The influence of SP6 RNA polymerase on nucleosome positioning near the transcription start site in the downstream direction

Zhongling Jiang
Department of Chemistry,
Massachusetts Institute of Technology,
Cambridge, MA 02139
Email: jiang97@mit.edu
(Dated: May 14, 2018)

A nucleosome is composed by eight histone proteins wrapped around about 1.65 times[1] with DNA sequence. Nucleosome positioning on DNA sequence is critical for any biological process related to DNA, which is determined by DNA sequence, nucleosome remodeling enzymes and so on. In this article, we focus on the influence of SP6 RNA polymerase, a nucleosome remodeler, on nucleosome positioning with our Gillespie model[2]. It turns out that this RNA polymerase functions as an active motor pushing nucleosomes moving faster in one direction and having some probability causing nucleosome evictions, which leads to nucleosome positioning oscillation pattern near the transcription start site (TSS) in the downstream direction[3].

I. INTRODUCTION

A nucleosome consists of a 147 base pairs (bp) long DNA segment wrapping around a histone octamer core, which includes two H2A-H2B dimers and one H3-H4 tetramer[1]. Nuclesome density in open reading frames (ORFs) is around 170 bp occupied by one nucleosome[4], i.e., around 85%, which is higher than that in intergenic regions[5].

Nucleosome positioning on genomic DNA is very important for gene regulation. For example, nucleosomes are believed to prevent transcription factors from binding to the promoter region; thus, the promoter region has much less nucleosome density than other regions on DNA sequence, and the promoter regions have been thought to be "nucleosome-free regions"(NFRs). Besides, nucleosome positioning also plays an important role in gene expression, chromatin packaging and so on[6]. The behavior of nucleosome positioning in equilibrium have been studied quite thoroughly, but the influence of nucleosome remodelers on nucleosome positioning still have many unsolved problems. One reason is that there are so many different types of remodelers, and here we focus on the non-equilibrium state with SP6 RNA polymerase remodeling nucleosome positioning.

The reason for RNA polymerase changing nucleosome positioning is that RNA polymerase only works on DNA sequence, but nucleosomes have DNA sequence wrapping around its histone octamer. As a result, if a RNA polymerase wants to pass a nucleosome successfully, it has to partially or totally uncoil the DNA sequence around the histone octamer, which influences the nuclesome in some way. Through experiments, we have found that different RNA polymerases have different effects on the nucleosome they encounter. For example, the way that RNA polymerase II goes through a nucleosome is 'kicking off' the whole octamer or just release one H2A-H2B dimer[7], whereas SP6 RNA polymerase translocates the nucleosome to the upstream of the DNA sequence without octamer leaving the DNA sequence or just loses the octamer during transcription[8].

In this article, we build a model based on the mechanism of SP6 RNA polymerase remodeling nucleosome positions during transcription to see its influence on nucleosome positioning, especially near TSS in the downstream direction.

II. MODEL

A. mechanism of SP6 RNA polymerase remodeling nucleosome positioning

During transcription, SP6 polymerase starts transcription from TSS. When it encounters a nucleosome, it has 90% probability to transfer it to the upstream DNA sequence through forming a DNA loop and 10% probability to lose it[8].

During transcription, the octamer stays on DNA sequence even when being translocated, so it probably does not overcome the neighboring nucleosome on its upstream DNA sequence. What's more, the experiment results show that, on a DNA template with only one nucleosome on one side and the SP6 polymerase on the other side, the longer DNA template leads to longer translocation distance of the nucleosome[3]. Combining the two features of SP6 polymerase translocating nucleosomes, we model this translocation as that SP6 polymerase actively transfer the nucleosome to the middle between its original site and its upstream neighbor's site (as shown in Fig.1).

The nucleosome eviction is simply removing the nucleosome from the DNA sequence. Here, we interpret the probability of translocation and eviction as the ratio of the rate of these two reactions in our Gillespie model[2], i.e., the rate for translocation is nine times that for eviction.

Apart from the influence of SP6 RNA polymerase, we also need to consider nucleosome adsorption given that

in biological system, transcription happens in the environment containing a certain concentration of octamers, H2A-H2B dimers and so on. Depending on different concentrations of these compositions for nucleosomes, we can have various nucleosome absorption rates. The only thing that needs concerns, regarding to the difference between adsorption and SP6 polymerase transcription, is that adsorption can happen on any possible sites along the DNA sequence, while transcription has a specific direction, thus each time, transcription can only happen on one specific nucleosome.

