

Bistability and state switching in computational dynamic histone modification models

Michel Krecké



UNIVERSITÄT
LEIPZIG

MASTERARBEIT

eingereicht am
Master of Science

Bioinformatik

in Leipzig

im Mai 2021

Primary advisor:
Prof. Dr. Sonja Prohaska

Second advisor:
Dr. Christian Arnold

© Copyright 2021 Michel Krecké

This work is published under the conditions of the *Creative Commons License Attribution-NonCommercial-NoDerivatives 4.0 International* (CC BY-NC-ND 4.0)—see <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Declaration

I hereby declare and confirm that this thesis is entirely the result of my own original work. Where other sources of information have been used, they have been indicated as such and properly acknowledged. I further declare that this or similar work has not been submitted for credit elsewhere.

Leipzig, May 10, 2021

Michel Krecké

Contents

Declaration	iv
1 Theoretical Background	1
1.1 Eukaryotic transcription regulation	1
1.1.1 Chromatin	1
1.1.2 Histone-modifying enzymes	2
1.1.3 Bivalency	3
1.2 Dynamic histone PTM models	4
1.2.1 Chemical master equation	4
1.2.2 Gillespie's algorithm	5
1.2.3 <i>EpiDynaST</i>	6
1.3 Epigenetic fitness landscapes	7
1.3.1 From landscape to vector field	7
1.3.2 Monostable landscapes	8
1.3.3 Multistable landscapes	9
1.4 Impact of this work	12
2 Methods	13
2.1 Chromatin Model	13
2.2 Enzyme models	14
2.2.1 Enzymes in <i>EpiDynaST</i>	14
2.2.2 Enzyme types	14
2.2.3 Enzyme rule sets	17
2.3 Simulation details	18
3 Results	20
3.1 Non-cooperative enzymes do not entail bistable systems	20
3.2 Bistability on a non-cyclic nucleosome string	24
3.2.1 Impact of cooperative enzymes	24
3.2.2 Bistable switching on a non-cyclic string	27
3.3 Bistable switching on a cyclic string	31
3.4 Influence of the dissociation rate	34
3.5 The boundaries of bistability	36
3.6 Bivalency	37

4 Discussion	40
4.1 Conditions for bistability	40
4.2 Proposed mechanism for bistable switching	40
4.2.1 Bistable switching: a biologically useful phenomenon	40
4.2.2 cyclic	40
4.2.3 non-cyclic	40
5 Outlook	41
5.1 Make <i>EpiDynaST</i> account for recent biological findings	41
5.2 Explore recently discovered non-cooperative edge case	41
A Specifications	43
A.1 Summary of enzyme types	43
A.2 Summary of simulation parameters	43
B Additional runs	45
B.1 Additional runs from section 3.1	45
B.2 Additional runs from section 3.2	46
References	47
Literature	47

1. Theoretical Background

1.1 Eukaryotic transcription regulation

1.1.1 Chromatin

Eukaryotic DNA is organized as chromatin in the cell nucleus, which consists of nucleosomes that mainly serve the increase of packing density and robustness of the DNA, but also play an important role in gene regulation. Nucleosomes are built out of DNA wrapped around an octamer of homologous, basic proteins, the histones. These proteins contain a great amount of the positively charged amino acids arginine (Arg, R) and lysine (Lys, K), which results in attracting the negatively charged DNA (phosphate backbone). The so-called histone tails on the amino end of the proteins stick out of the nucleosome core complex. Albeit not having a fixed secondary structure, the tails, as well as the rest of the histones are very well conserved throughout a large set of eukaryotes, from *Saccharomyces cerevisiae* all the way to *Homo Sapiens Sapiens* [1].

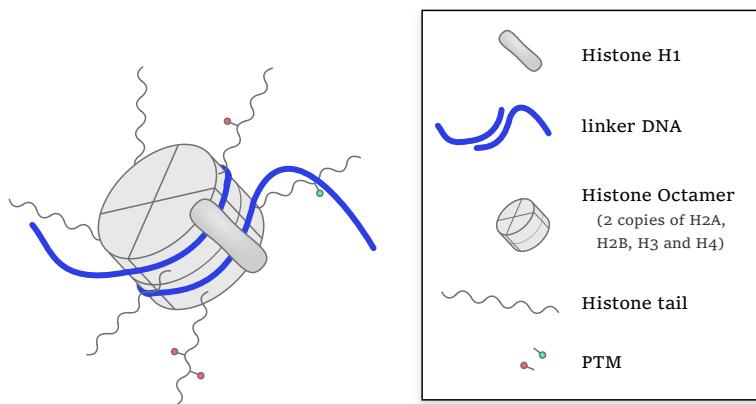


Figure 1.1: Schematic model of a nucleosome. It consists of DNA wrapped around a histone octamer. The histones themselves are organized as two homologous tetramers (H2A, H2B, H3, H4). The DNA is kept in place by histone H1, which also play a role in establishing higher order chromatin structures.

Chromatin can show higher order structure than the simple “beads on a string” variant. As such, nucleosomes that are not necessarily next neighbours along the DNA string can be in near proximity [1].

Chromatin structure plays an important role in eukaryotic gene expression. Some

DNA segments close to and on an actively transcribed gene are more easily accessible by proteins due to a locally more open chromatin structure. These chromatin regions are therefore called hypersensitive sites [2]. Logically, the location of hypersensitive sites depends on the set of active genes and is thus cell-type and age specific [1].

The complex process of chromatin remodelling which leads to activation or repression of gene transcription comprises a vast multitude of agents and their interaction network is still not fully understood.

1.1.2 Histone-modifying enzymes

This thesis is based on one aspect of the chromatin remodelling machinery, namely histone-modifying enzyme complexes (HME). These complexes are able to covalently bind or remove chemical groups on amino acids (mostly R and K) at very specific positions on the histone tail. These post-translational modifications (PTMs) are named according to the amino acid they have been bound to. H3K27ac, for instance, denotes an acetylation (ac) on lysine 27 (K27) of histone 3 (H3).

The presence of these chemical groups changes the charge or polarity of the modified amino acids and thus influences the histone's affinity to the DNA. Histone-acetyl-transferases (HATs), for instance, add an acetyl group to lysine, thus neutralizing the positive charge on the ammonium cation at neutral pH (see figure 1.2). This neutralization decreases the attraction to the negatively charged DNA backbone significantly resulting in the occurrence of a hypersensitive site. [1]

Apart from the acetyl group, a multitude of other markers have been found on histone tails, e.g. methylation (one-, two and three-fold), phosphorylation, ubiquitylation and so forth [3]. These groups can entail gene transcription activation, silencing (repressing), or fulfil completely different purposes [4, 5].

The enzymes' modification activities are not forcibly isolated processes. The different enzyme types as well as the modifications on specific amino acids on the histone tails are connected through complex interaction networks [6–8]. Such crosstalk between different modifications is achieved by the enzymes' ability to "read" modifications and "write" another one, if the reading process was successful. Reading is achieved by binding to specific modifications on the histone tail. A very popular example for this behaviour is the bromodomain, which specifically binds acetylated lysines [9]. The modification pattern that certain enzymes need in order to modify an amino acid will further on be called the enzyme's *context*. Upon successful binding, the enzyme can then catalyse the covalent modification of an unmodified amino acid.

In order to clarify the meaning of the vocabulary concerning the enzyme models, consider the following illustration:

For the HAT enzyme mentioned earlier, one would define the enzyme's context as an unmodified nucleosome, as this is needed in order for the enzyme to perform its specific chemical modification reaction on the nucleosome string, which is acetylation of an unmodified nucleosome. Possibly, the enzyme needs a specific reading pattern in order to bind to the string in proximity of the nucleosome to be written on. This is the case when concerning self-reinforcing enzymes, for instance, which read their own modification and are then able to write it to another unmodified nucleosome which results

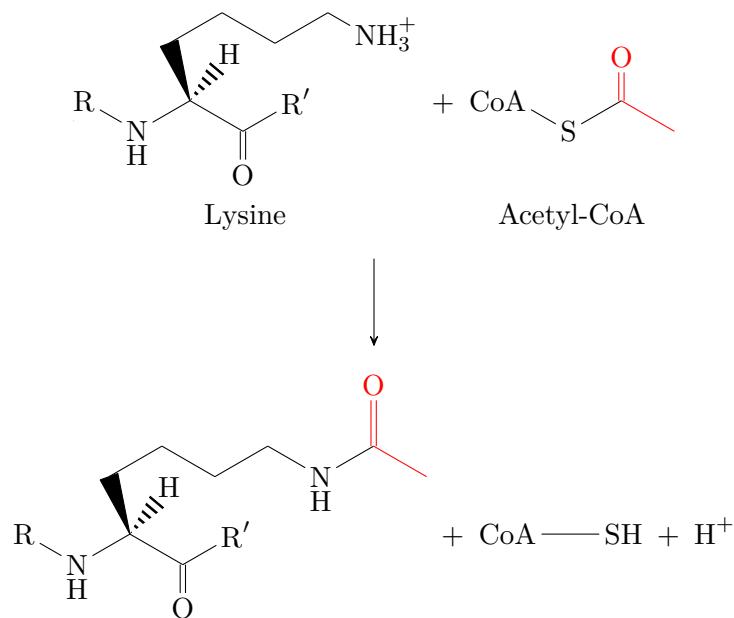


Figure 1.2: Acetylation of lysine. This reaction is catalysed by the HAT enzyme which by means of the cofactor acetyl coenzyme A (Acetyl-CoA) is able to trigger the transfer (substitution reaction in chemical terms) of an acetyl group onto the nitrogen atom in the side chain of lysine. The latter can be part of a histone tail.

in a positive feedback loop. Such specific reading/binding patterns are also contained in the context of this specific enzyme. Thus, the enzyme's context is a summary of rules that impose constraints on the presence or absence of modifications on the nucleosome string in order for the enzyme to gain reading/writing ability.

This thesis will only feature acetylation as an activating modification and methylation as a silencing modification. Accordingly, the “state” of a nucleosome is defined as one of the following:

- **unmodified:** There has been no PTM on any histone tail of the concerning nucleosome.
- **active:** Every PTM found on the concerning nucleosome enables gene activation. In this thesis, every one of these PTMs is an acetylation.
- **silent:** Every PTM found on the concerning nucleosome disables gene activation. In this thesis, every one of these PTMs is a methylation.
- **bivalent:** see 1.1.3

1.1.3 Bivalency

In pluripotent stem cells, nucleosomes have been found to contain both activating and silencing markers on histone tails of one and the same octamer. This bivalent state is

believed to maintain a “poised state”, being ready to induce a gene expression cascade as soon as the silencing marker is removed [10, 11]. Others believe, that this bivalent state is connected to cell division and the ability of inheriting the active gene set for one daughter cell to induce differentiation while the other daughter cell remains a pluripotent stem cell [12].

1.2 Dynamic histone PTM models

1.2.1 Chemical master equation

The chemical master equation (CME) is the differential equation underlying a chemical mixture that describes the time-dependent evolution of said mixture from a reactive point of view. Applied to the case at hand, we can establish the CME system made out of two equations for either the concentration of active (acetylated) and silent (methylated) nucleosome states.

In order to do this, one can establish the state concentration dependent differential equation [13] for each state type (i.e. active and silent) with respect to the HME types that modify the respective state. This was already done by Mayer in [14].

Eqns. 1.1 describe the non-cooperative (see 2.2.2) case for $a = \frac{A}{N}$ and $m = \frac{M}{N}$ with A the number of acetylated nucleosomes, M the number of methylated nucleosomes and N the total number of nucleosomes. α_i and β_i are coefficients taking into account the types, association and dissociation ratios of the enzymes in the system.

$$\frac{\partial a}{\partial t} = \underbrace{-\alpha_1 a}_{\text{ac removal}} + \underbrace{\alpha_2 a * (1 - a - m)}_{\text{ac addition}} \quad (1.1a)$$

$$\frac{\partial m}{\partial t} = \underbrace{-\beta_1 m}_{\text{me removal}} + \underbrace{\beta_2 m * (1 - a - m)}_{\text{me addition}} \quad (1.1b)$$

Obviously, this would only be a usable model if the neighbour relations of the nucleosomes could be neglected. Given that the context from the enzyme rule sets is an important aspect of *EpiDynaST* (see 1.2.3), an analytical solution of the CME would not be the best approximation. Also, *EpiDynaST*'s model strongly depends on discrete numbers such as the number of nucleosomes in the string, rendering mere state concentrations without positional information insufficient. Thus, the analytical solution of the CME as a continuous system, makes it even more unfitting as an approximation for the system at hand.

A more fitting model would be to establish and solve the CME for every nucleosome while the CME for nucleosome i would depend on the number of neighbours equal to the biggest enzyme context in the system's rule set. For instance, if the rule set contains at least one rule including the next neighbours of the modified nucleosome i , the CME

system would change to eqns. 1.2.

$$\frac{\partial a_i}{\partial t} = -\alpha_1(a_{i-1} + a_i + a_{i+1}) + \alpha_2(a_{i-1} + a_i + a_{i+1}) * (1 - a - m) \quad (1.2a)$$

$$\frac{\partial m_i}{\partial t} = -\beta_1(m_{i-1} + m_i + m_{i+1}) + \beta_2(m_{i-1} + m_i + m_{i+1}) * (1 - a - m) \quad (1.2b)$$

This differential equation system quickly increases in complexity and dimensionality with increasing number of nucleosomes in the system and higher reach of the enzyme models up to a point where an analytical solution is impossible to achieve. Mainly for this reason, it is convenient to numerically simulate the time-dependent evolution of the system.

1.2.2 Gillespie's algorithm

Gillespie's algorithm or *stochastic simulation algorithm*, as called by the author, simulates the evolution in time of a spatially homogenous molecular mixture with a discrete number of reactants under specification of the coupled reaction channels (i.e. association and enzymatic reaction with simultaneous dissociation) based on stochastic chemical kinetics [15, 16]. This is useful especially when the practice of solving the chemical master equation analytically is not ideal (see 1.2.1).

Reconsidering the HAT enzyme example from before, its two reaction channels would be the association for one and secondly the reaction and dissociation channel.

Contrarily to other stochastic simulations, Gillespie takes an event-based time step approach. This ensures that at every time step, exactly 1 event is taking place. This approach obviously reduces the overhead compared to an equidistant time step approach (see fig. 1.3) and can massively facilitate implementation and later interpretation of the simulation results.

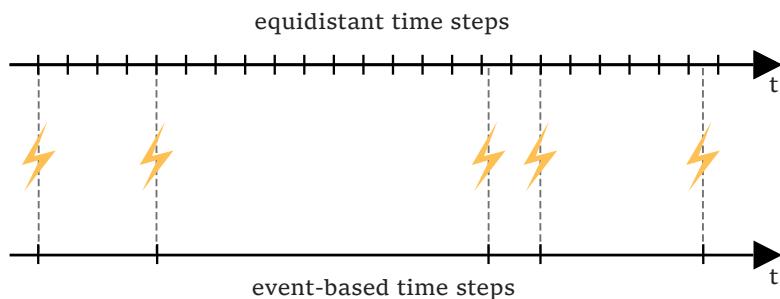


Figure 1.3: Schematic illustration of equidistant time steps vs. event-based time steps inspired by Mayer in [14].

The algorithm accounts for an appropriate choice of the time elapsed between two time steps by computing and comparing the number of legal associations and dissociations as defined by the enzyme rules given to the algorithm. The rule set in this setting consists of enzymes that covalently modify histone tails. The legality of such a reaction

is assessed by checking, if a certain modification pattern (context) is already present on the nucleosome to be modified and other neighbouring nucleosomes on the string in order to perform the reading and writing actions for the specific enzyme.

The algorithm then summarizes these legal channel possibilities in a propensity sum (normalized from 0 to 1) taking the given association and dissociation rates as well as the concentrations into account. An event is then chosen at random by selecting a random number between 0 and 1. If many events are contributing to the propensity sum at this time step, the time in between events is smaller. Conversely, if only few events are possible at this time step, the time in between events is larger.

In more systematical terms, Gillespie's algorithm can be summarized in the following 4 steps [14, 17]:

- **Step 0 (initialization):** Set an initial starting state X and a rule set. Set $t = 0$.
- **Step 1:** Calculate the propensity sum a_0 as the sum of all legal reactions and their occurrence possibility based on their association/dissociation rates and their concentration.
- **Step 2:** Based on random numbers, choose a specific reaction channel μ as the event to happen in this time step as well as the time τ elapsed from the previous event to the present one.
- **Step 3:** Update the time $t_{i+1} = t_i + \tau$ as well as the effect that μ has on the system $X_{i+1} = \mu(X_i)$. Continue with step 1.

1.2.3 *EpiDynaST*

The software used in this thesis is *EpiDynaST* (**E**pigenetic **D**ynamics **S**imulation **T**ool), which bases on *StoChDyn* by Arnold et al. [18] and was developed by N. Herbig et al. (unpublished results).

EpiDynaST is a nucleosome post-translational modification (PTM) simulation software which, at its core, uses Gillespie's algorithm in order to show the dynamic change on a nucleosome string as a function of time.

