THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

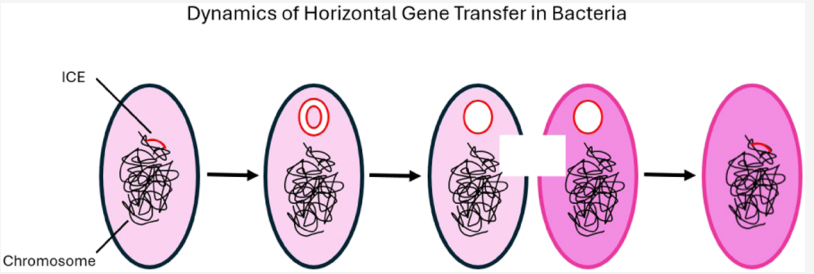
Article

Theoretical Understanding of Target Search Dynamics in Horizontal Gene Transfer in Bacteria

Natalie Collins, Yaakov Levy, and Anatoly B. Kolomeisky*

Cite This: J. Phys. Chem. B 2025, 129, 6828−6836

Read Online



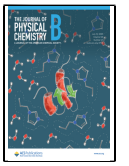
ABSTRACT: Horizontal gene transfer (HGT) is a fundamental process of increasing genetic diversity in microbial species. It allows bacterial cells to acquire new beneficial traits quickly by incorporating new genetic material into existing genomes. Despite the critical importance of HGT phenomena, the underlying molecular mechanisms are still poorly understood. Recent experiments investigated the dynamics of conjugation HGT processes in which DNA is transmitted directly from the donor to the recipient bacterial cell. It is accomplished by special mobile genetic particles known as integrative and conjugative elements (ICE). However, the molecular picture of how ICE can efficiently find the unique integration sites in a new genome is not yet clear. We present a novel theoretical model to explain the dynamic processes in HGT after ICE reaches the recipient cell. It is shown that the target search for integration sites can be viewed as a set of stochastic transitions between discrete states, allowing us to obtain an explicit description of the dynamic properties using analytical calculations supported by Monte Carlo computer simulations. Search times are found to depend on the location of integration sites, the size of the genome, the effective diffusion rate of mobile genetic elements, and the binding/unbinding transitions between ICE and DNA. Theoretical estimates for search times agree well with experimental observations for integration in *Bacillus subtilis* bacterial species. Physical-chemical arguments are presented to explain the dynamics of the ICE target search. This study clarifies some important mechanistic aspects of HGT phenomena.

INTRODUCTION

Horizontal gene transfer (HGT) is a unique biological process by which genetic information is shared between organisms that are not related as parent and offspring.[1,2] It has been critically important for the evolution of all living systems,[1−3] allowing for the exchange of genetic material between different cells that leads to genome modifications.[4−7] This is a dynamic phenomenon that might have immediate or delayed effects in the recipient host organism.[8,9] Although HGT events occur between cells in different kingdoms of life, they are dominating in bacterial species because of the lack of sexual methods of increasing genetic diversity in bacteria.[8,10,11] Furthermore, it is now well established that HGT events are crucial in minimizing bacterial genome damage and for acquiring critically important antibiotic resistance genes.[3,8]

Experimental studies show that there are multiple pathways for HGT events in bacteria.[2,8,11] However, one of the most common mechanisms is conjugation, which is the contact-dependent unidirectional transfer of DNA from a donor to a recipient cell via conjugating (mating) apparatus expressed in the donor cell.[11,12] It is accomplished by mobile protein-nucleic acid complexes called integrative and conjugative elements, also known as conjugative transposons.[12,13] integrative and conjugative elements (ICEs) already contain all the genes needed for their activities, namely, for the excision from the host genome, transferring between the cells, and for integration in the recipient genome. The main stages of the life cycle of ICE are shown schematically in Figure 1. Under normal circumstances, ICEs are integrated into the donor genome, but the conjugation genes are silent. When the potential host cell is nearby, these genes are somehow induced, and ICE is excised from the original bacterial chromosome, forming a separate circular DNA molecule (plasmid). There are other products of conjugative gene activation, such as the

Received: April 9, 2025
Revised: June 25, 2025
Accepted: June 26, 2025
Published: July 1, 2025



https://doi.org/10.1021/acs.jpcb.5c02436
J. Phys. Chem. B 2025, 129, 6828−6836

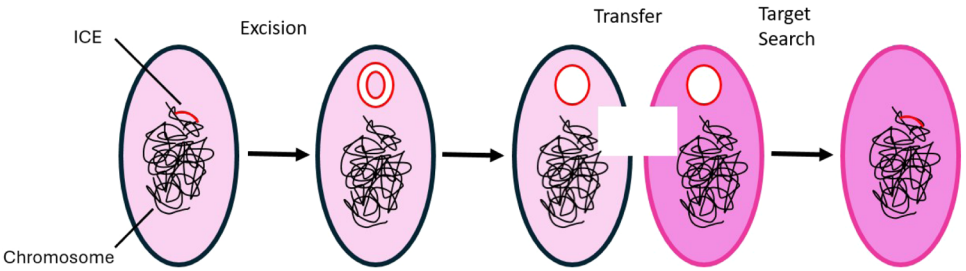


Figure 1. Schematic view of conjugative horizontal gene transfer. The process starts with ICE being incorporated in the donor cell genome. In the next step, after some external signal that a recipient cell is close, ICE is excised from DNA while still in the donor cell. Next, the donor and recipient cells are connected by the conjugating (mating) channel, allowing for ICE transfer. In the last stage, the ICE complex is searching for integration sites in the new genome.

