1 RPA dynamic binding to long ssDNA can be simulated by a generalized random sequential adsorption (RSA) model

To elucidate the biophysical mechanism of replication protein A (RPA) dynamic binding to long single-stranded DNA (ssDNA), we began by establishing a continuous-time discrete Markov chain model (Fig. 1a-b). Random sequential adsorption (RSA) models were initially proposed by Paul Flory to describe irreversible adsorption processes of large molecules to a liquid-solid interface[1]. If the interface is chosen to be one-dimensional (1D) and a finite length, the so-called 1D RSA model will provide an ideal basis for further modelling the processes of protein binding to DNA. Because this model precisely captures a key property, in which each nucleotide (nt) of ssDNA cannot be occupied by more than two protein molecules. This property differentiates DNA-relevant reactions from elementary chemical reactions described by the mass action law. And more importantly, it leads to incomplete occupation even if the proteins are oversaturated. Relevant theoretical consequences of applying RSA-like model to DNA-protein reaction kinetics are also known as the McGhee-von Hippel model[2, 3, 4]. This type of model has been applied to study similar biophysics settings, where a standard mean field approximation is employed to obtain accurate parameters of the kinetic parameters [5]. In order to reveal finer structures such as the gap distribution, we implemented an exact stochastic sampling approach to this model.

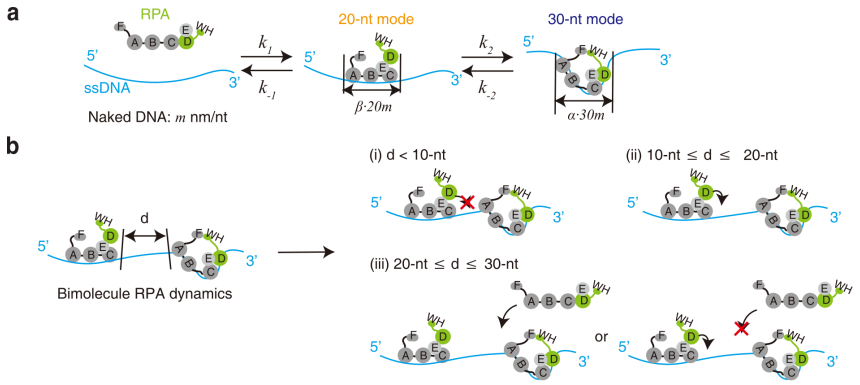

Figure 1: **A continuous-time discrete Markov chain model for multiple RPA molecules binding to long ssDNA.** (a) Schematic of building up continuous-time discrete Markov chain model. m represents absolute extension length of naked ssDNA per nt under a constant force in 5′-3′ direction. (b) All possible behaviors of bimolecular dynamics of RPA on ssDNA, which composed a complicated scenario of multiple RPA molecules binding. d represents the size of the naked ssDNA gap between two randomly loaded RPA molecules.

In our work of RPA, we adapted the model according to our current knowledge of RPA binding modes to obtain full element simulation results in which multiple binding modes and volume exclusion effects are all taken into consideration. Based on the previous references, we assumed that RPA has two binding modes, 20-nt mode and 30-nt mode, representing partial binding mode (PBM) and full-length binding mode (FLBM) (Fig. 1a). RPA initially binds to a 20-nt ssDNA with a rate of k_1 (unit s-1, the concentration of DNA and RPA incorporated), and dissociates to ssDNA with a rate of k_{-1} (unit s^{-1}). This is the 20-nt mode. k_1 is assumed to be constant in the model because ssDNA Curtains maintain a laminar flow of RPA, which allows the concentration of free RPA to be constant. One possible scenario for the 20-nt mode is the DBD-A, DBD-B, and DBD-C together binding to ssDNA. The DBD-D subsequent binding leads to the 30-nt mode (Fig. 1a). When an RPA molecule is in the 20-nt mode, DBD-D starts to bind to an extra 10-nt ssDNA with a rate of k_2 (unit s^{-1}), and dissociates to ssDNA with a rate of k_{-2} (unit s^{-1}). This is the 30-nt mode (FLBM). Here, we need to highlight the polarity of RPA binding to ssDNA. RPA always aligns along DNA in the same direction.

Due to the existence of multiple binding modes, RPA molecules exhibit interesting kinetic features such as facilitated exchange and desorption. With the advancement of computational power, we can now use the novel dynamic language Julia to carry out exact stochastic simulations that allow for all the possible behaviors of RPA-ssDNA interactions that are known to us. The relevant codes are available through link (`https://github.com/hsianktin/RPA_model`).

In this model, the state of ssDNA fragment is represented by a vector of length L, taking values in $\{0, 1\}$, where 0 represents that this nucleotide is not occupied and 1 represents the nucleotide being occupied. Each RPA must take up $\ell = 20$ nts for initial binding with DNA. And if the local state of DNA permits, it can further occupy another $\Delta\ell = 10$ nts in the $3'$ direction. In particular, one DNA site can be occupied by at most one RPA molecule, which modifies the available reactions for the RSA model, as shown in Fig. 1b.

