

RESEARCH ARTICLE | JUNE 22 2023

## Dynamics of single-base editing: Theoretical analysis

Vardan Hoviki Vardanyan ; Qian Wang ; Anatoly B. Kolomeisky  



*J. Chem. Phys.* 158, 245101 (2023)

<https://doi.org/10.1063/5.0157193>



06 June 2024 15:14:07



### The Journal of Chemical Physics

Special Topic: Algorithms and Software  
for Open Quantum System Dynamics

**Submit Today**



# Dynamics of single-base editing: Theoretical analysis

Cite as: J. Chem. Phys. 158, 245101 (2023); doi: 10.1063/5.0157193

Submitted: 5 May 2023 • Accepted: 6 June 2023 •

Published Online: 22 June 2023



View Online



Export Citation



CrossMark

Vardan Hoviki Vardanyan,<sup>1,2</sup>  Qian Wang,<sup>3</sup>  and Anatoly B. Kolomeisky<sup>1,2,4,5,a)</sup> 

## AFFILIATIONS

<sup>1</sup> Department of Chemistry, Rice University, Houston, Texas 77005, USA

<sup>2</sup> Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, USA

<sup>3</sup> Hefei National Laboratory for Physical Sciences at the Microscale and Department of Physics, University of Science and Technology of China, Hefei 230026, Anhui, China

<sup>4</sup> Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77005, USA

<sup>5</sup> Department of Physics and Astronomy, Rice University, Houston, Texas 77005, USA

<sup>a)</sup> Author to whom correspondence should be addressed: [tolya@rice.edu](mailto:tolya@rice.edu)

## ABSTRACT

Recent experimental advances led to the development of DNA base editors (BEs) with single-nucleotide precision, which is critical for future progress in various scientific and technological fields. The molecular mechanisms of single-base discrimination, however, remain poorly understood. Using a recently developed stochastic approach, we theoretically investigated the dynamics of single-base editing. More specifically, transient and mean times to edit “TC” motifs by cytosine BEs are explicitly evaluated for correct (target) and incorrect (bystander) locations on DNA. In addition, the effect of mutations on the dynamics of the single-base edition is also analyzed. It is found that for most ranges of parameters, it is possible to temporarily separate target and bystander products of base editing, supporting the idea of dynamic selectivity as a method of improving the precision of single-base editing. We conclude that to improve the efficiency of single-base editing, selecting the probability or selecting the time requires different strategies. Physical–chemical arguments to explain the observed dynamic properties are presented. The theoretical analysis clarifies some important aspects of the molecular mechanisms of selective base editing.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0157193>

## I. INTRODUCTION

One of the most striking recent developments that strongly affected multiple scientific and technological areas, ranging from chemistry and biology to engineering and medicine, is the discovery of experimental methods that allow for precise genome editing.<sup>1–3</sup> Originally based on utilizing clustered regularly interspaced short palindromic repeat (CRISPR) techniques, they have been significantly improved in recent years, revolutionizing multiple research fields.<sup>4–9</sup> While a wide range of gene editing tools that utilize the CRISPR-Cas9 methods has been proposed,<sup>4,10</sup> most of them require double-strand DNA breaks, frequently leading to unpredictable editing outcomes. Much more precise editing is achieved by base editors (BEs) that are constructed by fusing protein domains with specific enzymatic properties (for example, deaminases) and nickase Cas9 proteins that allow quickly locating the proper site on a

DNA.<sup>7,11–13</sup> In this case, the process is taking place without double-strand breaks in DNA, and this provides much more efficient and controllable genome editing.<sup>6,9,11,12</sup>

A wide spectrum of single-base editors has been reported in recent years.<sup>12–16</sup> These biologically engineered systems allow for better editing precision and higher purity of products.<sup>17,18</sup> At the same time, the main problem with BEs remains the discrimination of identical bases in the activity windows (typically, 4–10 nucleotides) of these enzymatic complexes. In other words, it is challenging for the editor to modify only the specific base at the given site if identical bases can be found close to the target in the region labeled as the activity window. As a result, the bystander or both target and bystander nucleotides might be modified, negatively impacting the efficiency and precision of genome editing. While some improvements have been made by utilizing beneficial mutations in enzymatic domains, the discrimination issue has not

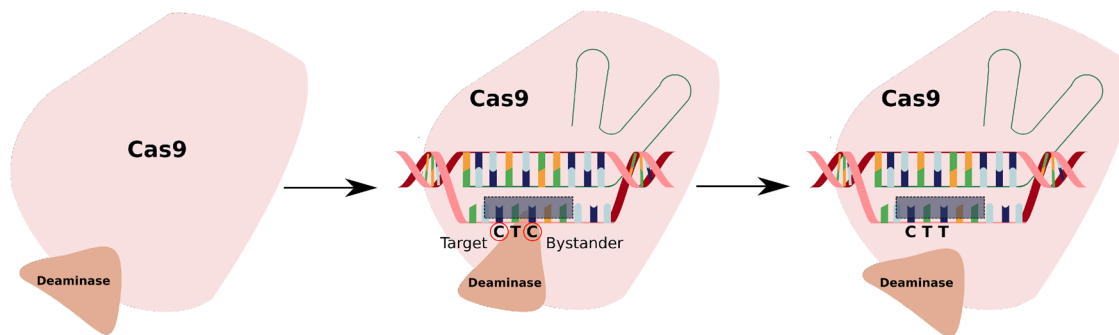


FIG. 1. A schematic view of the functioning of the cytosine single-base editor. The gray area indicates the activity window of this BE.

been fully resolved yet.<sup>14,19,20</sup> The main problem here is that the molecular mechanisms of the underlying processes remain poorly clarified.