The main working mechanism may be outlined as follows: First, one defines an enzyme rule set and a starting nucleosome string, which is defined as an array of nucleosomes reduced to their PTMs (see 2.1). Then, *EpiDynaST* simulates the stochastic time-dependent change of said modifications on the string, exactly one event at a time. The two events that can occur for each enzyme are either an association step or a reaction step that immediately entails the enzyme's dissociation from the nucleosome.

The enzyme rule sets simply describe a pattern (further on called "context") on the string, that is then changed according to the rule. For instance, a linear acetylation extender enzyme would look for a pattern with two neighbouring nucleosomes, one acetylated and the other one unmodified, and acetylate the latter (see 2.2.2 for details on the enzyme types and their reactions).

1.3 Epigenetic fitness landscapes

1.3.1 From landscape to vector field

The term of fitness landscape (referred to as landscape for the rest of this work) is mathematically defined as the triple (V, χ, f) where V is a set of configurations, χ refers to the neighbouring relationship or similarity among the configurations and f defines the fitness function of the landscape [19] (see fig. 1.4 for a simplified example).

In the case of epigenetic fitness landscapes, V comprises the entirety of active, silent and unmodified state distributions along the nucleosome string, χ determines the dissimilarity between two nucleosome strings, i.e. which nucleosome has to be changed on the string $v_1 \in V$ in order to turn it into another nucleosome string v_2 . f , which indicates the relative height or depth $f(v)$ of a specific configuration $v \in V$, is determined by all the association and dissociation rates, as well as the type of the enzymes in the system. In this specific landscape, the stochastic Gillespie's algorithm will always tend towards configurations (fix points) at the very base of landscape basins. These configurations are those, that are adopted in the majority of time steps throughout Gillespie's simulation algorithm.

The term of epigenetic fitness landscape defined in this work is not to be confounded with the very frequently referenced concept of “epigenetic landscape” introduced by C. H. Waddington around 1957 [20]. The epigenetic landscape describes a conceptual model which draws an analogy between cell differentiation with its influential factors and a ball on a rugged hill [21]. The ball inevitably rolls down the hill, but the exact track it is taking along its way down will fundamentally decide, where it is going to end up in the valley. Analogously, the observed cell’s fate will eventually be differentiation, but, according to Waddington, the pathway taken along the branched track is defined by the starting point, gene interaction (regulation) and the counterplay of inductive events, which push the cell in a distinct direction on the landscape, and the cell’s competence or ability to follow this path.

The idea of landscapes is very similar in the concept of Waddington’s cell differentiation model as well as in the epigenetic fitness landscape model used in this work. However, the underlying phenomenon that is explained is entirely different. Waddington aims at providing a graspable model for a stem cell’s journey through differentiation. The epigenetic fitness landscape model defined in this work however serves the purpose of explaining histone PTMs on a chromatin string. Furthermore, the fitness landscape is a mathematically graspable concept with numerical influential factors and a slope defined by f while Waddington’s model only offers a figurative model.

Nonetheless, in order to avoid confusion on the reader’s side, Waddington’s definition of the epigenetic landscape will not be featured from this point on in this work. Every mention of landscape will refer to the epigenetic fitness landscape mathematically defined above.

Depending on the landscape, i.e. if there are deep and steep basins, it gets more and more difficult for the system to exit the basin and adopt a configuration that lies outside. Each histone modifying enzyme thus induces an individual stability trend among the configurations which results in a vector field specific to each type of enzyme. In more

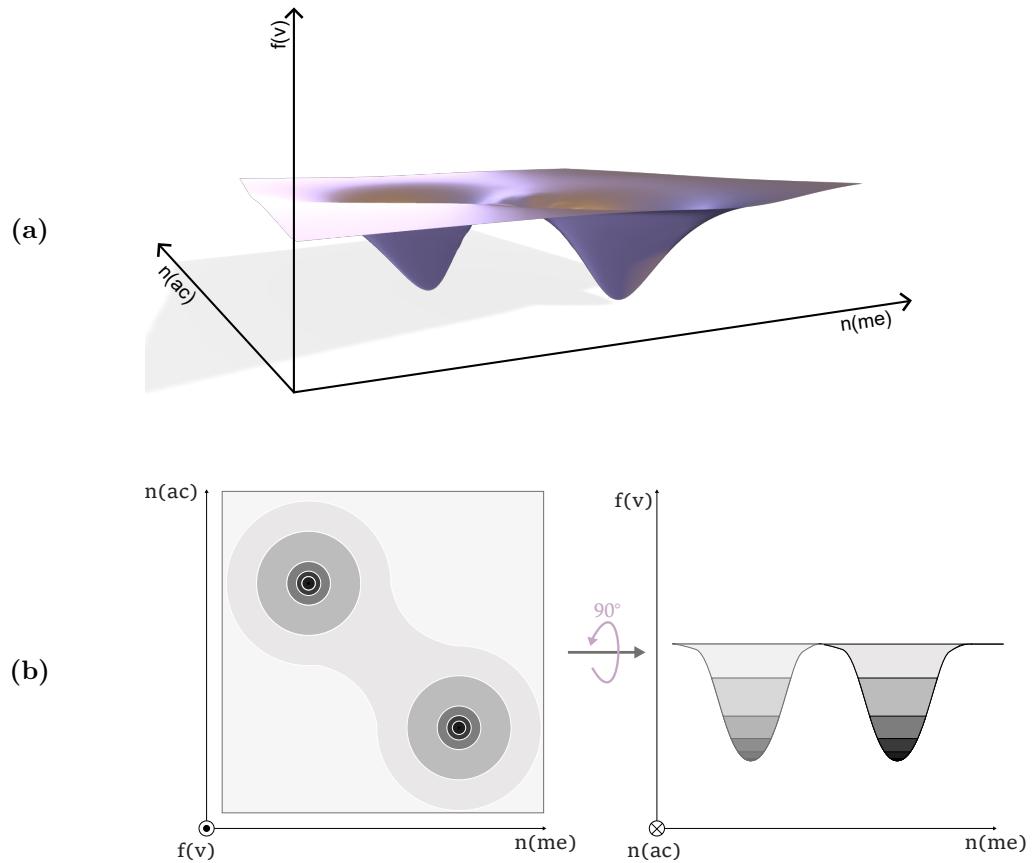


Figure 1.4: Simplified epigenetic fitness landscape example of a bistable system. **(a)** 3D illustration modelled with Blender [22]. **(b)** 2D-projections along the indicated axes. The fix points in the basins are the most frequent configuration throughout a simulation. In most bistable systems, the basins are to be found at the extrema of the $n(ac)$ and the $n(me)$ axes.

complex systems with a multitude of different enzyme types, these vector fields are combined by superposition and create the resulting landscape at hand containing zero or more fix points, depending on the un-, mono- or multistable nature of the system.

In this thesis, there won't be any numerical values given to $f(v)$. However, by modifying the relative enzyme rate ratios, one can easily see that f can change drastically resulting in variation of ease for the system to maintain a certain state or move along the shape of the landscape to switch from one state to another. Thus, even though f can not be numerically anchored, the influential factors that change $f(v)$ for one $v \in V$ as well as their impact on the landscape can very well be identified.

1.3.2 Monostable landscapes

The number of fix points in the landscape can be analytically derived from the differential equation system that can be established for a given set of enzyme types.

In [14], Mayer found 4 critical values for eqns. 1.1 in the Cartesian coordinate system

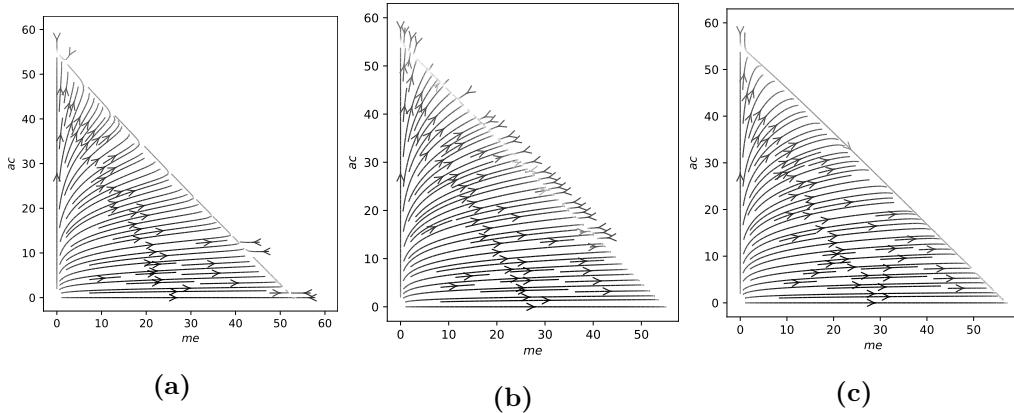


Figure 1.5: Vector fields describing the non-cooperative 60 nucleosome system with varying enzyme rates. These mainly define the parameters α_n and β_n if the enzyme types are constant. The axes describe in absolute numbers the occurrence of methylation (x-axis) and acetylation (y-axis). **(a)** describes the case for $\frac{\alpha_1}{\alpha_2} > \frac{\beta_1}{\beta_2}$ resulting in a separatrix with gradient towards $(0, N(1 - \frac{\beta_1}{\beta_2}))$. In **(b)**, the equality $\frac{\alpha_1}{\alpha_2} = \frac{\beta_1}{\beta_2}$ results in a zero-gradient separatrix. In **(c)**, $\frac{\alpha_1}{\alpha_2} < \frac{\beta_1}{\beta_2}$ results in a separatrix with gradient towards $(N(1 - \frac{\alpha_1}{\alpha_2}), 0)$. From Mayer in [14].

(m, a) with origin $(0, 0)$ and m the amount of methylated nucleosomes and a the amount of acetylated nucleosomes. 3 are fix points at $(0, 0)$, $(0, 1 - \frac{\beta_1}{\beta_2})$, $(1 - \frac{\alpha_1}{\alpha_2}, 0)$ and one is a separatrix whose gradient depends on the ratio between $\frac{\alpha_1}{\alpha_2}$ and $\frac{\beta_1}{\beta_2}$ and which connects the two non-trivial fix points. The separatrix always has a gradient towards one of the non-trivial fix points, except for the case $\frac{\alpha_1}{\alpha_2} = \frac{\beta_1}{\beta_2}$. Here, the separatrix has no gradient (see fig.1.5).

One could argue, that for the case of $\frac{\alpha_1}{\alpha_2} = \frac{\beta_1}{\beta_2}$, the zero-gradient separatrix interpreted as a set of points fulfilling a linear equation presents an infinite amount of stable points, thus rendering this specific system multistable. However, one point on the separatrix can hardly be described stable, because it is not resistant to small perturbations. In fact, even the most atomic displacement defined in the system by χ , namely the state change of one nucleosome, most likely leads to another point on the separatrix and not forcibly to the previous one. Thus, the separatrix as a whole, containing both non-trivial fix points can be described as stable and disqualifies every other point on the separatrix from individual stability, hence the monostable nature of all the systems in fig. 1.5.

1.3.3 Multistable landscapes

In order for a dynamic histone PTM system to be bistable, the enzymes have to show cooperativity [14, 23, 24]. According to Sneppen [25, p.48], “cooperative binding means that the probability of occupying a state increases more than linearly with the concentrations of the binding molecules”.

In [23], Dodd et al. specify the nature of cooperativity in order to reach ultrasensitivity.

tivity¹ and thus a bistable system as follows²:

“Cooperativity can be direct, where two modified nucleosomes act together to recruit an enzyme to modify a third nucleosome [26–28], or indirect, where each modified nucleosome catalyzes one of two separate modification reactions to fully convert a third nucleosome [26, 29]. A critical requirement for ultrasensitivity is that modified nucleosomes must act nonlocally, stimulating modification of nucleosomes located some distance away on the DNA. This long-range interaction is necessary for any nucleosome to be able to ‘sense’ the majority nucleosome type within the patch and cannot be provided by simple neighbor-to-neighbor contact [26, 28].”

In other words, cooperativity can only be achieved by allowing the enzymes to detect more than one nucleosome that, in order to reach bistability, must not be a direct neighbour of the nucleosome to be (un)modified. This allows the modification that is superiorly prevalent over the whole string at that moment to be accounted for and recognized by the enzymes.

On a sidenote, the definition of cooperativity might seem slightly counterintuitive from a biochemical point of view, where the notion of cooperativity is strongly associated to be an asset of the enzyme [30]. In contrast, according to Dodd et al., cooperativity is described as being a property of a set of nucleosomes being able to “cooperate” in order to recruit an enzyme and catalyse a reaction. A biochemist might be more comfortable with the enzymes being the active part in the system and most notably the catalyst of the occurring reactions instead of the nucleosomes, which most definitely do not catalyse any chemical reaction.

Even though this is not ideally put from a biochemical point of view, the mathematical implications are untouched from these imprecisions.

Mayer in [14] expressed the time dependent concentration of the modifications in the system in function of cooperative enzymes in the differential equation system in eqns. 1.3.

$$\frac{\partial a}{\partial t} = \underbrace{-\alpha_1 a}_{\text{ac rem}} + \underbrace{\alpha_2 \frac{1}{n} a^2 * (n - a - m)}_{\text{ac add}} \quad (1.3a)$$

$$\frac{\partial m}{\partial t} = \underbrace{-\beta_1 m}_{\text{me rem}} + \underbrace{\beta_2 \frac{1}{n} m^2 * (n - a - m)}_{\text{me add}} \quad (1.3b)$$

¹From Dodd et al. [23]: “Ultrasensitivity is a nonlinearity that magnifies any numerical advantage of one nucleosome type over another, allowing positive feedback to strongly push the system away from intermediate states and towards a large majority of one or other type.”

²citations inside the quote changed their appearance in order to remain functional and to stylistically fit this work

Given that the cubic cooperative terms show the highest potency, it is known that the differential equation system has a maximum of 9 critical values. Some of these critical values can quite easily be identified graphically.

Fig. 1.6 depicts the vector field of a system with random and cooperative enzymes. Some critical values can clearly be identified. They are summarized in tab. 1.1.

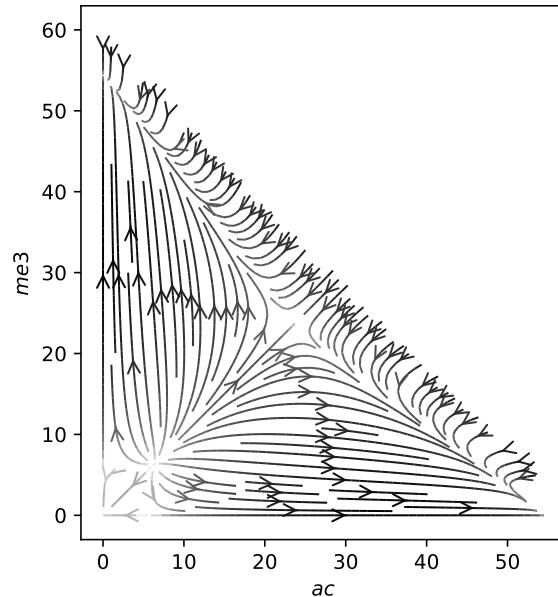


Figure 1.6: Vector fields describing the cooperative 60 nucleosome system with varying enzyme rates. These mainly define the parameters α_n and β_n if the enzyme types are constant. The axes describe in absolute numbers the occurrence of methylation (x-axis) and acetylation (y-axis). From [14].

Table 1.1: Critical values in (a, m) notation as identified graphically in fig. 1.6.

critical point	stability
$(0, 0)$	stable
$(8, 8)$	unstable
$(0, 5)$	unstable
$(5, 0)$	unstable
$(0, 55)$	stable
$(55, 0)$	stable
$(25, 25)$	saddle point

1.4 Impact of this work

Although the mathematical foundation on whether a system can and cannot show bistability was already established, the execution in terms of building a model and rigorously identifying the influence of different factors within the model on the dynamics of a bistable system have, to my knowledge, not been explored yet.

Also, to date, bistable systems have not yet been modelled by means of a software that takes limited enzyme reach into account like *EpiDynaST* does. This is an important distinction from other models, because neighbour-to-neighbour relations are not neglected, which marks an important step towards real-life chromatin systems.

This thesis will show the potential and limitations of a model with fixed neighbour relations concerning bistability and the switching of stable states throughout the simulation. Furthermore, this work will propose working mechanisms of bistability occurring on non-cyclic, as well as cyclic nucleosome strings.

2. Methods

2.1 Chromatin Model

The chromatin in *EpiDynaST* is modelled as an array of half-nucleosomes (for simplicity, the array of half-nucleosomes is referred to as nucleosome string in the rest of the work), meaning that they only contain one tetramer of H2A, H2B, H3 and H4. The nucleosomes hold their respective position on the string so that the neighbour relations are fixed. That way, every nucleosome has exactly two neighbouring nucleosomes (in the cyclic case, see below) that stay the same throughout the entirety of a simulation. Furthermore, the nucleosomes are reduced to presence or absence of PTMs on their tails.

