix, $\overline{W}$, our model predicts that the epigenetic profile within a TAD-like region is characterised by activating marks if the diagonal interactions in the rest of the domain are sufficiently strong. Experimentally obtained contact frequency maps (see Fig. 1d) indicate that diagonal interactions within individual genomic loci are more frequent than interactions with distant loci (away from the main diagonal), suggesting higher values of w_{diag} in our theoretical framework. Consequently, our model indicates that rugged pattern 2, with an acetylation peak at the TAD-like region, is more likely to be observed than rugged pattern 1. Additionally, since these results also hold for a constant interaction matrix, $\overline{W}$, we hypothesise that this behaviour is due to a combination of chromatin architecture and enzyme competition for binding at different genomic loci.

4 Discussion

In this study, we have presented a stochastic model for epigenetic regulation within the context of folded chromatin polymers and enzyme competition. These two effects, frequently overlooked in existing epigenetic models (Dodd et al., 2007; Sneppen et al., 2008; David-Rus et al., 2009; Sneppen and Dodd, 2012, 2016; Thalheim et al., 2017, 2018; Sneppen and Ringrose, 2019; Zhang et al., 2019; Nickels and Sneppen, 2023), are essential for understanding the dynamics of epigenetic modifications. Our approach specifically focuses on the post-translational covalent modifications of histone H3 tails, particularly H3K27me3, H3K4me3, and H3K27ac. By employing a mesoscopic framework, we were able to examine the dynamics of epigenetic marks at genomic loci comprising multiple nucleosomes and to explore the formation of epigenetic patterns on a larger scale of DNA compaction, such as loops and TADs.

Our primary focus has been to analyse the model in order to identify parameter regimes that drive the formation of rugged epigenetic profiles. The mathematical formulation of our model offers a significant advantage compared to purely computational approaches (Dodd et al., 2007; Sneppen et al., 2008; Sneppen and Dodd, 2012; Sneppen and Ringrose, 2019; Zhang et al., 2019; Nickels and Sneppen, 2023), allowing us to efficiently explore parameter space and characterise the qualitative behaviour it generates. One of our key findings involves the identification of conditions that lead to either uniform or rugged epigenetic patterns. Specifically, our analysis of a two-site system suggests that rugged patterns tend to form for intermediate enzyme concentrations. In contrast, low or high enzyme levels tend to produce more uniform epigenetic landscapes, with the specific outcome depending on the enzymes involved. Our analysis of the model's behaviour for different choices of the interaction map between genomic sites suggests that increasing cross-interaction generally leads to more uniform epigenetic profiles and that a higher frequency of self-interaction within genomic regions increases the robustness of rugged patterns.

Our model successfully captures the emergence of bivalent chromatin states characterised by a mixture of activating and repressive epigenetic marks. These bivalent states arise in bi- and tristable parameter regions, where they coexist with, and are less robust than, uniformly open and/or closed chromatin states. Notably, we observe that variations in parameter values (especially the strength of cross-interactions between genomic sites) can transform bivalent states into rugged states. Given the dynamic nature of chromatin and its folding (Dekker and Mirny, 2016), we hypothesise that the bivalent state may facilitate transitions between distinct epigenetic states in response to subtle changes in chromatin's local environment. We also emphasise that our mesoscopic modelling framework is consistent with the hypothesis that mixed levels of activating and repressive epigenetic marks observed in bivalent chromatin may arise from rapid bistable switching dynamics (Sneppen and Ringrose, 2019).

In this work, we performed steady-state and bifurcation analysis of a reduced system of equations obtained after applying a timescale separation based on the assumption that histone-modifying enzymes are significantly less abundant than unmodified lysine residues. This assumption is supported by experimental observations (Stafford et al., 2018) and previous modelling work (Owen et al., 2023). However, estimating enzyme abundance, particularly within spatially heterogeneous nuclear