To assist in understanding the microscopic features of base editing and to rationally design more efficient BE systems, a theoretical approach that combines a discrete-state stochastic model with all-atom molecular dynamics has recently been proposed by one of us.<sup>21</sup> It presented a minimal chemical-kinetic model that accounts for the most relevant chemical states of base editing and includes the possibility of target and bystander editing. More specifically, cytosine BE that converts cytosine (C) to thymine (T) has been considered.<sup>14</sup> This process is shown schematically in Fig. 1. Some transition rates for the stochastic model have been estimated using various experimental observations,<sup>14,20,22</sup> while the rest of the parameters have been obtained using all-atom molecular dynamics simulations.<sup>21</sup> The method of first-passage probabilities<sup>23–25</sup> has been utilized then to characterize the features of base editing. This theoretical approach clarified some important molecular aspects of the process, and it proposed a set of general rules for designing BEs with improved editing selectivity. Importantly, the theoretical predictions of decreased bystander effect for specific mutations have been successfully verified by experiments.<sup>21</sup>

Although this theoretical approach<sup>21</sup> has provided some crucial information on the mechanisms of base editing, it has some limitations. This method mostly concerns the thermodynamic aspects of base editing since only the probabilities of different editing outcomes have been considered. The dynamics of base editing have not been effectively evaluated at all. However, different products might appear in the system at different times, and in real experiments, the most optimal conditions from the point of view of probability might not be realized because it might take too long to achieve them. Different times for different products also suggest that the selectivity of BEs might be improved by temporarily separating the products. The idea of improving precision by collecting the products at different times is similar to the concept of dynamic selectivity.<sup>26</sup> These observations raise several important questions. What is the dynamics of transient processes during base editing? At what times do different products (target and bystander) appear in the system? Is it possible to separate them to enhance selectivity? If yes, is the range of parameters for which the optimal dynamic selectivity might be

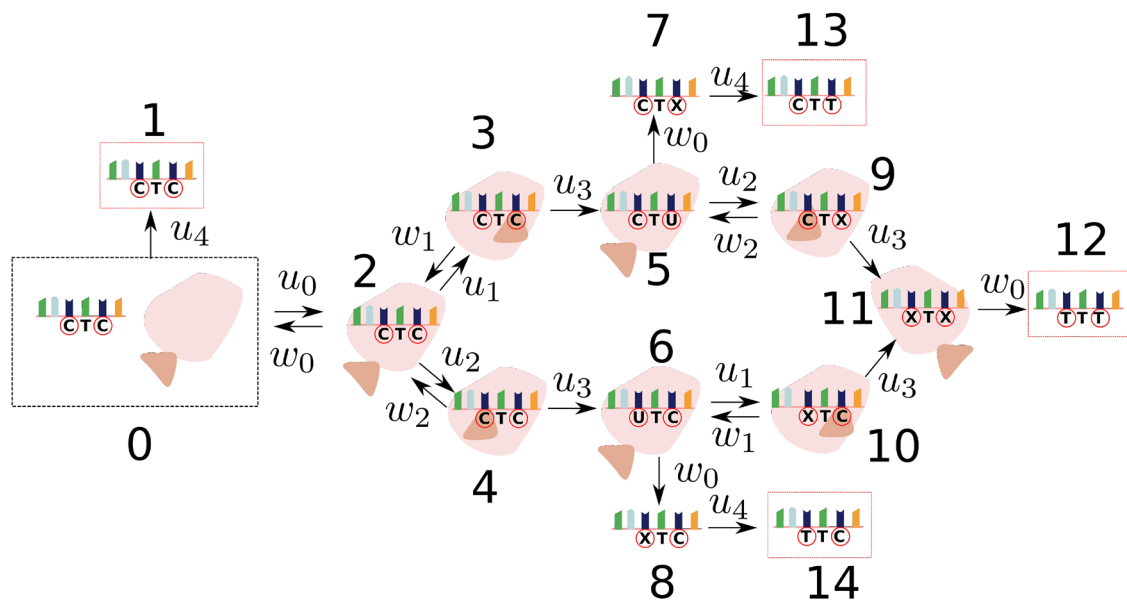
observed correlated with the range of parameters when the bystander effect is reduced?

To answer these questions, we extend the original theoretical approach<sup>21</sup> to explicitly evaluate the dynamics of transient processes and calculate the mean editing times for different outcomes of single-base discrimination. This allows us to better understand the complex dynamics of single-base editing. Our analysis suggests that there are parameters at which the different products of base editing can be temporarily separated, exhibiting dynamic selectivity. Interestingly, it is found that the best conditions to dynamically separate the target editing do not correlate with the situations in which the probability of bystander editing is small. Physical-chemical arguments to explain these observations are presented. Our theoretical method allows for a better understanding of the microscopic processes associated with single-base editing in DNA.

## II. THEORETICAL METHOD

To investigate the dynamics, we utilize a minimal chemical-kinetic model already developed to describe the process of genome editing of EGFP site, which codes for green fluorescent protein, 1 by A3A-BE3 cytosine BE.<sup>21</sup> It is schematically presented in Fig. 2. Note that although the total number of states in this chemical-kinetic description is relatively large (total of 15 states), they describe four major pathways in this system that cannot be neglected if one aims to correctly analyze genome editing by BEs.<sup>21</sup>

The base-editing process starts in state 0, and then there are two possible transitions out of this state. The BE can change into inactive conformation (transition  $0 \rightarrow 1$ ) with a rate of  $u_4$ , which corresponds to the situation when no editing events happened. However, if the BE stays active, its Cas9 domain can associate with the single-stranded DNA segment with a rate of  $u_0$  (transition  $0 \rightarrow 2$ ). This process is reversible ( $2 \rightarrow 0$ ), and the corresponding dissociation rate is  $w_0$ . Then there are two possibilities. The deaminase domain of the BE can bind to the target nucleotide with a rate of  $u_1$  ( $2 \rightarrow 3$ ), or it can bind to the bystander cytosine with a rate of  $u_2$  ( $2 \rightarrow 4$ ). Since the target and bystander nucleotides are chemically identical and spatially very close, it is reasonable to assume that  $u_1 = u_2$ .<sup>21</sup> Both of these processes are reversible, and the backward reaction rates are  $w_1$  ( $3 \rightarrow 2$ ) and  $w_2$  ( $4 \rightarrow 2$ ). In the next step, the enzymatic reaction converts cytidine (C) into uridine (U) with the same rate of



**FIG. 2.** Chemical-kinetic model to describe single-base transformations by cytosine BE. The details of the process are discussed in the text. The system starts in state 0, and there are four possible outcomes. State 1 corresponds to no editing events, state 12 describes the editing at both target and bystander sites, state 13 corresponds to the target-site editing (the desired outcome), and state 14 describes the bystander editing. The nucleotides labeled “C” or “T” correspond to cytosine or thymine, respectively, while the nucleotide labeled “X” corresponds to uridine or thymine.

