Protein sliding and hopping kinetics on DNA

Ben Cashen
NORTHEASTERN UNIVERSITY | PHYSICS DEPT.

Sec. I Introduction

DNA-binding proteins often reach their target binding sites through a facilitated diffusion process which consists of two main types of motion: sliding and hopping. As seen in Fig. 1, sliding allows the protein molecule to diffuse across the DNA strand without losing contact, while hopping involves the protein "jumping" off and undergoing 3D diffusion within the fluid medium of the cell before reassociating to a different segment of the DNA. Taken together, these two types of motion constitute the general mechanism by which a protein is transported along DNA. Understanding the time frames and distances involved in these processes is of importance when considering cellular response to external stimuli. Because DNA-binding proteins play such a large role in cellular processes, such as DNA replication and gene expression, quantifying the kinetics of protein transport along DNA, which leads to proper strand binding, is of significance and interest to many single-molecule researchers.

Using Monte Carlo techniques, we simulate and characterize protein hopping along a nonspecific strand of DNA for two different proteins: GFP-LacI (green fluorescence protein) and C-Ada (C-terminal domain of Ada). Using these results, along with experimental data, we quantify the effect of facilitated diffusion on the protein target binding rate through calculations of the mean sliding time $\langle t_1 \rangle$ and sliding diffusion coefficient D_1. In doing so, we gain an understanding of the role that these motions play in the transport of protein along DNA.

Sec. II Monte Carlo Simulations

Through the use of Monte Carlo simulations, we first calculate the hopping kinetics of both GFP-LacI and C-Ada proteins. By running 10^4 hopping simulations, we compute the mean number of sliding and hopping alternations within a single diffusion trajectory N, as well as the mean hopping time $\langle t_3 \rangle$. Combining these results with experimental values of D and t (diffusion coefficient and total time of trajectory, respectively) as well as the relations below, we compute the mean sliding time $\langle t_1 \rangle$ and sliding diffusion coefficient D_1.

$$t = N\langle t_1 \rangle + N\langle t_3 \rangle \qquad (1)$$

$$Dt = D_1 N\langle t_1 \rangle + D_3 N\langle t_3 \rangle \qquad (2)$$

where D_3 is the 3D diffusion coefficient of the protein.

Since protein hopping is driven by unbiased 3D diffusion, an inherently random process, Monte Carlo methods are particularly appropriate for capturing the physics of the problem. For each simulation, the protein's center of mass is initially positioned a step δ outside the capture radius of $R = r_{DNA} + r_{protein} + \Delta r$ where r_{DNA} is the DNA radius, $r_{protein}$ is the hydrodynamic radius of each protein, and Δr is the protein-DNA binding distance. The values used for these distances are outlined for both GFP-LacI and C-Ada in Table 1. The protein is then allowed to undergo unbiased 3D diffusion until either rebinding occurs, or the maximum number of simulation steps is reached. For rebinding to occur, one of the following criterions must be met:

either the positions of the protein within two consecutive steps must be within the capture radius, or, as seen in Fig. 2, the length of the perpendicular drawn from the DNA center to the line connecting two consecutive protein positions must be shorter than the capture distance. Once reassociation occurs, the protein binding location is recorded as the midpoint between these two positions and the simulation is stopped. If the maximum number of simulation steps is reached, the protein is assumed to have permanently dissociated from the DNA. For each type of protein, we model the DNA in a slightly different manner. While simulating GFP-LacI hopping we approximate the strand as an infinitely long, rigid cylinder with constant radius. This approximation is valid since the persistence length of DNA is much greater than the average protein hopping distance, and the diameter of GFP-LacI is greater than the pitch of any (A, B, or Z) DNA structure. In an attempt to account for the periodic structure of DNA, we artificially choose to model the geometry of Z-DNA while simulating the hopping of C-Ada. Because the pitch of Z-DNA (4.56 nm) is larger than the diameter of C-Ada, it's an appropriate geometry for considering variations in the DNA capture radius, potentially allowing the protein to nestle in between its helical structure. The binding radius of the C-Ada simulations is modeled as

$$R = \frac{r_{max} + r_{min}}{2} + \left(\frac{r_{max} - r_{min}}{2}\right)\cos\left(\frac{2\pi}{P}z\right)$$

where $P = 4.56\ nm$ is the pitch of Z-DNA, $r_{max} = r_{DNA} + r_{protein} + \Delta r$, and $r_{min} = r_{protein} + \Delta r$. We make the additional assumption that there is a 100% probability for association upon DNA-protein collision.