The model in this work can even be reduced further, so that a nucleosome is modelled as 1 (K_{27}) single amino acid which can be either acetylated (referred to as active) or monomethylated (referred to as silenced). Di- and trimethylation will not be featured in this work. Exceptionally, in 3.6, every nucleosome possesses 2 modifiable amino acids, called K_x and K_y . All of these amino acids are monoacetylatable as well as monomethylatable. These two modifications are mutually-exclusive one amino acid. However, K_x and K_y can very well contain opposing modifications in which case the concerning nucleosome will be referred to as bivalent.

The nucleosome string is either modelled to be non-cyclic (as done in 3.1 and 3.2) or cyclic (3.3 to 3.6). In the cyclic case, the enzymes' context can include nucleosomes from the start as well as the end of the string simultaneously. In the non-cyclic case, the first and last nucleosome logically only have one neighbour.

The non-cyclic models contain 60 nucleosomes whereas the cyclic ones only contain 40 nucleosomes in order to reduce computation time and storage space.

The starting state for every simulation in this work is a completely unmodified nucleosome string. As long as there are random adders in the system (which is the case for every system featured in this work), the completely unmodified string logically is an unstable configuration which does not offer any sort of bias in the direction of acetylation or methylation respectively, hence the choice of an entirely unmodified starting state. The time it takes the system to adjust to establishing a modified string is negligible compared to the total time of one simulation run.

2.2 Enzyme models

2.2.1 Enzymes in *EpiDynaST*

The enzymes in *EpiDynaST* are mainly described by their reaction nature, the context needed in order to perform the reaction, their association and their dissociation rate. The reaction nature defines the change that the enzyme is performing on the nucleosomes' modifications. Generally, the enzymes are either modification adders or modification removers (see 2.2.2 for details).

The enzyme's context can be defined as the set of one or more nucleosome PTMs that must be present in a precisely determined neighbour-relation to the nucleosome that is intended to be changed. If the enzyme finds the needed context to be unfitting, the reaction of this enzyme with the determined nucleosome is not taken into the propensity sum of that simulation step (see the explanation of Gillespie's algorithm in 1.2.2). Random enzymes have a context which exclusively contains the one nucleosome that is about to be modified by the enzyme (see 2.2.2 for details). The reaction nature and context are defined together by a specific set of (possibly multiple) rules for each enzyme.

The association rate together with the enzyme's concentration define the enzyme's affinity to its substrate. The dissociation rate in turn defines the enzyme's speed concerning reaction and diffusion away from the modified nucleosome. On a sidenote, Gillespie's algorithm and thus *EpiDynaST* offer the possibility to model concentration depletion effects. This was not used in this work. Accordingly, all enzymes were assumed to be equally and infinitely available in the simulation.

The enzyme rule sets as well as their rates are defined symmetrically throughout the entirety of the simulations that were done for this work. Thus, for instance, every rule defining the addition of an acetyl group to a nucleosome next to one that already has an acetyl group is defined in either direction on the string and has a methylation counterpart at equal rates.

2.2.2 Enzyme types

A short summary of all enzyme types featured in this work can be found in tab. A.1.

Linear enzymes

Linear enzymes are used to extending sites containing a specific modification (e.g. acetylation, see fig. 2.1) by either propagating said modification from nucleosome to neighbouring unmodified nucleosome or by deleting an opposing modification next to a nucleosome with the desired modification. They exclusively have next neighbour reach.

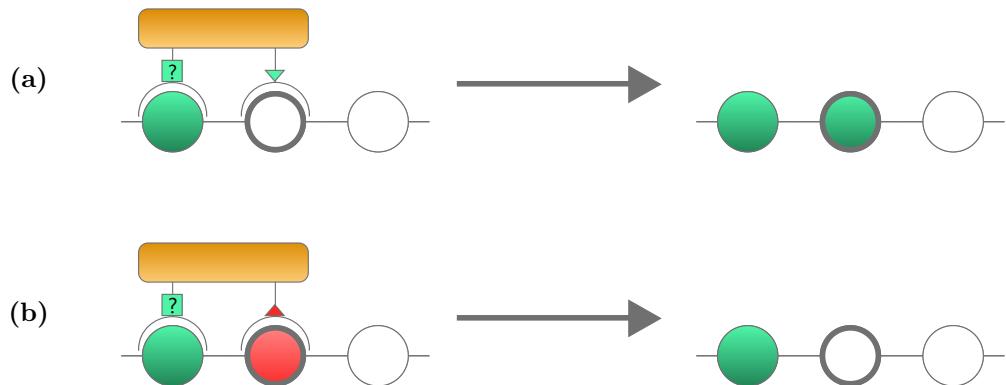


Figure 2.1: Simplified model of linear enzyme acetylation addition (a) and methylation removal (b) reactions. Acetylated nucleosomes are coloured in green, methylated nucleosomes are red and colourless ones are unmodified. The reactions shown are also defined in the rule set to occur in the opposite direction. Linear enzymes in favour of methylation extension (or acetylation deletion) work analogically.

Cooperative enzymes

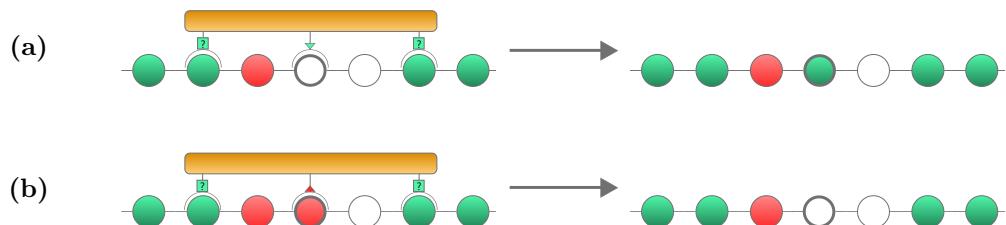


Figure 2.2: Simplified model of cooperative enzyme acetylation addition (a) and methylation removal (b) reactions. Acetylated nucleosomes are coloured in green, methylated nucleosomes are red and colourless ones are unmodified. The reactions shown are also defined in the rule set to occur in the opposite direction. Cooperative enzyme methylation addition and acetylation removal work analogically.

Cooperative enzymes generally read the state of two different nucleosomes on the string and write to or remove a modification from a third nucleosome. It is important to note that the two nucleosomes that are read do not need to have any next-neighbour relation to the one the enzyme is modifying. Thus, to some extent, these enzymes, unlike any other enzyme featured in this work, take the global modification trend on the string into account: if many nucleosomes are acetylated, then cooperative enzymes in favour of acetylation (meaning cooperative acetylation adders and cooperative methylation removers) are more active which, in turn, reinforces the acetylation distribution on the string.

The cooperative enzymes in this work are implemented in a way that they are always reading two nucleosomes that are equally far away from the nucleosome the enzyme wants to modify (see fig. 2.2).

The notion of 'space' of a cooperative enzyme defines the context reach of said

enzyme. For instance, a cooperative adder with a reach of 3 will read the nucleosome it wants to write on and ignore the 3 next neighbours of this nucleosome. It will only read the 4th nucleosomes situated to the left and the right of the first one. Accordingly, a cooperative enzyme which reads the next neighbours of the nucleosome to write on by definition has a reach of 0.

Random enzymes

Random adder and remover enzymes serve as noise in the system. As these enzymes act on one single nucleosome, they do not take into account any other nucleosomes on the string. As such, random adders are the only enzymes featured in this work that are able to modify a nucleosome on an otherwise completely unmodified string.

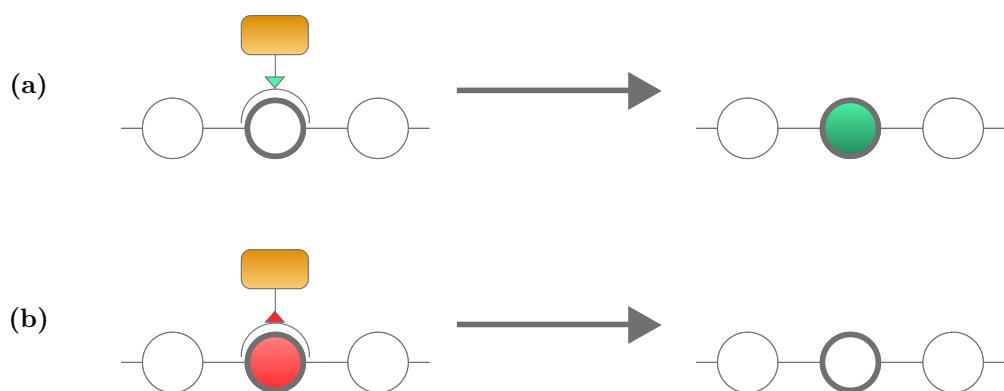


Figure 2.3: Simplified model of random enzyme acetylation addition (a) and methylation removal (b) reactions. Acetylated nucleosomes are coloured in green, methylated nucleosomes are red and colourless ones are unmodified. Random enzyme methylation addition and acetylation removal work analogically.

Random enzymes still have a context, as for example a random methylation adder cannot bind to any already modified nucleosome, only to an unmodified one. Given that any nucleosome can be targeted by a specific random enzyme at any moment, these enzymes' association rate should be smaller than the other ones' in the system by many orders of magnitude.

Random enzymes are not only “enzymes” in the true sense of the word when compared to the biological side of the model. They also exist in order to mimic the “noise” that is associated with such systems.

- Explain PRC2 with JARID2 DNA-binding TF subcomponent?

Completer enzymes

Completer enzymes are the only enzymes featured in this work which modify two different amino acids on one single nucleosome. These amino acids are exemplarily called K_x and K_y and, like the amino acid on the other nucleosomes, are both methylatable and acetylatable. Completer enzymes only act on one single nucleosome. They read one amino acid and write to or remove from the other one. As such, these enzymes are used

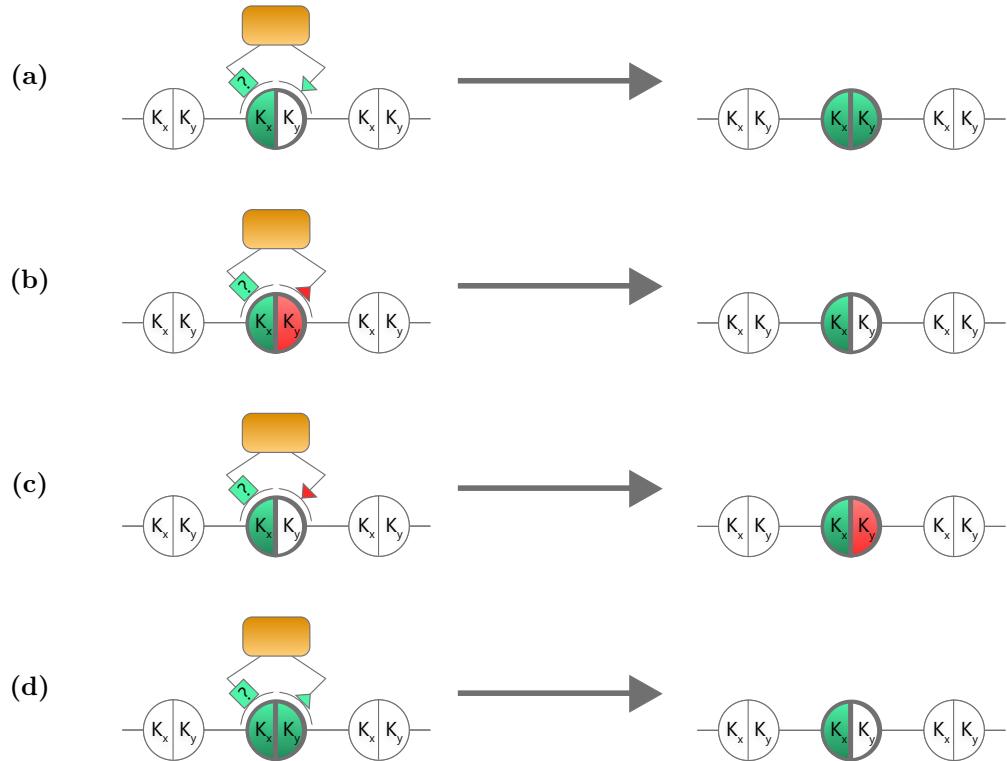


Figure 2.4: Simplified model of the enzyme acetylation addition and removal reactions. (a) and (b) show reactions in favour of “total” states, which means that they enforce one single type of modification on a nucleosome, while (c) and (d) respectively show addition and removal reactions favouring bivalent nucleosomes. The reactions shown are also defined in the rule set to occur in the opposite direction. Total enzyme methylation addition and acetylation removal work analogically.

in order to reinforce the generation of bivalent states or “total” (fully acetylated or fully methylated) states among the nucleosomes.

Completer enzymes have no known scientific background. They only serve the purpose of analysing the system dynamics when provoking bivalency or total states in the system.

2.2.3 Enzyme rule sets

This subsection provides an overview of the different enzyme rule sets used in the different simulations featured in this work. The rule sets are almost always kept constant throughout every section in the ‘Results’ chapter respectively. The enzyme rule sets used are the following (depending on the reader’s preference, the rule sets can also be looked up in tab. 2.1):

Table 2.1: Summary on whether an enzyme type was included in the designated experiment presented in the indicated 'Results' section. The **X** indicates that the enzyme type was featured, while the ~ indicates that there were some runs in this section that featured the referring enzyme type, while other runs in this same section did not. The explanation as to why this was done can be found in the respective sections.

	3.1	3.2	3.3	3.4	3.5	BivBist	3.6 Total	Bivalent
random adders	X	X						
random removers	X	X						
linear adders	X							
linear removers	X							
cooperative adders		X	X	X	X	X		
cooperative removers		X				~		
bivalent completer adders								X
bivalent completer removers								X
total completer adders							X	
total completer removers							X	
cyclic			X	X	X	X	X	X
non-cyclic	X	X						

2.3 Simulation details

The input files for every result section's *EpiDynaST* simulations can be found at [31]. In general, every *EpiDynaST* run needs 3 files: a *statefile*, a *rulefile* and a *paramfile*. The *statefile* contains the starting state of the nucleosome string. The *rulefile* contains all specifications to the enzymes: their contexts, the modification pattern, the association rate and the dissociation rate. The *paramfile* holds general information about the simulation itself with the most important one being the simulation time. This numeric parameter sets the exit condition for the algorithm.

The simulation time changes significantly between some runs. This is due to Gillespie's algorithm's event-based time approach. Changing the enzyme set can result in a significantly different number of possible reactions during the simulation which can lead to very short simulation step numbers. Thus, in order to grant statistically significant runs, the simulation time was increased for certain runs, where the overall run time was empirically found to be too short.

Furthermore, for reasons of performance and storage space, the graphs included in the results section were made from simulations with different simulation time. Preprocessing for the heatmaps was significantly more expensive than the other plots. Therefore, heatmaps were generated from *short* runs, whereas the other plots were generated from *long* runs. Meaning:

- *short*: Every step (event) is plotted and metadata are used to plot association numbers and relative binding time (resulting in heatmaps).
- *long*: Only every 1000th step is plotted. Given that the system is chaotic thanks to the random nature of the algorithm, regular plotting (f. ex. every 1000th data

point) results in a smoothing of the histogram because the chosen data points are more representative for the underlying distribution.

The different simulation parameters for the runs featured in this work are summarized in A.2.

3. Results

3.1 Non-cooperative enzymes do not entail bistable systems

In this section, we consider a system without cooperative enzymes. Thus, all the enzymes in the system read nucleosomes with a distance of 1 from the nucleosome that is written on. It is important to point out, that the nucleosome which is modified is included in the context as well.

Thanks to Mayer's pionous work [14], there were already some presumptions available that allowed to start from an educated guess. As pointed out before, bistability can not be achieved in such a system.

In this section, we will take a look at a system containing exclusively non-cooperative enzymes acting on a non-cyclic nucleosome string. The system contains random adder and remover enzymes, as well as linear adders and removers for methylation as well as acetylation respectively.

Looking at fig. 3.1, the histogram which counts the occurrence of active and silent nucleosomes respectively within the string shows a unimodal distribution throughout the simulation. Accordingly, the nucleosome string states achieved during the simulation indeed disclose a monostable system.

Fig. 3.1 exemplarily shows that the total number of active and silent states seems to ambulate around one same value. Although susceptible to some variance, this value can, throughout all simulations that were done, be approximated to 30, which is half of the totality of nucleosomes in the system.

On a sidenote, it should be mentioned that, in order to guarantee reproducibility and truthfulness of the statements derived from the plots, it was always made sure, that the state distribution indicated by the histograms in the plots were as close to each other as possible. However, in some cases, this goal could not be achieved, as can be seen in this case. Although always showing the same basic pattern, the histograms did not always look completely identical in additional simulations under the exact same conditions. Plots of the additional runs with the exact same starting files can be found in B.1. The impact of this issue will be evaluated in section 4.