$u_3$  for the target site ( $3 \rightarrow 5$ ) and the bystander site ( $4 \rightarrow 6$ ). After that, from the states 5 or 6, there are two possible chemical pathways. If the Cas9 domain dissociates from DNA, the uridine will be modified into thymidine by a DNA repair mechanism (transitions  $5 \rightarrow 7 \rightarrow 13$  or  $6 \rightarrow 8 \rightarrow 14$ ). State 13 corresponds to the product of correct editing (target CTT), while state 14 corresponds to the wrong product (bystander TTC). However, if the Cas9 domain stays longer on DNA, it can also modify the neighboring nucleotides (transitions  $5 \rightarrow 9 \rightarrow 11 \rightarrow 12$  or  $6 \rightarrow 10 \rightarrow 11 \rightarrow 12$ ) with the corresponding transition rates (see Fig. 2). These events result in the wrong product where both target and bystander nucleotides are edited (TTT, state 12).

The transition from state 11 to state 12 should also have two steps with the rates  $w_0$  and  $u_4$ , similar to  $5 \rightarrow 7 \rightarrow 13$  and  $6 \rightarrow 8 \rightarrow 14$ . However, because for the experimentally estimated rates, it was found that  $w_0 \ll u_4$ ,<sup>21</sup> this transition can be viewed as a single step with the limiting rate  $w_0$ . This approximation does not affect any of the results of our calculations.

One should also note that cytosine and thymine nucleotides are labeled as “C” and “T,” respectively (Fig. 2). At the same time, the nucleotide labeled “X” represents both uridine and thymine, reflecting the fact that it is not known when the chemical transformation of uridine into thymine is taking place.

To obtain the dynamic properties of single-base editing, we utilize a method of first-passage probabilities that have been widely explored for investigating stochastic processes in chemistry, physics, and biology.<sup>21,23–25,27–30</sup> More specifically, we explicitly evaluate not only the probabilities of different editing outcomes, which was already accomplished before,<sup>28</sup> but also the mean editing times to

obtain different products. To simplify our calculations, we notice that some of the transitions are irreversible (see Fig. 2), allowing us to consider the editing process (namely, reaching states 12, 13, and 14) as a sequence of two events. In the first event, the system reaches the intermediate states 5 or 6 starting from state 0, and in the second event, the final editing states are achieved after initiating in state 5 or 6 (see Fig. 2). This allows us to better understand transient processes in base editing.

Let us consider the first processes that start in state 0 and end in state 5. One can introduce a function  $F_j(t)$  defined as a probability density to reach the state 5 at time  $t$  for the first time if, at  $t = 0$ , the system started in the state  $j$  ( $j = 0, 2, 3, 4$ ) before reaching other exit states 1 and 6. Thus, these are conditional first-passage probability density functions. The temporal evolution of these functions is given by a set of backward master equations,<sup>21,23,24,29</sup>

$$\frac{dF_0(t)}{dt} = u_0 F_2(t) - (u_0 + u_4) F_0(t), \quad (1)$$

$$\frac{dF_2(t)}{dt} = w_0 F_0(t) + u_1 F_3(t) + u_2 F_4(t) - (w_0 + u_1 + u_2) F_2(t), \quad (2)$$

$$\frac{dF_3(t)}{dt} = w_1 F_2(t) + u_3 F_5(t) - (w_1 + u_3) F_3(t), \quad (3)$$

$$\frac{dF_4(t)}{dt} = w_2 F_2(t) - (w_2 + u_3) F_4(t). \quad (4)$$

In addition, there is a boundary condition  $F_5(t) = \delta(t)$ , which has a physical meaning that if the system starts in state 5, the process

is immediately accomplished. These equations also assume that if the system reaches states 1 or 6, these are unsuccessful events, i.e.,  $F_1(t) = F_6(t) = 0$  at all times.

The system of backward master equations can be solved by utilizing the Laplace transformation of the first-passage probability functions,  $\tilde{F}_j(s) \equiv \int_0^\infty e^{-st} F_j(t) dt$ , which allows us to modify the original differential equations into a system of algebraic equations,

$$\tilde{F}_0(s)(s + u_0 + u_4) = u_0 \tilde{F}_2(s), \quad (5)$$

$$\tilde{F}_2(s)(s + w_0 + u_1 + u_2) = w_0 \tilde{F}_0(s) + u_1 \tilde{F}_3(s) + u_2 \tilde{F}_4(s), \quad (6)$$

$$\tilde{F}_3(s)(s + w_1 + u_3) = w_1 \tilde{F}_2(s) + u_3, \quad (7)$$

$$\tilde{F}_4(s)(s + w_2 + u_3) = w_2 \tilde{F}_2(s). \quad (8)$$

These equations can be exactly solved to obtain the explicit expressions for  $\tilde{F}_j(s)$ , from which one can estimate the dynamic properties of the process. More specifically, we evaluate the probability of the event  $(P_{0 \rightarrow 5})$ , which is the splitting probability of reaching state 5, before reaching other exit states,

$$P_{0 \rightarrow 5} = \frac{u_3(u_3 + w_2)}{u_3(\gamma_1 + 1)(2u_3 + w_1 + w_2) + \gamma_1 \gamma_3(u_3 + w_1)(u_3 + w_2)}, \quad (9)$$

and the mean first-passage time for this event  $(T_{0 \rightarrow 5})$  is equal to

$$T_{0 \rightarrow 5} = \frac{u_1(u_3 + w_2)[\gamma_3(u_3 + w_1)(u_3 + w_2) + u_3(2u_3 + w_1 + w_2)]}{u_0 u_1(u_3 + w_2)[(\gamma_1 + 1)u_3(2u_3 + w_1 + w_2) + \gamma_1 \gamma_3(u_3 + w_1)(u_3 + w_2)]} + \frac{u_0[\gamma_1 \gamma_3 u_1(u_3 + w_2)^2 + (\gamma_1 + 1)[(u_3 + w_1)(u_3 + w_2)^2 + u_1[w_2(4u_3 + w_1) + 2u_3^2 + w_2^2]]]}{u_0 u_1(u_3 + w_2)[(\gamma_1 + 1)u_3(2u_3 + w_1 + w_2) + \gamma_1 \gamma_3(u_3 + w_1)(u_3 + w_2)]}, \quad (10)$$

where parameters  $\gamma_1$ ,  $\gamma_2$ , and  $\gamma_3$  are given by<sup>21</sup>

$$\gamma_1 = \frac{u_4}{u_0}, \quad \gamma_2 = \frac{w_1}{u_3}, \quad \gamma_3 = \frac{w_0}{u_1}. \quad (11)$$

