Statistical Mechanics on the Regulation of Gene Expression

Víctor Martín
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.**

Abstract: In recent years it has been shown that genetic information can be inherited through the transcriptional state of the DNA, which can be either active or silenced. This is controlled by chemical reactions of acetylation, which allows transcription, and methylation, which blocks it. We have reproduced a computational model for this description and found that, in the thermodynamic limit, a phase transition between the active and the silenced state should take place. Due to the lack of physical frameworks to this problem, in this article we propose (up to our knowledge) the first statistical mechanics description based on an Ising model. We identify which is the critical temperature of our system at which the phase transition occurs and discuss it in relation to experimental data.

I. INTRODUCTION

As may be known by the reader, macromolecules of *deoxyribonucleic acid* (DNA) encode genetic instructions used in development and functioning of cells [1]. Most of these molecules are double-stranded helices made of simpler units called nucleotides. Each nucleotide is composed of a nucleobase (guanine, adenine, thymine and cytosine) as well as a backbone made of alternating sugars (deosxyribose) and phosphate groups. Linear chains of DNA form larger units called genes which, in turn, organize themselves to create what we know as the cell genome. Cells translate the information stored in their genome and are able to pass that information to their descendants through replication.

However, many inheritable changes in gene function are not explained by changes in the DNA sequence but by which parts of that sequence are expressed (transcribed) and which are not. The information relying on the changes of cell gene expression is what we know as *epigenetic cell memory*. Let's allow ourselves to simplify the situation, and to imagine two alternative regulatory gene states, that are stable and inherited through cell division. One of them, the "silenced state", denotes that the gene is not transcribed. On the contrary, the "active state" will determine the transcription of the said gene. This leads to cells with identical genome maintaining completely different functional identities, regardless the surrounding conditions. Epigenetic cell memory is specially important among multi-cellular organisms, due to cells with distinct functional identities having the same genome.

In order to forbid or allow the expression of a DNA region, nucleosomes [1] are chemically modified by the action of an enzyme or the binding to a repressor molecule, which provoke them to be ignored when the DNA transcription takes place.

In prokaryotes, genes are commonly switched on and off by the interactions of regulatory proteins with specific DNA sequences [2], [3]. Let us now take a closer look to eukaryotic systems, and review a model based in chemical modification of nucleosomes by histone enzymes, where a nucleosome can adopt three different states [4], [5]:

- **Unmodified (U):** the nucleosome itself, without any chemical modification.
- **Acetylated (A)**: state that occurs when an acetyl terminal is binded to the lysine[2] residues of a histone. This chemical reaction is known as acetylation and is catalyzed by the enzyme *histone acetyltransferase* (HAT). The opposite process, deacetylation, removes the acetyl terminals and is catalyzed by the enzyme *histone deacetylase* (HDAC).
- **Methylated (M)**: state of the nucleosome when subjected to methylation. Methylation is similar to acetylation, but binding methyl terminals to the histones instead of acetyl ones. This process and its opposite (demethylation) are catalyzed by the enzymes *histone methyltransferase* (HMT) and *histone demethylasa* (HDM) respectively.

We should understand an acetylated nucleosome as a nucleosome where DNA chains wrapping the histone are less compact and, thereby, are more likely to bind to a regulatory protein of DNA transcription [6]. When almost every nucleosome of a region is acetylated we will say it presents the global active state, with an increased transcriptional activity. Otherwise, methylated nucleosomes are more compact and will define regions with the silenced global state, that will be ignored by the transcription proteins.

It has been proposed that nucleosomes are actively interconverted by modifying and demodifying enzymes that are recruited by the already modified nucleosomes [7]. By *feedback conversion* we mean a positive recruiting process, where M nucleosomes recruit HDMs and HAT, while A nucleosomes recruit HDACs and HMTs; once recruited, enzymes act on any other nucleosome within the region[3] and modify/demodify it. As the reader may suppose, these feedback conversions occur altogether with noisy reactions that can take place independently of the local nucleosomes, and are due to non-recruited or external enzymes.

*Electronic address: vmv9191@gmail.com

In the next section, we will reproduce a stochastic simulation based on this feedback conversion that was introduced in Ref. [7]. As it is a strictly computational model, we have proposed a distinct point of view, which tries to describe the process in a more physical way. Based on energy arguments explained in the next section, we have represented the three kinds of nucleosome (A, U and M) as the states (1, 0,- 1) of a three-state one-dimensional Ising Model with infinite-range interaction. We have both theoretically and computationally solved this Ising model, finding a second order transition and the critical temperature of the system.