As can be seen in fig. 3.2, every enzyme type (e.g. linear removers, random adders, etc.) shows similar activity, regardless of direction and modification subject (i.e. methylation or acetylation). However, compared to one another, the enzyme types show, in some cases tremendously, different degrees of activity. Interestingly, it seems that the linear adders were less active in adding modifications to the string than the random

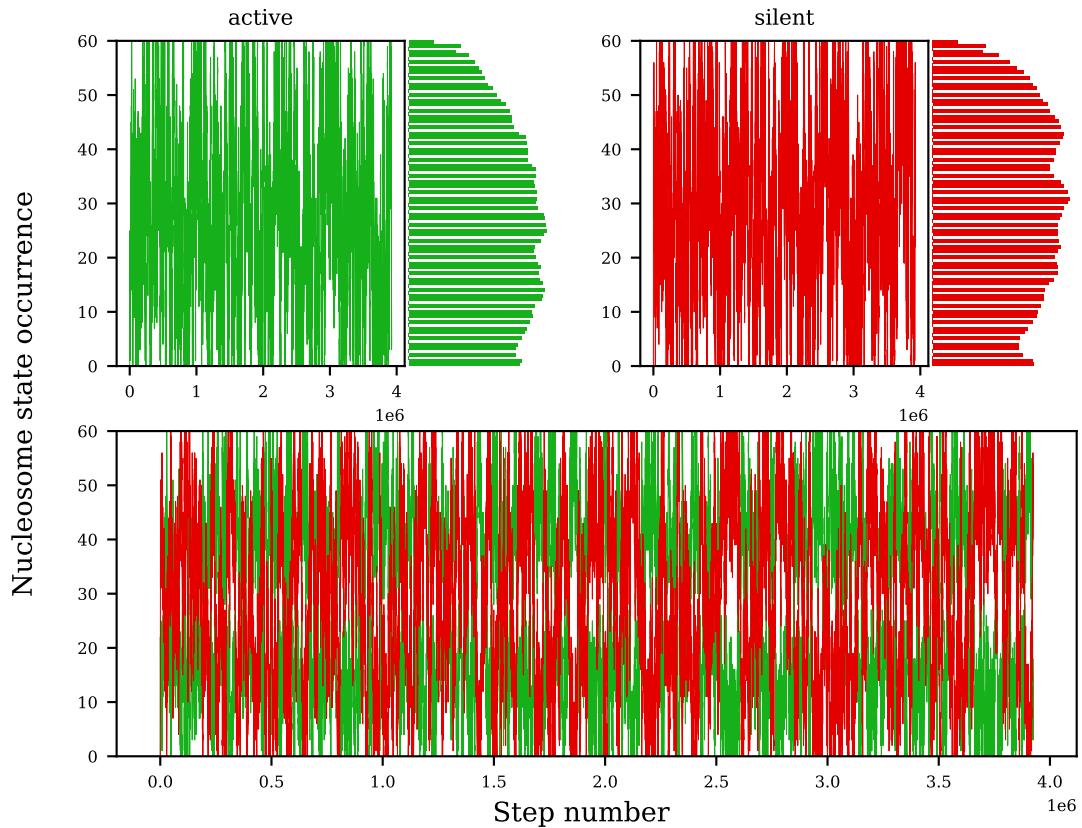


Figure 3.1: Absolute number of nucleosome states (active in green, silent in red) during the course of one long simulation (about 390 million reaction steps). Only every 1000th step is depicted. The enzyme rule set contains linear adders, linear removers, random adders and random removers. The rule set does not contain cooperative enzymes.

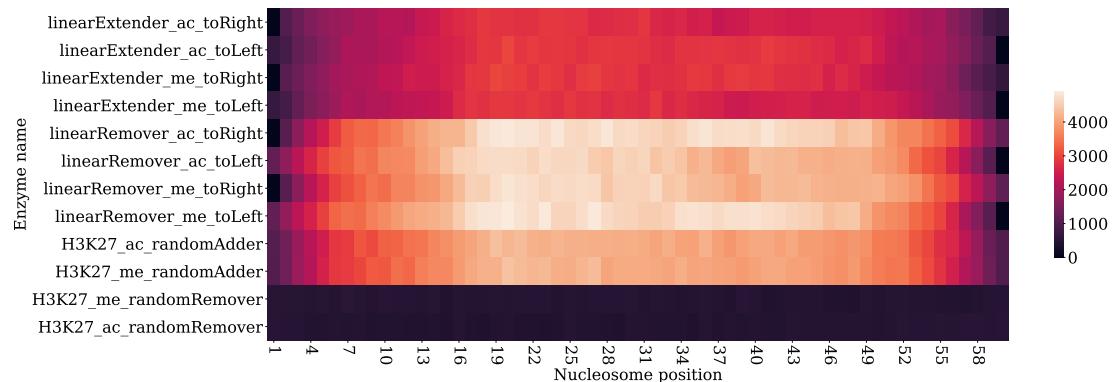


Figure 3.2: Heatmap depicting the absolute numbers of enzyme associations per enzyme (y-axis) and per nucleosome (x-axis) on the nucleosome string. The numbers were summed up over 5 simulations, where each run was simulated over 750,000 time steps on average.

adders, whereas the removers show the exact opposite picture. A system that largely depends on random adders is highly likely to show a significant amount of randomness or noise concerning the nucleosome state occurrence numbers on the string, which is exactly what can be seen in fig. 3.1.

Also, the distribution of the association events for one enzyme along the nucleosome string is not uniform. It can clearly be seen that for the linear adders, the linear removers and the random adders, the enzymatic activity is lower on the string borders. The fact that the very first nucleosome on the string does not bind any enzymes that have a reading context to the left (with convention of reading from nucleosome 1 on the left to nucleosome 60 on the right) naturally was to be expected. This phenomenon is also observable on the last nucleosome with enzymes which have a reading context to the right. Given that the linear adders and removers merely have a context reach of 1 (next-neighbour reach) and random adders do not have any reach at all, the lower border activity extending nucleosomes 1 and 60 is an interesting property of the overall system.

The small dominance around the state occurrence number of 30 could be an effect of equally strong linear adder enzymes which mainly build up exclusive methylation areas, as well as acetylation areas. The main activity in the system then is moved towards the centre of the string, where the areas collide. Such a phenomenon would also explain the high linear remover activity, as the colliding areas would present the fitting context for these enzymes.

Table 3.1: Enzyme types that are included in the rule set for the run illustrated in fig. 3.1 with their respective association rates. All enzymes' dissociation rates are at an equal rate of 100000. The enzyme rule set is symmetrical, which means that every enzyme type exists in favour of acetylation as well as methylation at equal rates respectively.

enzyme type	association rate
linear adder	20000
linear remover	20000
random adder	10000
random remover	2

Looking at tab. 3.1 which summarizes the enzymes' association rates in the system, it comes as a surprise that random enzymes are more active than the linear adders. This phenomenon could simply be attributed to the fact that linear adders have a more restrictive context and are thus inherently less likely to become active.

On a sidenote, the random removers' association rate is intentionally kept very low relatively in order to keep the string highly saturated with modifications. As there are many modifications in proximity to each other on the string this way, the concurrence between the enzyme types is much more analysable. Plus, as mentioned earlier, the random removers merely serve the purpose of adding noise to the system. Hence, it seems biologically justifiable to handle the random removers differently from the rest of the enzymes.

- If I change the random adder's rate to 20000 (same as linear enzymes), I get a system that holds one state. The big difference to cooperativity is that with cooperativity I get some variance around the states that occur most often. Include?

3.2 Bistability on a non-cyclic nucleosome string

3.2.1 Impact of cooperative enzymes

Contrarily to section 3.1, the simulations described in this section does contain cooperative adders and removers in the enzyme sets which, in theory, enables the system to show bistability. Other than these enzymes, the rule set only contains random adders and removers.

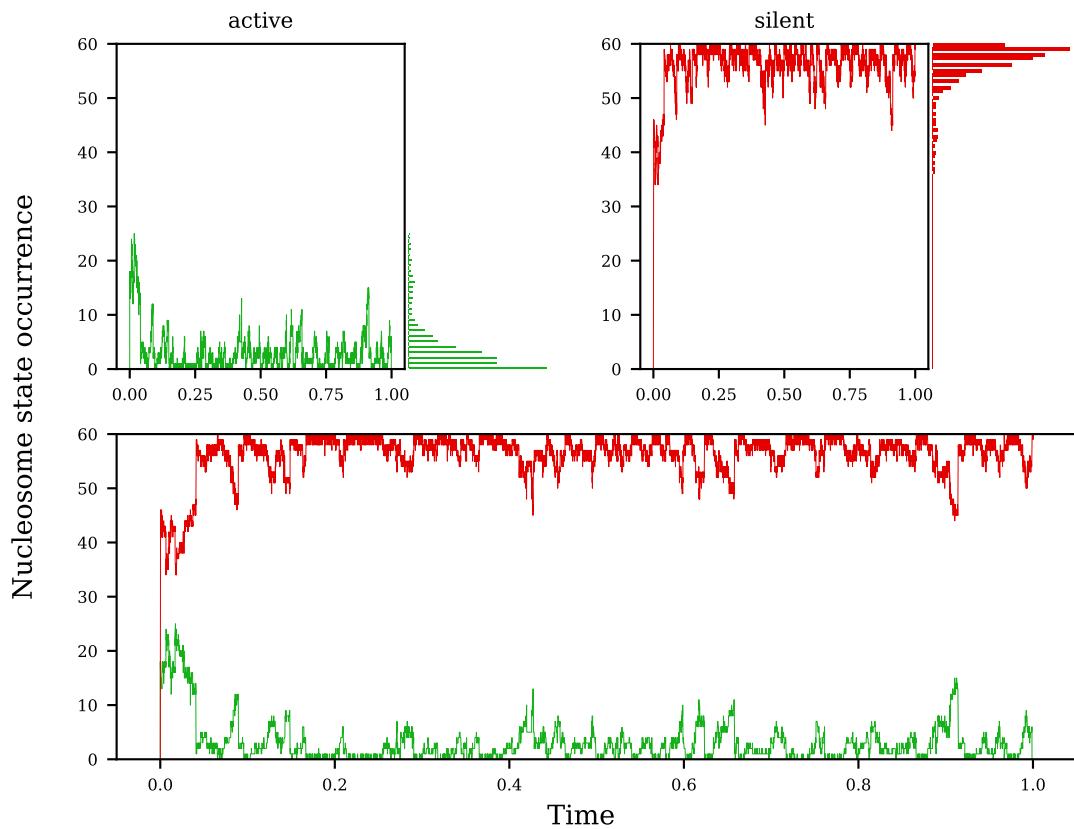


Figure 3.3: Absolute number of nucleosome states (active in green, silent in red) during the course of one long simulation (about 56000 reaction steps). The enzyme rule set contains cooperative adders, cooperative removers, random adders and random removers. The amount of reaction steps is significantly lower than in fig. 3.1 in order to be able to plot the heatmaps in figs. 3.4 and 3.5 with the data set from the same simulation.

Figs. 3.3, 3.4 and 3.5 are derived from a model containing the two cooperative enzyme types.

One can easily see in fig. 3.3 that the newly added enzymes seem to stabilize the predominant state (here the silencing one). It was found in additional simulation runs (see appendix B) that the active state could just as easily turn out as the predominant state for the entire run. This fact seems rather logical as the starting state is a completely unmodified string and the enzyme set that is used is purely symmetrical, which means

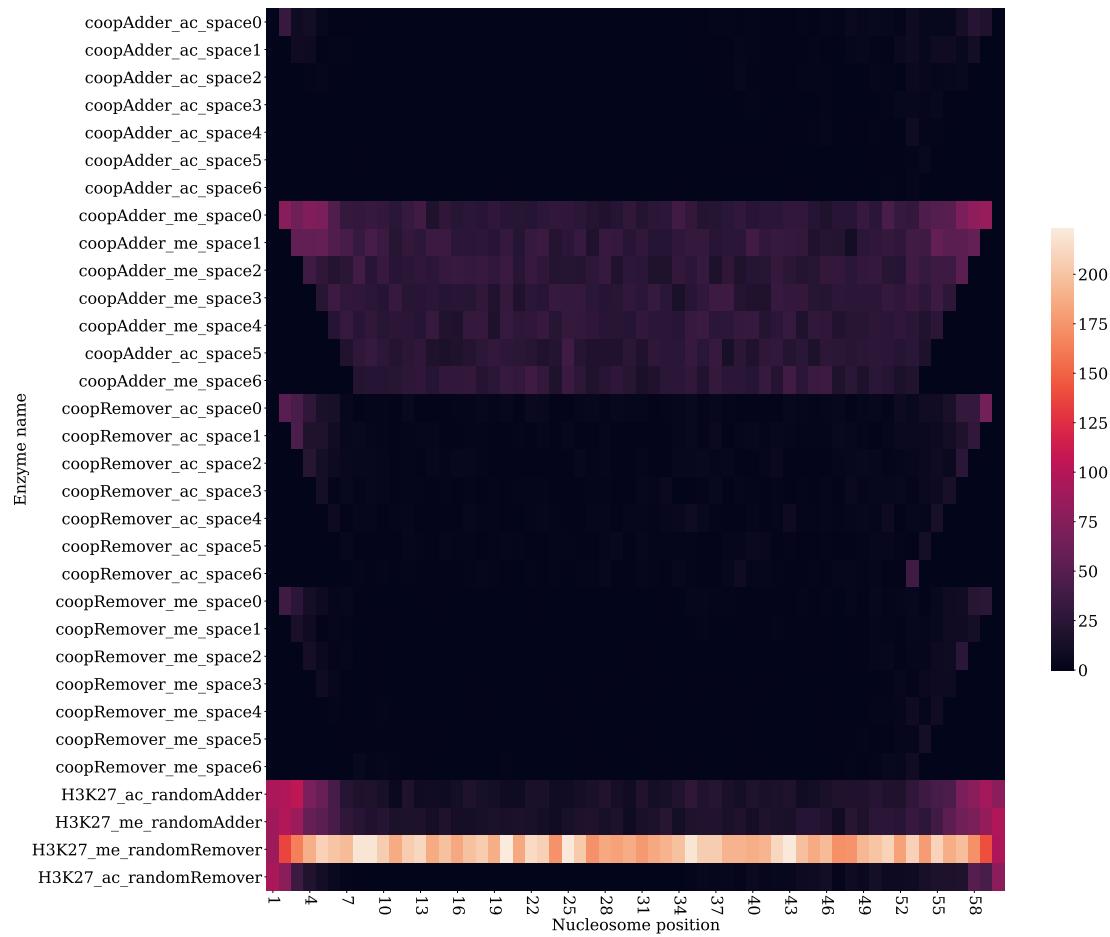


Figure 3.4: Heatmap depicting the absolute numbers of enzyme associations per enzyme and per nucleosome on the nucleosome string. The numbers originate from the same simulation plotted in fig. 3.3.

that it does not lean towards any of the two states. It is important to note, that the histogram in fig. 3.3 shows some variance around the states that occur most often which is an important condition for stability in the mathematical sense. In other words, the nucleosome string resists small perturbations and is able to return to a state that mostly contains one type of modifications, which will from now on be called a *complete* state.

These facts that the system can either stably find itself in a completely active as well as a completely silent state shows that bistability has been achieved with this enzyme set.

Figs. 3.4 and 3.5 provide more insights into the underlying mechanisms of the system. They originate from the same data as fig. 3.3 does. As for the notion of 'space' indicated for each cooperative enzyme, please refer to section 2.2.2. A few seemingly surprising factors are to be addressed here.