B. Gillespie model

In our model, we use some basic assumptions as followed.

First, we use the "hard core potential" model for nucleosome, which means that each nucleosome occupies 147 bp on DNA sequence and different nucleosomes cannot have any overlap.

Second, we assume the distance between neighboring SP6 RNA polymerases is long enough that the forward polymerase does not influence the following ones. Therefore, to simplify the model, we assume there is only one SP6 polymerase on the segment of DNA in one round. To model different SP6 RNA polymerases, we apply periodic boundary condition to it, which means the polymerase could go back to the origin and start a new round of transcription. This does not change anything except the speed of polymerase remodeling nucleosome positioning, and the only thing we care about in this model is the final pattern of nucleosome positioning after transcription; so this is a reasonable hypothesis in our model.

Third, we apply fixed boundary condition to the nucleosomes, because one of the two boundaries represents TSS, which probably can not be overcome by nucleosomes through transcription. Also, based on our second assumption that the SP6 RNA polymerase cycles on a finite length of DNA sequence to simulate different polymerases start transcription independently and each transcription must start from TSS, one boundary must be TSS, allowing no nucleosome passing in any direction. As a result, we use fixed boundary condition for nucleosome repositioning.

Finally, we assume there is no thermodynamic sliding or eviction of nucleosome, because the binding energy of nucleosome on DNA sequence is about several tens of k_BT, (e.g., about $42k_BT$ in *Xenopus* extracts)[9], which is so high that thermodynamic sliding or eviction are very rare and very slow compared to nucleosome repositioning by SP6 polymerase.

Our procedure of modeling:

1. Randomly generate a nucleosome positioning pattern on one dimenional DNA sequence with initial nucleosome density 85%, and bind SP6 polymerase to the first nucleosome;

2. Each step, choose one of the following reactions to happen according to the reaction rates:

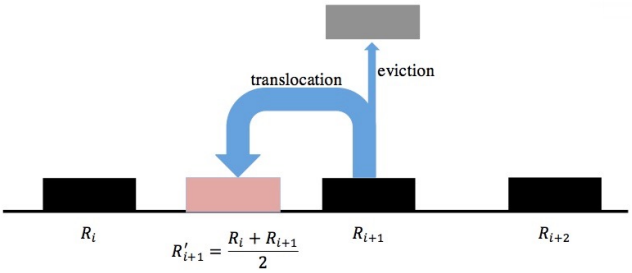

FIG. 1. Process of SP6 RNA polymerase remodeling nucleosome positioning. Black rectangle: nucleosomes on DNA sequence; pink rectangle: nucleosome after translocation of SP6 polymerase; grey rectangle: evicted nucleosome in the process of SP6 RNA polymerase remodeling. Width of arrows represents the probability of the reaction. R_i describes the positions of nucleosomes before remodeling, while $R_i^{'}$ describes after remodeling.

(1) adsorb one nucleosome on DNA sequence and then rearrange the numbering of all nucleosomes based on the new nucleosome order, where reaction rate: r_{on};

(2) remove the nucleosome bound by SP6 polymerase from DNA sequence and then bind the polymerase to the next downstream nucleosome, where reaction rate: r_{off} (Fig.1);

(3) translocate one nucleosome on DNA sequence to the middle of its original site and its neighboring upstream nucleosome site, and then bind the polymerase to the next downstream nucleosome, where reaction rate: r_{trans} (Fig.1);

After each step, calculate the available adsorption sites for the next reaction.

3. Repeat step 2 for enough long time, then calculate the radial distribution function $g(r)$, with r representing distance between a pair of nucleosomes and $g(r)$ showing the system density at radius r compared to the average system density. Radial distribution function $g(r)$ shows the same nucleosome positioning pattern as nucleosome density distribution (except $r = 0$, explained in Appendix), with r representing the distance to the +1 nucleosome. The details of the radial distribution function are introduced in appendix.