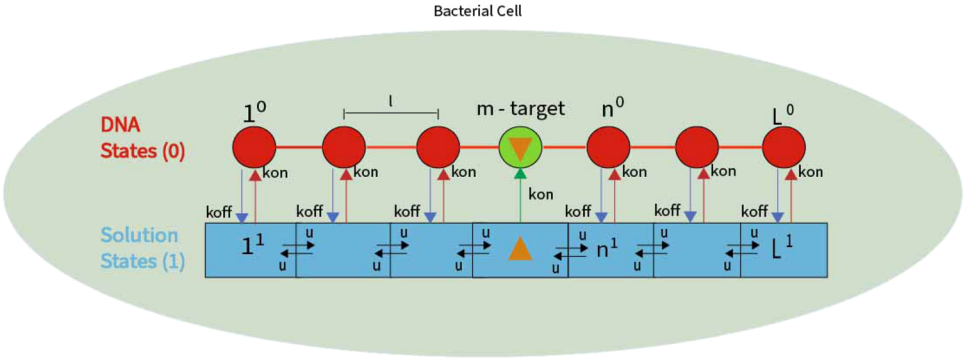


Figure 2. Discrete-state stochastic model for the search of integrations sites by ICEBs1. There are $L - 1$ nonspecific sites and one specific integration site at the position m on the DNA chain. The ICE particle can diffuse in the solution parallel to the DNA chain in with the rate u, or it can associate onto the DNA with the rate k_{on}. If attached to the DNA, ICEBs1 can dissociate back into the solution with the rate k_{off}. The states in the solution are labeled as $i = (1)$, and the states on DNA are labeled as $i = (0)$. Each segment in solution or on DNA is labeled with $n = 1, 2, \cdots, m, \cdots, L$.

assembly of a mating pore between the host cell and the recipient cell, tightly coupling them together for some time: see Figure 1. Then a DNA segment in ICE is transformed into a ssDNA−protein complex which is translocated by a mating pump from the host cell into the recipient cell. Finally, the ICE segment integrates into the new bacterial chromosome at a specific location(s), called the integration site. At the same time, the remaining strand of ICE in the host cell is incorporated back into the original chromosome. This unique set of events is robust, but the underlying microscopic picture remains poorly understood.[12]

Recent experimental studies focused on the last stage of HGT conjugation when the ICE is searching for an integration site.[11,14,15] By following the motion of the ICE and the chromosomal integration site, the dynamics of ICE integration into the recipient genome was investigated in the *Bacillus subtilis* bacterial species.[11,14,16] In these experiments, ICEBs1 species (∼20 kbp) were utilized which are known to enhance the host cell's ability to out-compete neighboring cells by inhibiting sporulation and biofilm formation, keeping the cell in an active growth and division phase. This prolonged growth increases opportunities for HGT to occur, allowing ICEBs1 to spread more effectively.[12,17] It should also be noted that in the recipient cell, ICEBs1 preferentially integrates near the t-RNA (trnS-leu2) gene into the site denoted attB (there are two such sites per genome), which is a highly conserved genomic region that provides both stability and relative protection against the hosts defense mechanisms.[18] In addition, it was found that successful genetic transfer typically occurs in about ∼1 h, and both ICE and the integration site in the recipient genome exhibited "sub-diffusive" behavior. This study highlights the important role of cellular crowding and chromosomal structure for HGT events. However, the molecular mechanisms of efficient target search for integration sites remain unclear.[11]

In this paper, we present a novel theoretical approach to investigate the last stage of the HGT process when ICE is searching for integration sites. The dynamics of protein target search has been intensively studied for various biological systems, and multiple theoretical methods have been proposed.[19−23] Our theoretical model is based on a discrete-state stochastic analysis of dynamic phenomena that explores the analogy of integration site target search with the homology recombination of RecA proteins, which search for sites of DNA damage.[19,24−26] This method allows us to explicitly evaluate the dynamic properties of the system, quantifying the ICE target search for integration sites in HGT. We identified that the main features that influence the search dynamics are the location of integration sites, size of the genome, nonspecific affinities between mobile genetic elements and DNA, and ICE diffusion rates in the volume around the DNA segment. Our analytical calculations are supplemented by Monte Carlo computer simulations, and our estimates of target search times are consistent with experimental observations, clarifying important dynamic aspects of HGT phenomena.

MATERIALS AND METHODS

Discrete-State Stochastic Model. To investigate the dynamics of ICE target search for integration sites (the last stage in the HGT process as shown in Figure 1), we propose a

https://doi.org/10.1021/acs.jpcb.5c02436
J. Phys. Chem. B 2025, 129, 6828−6836

discrete-state stochastic model presented in Figure 2. Our goal is to mimic the search of ICEBs1 particles for attB integration sites in the *B. subtilis* bacterial genome.[11,14] The bacterial genome has a size of 4.2 Mbp, and its projected length in the bacterial cell is ∼3 μM.[14,27] There are two integration sites that are located approximately 1/4 and 3/4 of vertical distance along the genome.[14] Because of this symmetry, it is reasonable to consider a DNA strand of half the length, ∼1.5 μM, with a single target in the middle of the DNA chain: see Figure 2. Although the bacterial chromosome is circular, experiments considered a projected view of the system,[14] such that in our theoretical model the DNA molecule can be approximated as a one-dimensional (1D) line of L segments each of length l, where $l \sim 100$ bp approximately corresponds to the size of the attB integration site.[28,29] This means that $L \simeq 15$, and the target is located in the middle of the DNA chain at $m \simeq L/2$. Additional arguments in support of our theoretical picture come from experiments exhibiting that the bacterial genome has a polymer bottle-brush structure, with the distance between neighboring branches being of the order of several hundred bps.[11,30] This distance in our theoretical picture corresponds to the distance l, since ICE can go between DNA genome branches, trying to find the integration site.