II. RESULTS

A. Computational model

We will now reproduce the model described above and introduced in [7], considering a DNA region of $N = 60$ nucleosomes, with the three kinds: Unmodified (U), Methylated (M) and Acetylated (A). Positive feedback transitions take place with probability α and non-recruited random transition with probability $1 - \alpha$. For completeness we indicate the program procedure as it is proposed in [7]:

- **Stage 1**: a random nucleosome n_1 is chosen among the 60 nucleosomes. With probability α a feedback transition will occur (go to stage 2A) or, with probability $1 - \alpha$ a noisy transition will take place (go to stage 2B).
- **Stage 2A**: Another nucleosome n_2 is selected among the 59 remaining, converting n_1 one step towards n_2. For example if $n_2 = A$ and $n_1 = U$, n_1 changes to the state A. When either recruiting nucleosome n_2 is U or both n_1 and n_2 are in the same state, n_1 remains unchanged.
- **Stage 2B**: The nucleosome n_1 is converted one step towards any of the other three states (with probability one third each). Nevertheless, direct $A \rightarrow M$ or $A \leftarrow M$ transitions are not allowed.

These steps are repeated successively. It is useful to define the *feedback-to-noise ratio* R as $R = \frac{\alpha}{1-\alpha}$. *Figure 1* shows the temporary evolution of the number of M nucleosomes and the probability distribution on long-time simulations. Two important conclusions can be extracted from the simulations performed [7]:

- For low values of the feedback-to-noise ratio, we perceive an almost constant one-third probability for each nucleosome type.
- For large values of feedback-to-noise ratio, we observe preference for the A or M nucleosome types.

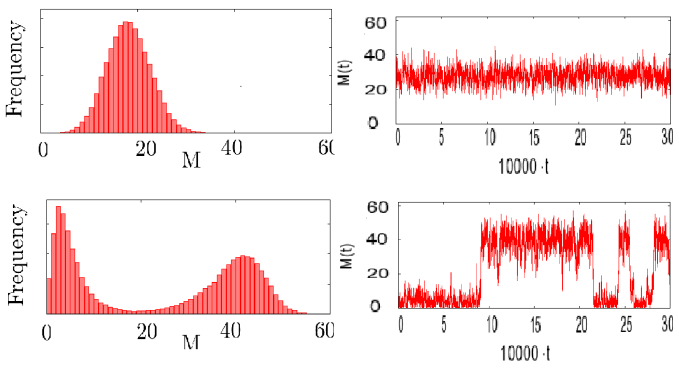

Figure 1: *The graphics on the left show the number of M nucleosomes for $R = 0.4$ (top) and $R = 2$ (bottom) as a function of the number of steps performed ($1t = 60$ steps). The graphics on the left show the probability distribution of M nucleosomes for the corresponding right simulation.*

Obtained results exactly reproduce and match conclusions in [7]. It is specially interesting the second conclusion, so it demonstrates that up to a certain α either the methylated or the acetylated nucleosome prevail in front of the unmodified, which means that the active or the silenced state will predominate. The fact that a DNA region is able to present these two states, and to be stable in both of them is what is known as *bistability*.

B. Description in terms of free energy

In the majority of the bibliography there is a lack of a physical description of this DNA expression process. In this section we have proposed one, supposing interaction of nucleosomes in pairs and acknowledging the change of energy related to a pair when a feedback transition takes place. As we have seen before, only for low values of the noisy transition probability the active state emerges, so we will make the assumption that feedback transition are due to the minimization of the energy and noisy transitions are due to the entropic factor of temperature.

Let us now consider a region of DNA with a number N of nucleosomes that can present either the acetylated (A), methylated (M) or unmodified (U) state. As we suppose pair interaction and we have three states, there will be 9 different combinations of the kind $\{n_2, n_1\}$, being n_1 and n_2 two any nucleosomes. We have classified the different pairs within three groups with different energies (as shows *figure 2*). Pairs with both nucleosomes modified, but in different states, will be the more energetic. Pairs with both modified nucleosomes and in the same state, will be the less energetic. Ultimately, pairs with at least one unmodified nucleosome will represent an intermediate energy level.