We assume that each consecutive unoccupied segment of length ℓ recruits one RPA at the rate of k_1 (unit s^{-1}, the concentration of RPA incorporated), and that the total binding rate of RPA

$$v_1 = k_1 \sum_{j=1}^{L-\ell+1} \delta_{\mathbf{0}^{\times\ell}}(\text{state}[j, j+1, \dots, j+\ell-1]), \tag{1}$$

where $\delta_a(b)$ stands for the Kronecker delta function, which equals 1 if $a = b$ and 0 otherwise, $\mathbf{0}^{\times\ell} \in \mathbb{R}^\ell$ is a vector of length ℓ with all elements being 0, and $\text{state}[j, j+1, \dots, j+\ell-1]$ stands for a vector of length ℓ with elements being the state of ssDNA at position $j, j+1, \dots, j+\ell-1$.

The summation $\sum_{j=1}^{L-\ell+1} 1_{0^{\times\ell}}(\text{state}[j, j+1, \dots, j+\ell-1])$ stands for the number of consecutive unoccupied segments of length ℓ on the ssDNA. For further occupying another 10nts, we assign a rate parameter k_2 and calculate the overall

rate by

$$v_2 = k_2 \sum_{q_j} \delta_{\mathbf{0}^{\times \Delta\ell}} \left(\text{state} \left[q_j + \ell, q_j + \ell + 1, \ldots, q_j + \ell + \Delta\ell - 1 \right] \right), \tag{2}$$

where q_j represents the leftmost position of the j^{th} bound RPA in the 20− nt mode.

In terms of the unbinding pathway, we assume that the 30-nt mode must be reopened into the $20-$ nt mode before detaching from DNA, and the rate is assumed to be k_{-2}. And the desorption rate for the 20− nt mode is denoted as k_{-1}. The overall rates v_{-2} and v_{-1} are calculated according to the following formula: let p_j iterate over all 30nt mode RPAs and let q_j iterate over all 20-nt mode RPAs as before,

$$\begin{aligned} v_{-2} &= k_{-2} \#\{\text{FLBM RPA}\} \\ v_{-1} &= k_{-1} \#\{\text{PBM RPA}\}. \end{aligned} \tag{3}$$

At each moment, the total possible reaction rate v_{tot} (s^{-1}) is:

$$v_{\text{tot}} = v_1 + v_{-1} + v_2 + v_{-2} \tag{4}$$

Summation is taken over all possible reactions under the specific configuration. All the underlying reactions occur stochastically according to their exponentially distributed waiting times whose distribution parameter being corresponding reaction rates. In other words, the waiting time between two successive reactions δt (s) follows an exponential distribution, which is a probability density function (pdf):

$$\text{pdf}(\delta t = t) = v_{\text{tot}} e^{-v_{\text{tot}} t}$$

And the mean value of δt (s) is:

$$\langle \delta t \rangle = \frac{1}{v_{\text{tot}}}$$

After each reaction step, the state of DNA is changed, and the number of possible reactions should be evaluated again based on the current state of DNA. Gillespie algorithm is applied to the system for sampling trajectories from this stochastic model.

We scanned the parameter space of $(k_1, k_{-1}, k_2, k_{-2})$ uniformly on the log scale, between 10^{-6} s^{-1} and 10^{-1} s^{-1}. For each combination of parameters, we simulated 100 trajectories comprized of the densities of 20-nt RPA and 30-nt RPA on DNA.

References

[1] Paul J. Flory. Intramolecular reaction between neighboring substituents of vinyl polymers. *Journal of the American Chemical Society*, 61(6):1518–1521, jun 1939.

[2] James D. McGhee and Peter H. von Hippel. Theoretical aspects of DNA-protein interactions: Co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *Journal of Molecular Biology*, 86(2):469–489, jun 1974.

[3] Javier Jarillo, José A. Morín, Elena Beltrán-Heredia, Juan P. G. Villaluenga, Borja Ibarra, and Francisco J. Cao. Mechanics, thermodynamics, and kinetics of ligand binding to biopolymers. *PLOS ONE*, 12(4):e0174830, apr 2017.

[4] Juan P. G. Villaluenga, Jules Vidal, and Francisco Javier Cao-García. Non-cooperative thermodynamics and kinetic models of ligand binding to polymers: Connecting McGhee–von hippel model with the tonks gas model. *Physical Review E*, 102(1), jul 2020.

[5] M Nabuan Naufer, Michael Morse, Guðfríður Björg Möller, James McIsaac, Ioulia Rouzina, Penny J Beuning, and Mark C Williams. Multiprotein e. coli SSB-ssDNA complex shows both stable binding and rapid dissociation due to interprotein interactions. *Nucleic Acids Research*, 49(3):1532–1549, jan 2021.