III. RESULTS AND DISCUSSION

A. Features of nucleosome positioning pattern

If there is no remodeler, just random adsorption of nucleosome under a fixed density, the pattern of nucleosome positioning must be randomly distributed on the DNA sequence with no specific feature. Here, through our model, we find nucleosome positioning oscillations near TSS in the downstream direction, and the closer it is to TSS, the stronger oscillation it shows, as shown

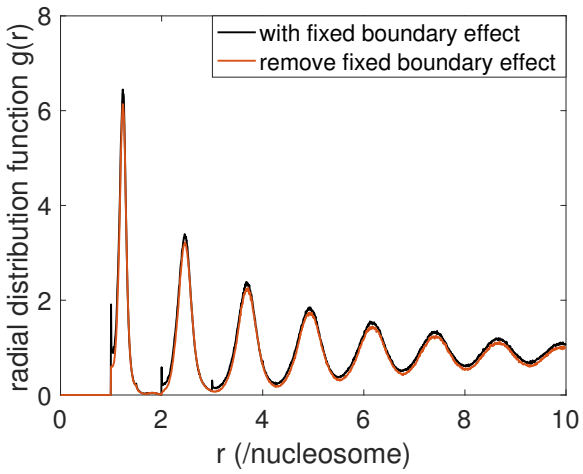

FIG. 2. Radial distribution of nucleosome positioning pattern with and without fixed boundary condition, up to 10 nucleosome length (10∗147bp) away. Model parameters: Intial nucleosome density = 0.85, total number of modeling steps = 1,000,000, $r_{on} = 1$,$r_{off} = 1$, $r_{trans} = 9$, total length of DNA sequence = 14,700 bps. The nucleosome positioning shows a decaying oscillation near TSS as r becomes larger in the downstream direction. The abnormal high values of g(r) at r = 1, 2, 3 in the black curve are caused by fixed boundary, which diminish when we remove the fixed boundary effect through dropping several nucleosomes near the boundary.

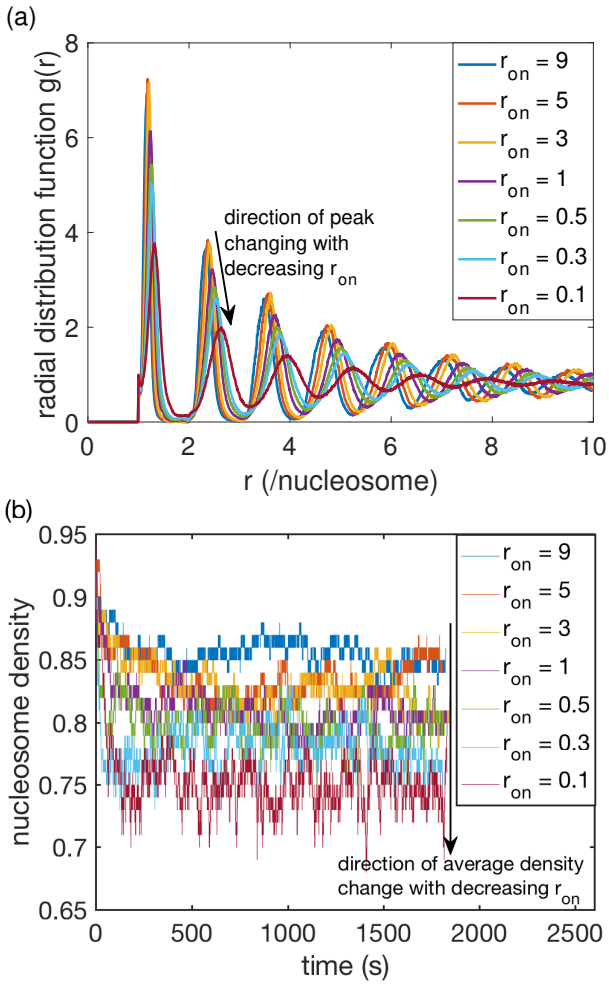

FIG. 3. g(r) of nucleosome positioning pattern and nucleosome density changing with time under different r_{on}, with $r_{off} = 1$, $r_{trans} = 9$, total length of DNA sequence = 14,700 bps. (a) g(r) in different r_{on} up to $r = 10$ nucleosome length away, with total modeling step = 1,000,000. As r_{on} decreasing, the peak-to-peak distance increases and peak magnitude decreases for nucleosome positioning. (b) Nucleosome density changes with time under different r_{on}, with total modeling step = 20,000. The average density for each steady state is decreasing with decreasing r_{on}

in Fig.2. What's more, we can see that the probability of finding a nucleosome within 147bp (length of one nucleosome) to the reference nucleosome is zero, which is consistent with our "hard core potential" model of nucleosome. In the corresponding physical model of Fig.2, +1 nucleosome, the first nucleosome downstream from NRF, serves as a reference particle, because in real biological system, it is directly positioned[10]; and we can see decay of oscillation magnitude as being away from the +1 nucleosome. This corresponds with what people have discovered in experiments, that on the downstream side, nucleosome positioning density shows decaying oscillatory pattern with respect to the distance from TSS[10].