Our simulations step size is chosen to be the mean free path δ of the protein within the aqueous solution of the cell. The instantaneous velocity of the protein is given by

$$\sqrt{\langle v_x^2 \rangle} = \sqrt{k_B T/m} = \delta/\tau$$

where k_B is the Boltzmann constant, $T = 294\ K$, m is the mass of the protein, and τ is the time between collisions. Using the Einstein-Stokes relation

$$D_3 = \delta^2/2\tau = k_B T/6\pi\eta r$$

where $\eta = 10^{-3}\ Ns/m^2$ is the viscosity of water and r is the protein radius, we have

$$\delta = 2D_3/\sqrt{\langle v_x^2 \rangle}$$

The simulation parameter values for each protein are shown in Table 2. Each simulation step in the x, y, and z dimensions is randomly drawn from a Gaussian distribution with a mean of zero and standard deviation of δ. We run each protein simulation for a maximum number of 1.1×10^8 steps, corresponding to an association time limit of $\approx 0.5\ ms$ for GFP-LacI and $\approx 0.2\ ms$ for C-Ada. These limits should be sufficiently long to allow all active hopping proteins to return and rebind to the DNA.

Sec. III Results

GFP-LacI

For 10^4 protein simulations, the distributions of the hopping distance and hopping height for GFP-LacI are shown in Figs. 3 and 4, respectively. It should be noted that the hopping height distribution has an artificial lower bound of δ rather than zero. This is due to the fact that for each simulation the protein's position was initialized at $\vec{r}_0 = (R + \delta, 0,0)$, just outside the DNA capture surface, giving a minimum hopping height of δ for each run. We find a mean hopping distance of 3.27 Å and a mean hopping height (maximum radial distance of protein from DNA during hop) of 4.99 Å. Out of 10^4 simulations, 29 resulted in the protein permanently dissociating from the DNA. We calculate the number of hops per diffusion trajectory $N = 344$ by taking the ratio of the number of rebinding events (9971) to the number of permanent dissociation events. The mean number of steps per hop is found to be $n = 2.83 \times 10^4$, yielding a hopping time of $n\tau_{GFP} = \langle t_3 \rangle = 0.13\ \mu s$. The distributions of the hopping time and number of hops per diffusion trajectory are shown in Figs. 5 and 6, respectively. We calculate a mean total hopping time per trajectory of $N\langle t_3 \rangle = 44.6\ \mu s$ and a root-mean-square (RMS) total hopping displacement per trajectory of $\sqrt{2D_3 N\langle t_3 \rangle} = 84.7\ nm$. Using relations (1) and (2), we have

$$\langle t_1 \rangle = \frac{t}{N} - \langle t_3 \rangle \approx \frac{t}{N}$$

$$D_1 = \frac{1}{\langle t_1 \rangle}\left[\frac{Dt}{N} - D_3\langle t_3 \rangle\right]$$

where $t = 10.4\ s$ and D has a range of $2.3 \times 10^2 - 1.3 \times 10^5\ \ nm^2/s$, giving a value of $\langle t_1 \rangle = 30.2\ ms$ and D_1 between -115.4 and $1.3 \times 10^5\ \ nm^2/s$. We set a lower bound on D by requiring that $D_1 \geq 0$, giving $D_{min} = D_3 N\langle t_3 \rangle / t = 345\ \ nm^2/s$. When $D > 2D_3 N\langle t_3 \rangle / t = 690\ \ nm^2/s$, the RMS total sliding displacement per trajectory becomes larger than the RMS total hopping displacement. Using a value of $\langle D \rangle \approx 2 \times 10^4\ \ nm^2/s$, we have $\langle D_1 \rangle \approx 1.97 \times 10^4\ \ nm^2/s$ allowing us to conclude that sliding is the dominant factor in "fast" diffusion of GFP-LacI across DNA. The RMS total sliding displacement per trajectory $x^S_{RMS} = \sqrt{2D_1 N\langle t_1 \rangle}$ has a range of $0.0 - 1.6 \times 10^3\ nm$, however, meaning, that while sliding is most often the dominant motion, during "slow" diffusion processes, hopping becomes the dominant factor.

C-Ada

The distributions of the hopping distance and hopping height for C-Ada are shown in Figs. 7 and 8, respectively. We find a mean hopping distance of 5.74 Å and a mean hopping height of 8.32 Å. Due to the periodic structure of the DNA molecule, these results are sensitive to the initial conditions of each simulation. The protein's position was initialized at $\vec{r}_0 = (r_{max} + \delta, 0,0)$ for each run, meaning that our findings are artificially higher than they should be, since