In our model, we assume that the ICE particle also has a size l and can be found in one of two possible sets of stochastic states: see Figure 2. The ICE particle can diffuse in the solution "parallel" to DNA with the diffusion rate u. The solution is divided into L volume segments, each of them surrounding the corresponding site on DNA: see Figure 2. ICE can also attach to DNA with the rate k_{on} to check if the site is the right target. If this is not the case, ICE dissociates back into the solution with the rate k_{off}. There are a total of $2L$ discrete states in the system (L states in the solution and L states when ICE is bound to DNA). ICE starts in the solution, and the process is completed when the target at site m is located for the first time. A single-molecule view of the target search dynamics is adopted, which is consistent with experimental setups where a single ICE species participated in integration.[14] It is important to note here that our theoretical method views the integration target search by ICE particles as an effective 1D process, which is stimulated by experimental geometry and the corresponding analysis in recent studies on *B. subtilis*.[14]

To analyze integration target search during HGT events, we notice that this process is similar to some degree to the homology search (by RecA proteins in bacteria or Rad51 proteins in eukaryotes), which is activated when a double-strand break in DNA appears.[31−34] The mechanisms of homology search have been investigated using a variety of theoretical tools, and it is reasonably well understood now.[26,35−37]

For the search of the integration site, it is convenient to adopt a first-passage method of calculating dynamic properties that connects diffusion and association/dissociation rates with the target search dynamics.[26] For this purpose, we introduce functions $F_n^{(0)}(t)$ and $F_n^{(1)}(t)$ defined as probability densities to reach the target (integration site) at time t for the first time if at $t = 0$ the system starts in the state $n^{(0)}$ (in the solution) or in the state $n^{(1)}$ (on DNA), respectively: See Figure 2. These probabilities evolve with time as described by a set of backward master equations[26]

$$\frac{dF_n^{(0)}(t)}{dt} = k_{off}F_n^{(1)}(t) - k_{off}F_n^{(0)}(t) \tag{1}$$

and

$$\frac{dF_n^{(1)}(t)}{dt} = u[F_{n+1}^{(1)}(t) + F_{n-1}^{(1)}(t)] + k_{on}F_n^{(0)}(t) - (2u + k_{on})F_n^{(1)}(t) \tag{2}$$

In addition, one should impose the initial condition[26]

$$F_m^{(0)}(t) = \delta(t) \tag{3}$$

which means that if the system starts with ICE on the integration site the target search process is immediately accomplished.

The set of backward master equations can be solved explicitly using the Laplace transformation approach, as explained in detail in the Supporting Information. More specifically, we introduce Laplace transforms of the first-passage probability functions

$$\widetilde{F_n}^{(0)}(s) = \int_0^\infty e^{-st}F_n^{(0)}(t)dt; \quad \widetilde{F_n}^{(1)}(s) = \int_0^\infty e^{-st}F_n^{(1)}(t)dt \tag{4}$$

and calculate them exactly: see the Supporting Information. Then, all of the dynamic properties of the integration target search can be obtained. Because we are mimicking experimental conditions,[14] we are interested in evaluating the mean search time to locate the integration site starting with equal probability anywhere in the solution

$$T = \frac{1}{L}\sum_{n=1}^{L}\tau_n^{(1)} \tag{5}$$

where

$$\tau_n^{(1)} = -\frac{d\widetilde{F_n}^{(1)}(s)}{ds}(s = 0) \tag{6}$$

is the mean first-passage time to locate the integration site if starting from the specific solution state $n^{(1)}$. As shown in the Supporting Information, the exact expression for the mean search time is given by

$$T = \frac{W}{6u\left(\frac{k_{off}}{k_{on}+k_{off}}\right)} + \left(\frac{L-1}{k_{off}} + \frac{L}{k_{on}}\right) \tag{7}$$

where the auxiliary function W is defined as

$$W = 1 + 3L + 2L^2 - 6m - 6Lm + 6m^2 \tag{8}$$

The expression for the mean target search time [eq 7] has a clear physical interpretation. The first term represents the time during which ICE is diffusing in solution and is not bound to DNA. The second term accounts for the total time the ICE spends trying to check if the given site on DNA is the integration site or not. It is done by sequentially associating and dissociating from the DNA chain. On average, there are L such dissociations and $L - 1$ such association transitions. The number of association events is smaller by one because the last association to the integration site successfully ends the target search process, and there is no need to dissociate.

To better understand the molecular picture for the search of the integration site, it is convenient to look at two limiting situations (assuming that the target is in the middle, $m \simeq L/2$). First, if the DNA genome is very long ($L \gg 1$) and the

https://doi.org/10.1021/acs.jpcb.5c02436
J. Phys. Chem. B 2025, 129, 6828−6836

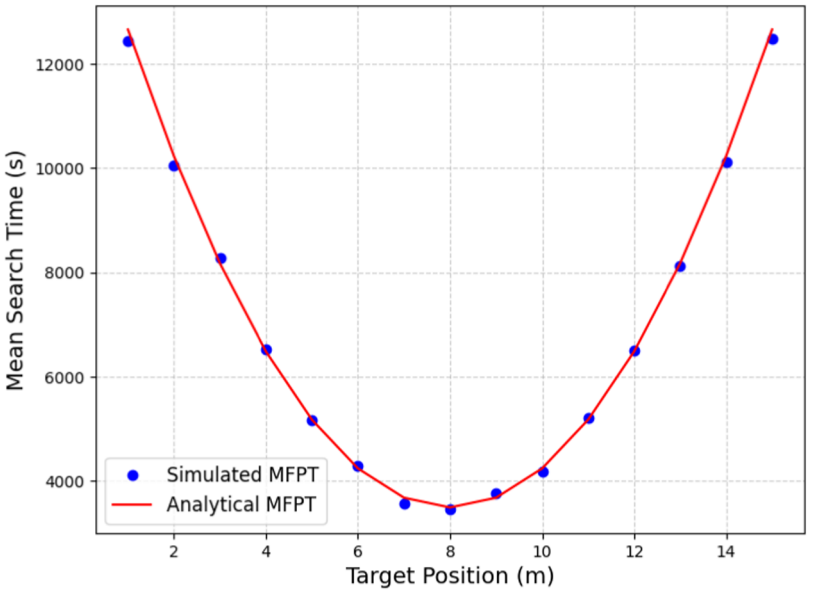


Figure 3. Mean search times to find the integration site as a function of the target position along the DNA segment. The red solid curve is from analytical predictions, and the blue symbols are Monte Carlo computer simulations. The following parameters have been used for calculations are $L = 15$, $k_{on} = 10^3\ s^{-1}$, $k_{off} = 10\ s^{-1}$, and $u = 0.54\ s^{-1}$.

