

Bistability and state switching in computational dynamic histone modification models

Michel Krecké



UNIVERSITÄT
LEIPZIG

MASTERARBEIT

eingereicht am
Master of Science

Bioinformatik

in Leipzig

im Februar 2021

Primary advisor:
Prof. Dr. Sonja Prohaska

Second advisor:
Prof. Dr. Peter Stadler

© 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, February 28, 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	3
1.2.1 (Solving the) chemical master equation	3
1.2.2 Gillespie’s algorithm	4
1.2.3 EpiDynast	4
1.3 Epigenetic fitness landscapes	5
1.3.1 From landscape to vector field	5
1.3.2 Enzyme types and bistability	6
1.4 Transition	7
2 Methods	8
2.1 Model	8
2.2 <i>EpiDynast</i> parameters	8
2.3 Enzyme models	8
2.3.1 linear enzymes	8
2.3.2 cooperative enzymes	8
2.3.3 random enzymes	8
2.3.4 completer enzymes	8
2.4 Enzyme rule sets	8
3 Results	9
3.1 Non-cooperative enzymes do not entail bistable systems	9
3.2 Bistability on a non-cyclic nucleosome string	9
3.3 Bistability on a cyclic nucleosome string	9
3.3.1 Achieving bistability/Probability of bivalency	9
3.3.2 Influence of dissociation rate on system’s noise	10
3.3.3 The boundaries of bistability	10
3.4 Bivalency	10

Contents	vi
4 Discussion	11
4.1 Conditions for bistability	11
4.2 Proposed mechanism for bistable switching	11
4.2.1 cyclic	11
4.2.2 non-cyclic	11
5 Outlook	12
A Run files and specifications	13
B Additional runs	14
References	15
Literature	15

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 plays 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].

Chromatin can show higher order structure than the simple “beads on a string” variant.

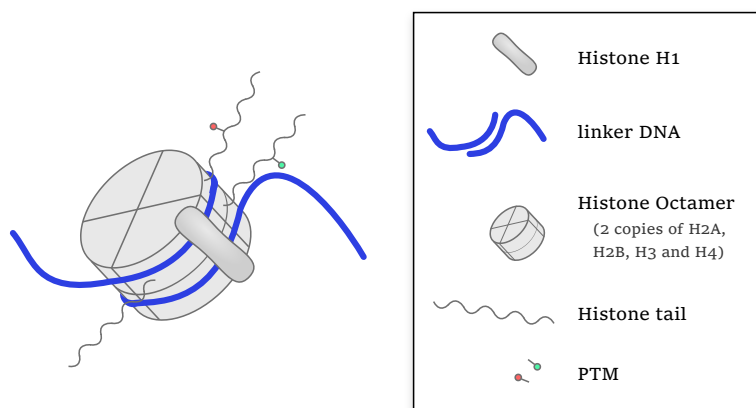


Figure 1.1: schematic model of a nucleosome

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 specific positions on the histone tail. The presence of these chemical groups changes the charge or polarity of the modified amino acids and thus influence the histone's affinity to the DNA. Histone-acetyltransferases (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

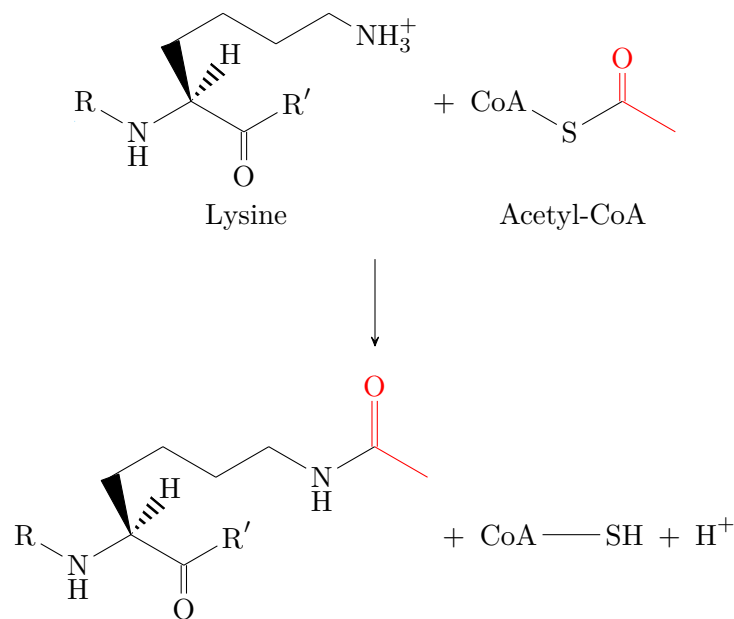


Figure 1.2: Acetylation of lysine. This reaction is catalysed by the HAT enzyme which by means of the cofactor acetyl coenzyme A (Acetyl-CoA) is able to trigger the transfer (substitution reaction in chemical terms) of an acetyl group onto the nitrogen atom in the side chain of lysine. The latter can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