the capture radius initially decreases along the z-direction on either side of the protein's initial position. The opposite would be true if the protein were initialized around r_{min}. To compensate for this fact, one should require that the protein's position be randomly initialized along the DNA capture surface. Unfortunately, this was realized too late, so our subsequent results for C-Ada should be understood with this in mind. Out of 10^4 simulations, 58 resulted in the protein permanently dissociating from the DNA, yielding a mean number of hops per trajectory of 171. The mean number of steps per hop is found to be $n = 9.2 \times 10^4$, giving a hopping time of $n\tau_{CA} = \langle t_3 \rangle = 0.17\ \mu s$. The distributions of the hopping time and number of hops per diffusion trajectory are shown in Figs. 9 and 10, respectively. We calculate a mean total hopping time per trajectory of $N\langle t_3 \rangle = 29.1\ \mu s$ and an RMS total hopping displacement per trajectory of $\sqrt{2D_3 N\langle t_3 \rangle} = 84.1\ nm$. Using relations (1) and (2), with values of $t \approx 10\ s$ and $D = 1.3 \times 10^6$, we find $\langle t_1 \rangle \approx 58.5\ ms$ and $D_1 \approx 1.3 \times 10^6$. Although this sliding diffusion coefficient is approximately two orders of magnitude smaller than D_3, the relatively large mean sliding time results in the majority of the protein displacement during a trajectory being dominated by sliding. We find an RMS total sliding displacement per trajectory of $x^s_{RMS} = 5.1 \times 10^3$ nm, reaffirming the notion that sliding is the dominant feature of facilitated diffusion for C-Ada.

Sec. IV Discussion

Our simulations have shown that DNA-binding proteins can, on average, make hundreds of alternating hops and slides during a single diffusion trajectory, allowing them a transport method with which they can potentially reach their target binding site. While it is unclear what cellular conditions lead to differences in the seconds-long diffusion coefficient D for GFP-LacI, our results indicate that for fast diffusion (large D) of GFP-LacI along DNA, the protein's overall displacement is dominated by sliding. Conversely, for slow facilitated diffusion (small D) processes, we find that hopping is the dominant factor. Although the maximum sliding diffusion rate D_{1max} is approximately two orders of magnitude smaller than that of hopping, sliding generally remains the dominant motion of transport due to the fact that the mean time per slide $\langle t_1 \rangle \sim 10^{-2}\ s$ is extremely large relative to the mean hopping time $\langle t_3 \rangle \sim 10^{-7}\ s$.

Accounting for potential variation in the capture radius due to the helical structure of DNA yields similar results to those found for GFP-LacI. Due to the fact that the sliding diffusion coefficient D_1 for C-Ada is approximately equal to the seconds-long diffusion coefficient D, we find that sliding is always the dominant form of transport along the DNA strand for this protein structure.

References

Berg, H. C. (1993). *Random Wlaks in Biology*. Princeton University Press.

DeSantis, M. C., Li, J.-L., & Wang, Y. M. (2011). *Protein sliding and hopping kinetics on DNA.* Phys. Rev. E.

Lin, Y., Zhao, T., Jian, X., & Farooqui, Z. (2009). *Using the bias from flow to elucidate single DNA repair protein sliding and interactions with DNA.* Biophys J.

Wang, Y. M., Austin, R. H., & Cox, E. C. (2006). *Single molecule measurements of repressor protein 1D diffusion on DNA.* Phys. Rev. Lett.

Protein	GFP-LacI	C-Ada
$\Delta r\ (nm)$	0.5	0.5
$r_{protein}\ (nm)$	2.68	1.77
$r_{DNA}\ (nm)$	1.0	0.0 – 1.0
$R_{capture}$	4.18	2.27 – 3.27

Table 1: Distance values used for implementation of capture

Protein	GFP-LacI	C-Ada
δ (Å)	0.267	0.214
$\tau\ (ps)$	4.46	1.88
$D_3\ (nm^2/s)$	8.03×10^7	12.16×10^7
$m\ (kDa)$	67.5	18.9
$r\ (nm)$	2.68	1.77