The little difference between the two distribution curves in Fig.2 is caused by the fixed boundary condition we introduced in the simulation. After getting rid of several nucleosomes near the boundary, we get the well-positioned pattern of nucleosomes near TSS on the downstream side in Fig.2(the red curve). There would be no such difference if we calculate the density distribution (details are in the appendix).

B. Nucleosome positioning pattern change with nucleosome density

In our simulation, the actual time scale of the Gillespie model[2] is meaningless and the actual time scale is closely linked with the sum of all reaction rates, so we just need to know the relative reaction rates, not the exact ones. Here, we have already known from experiments that translocations happen in 90% probability, while eviction occupied 10%[8]; thus, we can deduce that $r_{trans} = 9\, r_{off}$. As a result, we set $r_{trans} = 9$, $r_{off} = 1$ in our model. Now, the only rate we need to determine is r_{on}, which depends on the concentration of octamers, H2A-H2B dimers and other nucleosome compositions in the environmental solution.

In Fig.3, we show the nucleosome pattern changing with decreasing value of r_{on} ranging from 9 to 0.1 in Fig.3(a), and the corresponding nucleosome density changing with time in Fig.3(b). With fixed translocation

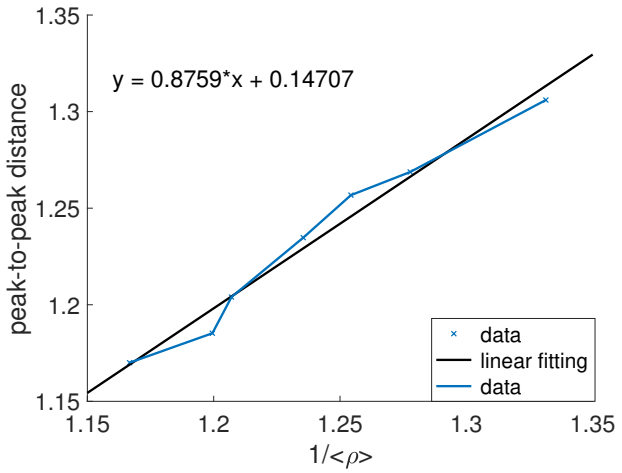

FIG. 4. Relationship between peak-to-peak distance in nucleosome positioning pattern and the reciprocal of mean nucleosome density. As shown in the figure above, the peak-to-peak distance is negatively correlated with mean nucleosome density, which we obtained from averaging the nucleosome density at steady state (data shown in Fig.3(b)).

rate, fixed eviction rate, and decreasing adsorption rate, the nucleosome density of the final steady state must be decreasing, as shown in Fig.3(b). Our system is somewhat different from usual ones in respect of density, that the nucleosome density in our system is always changing. From Fig.3(b), we can find that in long time scale, for a set of r_{on}, r_{off} and r_{trans}, the nucleosome density reaches a steady state quickly with some fluctuations. In addition, the adsorption rate cannot be too low; otherwise, eviction would be the leading factor in the fluctuation of the total number of nucleosome, resulting in no nucleosome in the end.

It is interesting to note that the pattern of nucleosome positioning changes with decreasing nucleosome adsorption rate in two aspects: peak-to-peak distance and peak magnitude. As we can see from Fig.3(a) and (b), nucleosome adsorption rate (r_{on}) decrease leads to lower nucleosome density, which results in larger peak-to-peak distance and smaller peak magnitude.

Specifically, the shift of peak position in high-order peak is larger than that in low-order peak, e.g., the positions of the first peak in different nucleosome densities are almost the same, while the positions of the seventh peak in different nucleosome densities diverge significantly. This is caused by lower average density, as shown in Fig.4, peak-to-peak distance shows a negative correlation with average nucleosome density. Moreover, the accumulation of the difference in peak-to-peak distance contributes to the peak divergence in high-order peaks.