tails, e.g. methylation (one-, two and three-fold), phosphorylation, ubiquitylation, etc. These groups can entail gene transcription activation, silencing (repressing), or completely different purposes.

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. 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 [13].

1.2 Dynamic histone PTM models

1.2.1 (Solving the) chemical master equation

The chemical master equation (CME) is the differential equation underlying the system that describes the time-dependent evolution of that system 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 Langevin stochastic differential equation [10] for each HME type, as was already done by Mayer in [11]. Eq. 1.1 describe the noncooperative (see) 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 can be neglected. Given that the context from the enzyme rule sets is an

important aspect of EpiDynast, an analytical solution of the CME would not be the best approximation. Also, depending on discrete numbers such as the number of nucleosomes in the string, the analytical solution of the CME as a continuous system, makes it even more unfitting as an approximation.

A more fitting system 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. So, if the rule set contains at least one rule including the next neighbours of the modified nucleosome i , the CME system would change to eqn. 1.2

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

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

1.2.2 Gillespie's algorithm

Gillespie's algorithm simulates the time evolution of a spatially homogenous molecular mixture under specification of the coupled reaction channels (i.e. association and enzymatic reaction with dissociation) based on stochastic chemical kinetics [7, 8]. This is useful especially when solving the chemical master equation analytically is not ideal.

- next-neighbours
- rule-based, (paper from Arnold, Prohaska is basis for Nico's work, also basis for my work)
- Gillespie's
 - explain event based
 - steps from his paper
- association and concentration are taken together.
- enzyme types

1.2.3 EpiDynast

The software used in this thesis is *EpiDynast*, developed by Herbig et al. The main working mechanism may be outlined as follows: After defining enzyme rule sets and a starting nucleosome string (concerning their PTMs), *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 dissociation of the enzyme 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.

1.3 Epigenetic fitness landscapes

1.3.1 From landscape to vector field

The term of epigenetic fitness landscape (referred to as landscape in the rest of this thesis) 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.

In this case, V comprises the entirety of active, silent and unmodified state distributions along the nucleosome string, χ determines the modification of one nucleosome state, e.g. from silent to unmodified, or vice versa and 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 algorithm simulation.

Depending on the landscape, e.g. if there are deep and steep basins, it gets more and

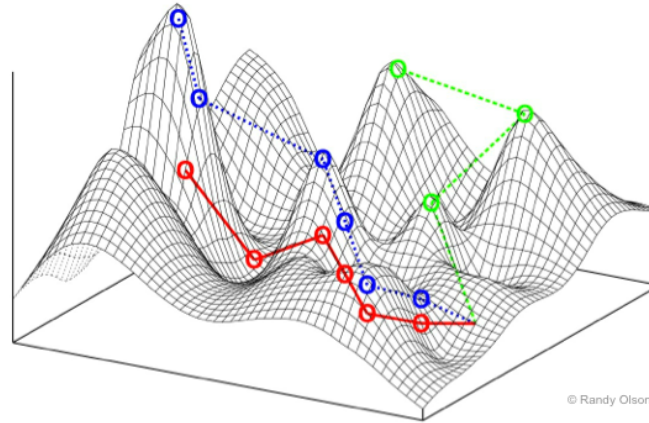


Figure 1.3: Fitness landscape example and 2D projection.

more difficult for the stochastic algorithm to exit the basin and adopt a configuration that lies outside. Each histone modifying enzyme thus induces a stability trend among the configurations which results in a vector field specific to each type of enzyme. In more complex systems with a multitude of different enzymes, these vector fields are combined and create the resulting landscape at hand containing one or more fix points.

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 it being much harder for the system to maintain a certain state or move along the shape of the landscape to switch from one state to another.