Table 2: Hopping simulation parameters

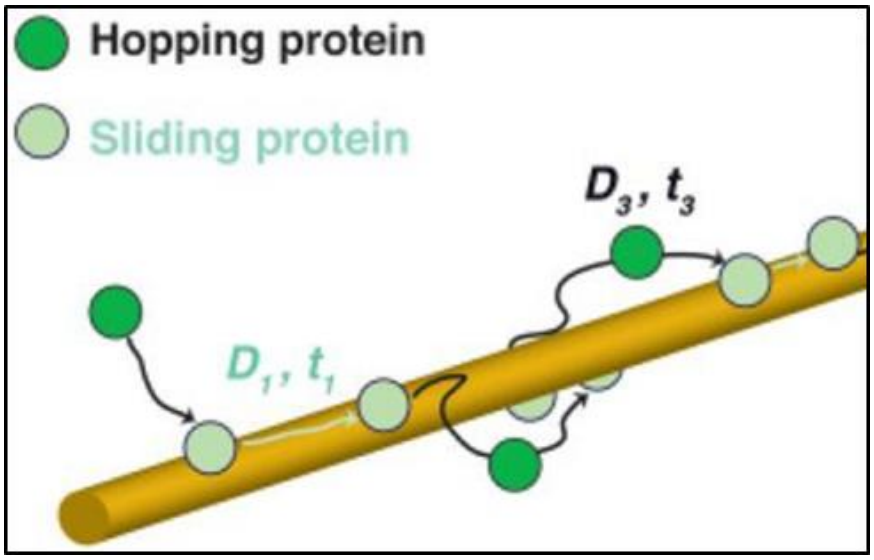

Fig. 1: Schematic of protein diffusion trajectory showing hopping (dark green) and sliding (light green)