A similar analysis can be done for another process in the first step of base editing,  $0 \rightarrow 6$  (see Fig. 2). One can derive that the

probability of the event  $(P_{0 \rightarrow 6})$  and the mean time before this happens  $(T_{0 \rightarrow 6})$  are given by the following expressions:

$$P_{0 \rightarrow 6} = \frac{u_3(u_3 + w_1)}{u_3(\gamma_1 + 1)(2u_3 + w_1 + w_2) + \gamma_1 \gamma_3(u_3 + w_1)(u_3 + w_2)} \quad (12)$$

and

$$T_{0 \rightarrow 6} = \frac{u_1(u_3 + w_1)[\gamma_3(u_3 + w_1)(u_3 + w_2) + u_3(2u_3 + w_1 + w_2)]}{u_0 u_1(u_3 + w_1)[(\gamma_1 + 1)u_3(2u_3 + w_1 + w_2) + \gamma_1 \gamma_3(u_3 + w_1)(u_3 + w_2)]} + \frac{u_0[\gamma_1 \gamma_3 u_1(u_3 + w_1)^2 + (\gamma_1 + 1)[(u_3 + w_2)(u_3 + w_1)^2 + u_1[4u_3 w_1 + 2u_3^2 + w_1(w_1 + w_2)]]]}{u_0 u_1(u_3 + w_1)[(\gamma_1 + 1)u_3(2u_3 + w_1 + w_2) + \gamma_1 \gamma_3(u_3 + w_1)(u_3 + w_2)]}. \quad (13)$$

In Fig. 3(a), we present the mean times for the first stage of base editing as a function of the binding rate of the deaminase domain to the nucleotide. As one can see, increasing the binding rate  $u_1$  lowers the time to reach these states, which is expected. However, what is less expected is that the events along the less probable pathway are faster (note that  $P_{0 \rightarrow 6} \ll P_{0 \rightarrow 5}$ ; see the caption of Fig. 3). This can be explained using the following arguments. Because  $w_2 > w_1$  (see Fig. 2), the system is rarely able to reach state 6 from state 4. Only fast transitions are able to do it. As a result, we have for the mean times  $T_{0 \rightarrow 6} < T_{0 \rightarrow 5}$ , although the probabilities of these events are quite small,  $P_{0 \rightarrow 6} \ll P_{0 \rightarrow 5}$ .

The second stage of the base editing process starts in the states of 5 or 6 (Fig. 2), and there are three final outcomes. Editing the target nucleotide leads to the state of 13, editing the bystander nucleotide produces the state of 14, and editing both target and

bystander nucleotides yields the state of 12. The first-passage analysis for these processes can be done in a similar way as done above for the processes in the first step of base editing. We then obtain the probabilities,

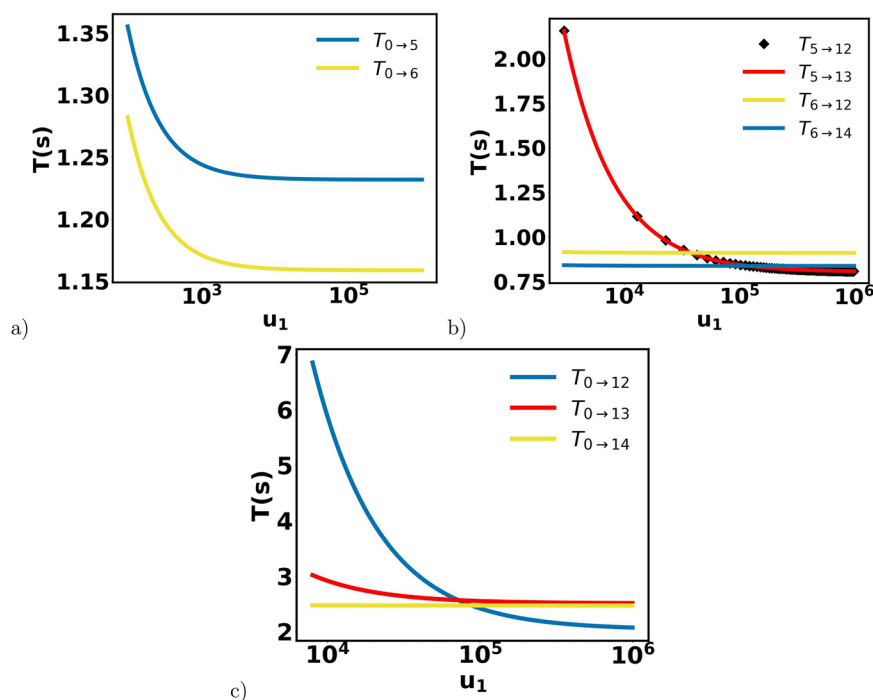
$$P_{5 \rightarrow 12} = \frac{u_1 u_3}{w_0(u_3 + w_2) + u_1 u_3}, \quad P_{5 \rightarrow 13} = 1 - P_{5 \rightarrow 12}, \quad (14)$$

and

$$P_{6 \rightarrow 12} = \frac{u_1 u_3}{w_0(u_3 + w_1) + u_1 u_3}, \quad P_{6 \rightarrow 14} = 1 - P_{6 \rightarrow 12}. \quad (15)$$

For the mean times for these events, the results are

$$T_{5 \rightarrow 13} = \frac{u_3^2 + 2u_3 w_2 + w_2(u_1 + w_2)}{(u_3 + w_2)[u_1 u_3 + w_0(u_3 + w_2)]}, \quad (16)$$