As to the peak magnitude, its decrease is also caused by the decreasing average density, since lower density means more flexibility in nucleosome positioning, resulting in wider but lower peaks. In fact, the decrease in peak magnitude is not very obvious for r_{on} larger than 1, but quite considerable for r_{on} smaller than 1. It might be linked with some threshold in density, beyond which the flexibility in positioning is not influenced significantly by density.

IV. CONCLUSION

In this article, we explore the influence of SP6 RNA polymerase on nucleosome positioning during transcription in the vicinity of TSS on the downstream side. SP6 polymerases propagate in the transcription direction all the time and has some specific rules of remodeling when it encounters a nucleosome, including 90% probability transferring it to upstream DNA sequence without leaving DNA sequence and 10% probability releasing it. Apart from Sp6 polymerase, we also include random adsorption of nucleosomes on available sites along DNA sequence.

Based on these three reactions, we use Gillespie[2] to model the system and get the nucleosome positioning pattern after transcription through radial distribution function. This pattern shows a decaying oscillatory nucleosome positioning, which means nucleosomes are well-positioned near TSS in the downstream direction and corresponds with experiment results[10].

Besides, we also show that the nucleosome positioning pattern changes with different average nucleosome densities, which is due to different adsorption rates under fixed eviction and translocation rate. We get the conclusion that the average nucleosome density is negatively correlated to the peak-to-peak distance in the pattern. Furthermore, lower nuclesome density leads to higher flexibility in nucleosome positioning, resulting in wider peaks and smaller peak magnitude, which is significant under some threshold density.

Nucleosome positioning is closely linked with many fundamental processes in biology, including gene expression, chromatin packaging and so on. Its positioning pattern can be influenced by many factors, partly because the complex process of transcription needs the cooperation of many different proteins, which influence nucleosome positioning in various ways. Similar modeling method as in this article can be applied to other types of nucleosome remodelers, which could help us understand nucleosome positioning better.

V. ACKNOWLEDGMENTS

I thank Prof. Mehran Kardar and Prof. Leonid Mirny for giving all of the interesting lectures and Ph.D. Johannes Nuebler for helpful discussion of my project, especially the modeling process.

VI. APPENDIX

We use radial distribution function g(r) to describe nucleosome positioning pattern, which is defined as:

$$g(r) = \rho(r)/ < \rho > \quad (1)$$

where $\rho(r)$ means the nucleosome density at distance r, $< \rho >$ means the average nucleosome density in the system. g(r) describes the distribution of the distance between a pair of nucleosomes, where r obviously cannot be zero. This is the place that g(r) is different from density distribution, because the nucleosome density at r=0 must be very high, which represents the density of the reference nucleosome.

Here, we calculated g(r), which treats each nucleosome as the reference nucleosome, but in our model, the fixed boundary condition of nucleosomes treats only the first nucleosome as the reference one. This causes the boundary effect in Fig.2, which can be removed by ignoring the first few nucleosomes as stated in section III A. There would be no such difference in density distribution, because density distribution also considers fixed boundary condition.

[1] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, Nature (1997), 10.1038/38444.

[2] R. Erban, S. J. Chapman, and P. K. Maini, (2007).

[3] V. M. Studitsky, D. J. Clark, and G. Felsenfeld, Cell (1995), 10.1016/0092-8674(95)90230-9.

[4] H. E. Peckham, R. E. Thurman, Y. Fu, J. A. Stamatoyannopoulos, W. S. Noble, K. Struhl, and Z. Weng, Genome Research (2007), 10.1101/gr.6101007.

[5] S. Ercan, M. J. Carrozza, and J. L. Workman, Genome Biology **5** (2004).

[6] M. Radman-Livaja and O. J. Rando, "Nucleosome positioning: How is it established, and why does it matter?" (2010).

[7] O. I. Kulaeva, F. K. Hsieh, H. W. Chang, D. S. Luse, and V. M. Studitsky, "Mechanism of transcription through a nucleosome by RNA polymerase II," (2013).

[8] V. M. Studitsky, D. J. Clark, and G. Felsenfeld, Cell (1994), 10.1016/0092-8674(94)90343-3.

[9] P. Ranjith, J. Yan, and J. F. Marko, Proceedings of the National Academy of Sciences (2007), 10.1073/pnas.0701459104.

[10] W. Möbius and U. Gerland, PLoS Computational Biology (2010), 10.1371/journal.pcbi.1000891.