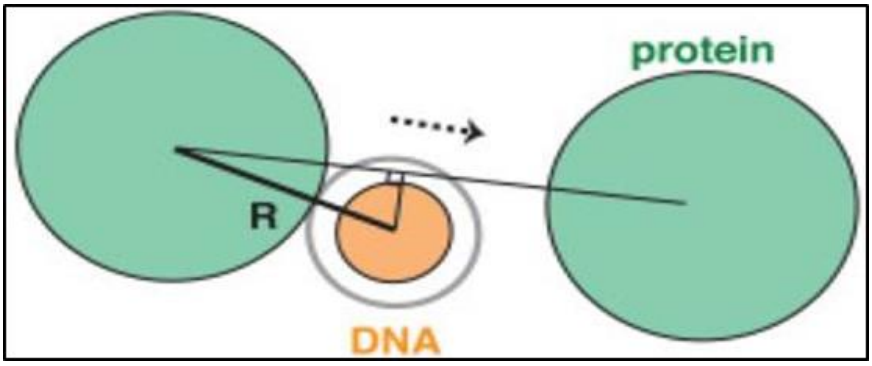

Fig. 2: Visual representation of second capture criteron

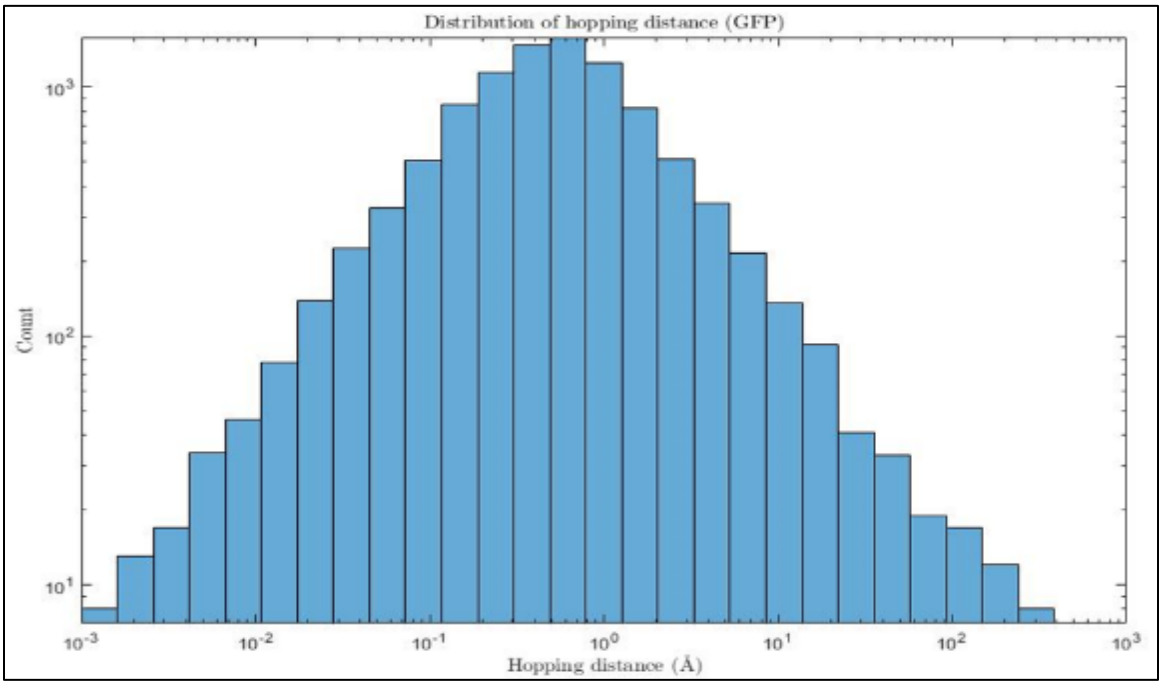

Fig. 3: Distribution of GFP-LacI hopping distance