**FIG. 3.** Dynamics of different base-editing processes as a function of the binding rate of the deaminase domain to the nucleotide  $u_1$ . The following parameters have been used in calculations:<sup>21</sup>  $u_0 = 1 \text{ s}^{-1}$ ,  $u_3 = 1.1 \text{ s}^{-1}$ ,  $u_4 = 2.1 \text{ s}^{-1}$ ,  $w_0 = 2.9 \times 10^{-5} \text{ s}^{-1}$ ,  $w_1 = 12.54 \text{ s}^{-1}$ , and  $w_2 = 5059 \text{ s}^{-1}$ . (a) Mean times for  $0 \rightarrow 5$  and  $0 \rightarrow 6$  processes in the first stage of base editing. The corresponding probabilities are  $P_{0 \rightarrow 5} = 0.321635$  and  $P_{0 \rightarrow 6} = 0.000867$ . (b) Mean times for  $5 \rightarrow 12$ ,  $5 \rightarrow 13$ ,  $6 \rightarrow 12$ , and  $6 \rightarrow 14$  processes in the second stage of base editing. The corresponding probabilities are  $P_{5 \rightarrow 12} = 0.882299$ ,  $P_{5 \rightarrow 13} = 0.117701$ ,  $P_{6 \rightarrow 12} = 0.999641$ , and  $P_{6 \rightarrow 14} = 0.000359$ . (c) Mean times for the overall target editing  $0 \rightarrow 13$ , bystander editing  $0 \rightarrow 14$ , and double editing  $0 \rightarrow 12$ . The corresponding probabilities are  $P_{0 \rightarrow 13} = 0.0378568$ ,  $P_{0 \rightarrow 14} = 3.11661 \times 10^{-7}$ , and  $P_{0 \rightarrow 12} = 0.284645$ .

$$T_{5 \rightarrow 12} = \frac{u_1 + u_3 + w_0 + w_2}{w_0(u_3 + w_2) + u_1 u_3}, \quad (17)$$

$$T_{6 \rightarrow 12} = \frac{u_1 + u_3 + w_0 + w_1}{w_0(u_3 + w_1) + u_1 u_3}, \quad (18)$$

$$T_{6 \rightarrow 14} = \frac{u_3^2 + 2u_3 w_1 + w_1(u_1 + w_1)}{(u_3 + w_1)[u_1 u_3 + w_0(u_3 + w_1)]}. \quad (19)$$

The results for dynamics in the second step of base editing are shown in Fig. 3(b). Increasing the binding rate  $u_1$  accelerates the processes that start in state 5, while there is a much smaller effect on the processes starting from state 6. Again, the least probable process  $0 \rightarrow 14$  is the fastest for most ranges of parameters, although for very fast binding rates, the processes that start in state 5 become slightly faster.

Now we can combine the analysis of the first and second steps of base editing to obtain the dynamic properties of the overall process. The probability and the mean time to edit the target nucleotide can be evaluated as

$$P_{0 \rightarrow 13} = P_{0 \rightarrow 5} P_{5 \rightarrow 13} \quad (20)$$

and

$$T_{0 \rightarrow 13} = T_{0 \rightarrow 5} + T_{5 \rightarrow 13}. \quad (21)$$

Similarly, one can derive for the editing of bystander nucleotides,

$$P_{0 \rightarrow 14} = P_{0 \rightarrow 6} P_{6 \rightarrow 14} \quad (22)$$

and

$$T_{0 \rightarrow 14} = T_{0 \rightarrow 6} + T_{6 \rightarrow 14}. \quad (23)$$

For the process of editing both target and bystander nucleotides, it can be shown that

$$P_{0 \rightarrow 12} = P_{0 \rightarrow 5} P_{5 \rightarrow 12} + P_{0 \rightarrow 6} P_{6 \rightarrow 12} \quad (24)$$

and

$$T_{0 \rightarrow 12} = \frac{P_{0 \rightarrow 5} P_{5 \rightarrow 12} (T_{0 \rightarrow 5} + T_{5 \rightarrow 12}) + P_{0 \rightarrow 6} P_{6 \rightarrow 12} (T_{0 \rightarrow 6} + T_{6 \rightarrow 12})}{P_{0 \rightarrow 5} P_{5 \rightarrow 12} + P_{0 \rightarrow 6} P_{6 \rightarrow 12}}. \quad (25)$$

The physical meaning of this expression is the following: There are two pathways to reach state 12 from state 0. The factor  $\frac{P_{0 \rightarrow 5} P_{5 \rightarrow 12}}{P_{0 \rightarrow 5} P_{5 \rightarrow 12} + P_{0 \rightarrow 6} P_{6 \rightarrow 12}}$  gives the probability for the system to choose pathway  $0 \rightarrow 5 \rightarrow 12$ , and the factor  $\frac{P_{0 \rightarrow 6} P_{6 \rightarrow 12}}{P_{0 \rightarrow 5} P_{5 \rightarrow 12} + P_{0 \rightarrow 6} P_{6 \rightarrow 12}}$  is the probability to follow the pathway  $0 \rightarrow 6 \rightarrow 12$ . The explicit expressions for all dynamic properties for the overall base editing can be easily obtained from Eqs. (12)–(19).



Figure 3(c) illustrates the overall dynamics of base editing for different processes. In all cases, increasing the rate of  $u_1$  accelerates the dynamics, although the amplitude is not the same for different pathways. The strongest effect is observed for the processes of editing both bystander and target nucleotides ( $0 \rightarrow 12$ ). The weaker effect is observed for editing the target nucleotide ( $0 \rightarrow 13$ ), and there is almost no effect for editing the bystander nucleotide ( $0 \rightarrow 14$ ). However, the important result is that the mean times to obtain different products of base editing in most ranges of parameters are not the same, suggesting that the products can be temporarily separated. This is the main reason to explore the idea of dynamic selectivity for improving the precision of single-base editing.