Based on this argument and if we take that these three energy states are equally separated, we can associate an energy of interaction J such that the pairs <AM, MA>

E_1	E_2	E_3
$< AM >$	$< UM >$	$< MM >$
$< MA >$	$< UA >$	$< AA >$
	$< UM >$	
	$< AU >$	
	$< UU >$	
	$< MU >$	

Figure 2: *All possible pair combinations of the three kinds of nucleosomes classified by the energy associated to each pair, where* $E_1 > E_2 > E_3$.

have energy $E_1 = J$, the pairs in the third column <AA, MM> have an energy $E_3 = -J$ and the remaining pairs have an energy $E_2 = 0$. It is readily seen that this description fits the following formula for the energy of the pairs:

$$E = -J \cdot S_i S_j \tag{1}$$

where S_i and S_j represent two nucleosomes with values $-1, 0, 1$ for the states M, U, A each. The expression on *Equation (1)* may remind the reader to a magnetic-like energy, which will be the matter of study from now on. Summarizing, we have a system (DNA region) of N particles (nucleosomes), being able to present three different states and experimenting an infinite-range pairwise interaction. Furthermore, within our model, the energies of the system fit a magnetic-like expression. Putting all this together, we rapidly see that a statistical mechanics description will be helpful to analyze the behaviour of the system. We have chosen an Ising model, specifically a one-dimensional [4], three-states and infinite-range Ising model. The hamiltonian of this model will be the sum of the energies from the interactions between all the pairs of nucleosomes, as follows:

$$\mathcal{H} = -J \sum_{<i,j>} S_i \cdot S_j \tag{2}$$

$$S_i = -1, 0, 1$$

where we have made the parallelism of magnetic spins with nucleosome chemical state. In the following sections this model will be solved both theoretically and computationally.

C. Theoretical analysis

We solved the infinite-range three-states Ising model through the mean field approximation. The procedure we followed to solve the hamiltonian in *equation (2)* relies on the mean field approach that the effect of the fluctuations is not significant, which means that the quantity $S_i - < S_i >$ is very small. Henceforth, the square of the fluctuations $(S_i - < S_i >)^2$ is minuscule and can be neglected [8], [9].

Starting from this assumption we define the order parameter as the mean value of S_i, $m =< S_i >$, and rewrite the previous hamiltonian by inserting it:

$$\mathcal{H} = -J \sum_{<i,j>} (S_i - m + m)(S_j - m + m) = \tag{3}$$

$$= -J \sum_{<i,j>} \left[m(S_i + S_j) - m^2 \right]$$

where $(S_i - m)(S_j - m)$ has been neglected. Knowing there is an infinite-range interaction, the number of all possible pair combinations is $\frac{N(N-1)}{2}$ and that

$$\sum_{<i,j>} S_i = \sum_{<i,j>} S_j = \frac{N-1}{2} \sum_{i=1}^{N} S_i \tag{4}$$

we can rewrite the hamiltonian as

$$\mathcal{H} = -J \left[(N-1) \sum_{i=1}^{N} S_i - \frac{N-1}{2} m^2 \right] \tag{5}$$

and then the partition function:

$$Z_N = e^{-\beta \mathcal{H}} = \tag{6}$$

$$= \exp\left(-\beta J \frac{N(N-1)}{2} m^2 \right) \exp\left(\beta J (N-1) \sum_{i=1}^{N} S_i \right)$$

which, summing over the three states of S_i gives

$$Z_N = e^{-\beta J \frac{N(N-1)}{2} m^2} \left\{ 1 + 2\cosh\left[\beta J (N-1) m \right] \right\}^N \tag{7}$$

It is necessary to write down the Gibbs free energy $G = -K_B T \ln Z_N$ and minimize it respect m, by applying $\left(\frac{dG}{dm}\right)_{m^*} = 0$. Having defined the model as such, the Gibbs free energy depends on N^2, and then the free energy per particle will not be extensive [5]. Once minimized the Gibbs free energy, we find the equation of state:

$$m = \frac{2\sinh[\frac{m}{T^*}]}{1 + 2\cosh[\frac{m}{T^*}]} \equiv f(m, T^*) \tag{8}$$

already in reduced coordinates, where $T^* = \frac{K_B T}{J(N-1)}$ is the reduced temperature.

The equation above is transcendent, yet their temperature limits can be studied easily. For temperatures near to zero ($T \to 0$), we develop in Taylor series the equation and we see the order parameter goes to 1 ($m \to \pm 1$),

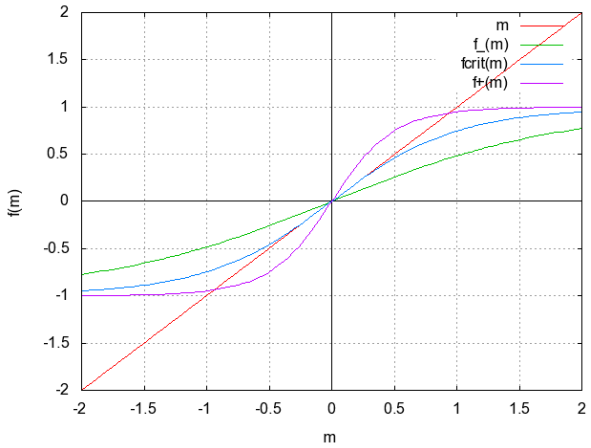

Figure 3: *Representation as a function of m of both sides of equation (9). $f_-(m)$ for $T^* > T_c^*$, $f_+(m)$ for $T^* < T_c^*$ and $f_{crit}(m)$ for $T^* = T_c^*$*