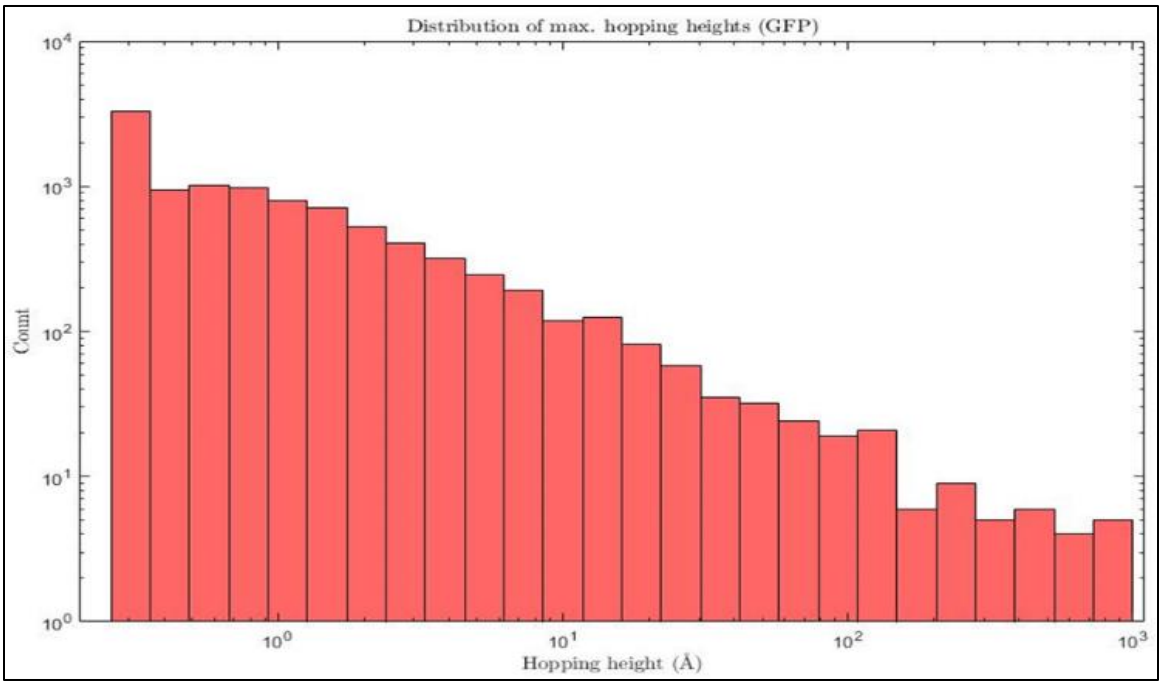

Fig. 4: Distribution of GFP-LacI hopping height

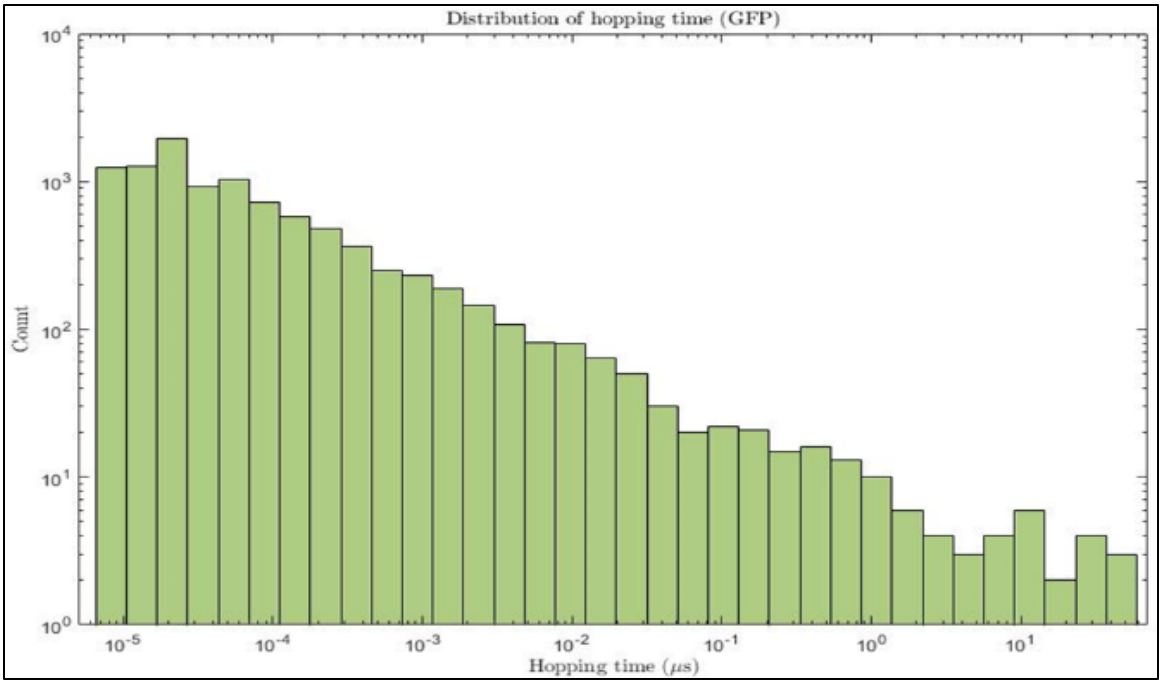

Fig. 5: Distribution of GFP-LacI hopping time

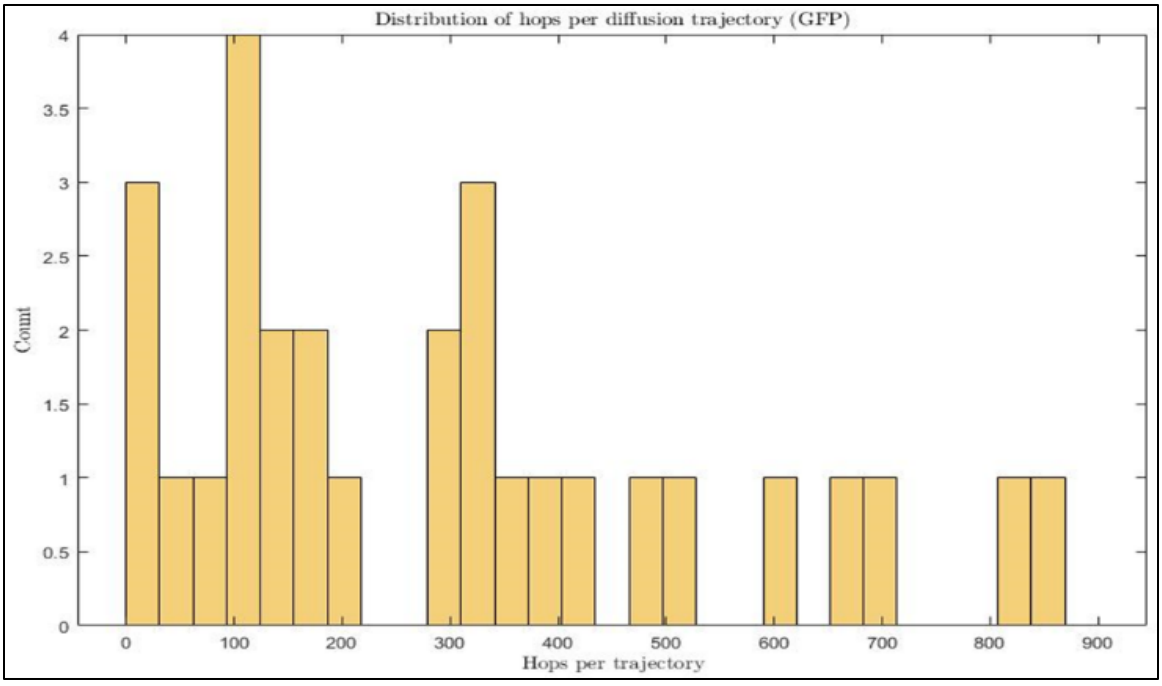