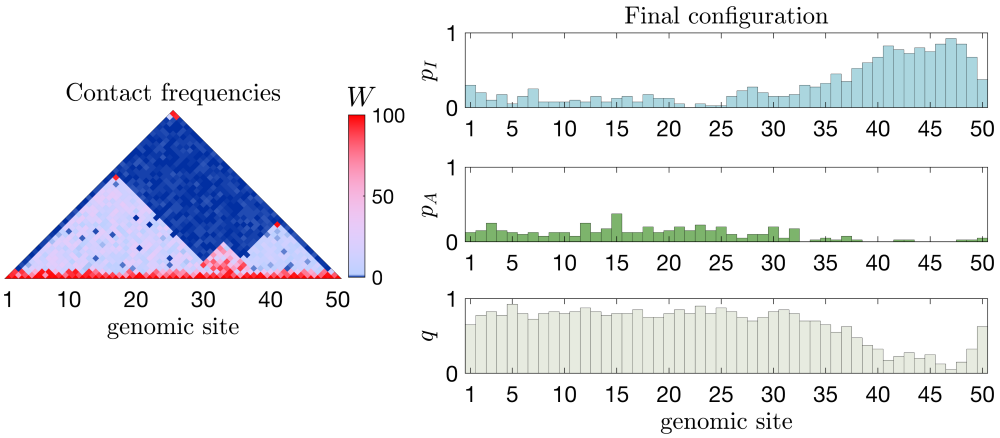

Fig. 8 A stochastic simulation of the multi-site system with $N = 50$ conducted for the chromatin architecture shown in Fig. 2e (and Graphical Abstract; see Supplementary information). Left panel: the upper-diagonal part of the contact frequency matrix, W. Right panel: final configuration of the epigenetic profile for this chromatin geometry. Here, p_I denotes levels of H3K27me3, p_A – H3K4me3, and q – H3K27ac. Different realisations yield similar epigenetic profiles regardless of the initial conditions. The final simulation time was set to $\tau_{final} = 30000$ (dimensionless). The remaining parameters are set to their baseline levels listed in Supplementary Table 1. For a movie of this numerical simulation, see Supplementary Movie 1

environments, remains a significant challenge. Current techniques, such as ChIP-seq (Park, 2009), provide spatial distributions of bound enzymes (representing complexes in our model), but quantifying the concentrations of soluble enzymes remains difficult.

From a mathematical perspective, the assumption of a low enzyme-to-substrate ratio aligns with the classical Briggs-Haldane framework (Briggs and Haldane, 1925), which justifies the application of the quasi-steady-state approximation (QSSA) for enzyme-substrate dynamics used in our model derivation. It is worth noting, however, that timescale separation can also be performed under alternative assumptions, such as the quasi-equilibrium approximation (QEA), leading to a different reduced system. The distinction between QSSA and QEA has been extensively explored in the literature (Holehouse et al., 2020; Schnoerr et al., 2017) and is crucial in determining the resulting system dynamics.

Importantly, the use of timescale separation techniques, such as the QSSA applied in this work to the non-reduced stochastic model of epigenetic mark dynamics, can lead to a loss of certain system behaviours. This occurs because these methods neglect the effects of intrinsic noise on the dynamics of transient fast variables. For example, depending on the parameter regime, stochastic QSSA may incorrectly approximate intrinsic noise or fail to capture noise-induced oscillations present in the original non-reduced system prior to timescale separation (Thomas et al., 2012). While our results are valid under the stated assumptions, the system's behaviour could differ under alternative assumptions or parameter regimes. Exploring these aspects in greater detail could uncover richer dynamical behaviours and would be a valuable avenue for future research.

Our approach enables us to shift the focus from the dynamics of specific genomic loci to larger-scale epigenetic patterns, as our model can be extended to account for an arbitrary number of genomic sites by suitable rescaling of the QSS probabilities for enzyme distributions. The limit on the number of sites that can be simulated with our modelling approach depends on the resolution of each site (its length in base pairs) and enzyme availability. Provided the assumption of enzyme competition holds, our model remains valid. In the simulations presented in this work, we assume that the total simulated chromatin region spans approximately 50 kb. This corresponds to a length of 1 kb (around 5 nucleosomes) per genomic site in the simulation shown in Fig. 8, or 5 kb (25 nucleosomes) per genomic site in Fig. 7. Most gene domains in the human genome typically range between 10 and 100 kb, while the length of TADs can extend from 100 kb to several megabases (Bickmore and Van Steensel, 2013). Therefore, our modelling framework is well-suited for simulating chromatin regions on biologically relevant scales, including gene domains and smaller TADs. We also note that our model can be extended to account for differential enzyme availability, which simplifies certain aspects of its formulation.