association rate is much faster than the dissociation rate ($k_{on} \gg k_{off}$), the mean search time behaves as

$$T \simeq \frac{L^2 K}{12u} \tag{9}$$

where $K = \frac{k_{on}}{k_{off}}$ is a nonspecific affinity of the ICE to the DNA chain. One can see that in this case, diffusion in the solution is a rate-limiting step. This situation is probably close to experimental conditions in *B. subtilis*.[14]

Another limit is when the diffusion rate in the solution is very fast ($u \gg 1$) and the length of the DNA genome is not too long. In this case, it can be shown that

$$T \simeq \frac{L-1}{k_{off}} + \frac{L}{k_{on}} \tag{10}$$

Here, the rate-limiting steps are DNA association and dissociation events, and the location of the target does not matter. This situation is probably not very realistic for *B. subtilis*.[14]

RESULTS AND DISCUSSION

For our theoretical approach, it is important first to estimate the kinetic parameters from the experimental data on integration sites search in *B. subtilis*.[14] In these experiments, ICEBs1 particles were observed to move "sub-diffusively" in the vertical direction, with mean-squared displacement scaling as

$$r^2 \sim D\tau^{\alpha} \tag{11}$$

with $D = 0.0077\ \mu m^2\ s^{1/\alpha}$ and $\alpha \simeq 0.43$. While this behavior likely arises due to finite-size or confinement effects and probably does not reflect true long-time anomalous diffusion, we use this relationship as a practical means to estimate the characteristic time to move a vertical step of size $r = l = 0.1\ \mu m$ between neighboring vertical volume segments. Equation 11 suggests that it will require a $\tau \simeq 2$ s. This allows us then to estimate the effective diffusion rate in our model, $u = 1/\tau$, yielding $u \simeq 0.5\ s^{-1}$. This approach suggests that we have effectively mapped our model to a normal 1D diffusion process: see Figure 2.

The binding/unbinding rates for ICE to attach and detach from DNA are not known experimentally. However, using the similarity of the integration site search with homology search, one might roughly estimate that $k_{on} \sim 10-1000\ s^{-1}$ and $k_{off} \sim 1-100\ s^{-1}$, suggesting that the nonspecific affinity $\left(K = \frac{k_{on}}{k_{off}}\right)$ is $K \sim 10-100$. The affinity cannot be too high because ICE would associate to nontarget sites on DNA too frequently, significantly slowing the target search. It also cannot be too low as ICE would infrequently bind to the DNA, preventing it from efficient scanning of DNA sites. One can also evaluate the length of the DNA segment per one integration site as $L = 15$ and $m = 8$, as was already explained above.

Substituting the kinetic parameters estimated from the experimental data into eq 7, one can calculate the mean search times T. We predict then [from eq 7] that $T \sim 400-4000$ s, i.e., finding the integration sites might take from 10 minutes up to 1 hour, and these estimates are in very good agreement with experimentally determined values.[14] These calculations suggest that our theoretical method likely captures the main aspects of the complex target search process in conjugative HGT events.

Our theoretical model allows us to evaluate the role of different factors during the ICE target search. First, let us understand the effect of varying the position of the integration site. The results of our analytical calculations supplemented by Monte Carlo computer simulations are presented in Figure 3. One can see that the location of the integration site does affect the dynamics of the target search. The fastest search times are observed when the integration site is in the middle of the DNA segment ($m \simeq L/2$), and varying the position might accelerate the search up to 4 times for parameters that are relevant for *B. subtilis*.

To explain these observations, we note that before ICEBs1 can integrate at its target site, it must first reach the volume segment surrounding the target. Since the initial position of