Fig. 6: Distribution of number of hops per trajectory for GFP-LacI

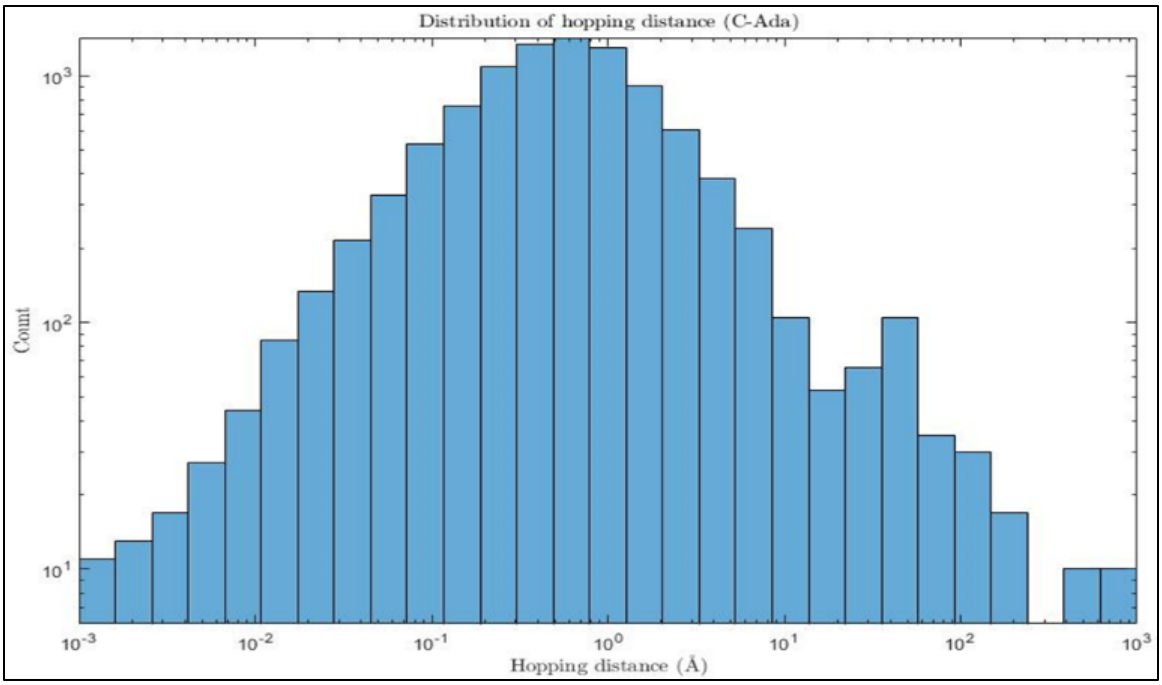

Fig. 7: Distribution of C-Ada hopping distance

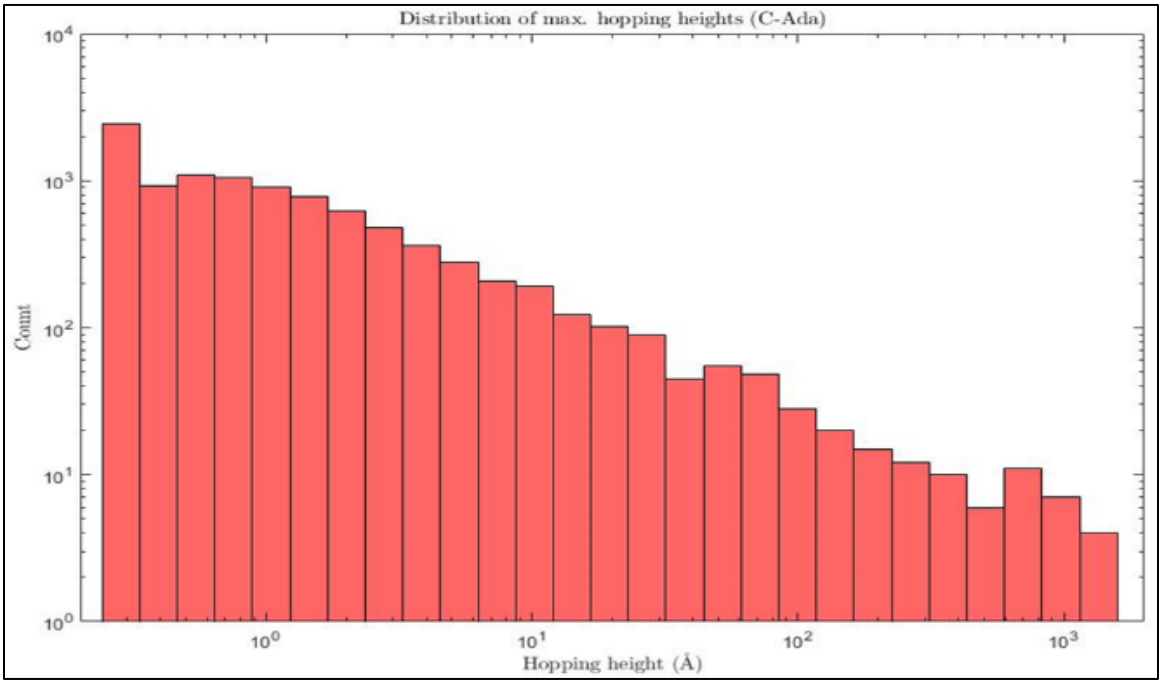

Fig. 8: Distribution of C-Ada hopping height