1.3.2 Enzyme types and bistability

The number of fix points in the landscape can be analytically derived from the enzymes and their dependence on the frequency of nucleosome states (e.g. acetylation, methylation, ...) over time (see eqns. 1.1).

Mayer solved the equations analytically and found 4 critical values [11]: 3 fix points at $(0,0)$, $(0, 1 - \frac{\beta_1}{\beta_2})$, $(1 - \frac{\alpha_1}{\alpha_2}, 0)$ and a separatrix that depends on the ratio between $\frac{\alpha_1}{\alpha_2}$ and $\frac{\beta_1}{\beta_2}$. The separatrix always has a gradient towards one of the non-trivial fix points. If $\frac{\alpha_1}{\alpha_2} = \frac{\beta_1}{\beta_2}$, the separatrix has no gradient.

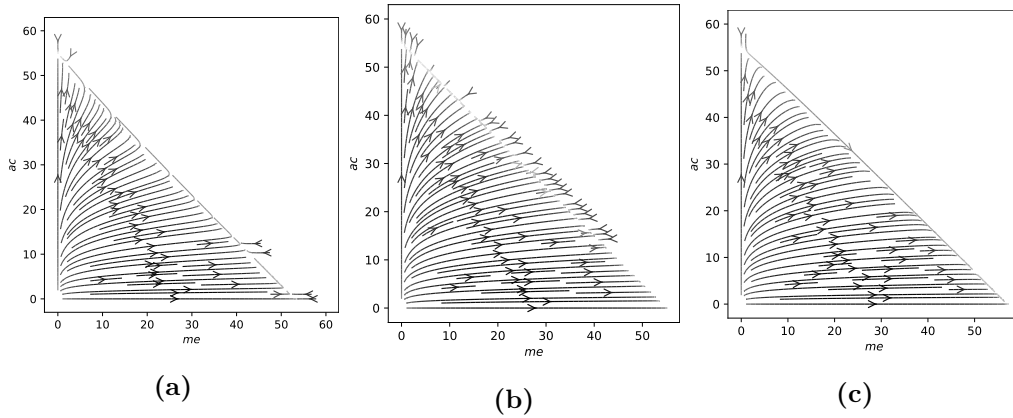


Figure 1.4: intervals.

- Can it be shown that the separatrix is only zero gradient if there is a null vector field?

In order for a dynamic histone PTM system to be bistable, the enzymes have to show cooperativity [5, 11, 16]. According to Sneppen [15, p.48], “cooperative binding means that the probability of occupying a state increases more than linearly with the concentrations of the binding molecules”.

In [5], Dodd et al. specify the nature of cooperativity in order to reach ultrasensitivity¹ 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 [6, 12, 14], or indirect, where each modified nucleosome catalyzes one of two separate modification reactions to fully convert a third nucleosome [4, 6]. A critical requirement for ultrasensitivity is that modified nucleosomes must act nonlocally, stimulating modification of nucleosomes located some distance away on the DNA. This

¹From Dodd et al. [5]: “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

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 [6, 12].”

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 [3]. 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. Even though this not ideally put from a biochemical point of view, the mathematical implications still remain valid.

- rephrase the biochemical side of cooperativity?

Mayer in [11] expressed the time dependent concentration of the modifications in the system in function of cooperative enzymes in the following differential equation system:

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

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

- Include Matilda’s vector graphs
 - Mark the fix points and separatrix

1.4 Transition

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 thesis will show exactly that.

2. Methods

2.1 Model

- How many nucleosomes? (Remember: This changed from 60 to 40 (at bistable switching))
- Half nucleosomes?
- fixed string, fixed neighbour relations
- limitations
- bistability impact on rule set

2.2 *EpiDynast* parameters

- How did I change it?
- Time setting etc
- refer to exemplary state, rule and param files in appendix

2.3 Enzyme models

2.3.1 linear enzymes

2.3.2 cooperative enzymes

2.3.3 random enzymes

2.3.4 completer enzymes

- Explain the sets later used? and enumerate them or sth so that the reader can easily refer to them. Explain them before.
- Types and vocabulary and spacing
- explain Kx Ky model

2.4 Enzyme rule sets

3. Results

3.1 Non-cooperative enzymes do not entail bistable systems

Considering a system without cooperative enzymes, that thus only contains enzymes with a context that includes the next neighbours at most, bistability can not be achieved. Indeed, fig. 3.1 discloses a monostable system, as the histogram shows an unimodal distribution throughout the simulation.