https://doi.org/10.1021/acs.jpcb.5c02436
J. Phys. Chem. B 2025, 129, 6828−6836

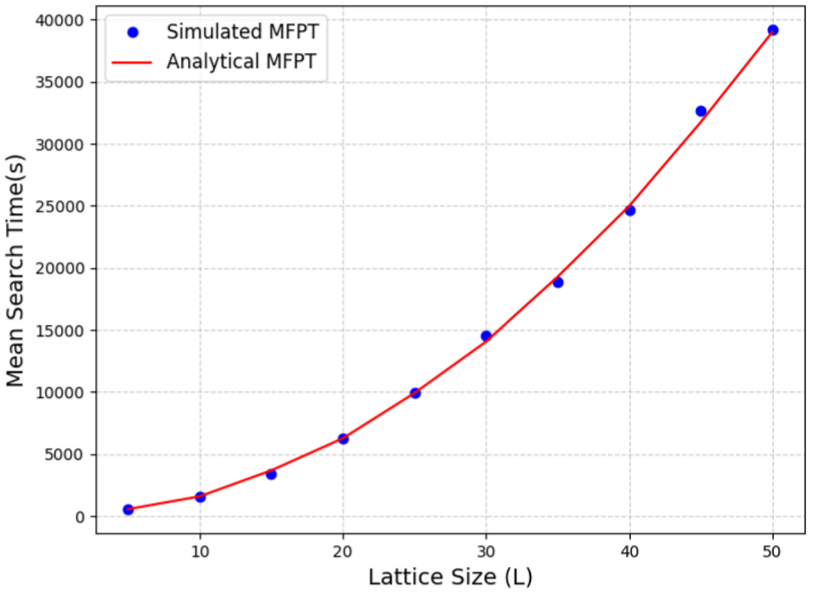


Figure 4. Mean search times as a function of the genome length L for the target located in the middle of the DNA segment ($m \simeq L/2$). The red solid curve is from theoretical predictions, and the blue symbols are from Monte Carlo computer simulations. The following parameters have been used for calculations: $k_{on} = 10^3$ s^{-1} and $k_{off} = 10$ s^{-1}, and $u = 0.54$ s^{-1}.

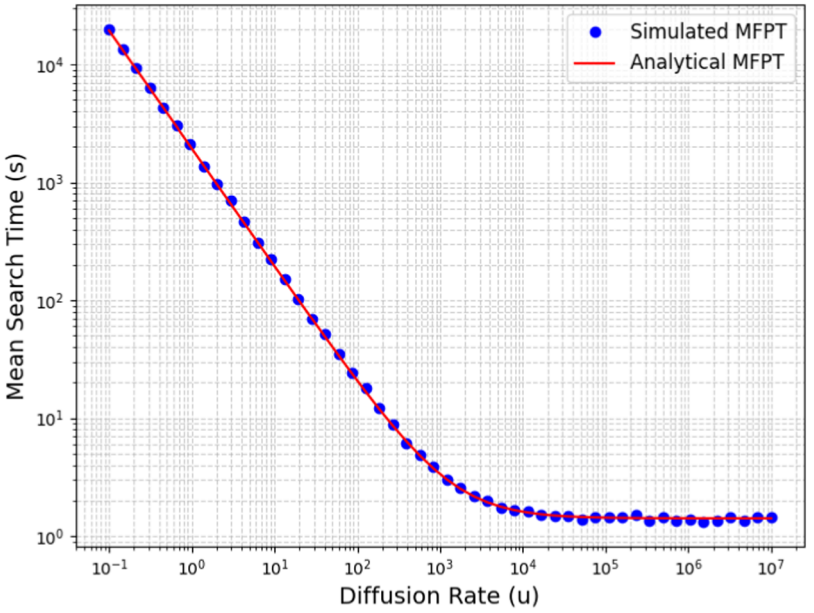


Figure 5. Mean search times to find the attachment site as a function of the diffusion rate u. The red solid curve is from analytical calculations, and the blue symbols are from Monte Carlo computer simulations. The following parameters have been used for calculations: $L = 15$, $m = 8$, $k_{on} = 10^3$ s^{-1} and $k_{off} = 10$ s^{-1}.

ICEBs1 in solution is equally likely to be anywhere along the DNA chain, the search process is most efficient when the target is centrally located. This is because, on average, ICEBs1 must diffuse approximately $L/4$ volume segments to reach a region next to the middle of the chain ($m \sim L/2$) since it can start with equal probability anywhere along the DNA segment. Whereas for the target at the end ($m \sim L$), it must traverse roughly $L/2$ volume segments. Since the diffusion in the solution is most probably a limiting factor at typical conditions in *B. subtilis*, one could estimate the mean search time as $T(m \simeq L/2) \sim (L/4)^2$ for the target in the middle, and $T(m \simeq L) \sim (L/2)^2$ for the target at the end, explaining the 4-fold acceleration in finding the target. Interestingly, the attB integration site in *B. subtilis* is located at the position at which the fastest search is observed. To be more precise, two integration sites are located at 1/4 and 3/4 of the genome length, which, due to symmetry, correspond to the middle of the DNA segment in our theoretical model. These arguments suggest that evolution may have optimized the position of targets so that the search dynamics for integration sites are the fastest possible. This is supported by observations that it is difficult to find the integration site due to molecular crowding in the cell, and any means to accelerate the search process are beneficial for the success of HGT.

One can also understand how the target search dynamics depend on the size of the genome. The results of our

https://doi.org/10.1021/acs.jpcb.5c02436
J. Phys. Chem. B 2025, 129, 6828−6836

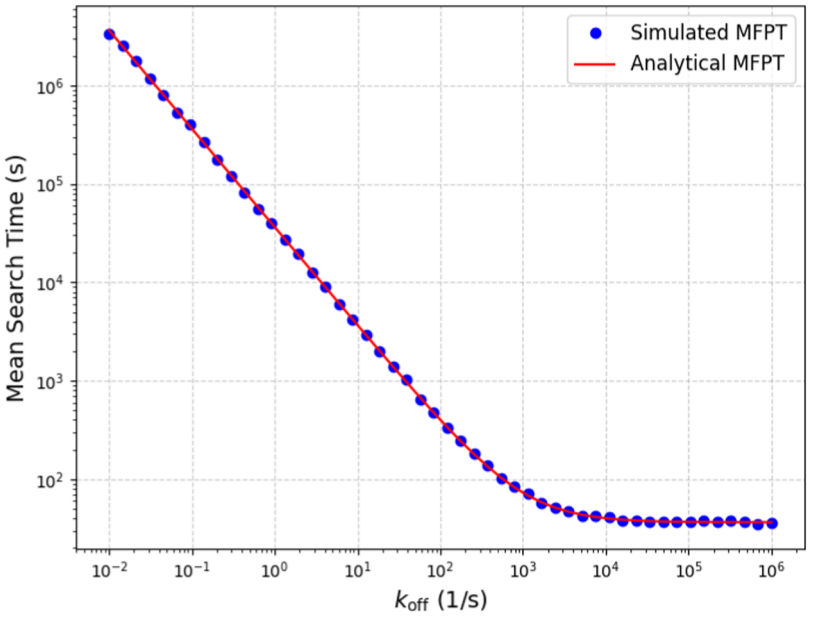


Figure 6. Mean search times to find the attachment site as a function of the dissociation rate k_{off}. Red solid curves are from analytical calculations, and blue symbols are from Monte Carlo computer simulations. The following parameters have been used for calculations: $L = 15$, $m = 8$, $k_{on} = 10^3$ s^{-1} and $u = 0.54$ s^{-1}.