To optimize the performance of single-base editors, the most common approach is to explore various mutations in the deaminase domain.<sup>14,19–21</sup> Then, the effect of a mutation can be viewed as a perturbation in the energy of the base-editing process. In the simplest approximation, which, however, was also supported by some recent experimental data,<sup>21</sup> it was argued that the effect will mostly occur in the dissociation steps of the deaminase from the DNA chain, leading to the following changes in unbinding transition rates:

$$\begin{aligned} w_{1,\text{mutation}} &= w_{1,\text{WT}} \exp \left[ \frac{\Delta E_m}{k_B T} \right], \\ w_{2,\text{mutation}} &= w_{2,\text{WT}} \exp \left[ \frac{\Delta E_m}{k_B T} \right], \end{aligned} \quad (26)$$

where  $\Delta E_m$  is the free-energy difference in the deaminase unbinding process for the system with a mutation in comparison with the wild-type (WT) case. When  $\Delta E_m$  is positive, the unbinding process is happening faster, while for  $\Delta E_m < 0$ , it is more suppressed. Recent investigations have shown that for the A3A-BE3 editing system, the mutations that lead to the perturbations  $\Delta E_m \approx 5 - 7 k_B T$  exhibit the decreased bystander effect, i.e., the probability of target editing in comparison with other outcomes is maximal at these conditions.<sup>21</sup>

The effect of mutations can be easily quantified in our theoretical approach. One can analyze the ratio of probabilities  $P_{0 \rightarrow 13}/P_{0 \rightarrow 14}$ , which specifies how more probable the target editing is over the bystander editing, and the ratio of probabilities  $P_{0 \rightarrow 13}/P_{0 \rightarrow 12}$ , which measures how the target editing is more probable in comparison with editing both target and bystander nucleotides. The results for different free-energy perturbations associated with different mutations are presented in Fig. 4. One can see that the target editing ( $0 \rightarrow 13$ ) is much more probable than the bystander editing ( $0 \rightarrow 14$ ) for most ranges of parameters, but the largest effect is achieved for the systems with  $\Delta E_m \approx 0$ . Mutations associated with strongly negative or strongly positive free-energy changes decrease the advantage of target editing. The situation is very different when we compare it with the double editing events. For the mutations with  $\Delta E_m < 0$ , it is more probable to obtain the products with editing in both target and bystander nucleotides. Only for strongly positive perturbations ( $\Delta E_m \approx 5 - 10 k_B T$ ), the target editing becomes more preferred. These are the conditions of the reduced bystander effect, as one can also see from the ratio  $P_{0 \rightarrow 13}/[P_{0 \rightarrow 12} + P_{0 \rightarrow 14}]$  presented in Fig. 4. This quantity describes the relative probability of obtaining the desired target editing in comparison with other undesired outcomes.

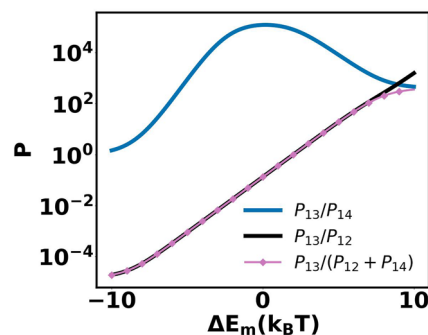


FIG. 4. Ratio of probabilities, or fractions of different outcomes of base editing as a function of free-energy perturbation due to mutations. The following parameters have been used in calculations:<sup>21</sup>  $u_0 = 1 \text{ s}^{-1}$ ,  $u_3 = 1.1 \text{ s}^{-1}$ ,  $u_4 = 2.1 \text{ s}^{-1}$ ,  $w_0 = 2.9 \times 10^{-5} \text{ s}^{-1}$ ,  $w_1 = 12.54 \text{ s}^{-1}$ , and  $w_2 = 5059 \text{ s}^{-1}$ .

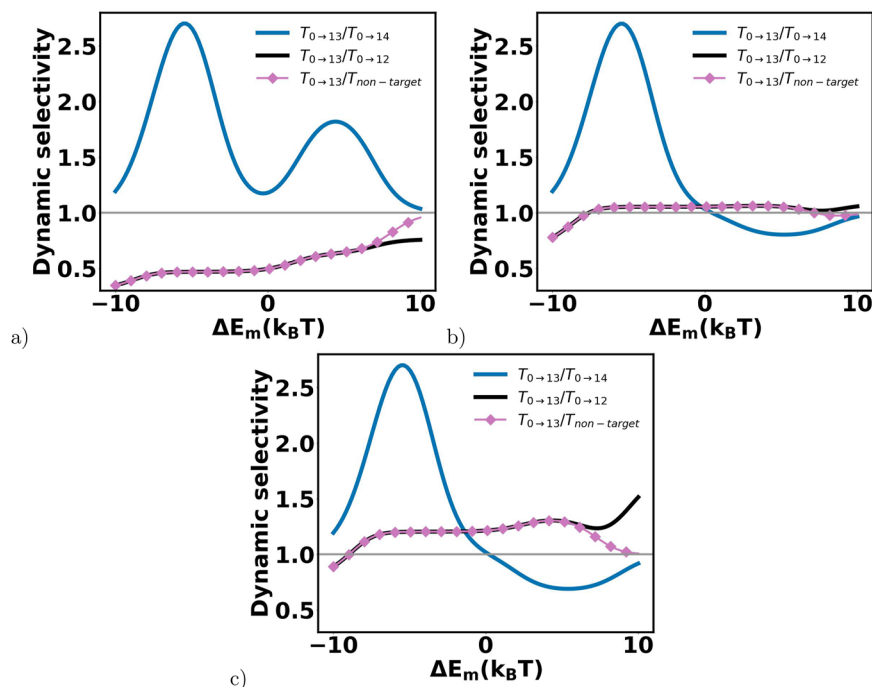
To understand which editing products are coming first, in Fig. 5, we present the results of explicit calculations for the ratios of mean editing times for different binding rates  $u_1$  as a function of free-energy perturbations associated with possible mutations. More specifically,  $T_{0 \rightarrow 13}/T_{0 \rightarrow 14}$ , which describes how faster the target nucleotide is edited in comparison with the bystander nucleotide, and  $T_{0 \rightarrow 13}/T_{0 \rightarrow 12}$ , which describes how faster the target nucleotide is edited in comparison with editing both nucleotides, are calculated (see Fig. 5). For slow binding rates  $u_1$  [Fig. 5(a)], it is usually faster to edit the bystander nucleotide, although the probability for such events is quite low (compare with Fig. 4). As we already argued, this is because only the fastest events will go along the pathway  $0 \rightarrow 14$ , but they are quite rare. Increasing the binding rate  $u_1$  [Figs. 5(b) and 5(c)] changes the situation. It is still faster to edit the bystander nucleotide for  $\Delta E_m < 0$ , while the desired target editing is faster for mutations associated with  $\Delta E_m > 0$ .