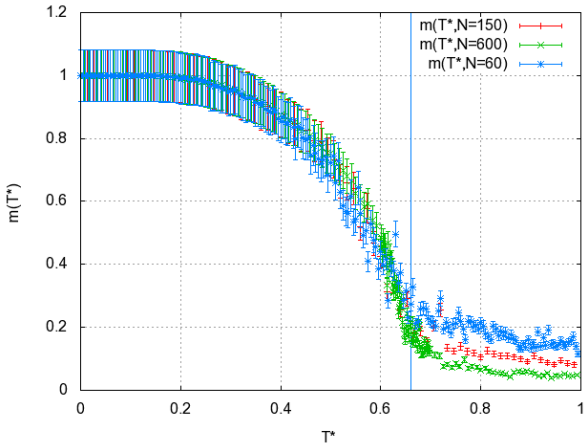

Figure 4: *Simulated data for the absolute value of the magnetization as a function of the reduced temperature for values of the number of nucleosomes $N = 60, 150, 600$. The magnetization and the error have been calculated as $\sum \frac{1}{M} \sum_{k=1}^{M} |m_k|$ and $\sqrt{\frac{Variance}{M}}$ respectively and simulated during 500 Monte-Carlo steps and averaged over the last $M = 300$ steps; and finally averaged over 5 distinct simulations. Blue vertical line at $T = T_c^* = \frac{2}{3}$.*

which are the saturation magnetizations of the system, when all spins are in the state 1 or -1 . Rethinking it in terms of the DNA, it means that for temperatures tending to zero, the global state could be either A or M, so the system presents the bistability mentioned before.

On the other hand, for temperatures going to infinite ($T \to \infty$), the sinh term tends to zero, and the magnetization of the system becomes also zero ($m \to 0$). This means equiprobability for every state and ,thus, equiprobability for every kind of nucleosome; due to the high degree of disorder at elevated temperatures.

Furthermore, the behavior described for the system on both temperature limits implies that there is a phase transition. In order to characterize the critical point at which the phase transition occurs, we will calculate the critical temperature T_c^* of the system. To do it, both sides of *equation (9)* should be represented in a graphic, each as a function of the order parameter m, as shows *figure 3*.

For temperatures below the critical temperature, there is only one solution for the magnetization: $m = 0$. For temperatures above the critical point, there are three solutions: $m = 0$, $m < 0$ and $m > 0$. Henceforth, in exactly the critical point, infinite solutions must arise, which graphically is traduced to an equal slope of m and $f(m)$ at the point $m = 0$. Up to this point, we can apply the condition for the slopes at $m = 0$ into *equation (9)* and find the exact critical temperature by:

$$\left(\frac{df(m, T_c^*)}{dm}\right)_{m=0} = \pm 1. \qquad (9)$$

Through this procedure, the value obtained for the critical temperature is $T_c^* = \frac{2}{3}$.

Computational analysis

To test the validity of the theoretical solution obtained in last section and to find the mean values of every nucleosome state, we have developed a stochastic simulation based on a *Metropolis* [6] algorithm. The aim of the program is to solve the initial hamiltonian (*equation (2)*) and here is the basic course of action of the algorithm:

- We define a linear chain of N spins (nucleosomes) randomly filled with values $\pm 1, 0$.
- We propose the change of a randomly chosen spin and we evaluate the energy change ΔE it does provokes [7]. If $\Delta E < 0$, we accept the change and return to the first step. If the change increases the energy of the system, $\Delta E > 0$, we jump to the next step.
- We calculate the *Boltzmann factor* [8] associated to the previous ΔE and sort a random number. If the Boltzmann factor is greater than the random number we reject the spin change, otherwise, we accept it. Either way we return to the first step.

In *figure 4* we present the results obtained for the magnetization after long simulations, for different values of the number of nucleosomes N.

As it was expected, the magnetization presents the same behavior than the described theoretically, tending to ± 1 at low temperatures and to zero at high temperatures. Beside, it also experiments a second order phase transition, at a temperature near the $T_c^* = \frac{2}{3}$ calculated before. The only relevant effect caused while increasing

the number of nucleosome is the improvement of the magnetization curve, which will be more likely to represent the thermodynamic limit behavior ($N \to \infty$).