Every single enzyme type shows a trapeze-like shape originating from a decreasing writeable area on the string with increasing context reach. This observation is tied to the

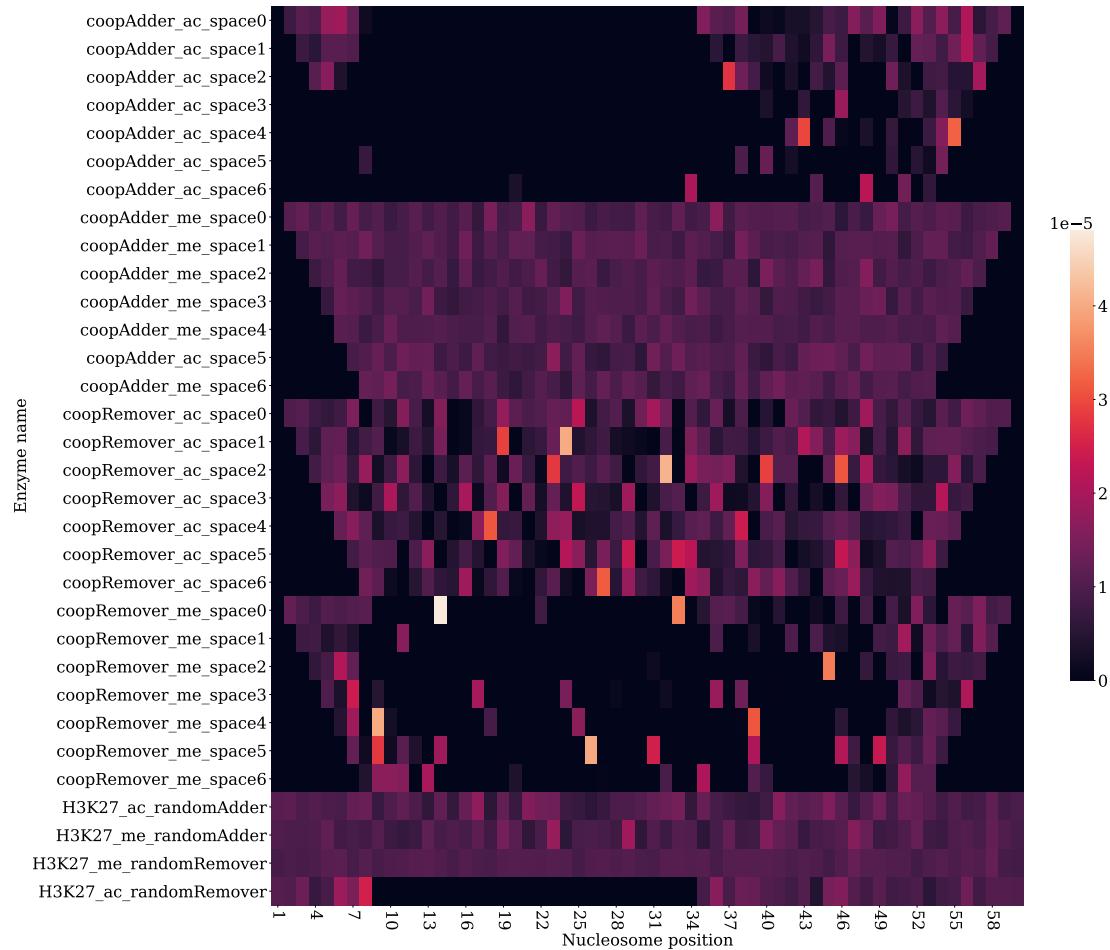


Figure 3.5: Heatmap depicting the average enzyme binding duration per enzyme and per nucleosome on the nucleosome string. The duration is defined as the number of time steps the enzyme has bound to one and the same nucleosome without dissociation event taking place in between. The numbers originate from the same simulation plotted in fig. 3.3.

inability of the enzymes to look beyond the string borders. As the context size increases, the enzyme is more and more unable to write onto a nucleosome increasingly further away from the border. This is an unwanted effect, as it does not reflect any behaviour that could be found in nature.

Fig. 3.5 reveals that there is a big discrepancy between the binding time of some enzymes compared to others. As the dissociation rates of the enzymes are very high and, importantly, precisely equal for every enzyme, completely dark spots must mean, that no enzyme has been active on that specific nucleosome at all. For instance, the cooperative acetylation adders only seem to have been active on some nucleosomes close to both borders. The rest of the variance must originate from the stochastic nature of the simulation. Thus, this kind of plot is very useful for determining, if an enzyme type has been active at all across the nucleosome string.

Referring to fig. 3.3, it is clear to see that the absolute number of associations of the random methylation remover is quite constant on most of the nucleosomes. As the random methylation remover's context is one single methylated nucleosome and, as *EpiDynaST* was used as simulation tool, it is made sure that there are no associations without a following reaction and dissociation, the association number of the random removers can be taken as a locator of the modification (i.e. methylation or acetylation) they are removing with quite acceptable efficiency. In other words, this fig. that shows the amount of enzyme associations per enzyme and per nucleosome directly indicates, where a modification type was most prevalent on the string.

The methylation mark was predominant throughout the entirety of the simulation. Given that the mentioned adders and removers are most active the more acetylation marks are already on the string, it does not seem surprising that the cooperative acetylation adders and the cooperative methylation removers show relatively low activity. The interesting activity pattern of those two enzyme types suggests that the acetylation subpopulations that can be seen in fig. 3.3 must have been prevalent at the borders of the string almost exclusively. This is supported by the fact that the random methylation remover shows reduced activity at the borders.

Thus can be concluded that it is very unlikely for any acetylation area to occur in the middle of the string, because empty spaces originating from random methylation removers are immediately filled up by cooperative methylation adders. The borders of the string, however, have a blocking effect on the cooperative methylation adders because with increasing space value, the nearest nucleosome that can be modified by them moves further and further away from the string border. The very first and very last nucleosomes are not at all modifiable by cooperative enzymes, hence the generation of acetylation subpopulations on the string borders.

3.2.2 Bistable switching on a non-cyclic string

Considering the biological implications of the system, it would be much more useful if the system could effectively switch from one state to another within the course of one simulation, i.e. without changing the enzyme types involved or their rates and without resetting the nucleosome string to a completely unmodified string.

This phenomenon, from here on called *bistable switching*, although rare, can be observed with the rule set described above (see fig. 3.6). With the rule set containing random adders, removers and cooperative adders and removers, the system shows a bimodal distribution originating from one single simulation. Bistable switching can be observed about once every 10th simulation. Among hundreds of simulations, none showed more than one bistable switching.

As can be seen in fig. 3.6 between the times 0.2 and 0.4, the system undergoes a predominance change rather progressively. In other simulation with the same settings (i.e. same enzyme rule set, starting state and simulation parameters), the system stayed in this middle range for a rather long time, sometimes progressing further and performing bistable switching, sometimes returning to the old state without undergoing bistable switching. It is quite hard to determine any points of no return on the upper or lower border of this saddle point.

In fig. 3.7, the main difference from fig. 3.5 in the previous subsection is that both

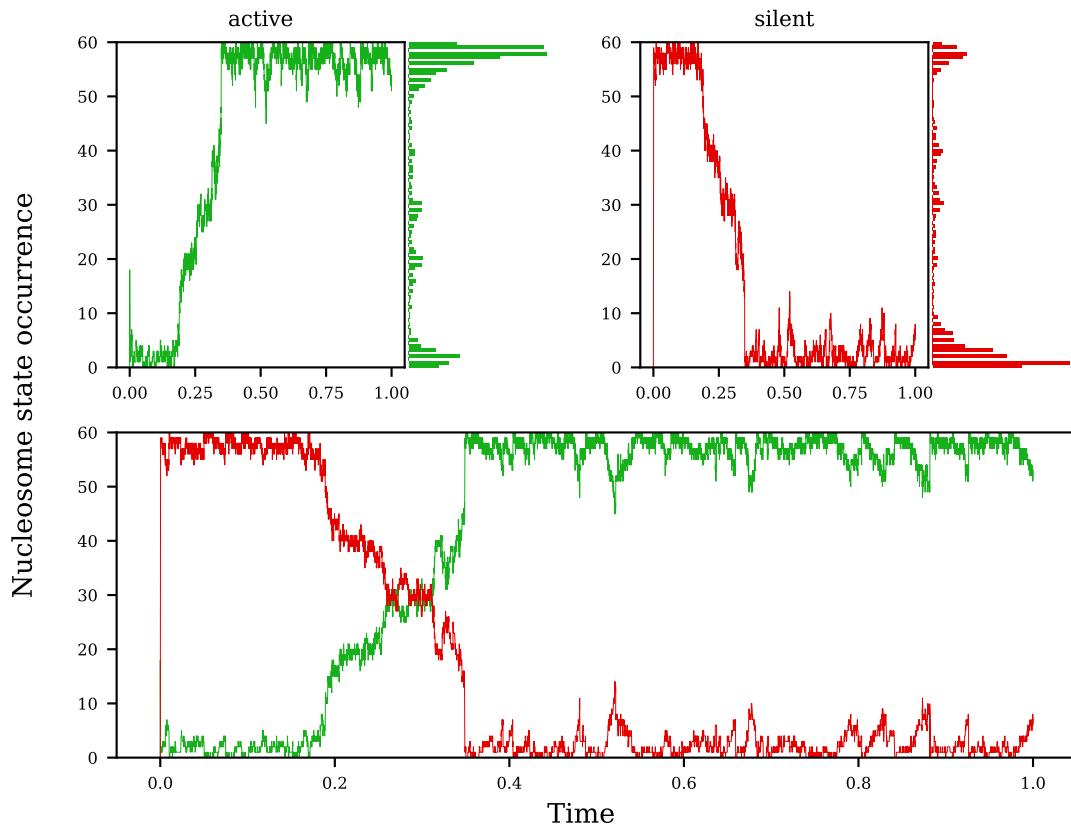


Figure 3.6: Absolute number of nucleosome states (active in green, silent in red) during the course of one simulation (about 56000 reaction steps). Every reaction step is depicted in the plot. The x-axis shows the reaction time given by Gillespie’s algorithm. The enzyme rule set contains cooperative adders, cooperative removers, random adders and random removers. Bistable switching, as seen in the plot, is only observed in about every 10th simulation of comparable reaction step size. The amount of reaction steps is significantly lower than in fig. 3.1 in order to be able to plot the heatmaps in figs. 3.7 and 3.8 with the data set from the same simulation.

cooperative adders have been active across the entirety of the string in the present simulation, which makes sense because both modifications were predominant at one point in time.

As can be seen in fig. 3.8, the random remover enzymes are by far the most active enzymes from the lot in this simulation, followed by the random adder enzymes which mainly show activity at the string borders. Apart from the random enzymes, the cooperative acetylation adders were active across the whole string while the other cooperative enzymes only were most active at the borders. Another interesting detail is that the cooperative methylation adders’ activity looks a little stronger from nucleosome 30 onward than on the other half of the string.

As already pointed out in 3.2.1, the acetylation mark must have taken over starting from the string borders, hence the intense enzymatic activity at these areas. The split

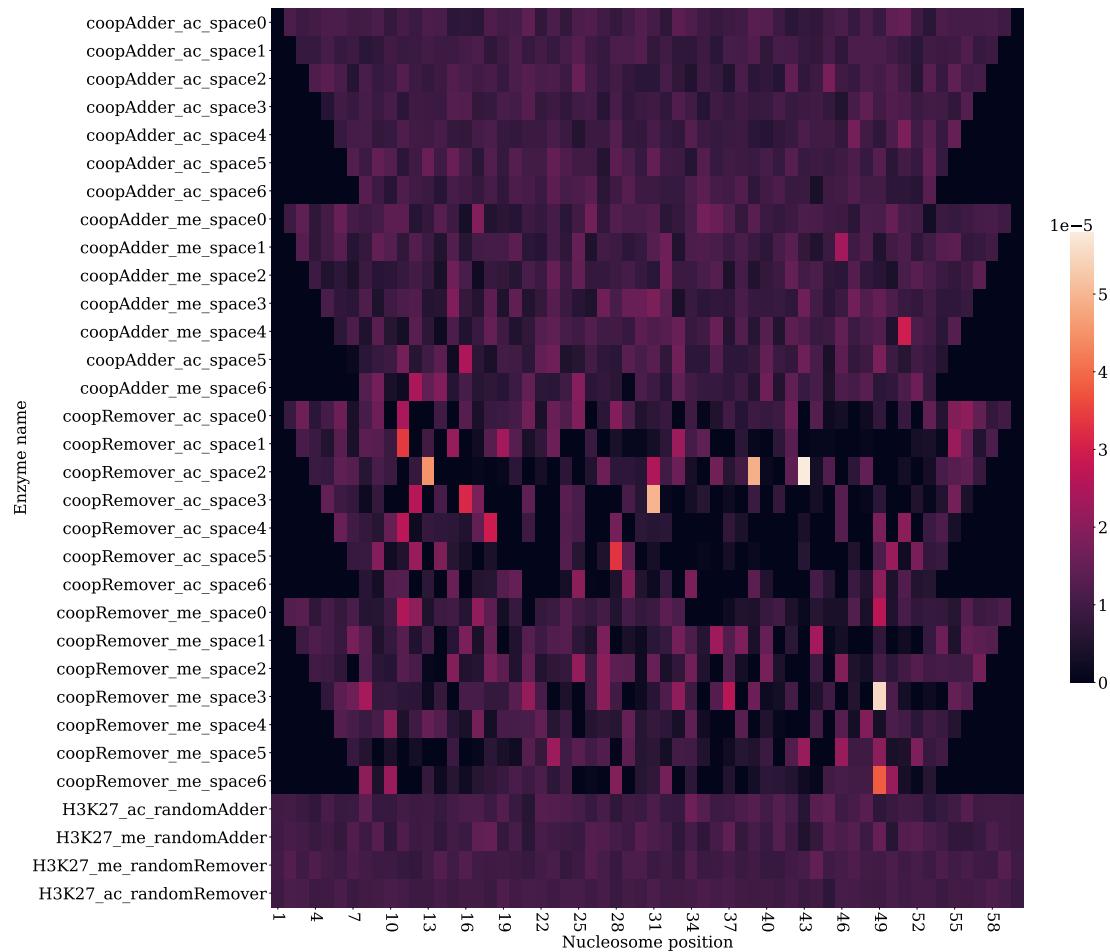


Figure 3.7: Heatmap depicting the average enzyme binding duration per enzyme and per nucleosome on the nucleosome string. The duration is defined as the number of time steps the enzyme has bound to one and the same nucleosome without dissociation event taking place in between. The numbers originate from the same simulation plotted in fig. 3.6.

activity level of the cooperative methylation adders is particularly indicative of how and from where exactly the acetylation mark must have taken over. As the system reaches the saddle point, the acetylation area must probably have progressed all the way to the centre of the string from the left (i.e. nucleosome 1) rendering the cooperative methylation adder particularly inactive in this area for $1/5$ of the simulation time. It is likely that the opposite effect was not seen on the cooperative acetylation adders' activity pattern because the acetylation modification is prevalent for a longer time span (about $3/5$), whereas the predominance period of the methylation modification is about as long as the system's prevalence periods in proximity of the saddle point (about $1/5$). These fraction indications are consistent even when looking at the absolute reaction step numbers. They are not distorted by the event-based time incrementation of Gillespie's algorithm.

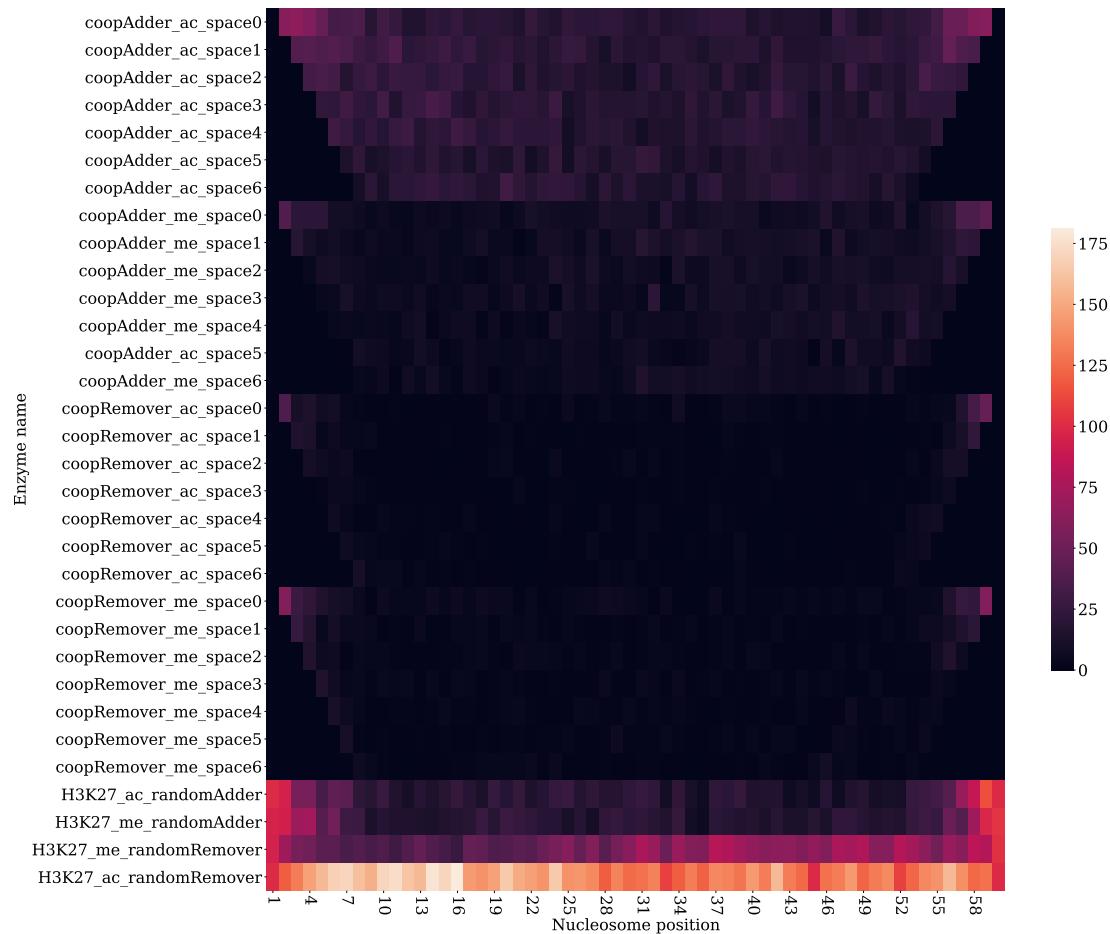


Figure 3.8: Heatmap depicting the absolute numbers of enzyme associations per enzyme and per nucleosome on the nucleosome string. The numbers originate from the same simulation plotted in fig. 3.6.