Epigenetic landscapes are influenced by many factors, including chromatin folding, which can significantly vary along the genome and across cell types and states. Although exploring characteristic chromatin geometries is beyond the current scope of our work, we highlight the potential of our modelling framework with a simulation based on a chromatin conformation shown in Fig. 2e (and also in Graphical Abstract; see Supplementary information). Figure 8 illustrates the resulting epigenetic landscape for this simulation, demonstrating the model's applicability to more realistic chromatin conformations.

A limitation of our current model is the assumption that the matrix of contact frequencies remains fixed. In biological systems, epigenetic modifications can influence chromatin folding, thus changing the map of contact frequencies.

Looking ahead, we plan to explore the formation of epigenetic landscapes for characteristic chromatin architectures such as loops, TADs, and architectural stripes. Additionally, we aim to extend our model to account for the dynamic coupling between the epigenetic landscape and interaction frequencies between genomic sites. By addressing these aspects, we hope to provide a more comprehensive understanding of epigenetic regulation in the context of 3D chromatin organisation.

Supplementary information.

Graphical Abstract.png Graphical Abstract summarising the modelling framework presented in this paper.

Supplementary Material.pdf Supplementary material containing a more detailed description of our model and additional figures and tables.

Supplementary Movie 1.mp4 A movie of a stochastic simulation of the multi-site system with $N = 50$ conducted for the chromatin architecture shown in Fig. 2E (and Graphical Abstract). The final configuration of this simulation is shown in Fig. 8. Left panel: the upper-diagonal part of the contact frequency matrix, W. Right panel: final configuration of the epigenetic profile for this chromatin geometry. Here, p_I denotes levels of H3K27me3, p_A – H3K4me3, and q – H3K27ac. Different realisations

yield similar epigenetic profiles regardless of the initial conditions. The final simulation time was set to $\tau_{final} = 30000$ (dimensionless). The remaining parameters are set to their baseline levels listed in Supplementary Table 1.

Acknowledgements. We are grateful to James Davies from the MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, for insightful discussions on the biological interpretation of our modelling results. This work has been funded by the Spanish Research Agency (AEI), through the Severo Ochoa and Maria de Maeztu Program for Centers and Units of Excellence in R&D (CEX2020-001084-M). D.S. and T.A. thank CERCA Program/Generalitat de Catalunya for institutional support. D.S. and T.A. have been funded by grant PID2021-127896OB-I00 funded by MCIN/AEI/10.13039/501100011033 'ERDF A way of making Europe'.

Declarations

Declaration of competing interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement. The detailed description of our model provided in this manuscript and/or in the supplementary material allows for the reproducibility of all the presented results. The code for model simulations used in this work is available upon request.

References

Alarcón, T., Sardanyés, J., Guillamon, A., Menendez, J.A.: Bivalent chromatin as a therapeutic target in cancer: An in silico predictive approach for combining epigenetic drugs. PLoS Computational Biology **17**(6), 1008408 (2021)

Buckle, A., Brackley, C.A., Boyle, S., Marenduzzo, D., Gilbert, N.: Polymer simulations of heteromorphic chromatin predict the 3d folding of complex genomic loci. Molecular Cell **72**(4), 786–797 (2018)

Briggs, G.E., Haldane, J.B.S.: A note on the kinetics of enzyme action. Biochemical Journal **19**(2), 338 (1925)

Ball, K., Kurtz, T.G., Popovic, L., Rempala, G.: Asymptotic analysis of multiscale approximations to reaction networks. The Annals of Applied Probability **16**(4) (2006)

Bickmore, W.A., Van Steensel, B.: Genome architecture: Domain organization of interphase chromosomes. Cell **152**(6), 1270–1284 (2013)

Chen, T., Ali Al-Radhawi, M., Sontag, E.D.: A mathematical model exhibiting the effect of dna methylation on the stability boundary in cell-fate networks. Epigenetics **16**(4), 436–457 (2021)

Cortini, R., Barbi, M., Caré, B.R., Lavelle, C., Lesne, A., Mozziconacci, J., Victor, J.-M.: The physics of epigenetics. Reviews of Modern Physics **88**(2), 025002 (2016)