Referring to the occupational number for every nucleosome state, the simulation gives an equiprobability for each one at high temperatures. At low temperatures, as magnetization tends to ± 1, it occurs a decrease of the unmodified nucleosomes.

III. CONCLUSIONS

As we have explained, in eukaryotic cells, bistability is needed for the correct operation of the epigenetic memory. We have found that, in order to present bistability, recruited transitions must be much more probable than noisy ones, so that the acetylated or the methylated nucleosome prevails among the unmodified (*figure 1*).

Through the Ising model description used we have realized that a second order phase transition takes place at the reduced temperature $T_c^* = \frac{2}{3}$. From the expression of the reduced temperature we can isolate the pairwise interaction constant:

$$J_c = \frac{K_B T}{T_c^*(N-1)}. \tag{10}$$

Knowing that the DNA temperature T for the majority of the eukaryotic organisms is around 310 K, the number of nucleosomes N per chromosome is of the order of 10-100, substituting the calculated value for the T_c^* and the known value for the Boltzmann constant K_B, we obtain that $J_c \sim 10^{-21}$ Joules. We can, thereby, suppose that for $J > J_c$ a DNA region will present bistability.

Although the interaction energy between histone enzymes and nucleosomes has not been experimentally calculated, it does have been found some values for the interaction between different kind of nucleosomes [10]. These values are between $0.1 - 10$ Kcal/mol, or, in Joules per nucleosome, $10^{-22} - 10^{-20}$ J. Nevertheless, despite the values seem quite similar, no value has been calculated yet for this specific interaction and no strict comparison can be made.

We are aware of the limitations of the models proposed through this text (such as the fact of having a finite system far from the thermodynamic limit) but we think that it may be useful to focus more investigations into this direction. Going beyond these approximations and the scope of the article, the model proposed in section II.B can be improved by defining a pairwise interaction constant J^* that depends on the distance. It would be valid because it seems logical that nucleosomes being far from each other are less probable to be recruited. Another point could be to redefine the Ising model into a two-dimensional model in order to make it more similar to reality.

Acknowledgments

I would specially like to thank my adviser Marta Ibañes for all the moments dedicated to this project, whether they have been in our meetings or her alone to revise it. I would also like to thank my family and my friends for the moments of company provided while working on this article. To all of them, thank you.

Notes

[1] A nucleosome is a structural unit formed by a histone and the DNA chain wrapping it

[2] The lysine is an essential amino-acid which is an important component of histones.

[3] There is no need that the recruiting nucleosome is a neighbor of the converted one.

[4] We chose one-dimensional because our system is a linear DNA chain

[5] This problem can be solved just by defining a new interaction constant $J' \equiv \frac{J}{N}$

[6] It uses a Monte-Carlo method working with Markov chains.

[7] All possible spin changes $S_{old \to S_{new}}$ are allowed, and the change of energy related is $\Delta E = 2(S_{new} - S_{old}) \sum_{i=1}^{N} S_i$.

[8] The Boltzmann factor is $e^{-\frac{\Delta E}{K_B T}}$

[1] R. Philips, J. Kondev, J.Theriot, *Physical Biology of the Cell* (2008)

[2] Gary K. Ackers, Alexander D. Johnson and Madeline A. Shea. "Quantitative model for gene regulation by λ phage repressor". *Proc. Natl Acad. Sci. USA* (1982).

[3] Schwarz, G. *Biophys. Chem.* 6, 65-76. (1977).

[4] Struhl, K. "Histone acetylation and transcriptional regulatory mechanisms", *Genes Dev.*, 12: 599-606 (1998).

[5] Bird A. "Perceptions of epigenetics", *Nature*, 447: 396-402 (2007).

[6] Hong, L., G.P. Schroth, H.R. Matthews, P. Yau, and E.M. Brad-bury. 1993. *Studies of the DNA binding properties of histone H4 amino terminus.* J. Biol. Chem **268**: 305314 (1993).

[7] Ian B. Dodd et al. "Theoretical Analysis of Epigenetic Cell Memory by Nucleosome Modification" *Cell* 129, 813-822 (2008a).

[8] K. Christensen, N.R. Moloney, *Complexity and Criticality*, London (2005). ;

[9] J.M. Sancho, *Fsica Estadstica: Sistemas en Interaccin*, Barcelona (2011)

[10] H.G. Garcia et al. "Energy of DNA in curved loop of nucleosome" *Biopolymers* 585(2): 115-30 (2009).