- elaborate a bit?

Many of the phenomena discussed in this section originated from the linearity of the nucleosome string. As such, it presents two borders which, as was pointed out, have massive impacts on the enzymes' activity and also bistability. As these borders hardly seem justifiable from a biological point of view, it would be interesting and much more relevant biologically to achieve bistable switching on a nucleosome string without borders. This will be examined in the next section.

- tell about the existence of runs where it is stuck in the middle state (“saddle point”) as can be seen in fig.

3.3 Bistable switching on a cyclic string

As pointed out before, a linear nucleosome string with a finite number of nucleosomes that is reduced of length because of computational feasibility might not be the best when aiming at designing a most general model that shows bistable switching. The border effects of such a non-cyclic chain are evident as of the previous subsection.

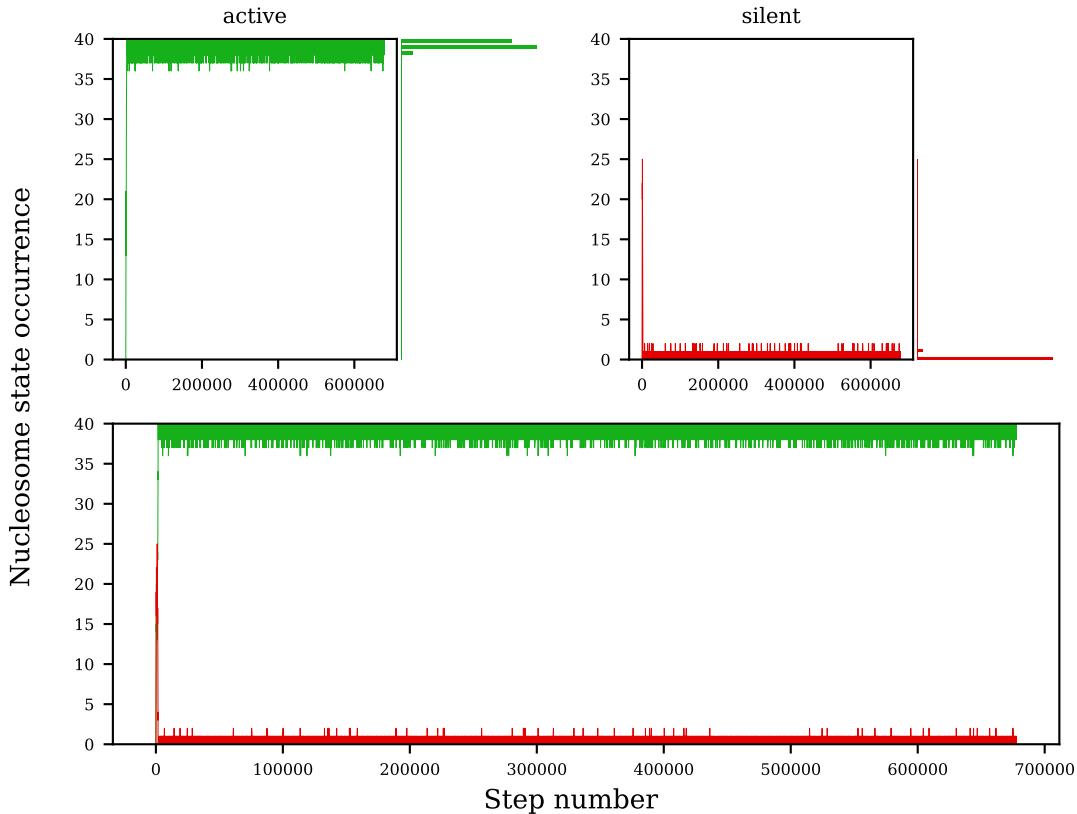


Figure 3.9: Absolute number of nucleosome states (active in green, silent in red) during the course of one long simulation (about 678 million reaction steps). Only every 1000th step is depicted. The enzyme rule set contains random adders, random removers and cooperative adders (no cooperative removers).

EpiDynaST provides a means to work with circular nucleosome chains that formally establishes a direct next-neighbour relationship between the very first and very last nucleosome in the string. This method was used from this point on in the rest of the work. As border effects could now be effectively excluded, the nucleosome string length was reduced from 60 to 40 in order to reduce computation expenses. It was made sure, that the enzymes' maximum reach in the system were always low enough in order for the reduced string length not to trigger interlacing effects or the like.

With the borders eliminated, no bistable switching could be observed any more (see fig. 3.9). After one of the 2 modification types turns out predominant, it stays like this with very little variance. This was to be expected, as the string borders helped the non-

predominant modification to stay on the string end in the first place before growing its area across up to the other end of the string. As was pointed out in the previous section, the cooperative removers are much too strong for a non-prevalent modification to survive long enough without the protecting borders.

The logical next step was to remove the cooperative removers from the rule set altogether. As can be seen in figs. 3.10 and 3.11, this step was decisive for obtaining bistable switching.

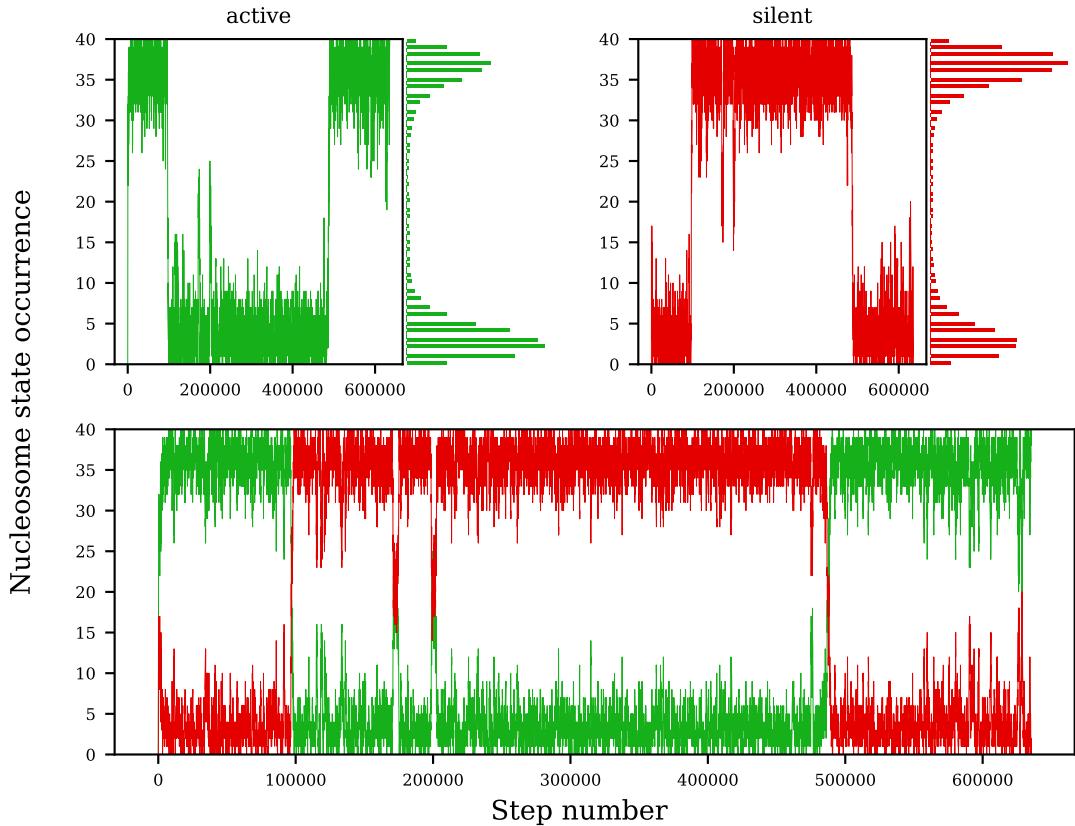


Figure 3.10: Absolute number of nucleosome states (active in green, silent in red) during the course of one long simulation (about 634 million reaction steps). Only every 1000th step is depicted. The enzyme rule set contains random adders, random removers and cooperative adders (no cooperative removers).

As can be seen in these two figs., the removal of the cooperative removers not only entails the reappearance of bistable switching. Additionally, bistable switching is much more frequent in the simulations so much so that even multiple switchings in one simulation is not a rare event. One can also see that the histogram peaks moved a bit closer towards the axis centre of 20 which means that the most stable states are not complete states any more. Instead, there is always a considerable subpopulation of about $15\% \pm 10\%$ of the opposite modification.

Although the system passes the saddle point around 20 nucleosomes quite often, it

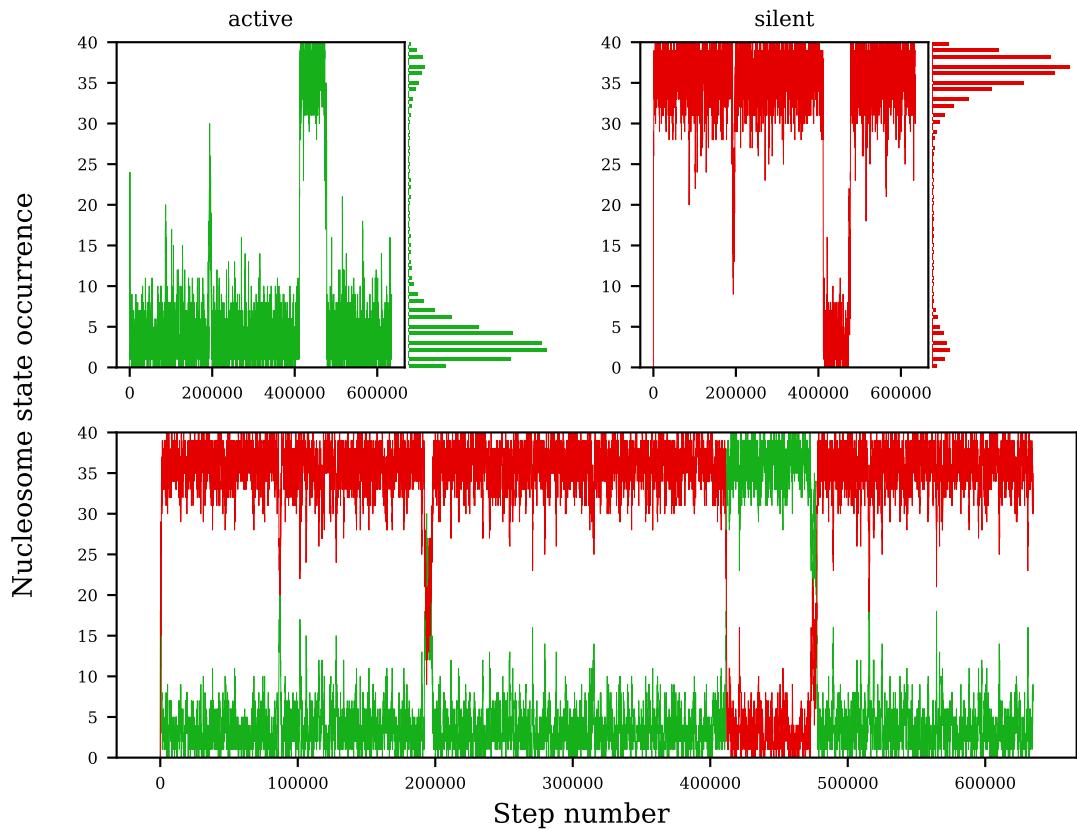


Figure 3.11: Absolute number of nucleosome states (active in green, silent in red) during the course of one long simulation (about 634 million reaction steps). Only every 1000th step is depicted. The enzyme rule set contains random adders, random removers and cooperative adders (no cooperative removers).

seems that this state is never populated for a very long time, as is shown by figs. 3.10 and 3.11. Accordingly, the point of no return seems much more narrow than in the linear case with cooperative removers.

- describe the low switching frequency
- describe the high variance in state length (that's why fig. 3.11 is asymmetric)
- maybe determine the frequency (cutoff at 30 etc.) with statistics?
- produce a plot that shows which modification follows which per nucleosome and calculate the methylation/acetylation probability. Maybe do the same with the enzymes ($p(\text{methylation}) = p(\text{coop_me_adder}) + p(\text{random_me_adder})$) while in the same spot $p(\text{acetylation}) = p(\text{random_ac_adder})$)

3.4 Influence of the dissociation rate

In the previous sections, the parameter that was most focussed on were the enzyme types in the rule set. The dissociation rate, however, was kept at a very high rate. This section serves the purpose of examining, what happens if the dissociation rate enters in concurrence with the associations throughout the simulation.

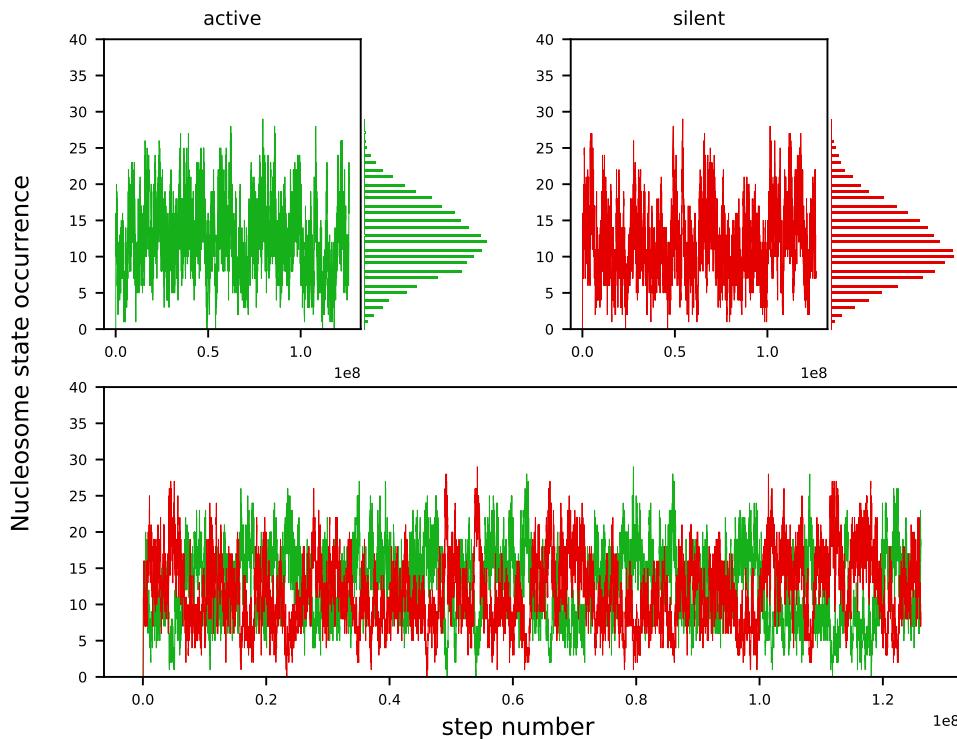


Figure 3.12: Absolute number of nucleosome states (active in green, silent in red) during the course of one long simulation (about 639 million reaction steps). Only every 1000th step is depicted. The enzyme rule set contains random adders, random removers and cooperative adders (no cooperative removers).

In the previous simulations, enzymes were bound for an average of 2 or 3 events before writing and dissociating from the bound nucleosome. Decreasing the dissociation rate inherently means a longer binding time on the nucleosome. Fig. 3.12 shows that this longer binding time because of a lower dissociation rate (100) has severe effects on the whole simulation outcome compared to a simulation with exactly the same parameters except for a 1000 times higher dissociation rate (100000) (see fig. 3.13).