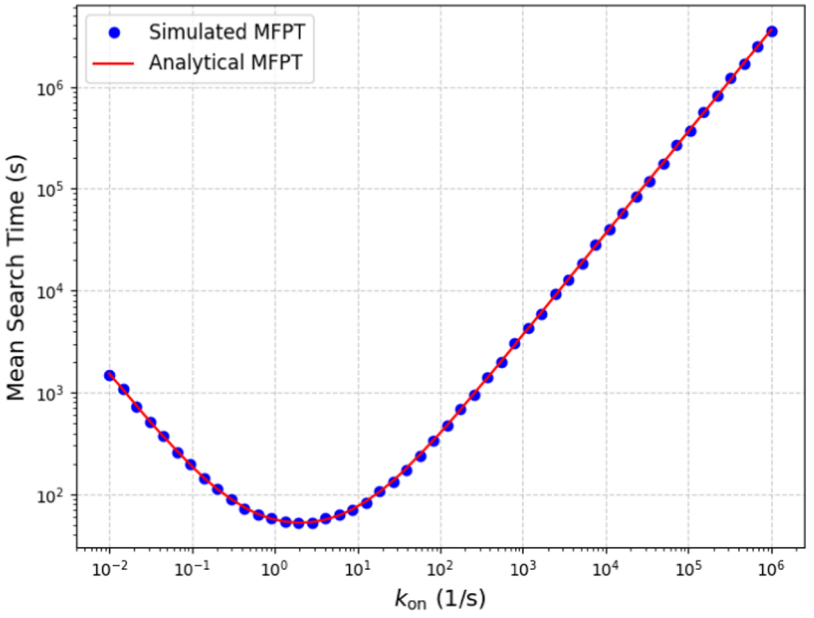


Figure 7. Mean search times to find the attachment site as a function of the association rate k_{on}. The red solid curve is from analytical calculations, and the blue symbols are from Monte Carlo computer simulations. The following parameters have been used for calculations: $L = 15$, $m = 8$, $k_{off} = 10$ s^{-1} and $u = 0.54$ s^{-1}.

theoretical analysis are presented in Figure 4. One can see that the search times grow mostly quadratically as a function of L. This is because the slowest process in the system is diffusion parallel to the DNA segment, which scales quadratically with the distance. Thus, it will take much longer for HGT to occur for longer DNA chains ($T \sim L^2$), if the same mechanism as considered here is taking place in other living systems. Since in bacterial species the size of DNA genomes is relatively short, it seems that horizontal gene transfer might be an efficient process of getting new genetic material in microbes, potentially explaining why HGT dominates in bacterial systems in comparison with organisms from other kingdoms of life that have much longer genomes.[3,8]

Next, let us investigate the role of the effective diffusion rate in the target search dynamics. The results are presented in Figure 5. As expected, increasing the rate u first lowers the search time, but for very fast diffusion rates the acceleration effect disappears and the search times reach a plateau: See Figure 5. The observations can be explained using the following arguments. The relationship between diffusion rate and search time highlights the importance of the molecular mobility of ICE particles. A higher diffusion rate allows ICEBs1 to transition between solution and DNA-bound states more

https://doi.org/10.1021/acs.jpcb.5c02436
J. Phys. Chem. B 2025, 129, 6828−6836

rapidly, thereby accelerating the search process. However, at sufficiently high u, the system enters a regime where search time is no longer limited by diffusion, but rather by the kinetics of DNA binding and unbinding. A further increase in the diffusion rate u will not affect the target search dynamics. This explains the saturation for high u: see Figure 5.

The dependence of the integration site target search on binding and unbinding rates between ICE and the DNA chain is analyzed in Figures 6 and 7. As one can see, increasing the rate k_{off} accelerates the search dynamics (Figure 6) because unbinding prevents ICE particles from being trapped at nonintegration sites. However, further increase in the dissociation rate will have no effect on the target search dynamics since other transitions (association or diffusion) will become rate-limiting: see Figure 6. Our estimates of dissociation rates ($k_{off} \sim 1-100\ s^{-1}$) also suggest that the integration search dynamics is probably already operating close to the most efficient conditions (see Figure 6).

More interesting dynamic behavior is found when we vary the association rate, as shown in Figure 7. A nonmonotonic dependence of the mean search times as a function of k_{on} is observed. This can be explained by noting that increasing the association rate first makes the ICE particle more mobile, allowing it to explore more regions on DNA due to lowering the time in the volume around the DNA segment. However, when the rate k_{on} is too fast, ICE will be trapped more frequently at DNA sites that are not the final target, which would slow down the search dynamics. As a result, there is a minimum in the mean search times for integration sites as shown in Figure 7. In addition, from our estimates of the association rates ($k_{on} \sim 10-1000\ s^{-1}$), one might conclude that the search dynamics is also operating close to the most optimal conditions.

SUMMARY AND CONCLUSIONS

In this paper, we developed a novel theoretical method for analyzing the dynamic aspects of HGT phenomena in bacteria. More specifically, we concentrated on the last stage of HGT when the ICE particle, which delivers the novel genetic material, is searching for integration sites in the genome of the recipient cell. The proposed theoretical approach views the target search as a set of stochastic transitions between different states of the system, specified by the location and association with DNA by the ICE particle. This discrete-state stochastic framework provides a comprehensive description of all dynamic processes during the search for integration sites, allowing for analytical calculations supported by Monte Carlo computer simulations and providing a clearer microscopic picture of HGT integration processes.