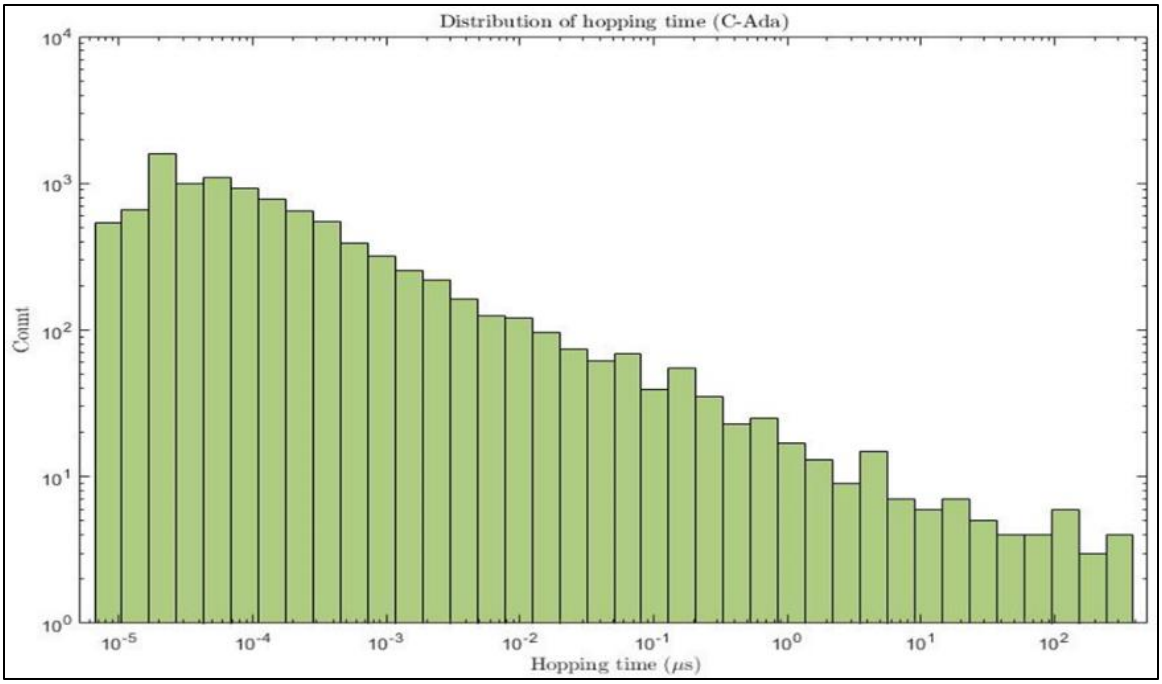

Fig. 9: Distribution of C-Ada hopping time

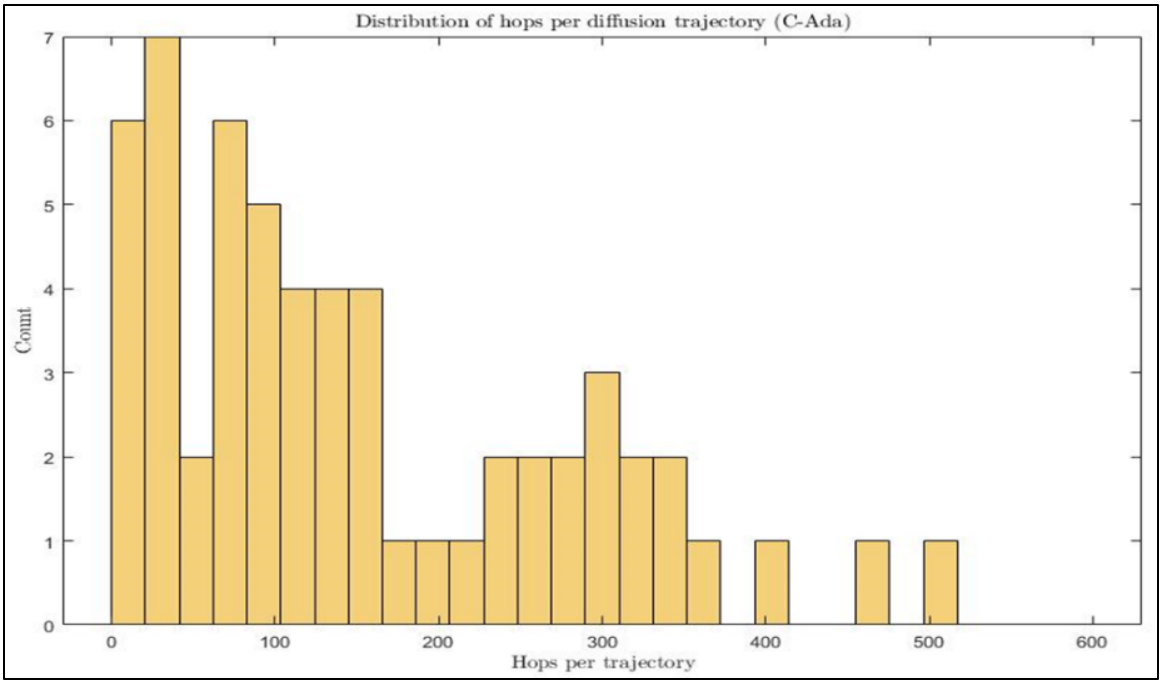

Fig. 10: Distribution of number of hops per trajectory for C-Ada