Comparing the dynamics of target editing with editing in both locations, as quantified by the ratio  $T_{0 \rightarrow 13}/T_{0 \rightarrow 12}$ , one can see that for the slow binding rates  $u_1$ , it is faster to edit the target nucleotide [Fig. 5(a)]. However, increasing the rate  $u_1$  makes both time scales comparable [Fig. 5(b)], and the target editing becomes even slightly slower for faster binding rates  $u_1$  [see Fig. 5(c)].

Different editing times suggest that the precision of the single-base discrimination process can be improved by collecting the products at different times. We can quantify the conditions under which such dynamic selectivity can lead to the most optimal performance. For this purpose, we define the mean time to obtain non-target undesired products,

$$T_{\text{non-target}} = \frac{P_{0 \rightarrow 12} T_{0 \rightarrow 12} + P_{0 \rightarrow 14} T_{0 \rightarrow 14}}{P_{0 \rightarrow 12} + P_{0 \rightarrow 14}}. \quad (27)$$

Then, the ratio  $T_{0 \rightarrow 13}/T_{\text{non-target}}$  will provide a measure of dynamic selectivity in the single-base editing process. This quantity is plotted in Fig. 5. It is clear that the most efficient single-base discrimination occurs when  $T_{0 \rightarrow 13}/T_{\text{non-target}}$  deviates significantly from unity. One can see that the best selectivity can be achieved for slow binding rates,  $u_1$  [Fig. 5(a)], and it is slightly better for the mutations with  $\Delta E_m < 0$ . However, increasing the binding rate  $u_1$  effectively eliminates the possibility of temporal separation [Fig. 5(b)].



**FIG. 5.** Ratios of mean editing times as a function of free-energy perturbation due to mutations. The following parameters have been used in calculations:<sup>21</sup>  $u_0 = 1 \text{ s}^{-1}$ ,  $u_3 = 1.1 \text{ s}^{-1}$ ,  $u_4 = 2.1 \text{ s}^{-1}$ ,  $w_0 = 2.9 \times 10^{-5} \text{ s}^{-1}$ ,  $w_1 = 12.54 \text{ s}^{-1}$ , and  $w_2 = 5059 \text{ s}^{-1}$ . (a) For unbinding rate,  $u_1 = 10^4 \text{ s}^{-1}$ . (b) For unbinding rate,  $u_1 = 10^5 \text{ s}^{-1}$ . (c) For unbinding rate,  $u_1 = 10^6 \text{ s}^{-1}$ .

However, a further increase in the rates  $u_1$  start to improve the selectivity, but not too much. One could also observe that the mean time to obtain any non-target products for most ranges of parameters is close to the mean time to edit both target and bystander nucleotides ( $T_{\text{non-target}} \approx T_{0 \rightarrow 12}$ ). This is because the editing of only the bystander nucleotide is quite fast and has a very low probability.

Another important result of our theoretical calculations is that the conditions of the highest probability and the best dynamic selectivity for the desired target single-base editing generally do not correlate. In other words, the temporal separation at the conditions of the lowered bystander effect ( $\Delta E_m \approx 5 - 10 k_B T$ ) cannot be efficiently utilized. This important observation underlies the difference between thermodynamic and kinetic controls of single-base editing processes.

### III. SUMMARY AND CONCLUSIONS

In this paper, we presented a theoretical analysis of the processes that take place during single-base editing. A specific system of cytosine BE has been considered. Using a minimal chemical-kinetic model together with the first-passage probabilities method, we explicitly evaluated the dynamic properties, precision, and selectivity of base editing. Our theoretical method allowed us to evaluate the probabilities and mean times of different editing outcomes. In addition, the role of mutations in optimizing the performance of

single-base editors is quantitatively estimated for experimentally relevant ranges of parameters. It is found that increasing the association rate of the enzymatic domain accelerates all editing processes, but it also lowers the degree of dynamic selectivity. Our calculations also show that mutations that increase the dissociation rates of enzyme domains provide the best conditions for the reduced bystander effect at which the probability of target editing is maximal. However, the best temporal separation can be achieved for mutations that lower these dissociation rates. Furthermore, we presented a detailed description of transient processes during base editing, allowing us to better understand the molecular mechanisms of underlying processes. Such information can help to rationally develop the most efficient single-base editing systems for various applications. The most important conclusion from our theoretical analysis is that there are different requirements to optimize the functioning of single-base editors from probabilistic or dynamic aspects of the process.

Although our theoretical approach has been able to obtain a comprehensive, dynamic picture of base editing processes, it is important to discuss its limitations. The utilized chemical kinetic model is rather very simplified, and many more biochemical transitions are neglected. In addition, some chemical transitions are assumed to be irreversible, while in reality, this might not be the case. However, despite these limitations, the most important advantage of our theoretical method is that it can provide specific quantitative predictions that can be experimentally tested. It will be important to



investigate more advanced models and do more experimental studies in order to clarify the microscopic picture of precise single-base discrimination.

## ACKNOWLEDGMENTS

A.B.K. acknowledges the support from the Welch Foundation (Grant No. C-1559), the NIH (Grant No. R01 HL157714-02), the NSF (Grant No. CHE-1953453), and the Center for Theoretical Biological Physics sponsored by the NSF (Grant No. PHY-2019745). Q.W. acknowledges the funding support from the National Natural Science Foundation of China (Grant No. 32000882).

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

## Author Contributions

**Vardan Hoviki Vardanyan:** Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Writing – original draft (equal). **Qian Wang:** Conceptualization (equal); Methodology (equal); Writing – review & editing (equal). **Anatoly B. Kolomeisky:** Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Project administration (equal); Resources (equal); Supervision (equal); Validation (equal); Writing – review & editing (equal).

## DATA AVAILABILITY

The data that support the findings of this study are available within the article.