It is found that theoretical predictions for the mean search times agree well with experimental observations for *B. subtilis*. Theoretical analysis indicates that the target search dynamics is strongly influenced by the location of the integration sites, the size of the DNA genome, the diffusion rate in the volume parallel to the DNA segment, and the kinetic rates of association to DNA and dissociation from DNA by ICE particles. Physical-chemical arguments to explain these quantitative trends are presented. It is also suggested that the ICE search in *B. subtilis* probably operates close to the most efficient conditions that support the fastest dynamics, explaining the high success of HGT phenomena in these bacterial species.

It is important to discuss why our theoretical model utilizes a simplified effective 1D description, while it is known that target search phenomena employ the facilitated-diffusion mechanisms that combine 1D and 3D searching modes.[19,20] There are several arguments to explain it. First, there is very little information about ICEBs1 complexes, and their shapes, sizes, and possible association with protein molecules remain unknown. Since there are some similarities between the search for the integration site by ICE and the RecA homology search that does not exhibit sliding, we assumed that there is also no sliding in our system. Second, the experimental setups for the ICEBs1 target search were performed in an effective 1D setup, with the vast majority of movements parallel to the DNA strand.[14] While ICEBs1 might also utilize 1D sliding or jumping along DNA, there is currently no experimental evidence to support this. Furthermore, our simplified model that incorporates 1D diffusion and occasional sampling of DNA seems to predict search times reasonably well as compared with experimental observations, suggesting that these more detailed considerations are probably already partially accounted for in our theoretical method.

While the proposed theoretical framework offers a straightforward and physically intuitive approach to analyzing the dynamics of HGT events, which is also consistent with available experimental data, it is essential to discuss its limitations. The weakest point of our theoretical method is the assumption of normal diffusion by ICE particles, as experimental observations clearly show the strongly "sub-diffusive" dynamics of both mobile genetic elements and integration sites due to large molecular crowding.[14] While we mapped the "sub-diffusive" dynamics into effectively normal diffusion, allowing us to obtain reasonable estimates of target search dynamics, it remains unclear what important aspects of these processes are not accounted for in our theoretical method. Additionally, our theoretical approach views DNA as a fixed linear object, not accounting for the complex structure (polymer bottle-brush) and conformational changes that DNA undergoes in in vivo conditions. Also, the integration process, when ICE has already found the right target, might consist of several steps, but we effectively modeled it as a single-step process. Furthermore, while our model assumes nonspecific DNA bindings, it does not account for sequence-specific interactions between ICE particles and DNA. Specific DNA sequences might influence the search dynamics by affecting association/dissociation dynamics, which will modify the underlying effective free-energy landscapes. However, despite these limitations, the proposed theoretical method provides a convenient quantitative tool for analyzing HGT phenomena that should help to clarify various aspects of these complex biological phenomena. Importantly, it also gives quantitative predictions that can be tested in experimental studies, and this should stimulate more studies of these critical biological processes.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.5c02436.

Complete derivation of the backward master equation solution for the model (PDF)

https://doi.org/10.1021/acs.jpcb.5c02436
J. Phys. Chem. B 2025, 129, 6828−6836

AUTHOR INFORMATION

Corresponding Author

Anatoly B. Kolomeisky − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States; Department of Chemistry, Rice University, Houston, Texas 77005, United States; Department of Chemical and Biomolecular Engineering and Department of Physics and Astronomy, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0001-5677-6690; Email: tolya@rice.edu

Authors

Natalie Collins − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States; Department of Chemistry, Rice University, Houston, Texas 77005, United States*

Yaakov Levy − *Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel;* orcid.org/0000-0002-9929-973X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.5c02436

Author Contributions

Y.L. and A.B.K. designed research; A.B.K. and N.C. performed research; N.C. and A.B.K. analyzed data; and N.C. and A.B.K. wrote the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

A.B.K. was supported by the Welch Foundation (C-1559), by the NIH (R01GM148537), and by the Center for Theoretical Biological Physics sponsored by the NSF(PHY-2019745). The authors thank J. Stavans and R. Arbel-Goren for useful comments and suggestions. A.B.K. also thanks to Rosi and Max Varon Visiting Professorship at the Weizmann Institute for support.

REFERENCES

(1) Koonin, E. V. Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000Research* **2016**, *5*, No. 1805.

(2) Soucy, S. M.; Huang, J.; Gogarten, J. P. Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* **2015**, *16*, 472−482.

(3) Arnold, B. J.; Huang, I.-T.; Hanage, W. P. Horizontal gene transfer and adaptive evolution in bacteria. *Nat. Rev. Microbiol.* **2022**, *20*, 206−218.

(4) Ochman, H.; Lawrence, J. G.; Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **2000**, *405*, 299−304.

(5) Frost, L. S.; Leplae, R.; Summers, A. O.; Toussaint, A. Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* **2005**, *3*, 722−732.

(6) Thomas, C. M.; Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **2005**, *3*, 711−721.

(7) Wozniak, R. A. F.; Waldor, M. K. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* **2010**, *8*, 552−563.

(8) Emamalipour, M.; Seidi, K.; Vahed, S. Z.; Jahanban-Esfahlan, A.; Jaymand, M.; Majdi, H.; Amoozgar, Z.; Chitkushev, L.; Javaheri, T.; Jahanban-Esfahlan, R.; Zare, P. Horizontal gene transfer: from evolutionary flexibility to disease progression. *Fron. Cell Dev. Biol.* **2020**, *8*, No. 229.