Dekker, J., Mirny, L.: The 3d genome as moderator of chromosomal communication. Cell **164**(6), 1110–1121 (2016)

Dodd, I.B., Micheelsen, M.A., Sneppen, K., Thon, G.: Theoretical analysis of epigenetic cell memory by nucleosome modification. Cell **129**(4), 813–822 (2007)

David-Rus, D., Mukhopadhyay, S., Lebowitz, J.L., Sengupta, A.M.: Inheritance of epigenetic chromatin silencing. Journal of Theoretical Biology **258**(1), 112–120 (2009)

Finogenova, K., Bonnet, J., Poepsel, S., Schäfer, I.B., Finkl, K., Schmid, K., Litz, C., Strauss, M., Benda, C., Müller, J.: Structural basis for prc2 decoding of active histone methylation marks h3k36me2/3. eLife **9**, 61964 (2020)

Folguera-Blasco, N., Pérez-Carrasco, R., Cuyàs, E., Menendez, J.A., Alarcón, T.: A multiscale model of epigenetic heterogeneity-driven cell fate decision-making. PLoS Computational Biology **15**(4), 1006592 (2019)

Felsenfeld, G.: A brief history of epigenetics. Cold Spring Harbor Perspectives in Biology **6**(1), 018200 (2014)

Gross, D.S., Chowdhary, S., Anandhakumar, J., Kainth, A.S.: Chromatin. Current Biology **25**(24), 1158–1163 (2015)

Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics **22**(4), 403–434 (1976)

Grubert, F., Srivas, R., Spacek, D.V., Kasowski, M., Ruiz-Velasco, M., Sinnott-Armstrong, N., Greenside, P., Narasimha, A., Liu, Q., Geller, B., *et al.*: Landscape of cohesin-mediated chromatin loops in the human genome. Nature **583**(7818), 737–743 (2020)

Holehouse, J., Sukys, A., Grima, R.: Stochastic time-dependent enzyme kinetics: Closed-form solution and transient bimodality. The Journal of Chemical Physics **153**(16) (2020)

Ingalls, B.P.: Mathematical Modeling in Systems Biology: an Introduction. MIT press, Cambridge (2013)

Keener, J., Sneyd, J.: Mathematical Physiology: II: Systems Physiology. Springer, New York, NY (2009)

Kim, D.-H., Tang, Z., Shimada, M., Fierz, B., Houck-Loomis, B., Bar-Dagen, M.,

Lee, S., Lee, S.-K., Muir, T.W., Roeder, R.G., *et al.*: Histone h3k27 trimethylation inhibits h3 binding and function of set1-like h3k4 methyltransferase complexes. Molecular and Cellular Biology **33**(24), 4936–4946 (2013)

Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.*: Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science **326**(5950), 289–293 (2009)

Lee, C.-H., Holder, M., Grau, D., Saldaña-Meyer, R., Yu, J.-R., Ganai, R.A., Zhang, J., Wang, M., LeRoy, G., Dobenecker, M.-W., *et al.*: Distinct stimulatory mechanisms regulate the catalytic activity of polycomb repressive complex 2. Molecular Cell **70**(3), 435–448 (2018)

Loubiere, V., Martinez, A.-M., Cavalli, G.: Cell fate and developmental regulation dynamics by polycomb proteins and 3d genome architecture. Bioessays **41**(3), 1800222 (2019)

Macrae, T.A., Fothergill-Robinson, J., Ramalho-Santos, M.: Regulation, functions and transmission of bivalent chromatin during mammalian development. Nature Reviews Molecular Cell Biology **24**(1), 6–26 (2023)

Miller, J.L., Grant, P.A.: The role of dna methylation and histone modifications in transcriptional regulation in humans. Epigenetics: Development and Disease, 289–317 (2012)

Newar, K., Abdulla, A.Z., Salari, H., Fanchon, E., Jost, D.: Dynamical modeling of the h3k27 epigenetic landscape in mouse embryonic stem cells. PLoS Computational Biology **18**(9), 1010450 (2022)

Nickels, J.F., Sneppen, K.: Confinement mechanisms for epigenetic modifications of nucleosomes. PRX Life **1**(1), 013013 (2023)