## REFERENCES

- <sup>1</sup>M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” *Science* **337**, 816–821 (2012).
- <sup>2</sup>J. A. Doudna and E. Charpentier, “The new Frontier of genome engineering with CRISPR-Cas9,” *Science* **346**, 1258096 (2014).
- <sup>3</sup>G. Gasiunas, R. Barrangou, P. Horvath, and V. Siksnys, “Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria,” *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2579–E2586 (2012).
- <sup>4</sup>J. Y. Wang and J. A. Doudna, “CRISPR technology: A decade of genome editing is only the beginning,” *Science* **379**, eadd8643 (2023).
- <sup>5</sup>P. D. Hsu, E. S. Lander, and F. Zhang, “Development and applications of CRISPR-Cas9 for genome engineering,” *Cell* **157**, 1262–1278 (2014).
- <sup>6</sup>R. Strack, “Precision genome editing,” *Nat. Methods* **16**, 21 (2019).
- <sup>7</sup>H. A. Rees and D. R. Liu, “Base editing: Precision chemistry on the genome and transcriptome of living cells,” *Nat. Rev. Genet.* **19**, 770–788 (2018).
- <sup>8</sup>A. V. Anzalone, L. W. Koblan, and D. R. Liu, “Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors,” *Nat. Biotechnol.* **38**, 824–844 (2020).
- <sup>9</sup>P. J. Chen and D. R. Liu, “Prime editing for precise and highly versatile genome manipulation,” *Nat. Rev. Genet.* **24**, 161 (2022).
- <sup>10</sup>B. A. Adler, T. Hessler, B. F. Cress, A. Lahiri, V. K. Mutalik, R. Barrangou, J. Banfield, and J. A. Doudna, “Broad-spectrum CRISPR-Cas13a enables efficient phage genome editing,” *Nat. Microbiol.* **7**, 1967 (2022).
- <sup>11</sup>N. M. Gaudelli, A. C. Komor, H. A. Rees, M. S. Packer, A. H. Badran, D. I. Bryson, and D. R. Liu, “Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage,” *Nature* **551**, 464–471 (2017).
- <sup>12</sup>A. C. Komor, A. H. Badran, and D. R. Liu, “CRISPR-based technologies for the manipulation of eukaryotic genomes,” *Cell* **168**, 20–36 (2017).
- <sup>13</sup>K. Nishida, T. Arazoe, N. Yachie, S. Banno, M. Kakimoto, M. Tabata, M. Mochizuki, A. Miyabe, M. Araki, K. Y. Hara *et al.*, “Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems,” *Science* **353**, aaf8729 (2016).
- <sup>14</sup>S. Lee, N. Ding, Y. Sun, T. Yuan, J. Li, Q. Yuan, L. Liu, J. Yang, Q. Wang, A. B. Kolomeisky *et al.*, “Single C-to-T substitution using engineered APOBEC3G–nCas9 base editors with minimum genome- and transcriptome-wide off-target effects,” *Sci. Adv.* **6**, eaba1773 (2020).
- <sup>15</sup>E. M. Porto, A. C. Komor, I. M. Slaymaker, and G. W. Yeo, “Base editing: Advances and therapeutic opportunities,” *Nat. Rev. Drug Discovery* **19**, 839–859 (2020).
- <sup>16</sup>A. Lapinaite, G. J. Knott, C. M. Palumbo, E. Lin-Shiao, M. F. Richter, K. T. Zhao, P. A. Beal, D. R. Liu, and J. A. Doudna, “DNA capture by a CRISPR-Cas9-guided adenine base editor,” *Science* **369**, 566–571 (2020).
- <sup>17</sup>A. C. Komor, K. T. Zhao, M. S. Packer, N. M. Gaudelli, A. L. Waterbury, L. W. Koblan, Y. B. Kim, A. H. Badran, and D. R. Liu, “Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity,” *Sci. Adv.* **3**, eaao4774 (2017).
- <sup>18</sup>L. W. Koblan, J. L. Doman, C. Wilson, J. M. Levy, T. Tay, G. A. Newby, J. P. Maiani, A. Raguram, and D. R. Liu, “Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction,” *Nat. Biotechnol.* **36**, 843–846 (2018).
- <sup>19</sup>Y. B. Kim, A. C. Komor, J. M. Levy, M. S. Packer, K. T. Zhao, and D. R. Liu, “Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions,” *Nat. Biotechnol.* **35**, 371–376 (2017).
- <sup>20</sup>J. M. Gehrke, O. Cervantes, M. K. Clement, Y. Wu, J. Zeng, D. E. Bauer, L. Pinello, and J. K. Joung, “An APOBEC3A–Cas9 base editor with minimized bystander and off-target activities,” *Nat. Biotechnol.* **36**, 977–982 (2018).
- <sup>21</sup>Q. Wang, J. Yang, Z. Zhong, J. A. Vanegas, X. Gao, and A. B. Kolomeisky, “A general theoretical framework to design base editors with reduced bystander effects,” *Nat. Commun.* **12**, 6529 (2021).
- <sup>22</sup>I.-J. L. Byeon, J. Ahn, M. Mitra, C.-H. Byeon, K. Hercík, J. Hritz, L. M. Charlton, J. G. Levin, and A. M. Gronenborn, “NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity,” *Nat. Commun.* **4**, 1890 (2013).
- <sup>23</sup>N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, 2007).
- <sup>24</sup>A. B. Kolomeisky, *Motor Proteins and Molecular Motors* (CRC Press, 2015).
- <sup>25</sup>T. Chou and M. R. D’Orsogna, “First passage problems in biology,” in *First-Passage Phenomena and Their Applications* (World Scientific, 2014), pp. 306–345.
- <sup>26</sup>A. R. Poggioli, A. Siria, and L. Bocquet, “Beyond the tradeoff: Dynamic selectivity in ionic transport and current rectification,” *J. Phys. Chem. B* **123**, 1171–1185 (2019).
- <sup>27</sup>A. A. Shvets and A. B. Kolomeisky, “Mechanism of genome interrogation: How CRISPR RNA-guided Cas9 proteins locate specific targets on DNA,” *Biophys. J.* **113**, 1416–1424 (2017).
- <sup>28</sup>Q. Wang, M. R. Diehl, B. Jana, M. S. Cheung, A. B. Kolomeisky, and J. N. Onuchic, “Molecular origin of the weak susceptibility of kinesin velocity to loads and its relation to the collective behavior of kinesins,” *Proc. Natl. Acad. Sci. U. S. A.* **114**, E8611–E8617 (2017).
- <sup>29</sup>S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, 2001).
- <sup>30</sup>D. E. Makarov, *Single Molecule Science: Physical Principles and Models* (CRC Press, 2015).