Fig. 3.12 discloses that bistability, although possible from a rule set point of view, can be eliminated by a low dissociation rate. During binding time, the enzymes are blocking the ability for other enzymes to bind to the same nucleosome and, in a way, act as protective groups. As a result, the number of possible events for each respective time step is massively reduced. This is also shown by the much lower number of total

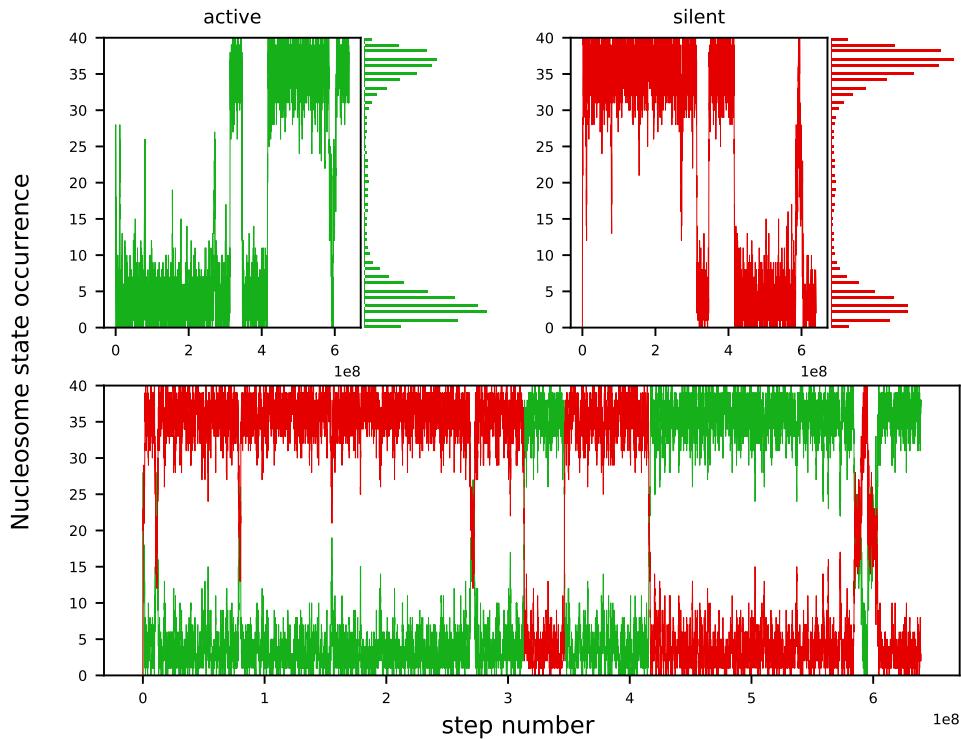


Figure 3.13: Absolute number of nucleosome states (active in green, silent in red) during the course of one long simulation (about 126 million reaction steps). Only every 1000th step is depicted. The enzyme rule set contains random adders, random removers and cooperative adders (no cooperative removers).

reaction events during the simulation from fig. 3.12 compared to fig. 3.13, as the Gillespie algorithm increases the time elapsed between events the lower the number of possible events is (see 1.2.2).

The protective behaviour of enzymes with low dissociation rate take away the velocity of the system and

Thus, it seems that the dissociation rate is an important influential factor that has to be taken into account or “eliminated” by raising its value high enough when trying to achieve bistable switching.

- high vs. low dissociation rate
- explain “protective groups” phenomenon
- verify: Is it always the same nucleosomes that are modified?
- discussion: discuss the biological significance of this

3.5 The boundaries of bistability

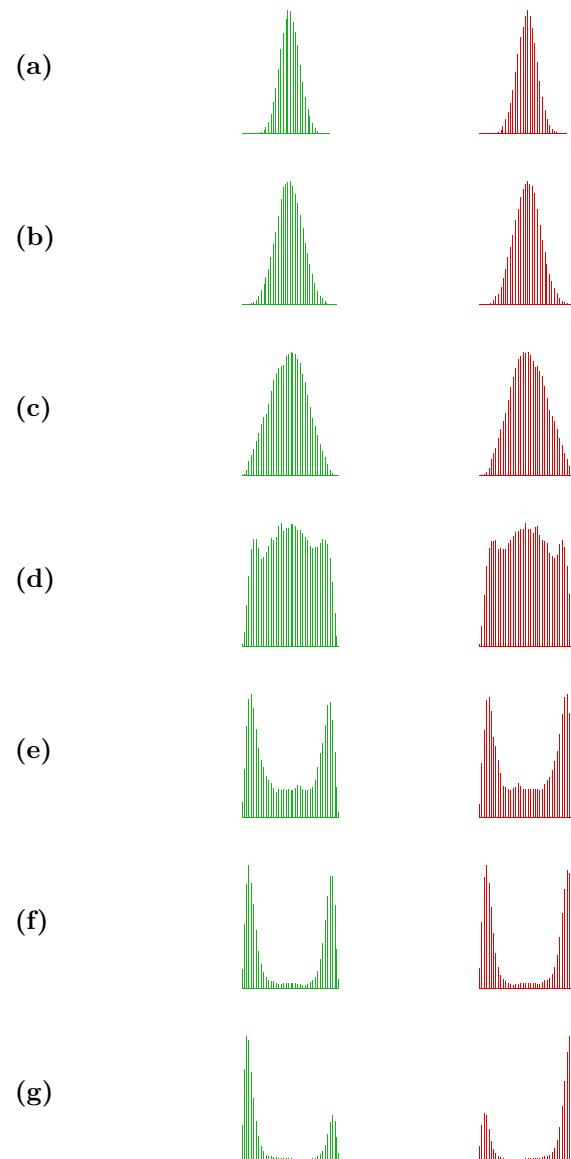


Figure 3.14

- describe loss of reach
- As can be seen in (g), the symmetric case becomes more and more improbable because switching gets more rare and ?

3.6 Bivalency

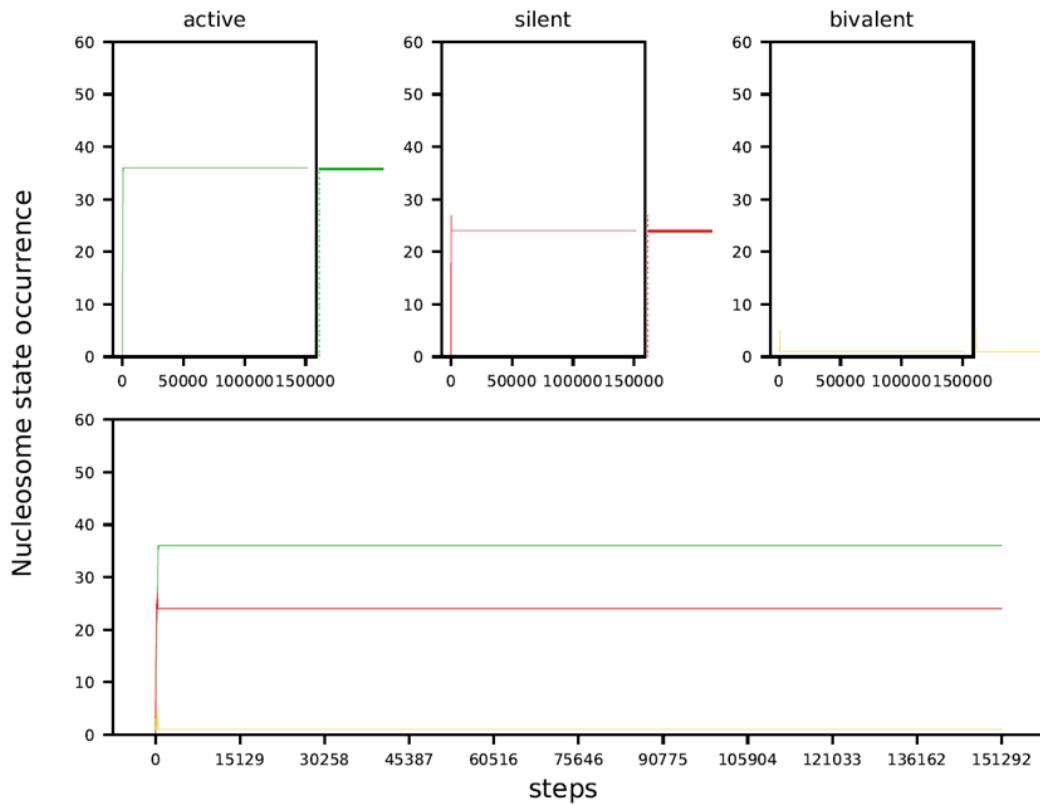


Figure 3.15

- Here, we are at K_x+K_y
- Two systems that either favour bivalence or total active/silent states as an introduction to bivalence
- Frequent switching and bivalence

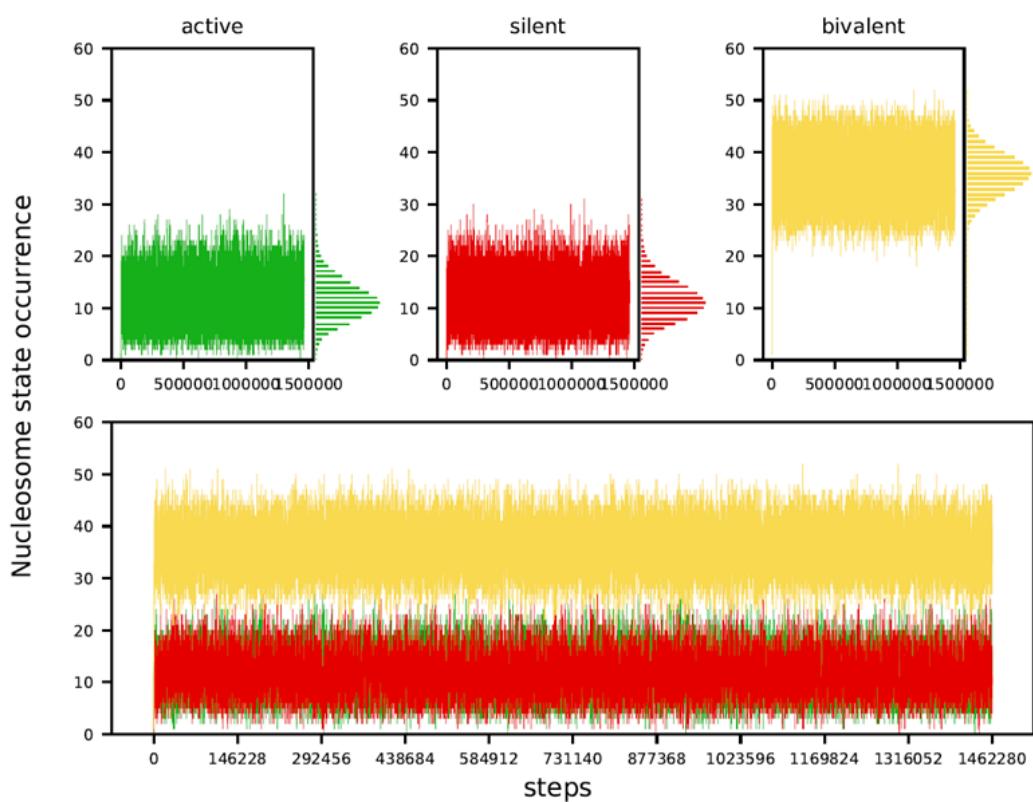


Figure 3.16

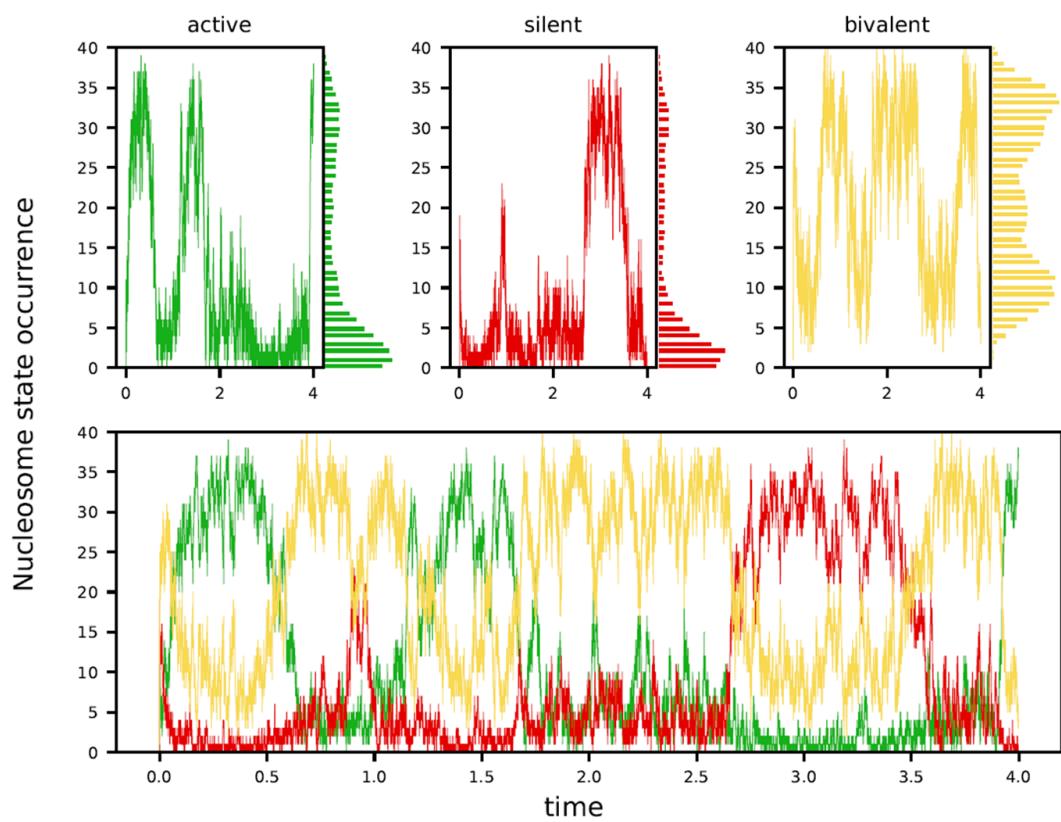


Figure 3.17

4. Discussion

- Summarize the findings.

4.1 Conditions for bistability

Hence, even though, there is evidence about the occurrence of fully activated and fully silenced states, this is only a statistical effect. The probability of the system being in an intermediate state is way higher, which disqualifies this system of being an effective “switch” between two extreme states.

- Bistability (condition: cooperativity) (Explain difference between biological and Sneppen coop.) (Errors in Sneppen’s insights on the topic)
- Impact of dissociation rate: The dissociation rate was not taken into account for most of the simulations. Setting one and the same high dissociation rate that hardly competes with the association rates of the enzymes smoothens the nucleosome state histograms and thus reduces noise in the system. In terms of epigenetic landscapes, a variable dissociation rate would introduce much complexity into the system, as the number of relevant states would increase. This is the case because every state now would not only be sufficingly described by its modification states but also by the number and type of enzymes bound to the nucleosomes. However, even though the complexity increases significantly, there would not be any inherent biological experience gain from this.

4.2 Proposed mechanism for bistable switching

4.2.1 Bistable switching: a biologically useful phenomenon

4.2.2 cyclic

4.2.3 non-cyclic

- A state seed will immediately be extinguished by an opposing cooperative remover. In order to prevent this, there are two possibilities:
 - Build the seed in a corner where the coop Remover cannot be active
 - Remove coop Removers

5. Outlook

5.1 Make *EpiDynaST* account for recent biological findings

Schuettengruber et al. in [12] (p44f) explain that PREs (Polycomb response elements) are responsible for the chromatin forming loops (TADs = topologically associating domains) and are thus able to form large silenced areas of condensed chromatin. These 3D-formations are critical for HOX gene regulation. It is also known that many active gene promoters interact with their enhancers and other promoters in a 3d-fashion. [32] These findings suggest that it might be beneficial for *EpiDynaST* simulations to further explore possibilities to emulate fixed and dynamic 3D interactions within the nucleosome string.

5.2 Explore recently discovered non-cooperative edge case

A quite unexpected result was found in a system of exclusively non-cooperative enzymes at a very specific association rate ratio summarized in table 5.1. The state adoption in this system resembles bistable behaviour as can be seen in fig. 5.1. One can clearly see that there are two very distinct peaks at complete activation and complete silencing of the nucleosome string. Even though this system seems to be bistable, it shows one very important difference to the bistable systems explored earlier in this work: the lack of variance around the peaks in the histogram.

Table 5.1: Enzyme types that are included in the rule set for the run illustrated in fig. 5.1 with their respective association rates. All enzymes' dissociation rates are at an equal rate of 100000. The enzyme rule set is symmetrical, which means that every enzyme type exists in favour of acetylation as well as methylation at equal rates respectively.

enzyme type	association rate
random adder	100
random remover	2
linear extender	20000
linear remover	20000

As stated earlier, the histograms' peak variance can be understood as the area where the system has a more or less strong tendency to return to the peak's state. Or,

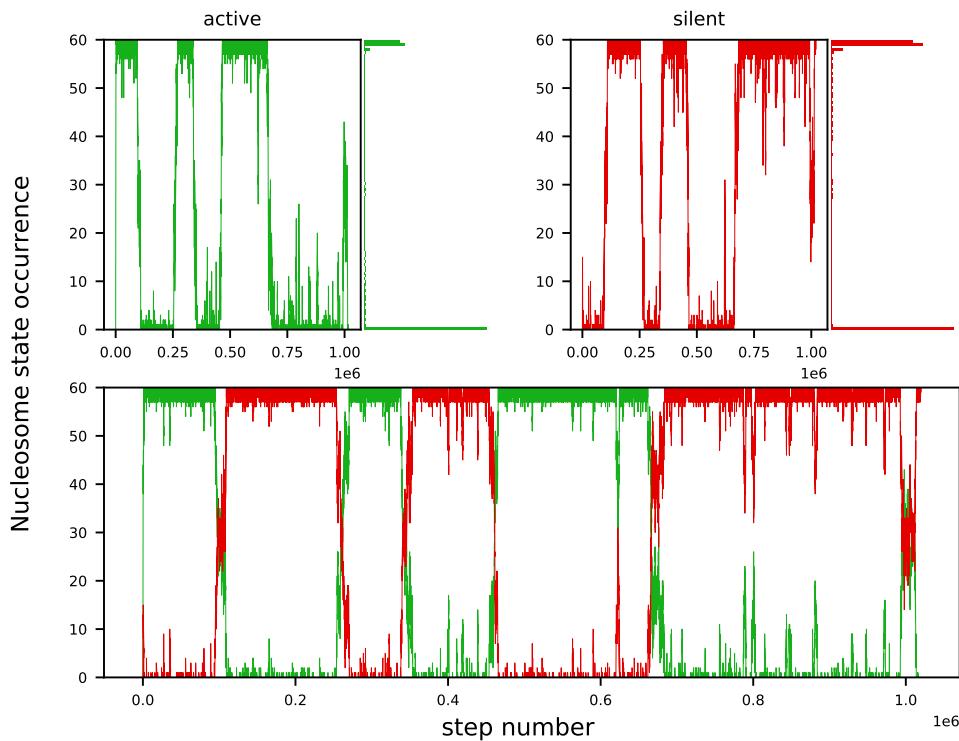


Figure 5.1: Absolute number of nucleosome states (active in green, silent in red) during the course of one long simulation (about 3.9 million reaction steps). The enzyme rule set contains linear extenders, linear removers, random adders and random removers. The rule set does not contain cooperative enzymes.

in mathematical terms, the variance is the area in the system's underlying vector field that points in the critical point's direction.