Owen, J.A., Osmanović, D., Mirny, L.: Design principles of 3d epigenetic memory systems. Science **382**(6672), 3053 (2023)

Park, P.J.: Chip–seq: advantages and challenges of a maturing technology. Nature Reviews Genetics **10**(10), 669–680 (2009)

Park, S.-Y., Kim, J.-S.: A short guide to histone deacetylases including recent progress on class ii enzymes. Experimental & Molecular Medicine **52**(2), 204–212 (2020)

Ringrose, L., Howard, M.: Dissecting chromatin-mediated gene regulation and epigenetic memory through mathematical modelling. Current Opinion in Systems Biology **3**, 7–14 (2017)

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson,

J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., *et al.*: A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell **159**(7), 1665–1680 (2014)

Strahl, B.D., Allis, C.D.: The language of covalent histone modifications. Nature **403**(6765), 41–45 (2000)

Szabo, Q., Bantignies, F., Cavalli, G.: Principles of genome folding into topologically associating domains. Science Advances **5**(4), 1668 (2019)

Sati, S., Cavalli, G.: Chromosome conformation capture technologies and their impact in understanding genome function. Chromosoma **126**, 33–44 (2017)

Sneppen, K., Dodd, I.B.: A simple histone code opens many paths to epigenetics. PLoS Computational Biology **8**(8), 1002643 (2012)

Sneppen, K., Dodd, I.B.: Nucleosome dynamics and maintenance of epigenetic states of cpg islands. Physical Review E **93**(6), 062417 (2016)

Stafford, J.M., Lee, C.-H., Voigt, P., Descostes, N., Saldaña-Meyer, R., Yu, J.-R., Leroy, G., Oksuz, O., Chapman, J.R., Suarez, F., *et al.*: Multiple modes of prc2 inhibition elicit global chromatin alterations in h3k27m pediatric glioma. Science Advances **4**(10), 5935 (2018)

Sneppen, K., Micheelsen, M.A., Dodd, I.B.: Ultrasensitive gene regulation by positive feedback loops in nucleosome modification. Molecular Systems Biology **4**(1), 182 (2008)

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., *et al.*: Histone methylation by prc2 is inhibited by active chromatin marks. Molecular Cell **42**(3), 330–341 (2011)

Sneppen, K., Ringrose, L.: Theoretical analysis of polycomb-trithorax systems predicts that poised chromatin is bistable and not bivalent. Nature Communications **10**(1), 2133 (2019)

Schnoerr, D., Sanguinetti, G., Grima, R.: Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. Journal of Physics A: Mathematical and Theoretical **50**(9), 093001 (2017)

Thalheim, T., Hopp, L., Binder, H., Aust, G., Galle, J.: On the cooperation between epigenetics and transcription factor networks in the specification of tissue stem cells. Epigenomes **2**(4), 20 (2018)

Thalheim, T., Herberg, M., Loeffler, M., Galle, J.: The regulatory capacity of bivalent genes—a theoretical approach. International Journal of Molecular Sciences **18**(5), 1069 (2017)

Tollefsbol, T.O.: Epigenetics of human disease. In: Epigenetics in Human Disease, pp. 3–10. Elsevier, Academic Press (2018)

Thomas, P., Straube, A.V., Grima, R.: The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. BMC Systems Biology **6**, 1–23 (2012)

Waddington, C.H.: The Strategy of the Genes. Routledge, London (1957)

Zoghbi, H.Y., Beaudet, A.L.: Epigenetics and human disease. Cold Spring Harbor Perspectives in Biology **8**(2), 019497 (2016)

Zhang, L., Lu, Q., Chang, C.: Epigenetics in health and disease. Epigenetics in Allergy and Autoimmunity, 3–55 (2020)

Zhang, Y., Liu, N., Lin, W., Li, C.: Quantifying the interplay between genetic and epigenetic regulations in stem cell development. New Journal of Physics **21**(10), 103042 (2019)

Zhao, W., Qiao, L., Yan, S., Nie, Q., Zhang, L.: Mathematical modeling of histone modifications reveals the formation mechanism and function of bivalent chromatin. IScience **24**(7) (2021)