(9) Husnik, F.; McCutcheon, J. P. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat. Rev. Microbiol.* **2018**, *16*, 67−79.

(10) Rancurel, C.; Legrand, L.; Danchin, E. G. Alienness: rapid detection of candidate horizontal gene transfers across the tree of life. *Genes* **2017**, *8*, No. 248.

(11) Arbel-Goren, R.; Stavans, J. *Target Search Problems*; Springer, 2024; pp 675−686.

(12) Johnson, C. M.; Grossman, A. D. Integrative and conjugative elements (ICEs): what they do and how they work. *Annu. Rev. Genet.* **2015**, *49*, 577−601.

(13) Delavat, F.; Miyazaki, R.; Carraro, N.; Pradervand, N.; van der Meer, J. R. The hidden life of integrative and conjugative elements. *FEMS Microbiol. Rev.* **2017**, *41*, 512−537.

(14) Arbel-Goren, R.; McKeithen-Mead, S. A.; Voglmaier, D.; Afremov, I.; Teza, G.; Grossman, A. D.; Stavans, J. Target search by an imported conjugative DNA element for a unique integration site along a bacterial chromosome during horizontal gene transfer. *Nucleic Acids Res.* **2023**, *51*, 3116−3129.

(15) Babic, A.; Berkmen, M. B.; Lee, C. A.; Grossman, A. D. Efficient gene transfer in bacterial cell chains. *mBio* **2011**, *2*, 10−1128.

(16) Couturier, A.; Virolle, C.; Goldlust, K.; Berne-Dedieu, A.; Reuter, A.; Nolivos, S.; Yamaichi, Y.; Bigot, S.; Lesterlin, C. Real-time visualisation of the intracellular dynamics of conjugative plasmid transfer. *Nat. Commun.* **2023**, *14*, No. 294.

(17) Jones, J. M.; Grinberg, I.; Eldar, A.; Grossman, A. D. A mobile genetic element increases bacterial host fitness by manipulating development. *eLife* **2021**, *10*, No. e65924.

(18) Menard, K. L.; Grossman, A. D. Selective pressures to maintain attachment site specificity of integrative and conjugative elements. *PLoS Genet.* **2013**, *9*, No. e1003623.

(19) Shvets, A. A.; Kochugaeva, M. P.; Kolomeisky, A. B. Mechanisms of protein search for targets on DNA: theoretical insights. *Molecules* **2018**, *23*, No. 2106.

(20) Mirny, L.; Slutsky, M.; Wunderlich, Z.; Tafvizi, A.; Leith, J.; Kosmrlj, A. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A: Math. Theor.* **2009**, *42*, No. 434013.

(21) Koslover, E. F.; de la Rosa, M. A. D.; Spakowitz, A. J. Theoretical and computational modeling of target-site search kinetics in vitro and in vivo. *Biophys. J.* **2011**, *101*, 856−865.

(22) Kolomeisky, A. B. Physics of protein−DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2088−2095.

(23) Vuzman, D.; Levy, Y. DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 21004−21009.

(24) Del Val, E.; Nasser, W.; Abaibou, H.; Reverchon, S. RecA and DNA recombination: a review of molecular mechanisms. *Biochem. Soc. Trans.* **2019**, *47*, 1511−1531.

(25) Bell, J. C.; Kowalczykowski, S. C. Mechanics and single-molecule interrogation of DNA recombination. *Annu. Rev. Biochem.* **2016**, *85*, 193−226.

(26) Kochugaeva, M. P.; Shvets, A. A.; Kolomeisky, A. B. On the mechanism of homology search by RecA protein filaments. *Biophys. J.* **2017**, *112*, 859−867.

(27) Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.; Alloni, G.; Azevedo, V.; Bertero, M.; Bessières, P.; Bolotin, A.; Borchert, S.; et al. The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature* **1997**, *390*, 249−256.

(28) Groth, A. C.; Olivares, E. C.; Thyagarajan, B.; Calos, M. P. A phage integrase directs efficient site-specific integration in human cells. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5995−6000.

(29) Abe, K.; Takamatsu, T.; Sato, T. Mechanism of bacterial gene rearrangement: SprA-catalyzed precise DNA recombination and its directionality control by SprB ensure the gene rearrangement and stable expression of spsM during sporulation in Bacillus subtilis. *Nucleic Acids Res.* **2017**, *45*, 6669−6683.

(30) Le, T. B. K.; Imakaev, M. V.; Mirny, L. A.; Laub, M. T. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **2013**, *342*, 731−734.
(31) Bell, J. C.; Plank, J. L.; Dombrowski, C. C.; Kowalczykowski, S. C. Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA. *Nature* **2012**, *491*, 274−278.
(32) Forget, A. L.; Kowalczykowski, S. C. Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search. *Nature* **2012**, *482*, 423−427.
(33) Hu, J.; Crickard, J. B. All who wander are not lost: the search for homology during homologous recombination. *Biochem. Soc. Trans.* **2024**, *52*, 367−377.
(34) Renkawitz, J.; Lademann, C. A.; Jentsch, S. Mechanisms and principles of homology search during recombination. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 369−383.
(35) Dorfman, K. D.; Fulconis, R.; Dutreix, M.; Viovy, J.-L. Model of RecA-mediated homologous recognition. *Phys. Rev. Lett.* **2004**, *93*, No. 268102.
(36) Kornyshev, A. A.; Wynveen, A. The homology recognition well as an innate property of DNA structure. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 4683−4688.
(37) Klapstein, K.; Chou, T.; Bruinsma, R. Physics of RecA-mediated homologous recognition. *Biophys. J.* **2004**, *87*, 1466−1477.

https://doi.org/10.1021/acs.jpcb.5c02436
J. Phys. Chem. B 2025, 129, 6828−6836