This behaviour is not observed in the non-cooperative case described in fig. 5.1. Thus, one single opposite modification added randomly to the string entails the system's return to a random walk. As the term "stability" describes a state's resistance against small deviations and the addition of one single modification is nothing more than an atomic step for this system one can say, that this non-cooperative case does not qualify as being bistable in the true sense of the word.

Nonetheless, it is an interesting phenomenon worth exploring, as it remains debatable if "true bistability" is really needed in the biological context.

A. Specifications

A.1 Summary of enzyme types

Table A.1: Summary of all the enzyme types that are referenced in this work. It is important to mention, that, for reasons of comprehensiveness, this table exemplarily only includes the rules which are in favour of acetylation (i.e. acetylation adders and methylation removers). However, every enzyme set used in this work is completely symmetrical. This means that if an enzyme set contains a linear acetylation adder it also contains a linear methylation adder at equal association and dissociation rates respectively and so forth.

type	operation	target state	final state
random	adder	-○-	-●-
	remover	-●-	-○-
linear	adder	-●○○-	-●●○○-
	remover	-●○●○-	-●●○○-
cooperative	adder	-●○(○) _n ○(○) _n ●-	-●○(○) _n ●○(○) _n ●-
	remover	-●○(○) _n ●○(○) _n ●-	-●○(○) _n ○○(○) _n ●-
bivalent completer	adder	-K ₁ K ₂ -	-K ₁ K ₂ -
	remover	-K ₁ K ₂ -	-K ₁ K ₂ -
total completer	adder	-K ₁ K ₂ -	-K ₁ K ₂ -
	remover	-K ₁ K ₂ -	-K ₁ K ₂ -

A.2 Summary of simulation parameters

- Explain coop?
- Decide on whether to insert cyclic?

Table A.2: Summary of the simulation parameters explained in 2.3 sorted by subsections in the 'Results' section.

		coop?	cyclic?	diss?	simulation time
3.1	long	no coop	no	100000	5
	short	no coop	no	100000	1
3.2	long	no coopRem	no	100000	5
	long	w coopRem	no	100000	5
	short	no coopRem	no	100000	1
	short	w coopRem	no	100000	1
3.3	long	no coopRem	yes	100000	20
	long	w coopRem	yes	100000	20
	short	no coopRem	yes	100000	8
	short	w coopRem	yes	100000	8
3.4	long	no coopRem	yes	100000	20
	long	no coopRem	yes	100	20
	short	no coopRem	yes	100000	6
	short	no coopRem	yes	100	6
3.5	maxReach0 (long)	no coopRem	yes	100000	20
	maxReach1	no coopRem	yes	100000	20
	maxReach2	no coopRem	yes	100000	20
	maxReach3	no coopRem	yes	100000	20
	maxReach4	no coopRem	yes	100000	20
	maxReach5	no coopRem	yes	100000	20
	maxReach6	no coopRem	yes	100000	20
3.6	BivalentBistability (short)	no coopRem 2 K	yes	100000	10
	FavBivalency (short)	w coopRem	yes	10000	2
	FavTotal (short)	w coopRem	yes	10000	2

- Add subsubsections to table

B. Additional runs

B.1 Additional runs from section 3.1

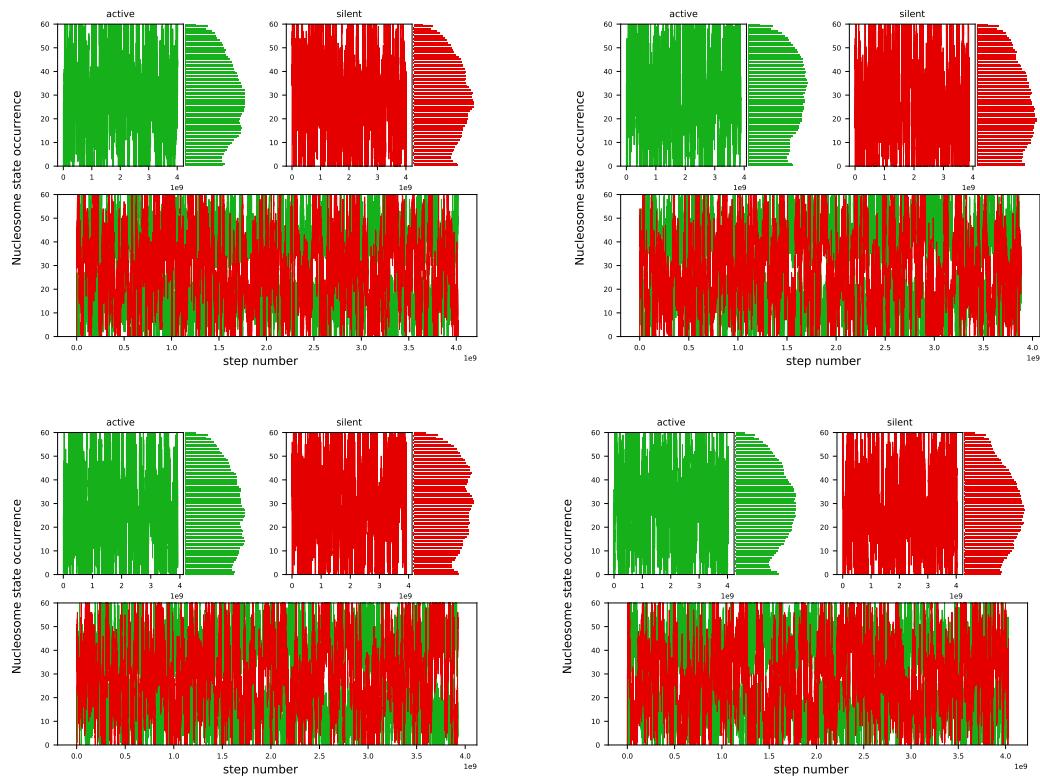


Figure B.1: Runs from 3.1

B.2 Additional runs from section 3.2

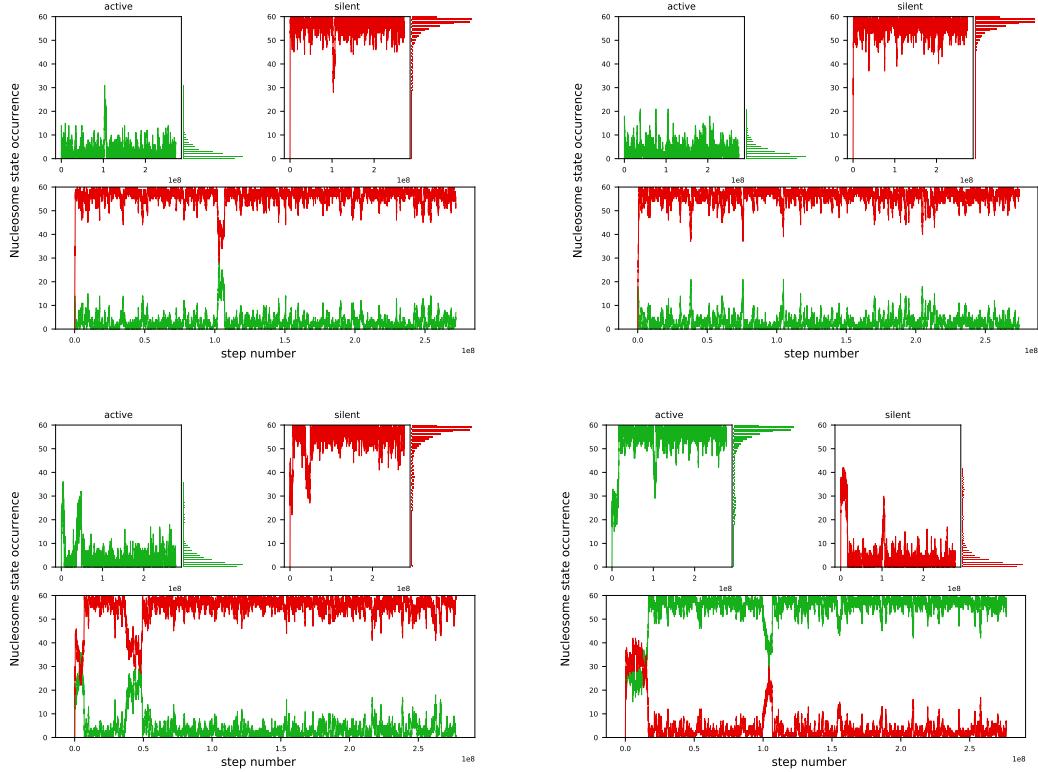


Figure B.2: Runs from 3.2

References

Literature

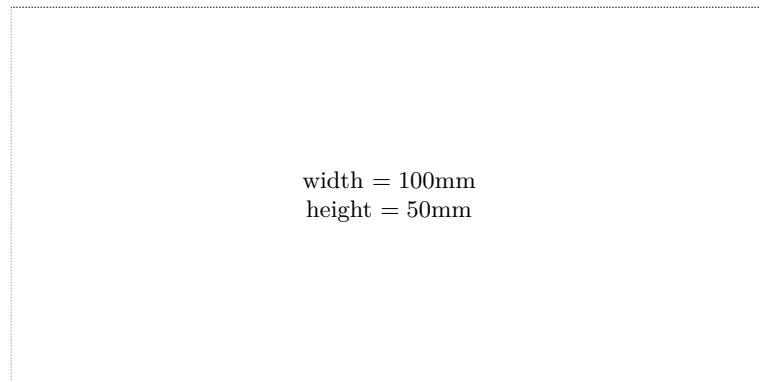
- [1] Jeremy M Berg, Lubert Stryer, and John L Tymoczko. *Stryer Biochemie*. Springer-Verlag, 2015, pp. 950–964 (cit. on pp. 1, 2).
- [2] James Cooper et al. “Genome-wide mapping of DNase I hypersensitive sites in rare cell populations using single-cell DNase sequencing”. *nature protocols* 12.11 (2017), p. 2342 (cit. on p. 2).
- [3] Andrew J Bannister and Tony Kouzarides. “Regulation of chromatin by histone modifications”. *Cell research* 21.3 (2011), pp. 381–395 (cit. on p. 2).
- [4] Dorine Rossetto, Nikita Avvakumov, and Jacques Côté. “Histone phosphorylation: a chromatin modification involved in diverse nuclear events”. *Epigenetics* 7.10 (2012), pp. 1098–1108 (cit. on p. 2).
- [5] Hengbin Wang et al. “Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage”. *Molecular cell* 22.3 (2006), pp. 383–394 (cit. on p. 2).
- [6] Tianyi Zhang, Sarah Cooper, and Neil Brockdorff. “The interplay of histone modifications-writers that read”. *EMBO reports* 16.11 (2015), pp. 1467–1481 (cit. on p. 2).
- [7] Catherine A Musselman et al. “Perceiving the epigenetic landscape through histone readers”. *Nature structural & molecular biology* 19.12 (2012), p. 1218 (cit. on p. 2).
- [8] Eva J Ge et al. “Nucleation and propagation of heterochromatin by the histone methyltransferase PRC2: Geometric constraints and impact of the regulatory subunit JARID2”. *Journal of the American Chemical Society* 141.38 (2019), pp. 15029–15039 (cit. on p. 2).
- [9] Lei Zeng and Ming-Ming Zhou. “Bromodomain: an acetyl-lysine binding domain”. *FEBS letters* 513.1 (2002), pp. 124–128 (cit. on p. 2).
- [10] Bluma J Lesch and David C Page. “Poised chromatin in the mammalian germ line”. *Development* 141.19 (2014), pp. 3619–3626 (cit. on p. 4).
- [11] Stephan H Bernhart et al. “Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer”. *Scientific reports* 6.1 (2016), pp. 1–18 (cit. on p. 4).

- [12] Bernd Schuettengruber et al. “Genome regulation by polycomb and trithorax: 70 years and counting”. *Cell* 171.1 (2017), pp. 34–57 (cit. on pp. 4, 41).
- [13] DS Lemons, A Gythiel, and Paul Langevin’s. “paper “Sur la théorie du mouvement brownien [On the theory of Brownian motion]””. *CR Acad. Sci. (Paris)* 146 (1908), pp. 530–533 (cit. on p. 4).
- [14] Daniel E Mayer. “Histone Driven Cell Differentiation”. MA thesis. Institute of Computer Science, University of Leipzig, 2020 (cit. on pp. 4–6, 8–11, 20).
- [15] Daniel T Gillespie. “A general method for numerically simulating the stochastic time evolution of coupled chemical reactions”. *Journal of computational physics* 22.4 (1976), pp. 403–434 (cit. on p. 5).
- [16] Daniel T Gillespie. “A rigorous derivation of the chemical master equation”. *Physica A: Statistical Mechanics and its Applications* 188.1-3 (1992), pp. 404–425 (cit. on p. 5).
- [17] Daniel T Gillespie. “Exact stochastic simulation of coupled chemical reactions”. *The journal of physical chemistry* 81.25 (1977), pp. 2340–2361 (cit. on p. 6).
- [18] Christian Arnold, Peter F Stadler, and Sonja J Prohaska. “Chromatin computation: Epigenetic inheritance as a pattern reconstruction problem”. *Journal of theoretical biology* 336 (2013), pp. 61–74 (cit. on p. 6).
- [19] Peter F. Stadler. “Fitness landscapes”. In: *Biological Evolution and Statistical Physics*. Ed. by Michael Lässig and Angelo Valleriani. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 183–204. URL: https://doi.org/10.1007/3-540-45692-9_10 (cit. on p. 7).
- [20] Conrad Hal Waddington. *The strategy of the genes*. Routledge, 2014 (cit. on p. 7).
- [21] *Epigenetic Landscape*. In: *Embryo Project Encyclopedia*. Navis, Adam R., 2007. URL: <http://embryo.asu.edu/handle/10776/1687> (visited on 03/16/2021) (cit. on p. 7).
- [22] Blender Online Community. *Blender — a 3D modelling and rendering package*. Blender Foundation. Stichting Blender Foundation, Amsterdam, 2018. URL: <http://www.blender.org> (cit. on p. 8).
- [23] Ian B Dodd and Kim Sneppen. “Barriers and silencers: a theoretical toolkit for control and containment of nucleosome-based epigenetic states”. *Journal of molecular biology* 414.4 (2011), pp. 624–637 (cit. on pp. 9, 10).
- [24] Kim Sneppen and Leonie Ringrose. “Theoretical analysis of Polycomb-Trithorax systems predicts that poised chromatin is bistable and not bivalent”. *Nature communications* 10.1 (2019), pp. 1–18 (cit. on p. 9).
- [25] Kim Sneppen. *Models of life*. Cambridge University Press, 2014 (cit. on p. 9).
- [26] Ian B Dodd et al. “Theoretical analysis of epigenetic cell memory by nucleosome modification”. *Cell* 129.4 (2007), pp. 813–822 (cit. on p. 10).
- [27] Mohammad Sedighi and Anirvan M Sengupta. “Epigenetic chromatin silencing: bistability and front propagation”. *Physical biology* 4.4 (2007), p. 246 (cit. on p. 10).

- [28] Mille A Micheelsen et al. “Theory for the stability and regulation of epigenetic landscapes”. *Physical biology* 7.2 (2010), p. 026010 (cit. on p. 10).
- [29] Diana David-Rus et al. “Inheritance of epigenetic chromatin silencing”. *Journal of theoretical biology* 258.1 (2009), pp. 112–120 (cit. on p. 10).
- [30] *Cooperativity*. In: Encyclopaedia Britannica, 2019. URL: <https://www.britannica.com/science/cooperativity> (visited on 01/12/2021) (cit. on p. 10).
- [31] Michel Krecké. *EpiDynaST simulation files for Masters Thesis*. https://github.com/mk36fyvy/MastersThesis_Calculations. 2021 (cit. on p. 18).
- [32] Biola M Javierre et al. “Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters”. *Cell* 167.5 (2016), pp. 1369–1384 (cit. on p. 41).

Check Final Print Size

— Check final print size! —



width = 100mm
height = 50mm

— Remove this page after printing! —