Fig. 3.1 exemplarily shows that the total number of active and silent states ambulate

Figure 3.1: Boring monostable case. The rule sets used the following enzymes: ? (Rates and additional runs can be found in Appendix.)

around one same value. In fig. , one can see that even though the overall number of active and silent states changes over time, the overall order in the system stays invariant as the only enzymatic actions are taking place in between the two modification areas at their border.

- Include a boring graph that does not change much. Emphasis on the histogram.
- Renew the graphs
 - Gillespie time instead of step number on x-axis
 - remove bivalency window where it isn't needed
 - print (not plot) the total step number and indicate it in the caption
 - remove title
- run the heat map plots on these in order to see if there are to areas or if they are homogeneously mixed

3.2 Bistability on a non-cyclic nucleosome string

- Explain bistability and low frequency switching

3.3 Bistability on a cyclic nucleosome string

3.3.1 Achieving bistability/Probability of bivalency

- case without switchings
- case with frequent switchings

3.3.2 Influence of dissociation rate on system's noise

- high vs. low dissociation rate

3.3.3 The boundaries of bistability

- reduce space (=reach) of the cooperative enzymes
- run without unmodified state

3.4 Bivalency

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

4. Discussion

4.1 Conditions for bistability

- Bistability (condition: cooperativity) (Explain difference between biological and Sneppen coop.) (Errors in Sneppen's insights on the topic)

4.2 Proposed mechanism for bistable switching

4.2.1 cyclic

4.2.2 non-cyclic

5. Outlook

- Schuettengruber et al. in [13] (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. [9] 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.

A. Run files and specifications

B. Additional runs

References

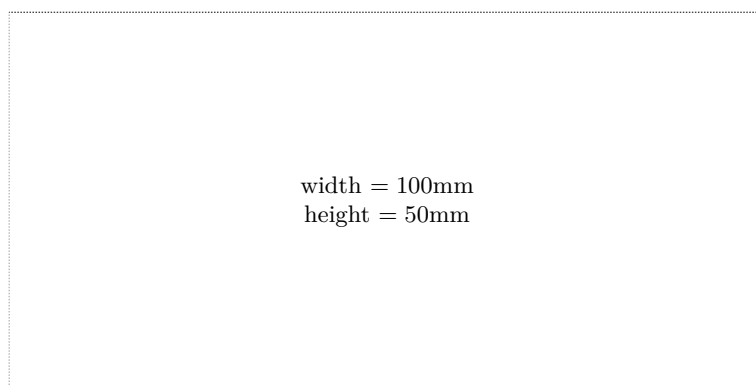
Literature

- [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. 1).
- [3] *Cooperativity*. In: Encyclopaedia Britannica, 2019. URL: <https://www.britannica.com/science/cooperativity> (visited on 01/12/2021) (cit. on p. 7).
- [4] Diana David-Rus et al. “Inheritance of epigenetic chromatin silencing”. *Journal of theoretical biology* 258.1 (2009), pp. 112–120 (cit. on p. 6).
- [5] 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 p. 6).
- [6] Ian B Dodd et al. “Theoretical analysis of epigenetic cell memory by nucleosome modification”. *Cell* 129.4 (2007), pp. 813–822 (cit. on pp. 6, 7).
- [7] 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. 4).
- [8] 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. 4).
- [9] 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. 12).
- [10] 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. 3).
- [11] Daniel E Mayer. “Histone Driven Cell Differentiation”. MA thesis. Institute of Computer Science, University of Leipzig, 2020 (cit. on pp. 3, 6, 7).
- [12] Mille A Micheelsen et al. “Theory for the stability and regulation of epigenetic landscapes”. *Physical biology* 7.2 (2010), p. 026010 (cit. on pp. 6, 7).

- [13] Bernd Schuettengruber et al. “Genome regulation by polycomb and trithorax: 70 years and counting”. *Cell* 171.1 (2017), pp. 34–57 (cit. on pp. 3, 12).
- [14] Mohammad Sedighi and Anirvan M Sengupta. “Epigenetic chromatin silencing: bistability and front propagation”. *Physical biology* 4.4 (2007), p. 246 (cit. on p. 6).
- [15] Kim Sneppen. *Models of life*. Cambridge University Press, 2014 (cit. on p. 6).
- [16] 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. 6).

Check Final Print Size

— Check final print size! —



— Remove this page after printing! —