Featured in Physics

Search and Localization Dynamics of the CRISPR-Cas9 System

Qiao Lu (路桥),[1] Deepak Bhat,[1] Darya Stepanenko,[1,2,3] and Simone Pigolotti[1,*]

[1]Biological Complexity Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan
[2]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, USA
[3]Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York 11794, USA

(Received 19 March 2021; accepted 30 September 2021; published 10 November 2021)

The CRISPR-Cas9 system acts as the prokaryotic immune system and has important applications in gene editing. The protein Cas9 is one of its crucial components. The role of Cas9 is to search for specific target sequences on the DNA and cleave them. In this Letter, we introduce a model of facilitated diffusion for Cas9 and fit its parameters to single-molecule experiments. Our model confirms that Cas9 search for targets by sliding, but shows that its sliding length is rather short. We then investigate how Cas9 explores a long stretch of DNA containing randomly placed targets. We solve this problem by mapping it into the theory of Anderson localization in condensed matter physics. Our theoretical approach rationalizes experimental evidence on the distribution of Cas9 molecules along the DNA.

DOI: 10.1103/PhysRevLett.127.208102

The discovery of the CRISPR-Cas9 system has revealed the functioning of the bacterial immune response and has opened previously unimaginable possibilities for gene editing [1]. The protein Cas9 is a central actor in this system. Cas9 is an endonuclease that is able to load a guide RNA strand. Its target is a sequence on the DNA complementary to the guide RNA, which Cas9 can identify and cleave. This recognition process has attracted considerable attention, including from the modeling side [2–4]. Recognition is triggered by a three-base sequence called protospacer adjacent motif (PAM), which precedes the target. The first two bases of a PAM are guanine, whereas the third can be any base. Cas9 can transiently bind to a PAM even in the absence of a neighboring target [5,6].

Other proteins such as the lac repressor in *Escherichia coli* [7] find their targets along the DNA by a mechanism termed facilitated diffusion—an alternance of 3D diffusion in the cytosol and one-dimensional diffusive sliding along the DNA chain [8]. This mechanism can significantly improve search efficiency [8,9]. The theory of facilitated diffusion has been extended to take into account the energetics of target search along the DNA [10–12] and other processes such as hopping, i.e., the possibility for proteins to briefly detach from DNA and then reattach at short distance [13]. This notion has stimulated experimental efforts to determine whether Cas9 finds its target by facilitated diffusion as well [6,14,15].

However, experimental single-molecule studies using DNA curtains [14] and fluorescence resonance energy transfer (FRET) [15] did not find evidence of sliding. They however found that the lifetime of Cas9 binding events is well fitted by a double exponential even in the absence of targets, suggesting a complex binding mechanism. In contrast, a more recent FRET experimental study provides evidence that Cas9 can slide [6]. This study found that, in a DNA sequence containing multiple PAMs without targets, the two exponential constants characterizing the binding lifetime distribution depend on the number of PAMs and the distance between them. In the absence of PAMs, this distribution reduces to a single exponential. The dependence of the exponential constants on the distance between PAMs suggests that the characteristic sliding length of Cas9 falls below the spatial resolution of previous experiments [14], potentially explaining why sliding was not previously observed.

An alternative way of probing the search dynamics of Cas9 is to experimentally measure the distribution of Cas9 molecules bound along the DNA. For example, an experiment based on DNA curtains shows that Cas9 is localized in regions that extend for hundreds of base pairs length around targets [14]. This length scale is much larger than the sliding length suggested by Ref. [6].

This contrasting experimental evidence calls for a theoretical explanation. To this aim, it is useful to think about a bacterial genome as a long stretch of DNA in which a large number of PAMs are disorderly distributed. For comparison, the *E. coli* genome is 4.6×10^6 base pairs long and contains about 0.5×10^6 PAMs [5]. We want to estimate the typical localization length of Cas9 on such DNA sequences. This problem bears an analogy with the theory of Anderson localization [16]. This theory predicts that, under general conditions, eigenvectors of disordered one-dimensional diffusive systems are localized, with profound consequences for fields of physics ranging from condensed matter to disordered and chaotic systems [17]. In this analogy, PAMs play the role of defects in one dimensional lattices.

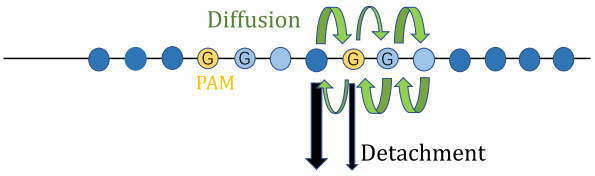

FIG. 1. Scheme of the model. PAM sites and nonspecific sites are shown in yellow and blue, respectively. The second and third bases of PAM sequences are considered as nonspecific sites (light blue). Green arrows represent sliding rates and black arrows represent unbinding rates, see Eq. (2). Thicker arrows correspond to larger rates.

In this Letter, we show that a facilitated diffusion model quantitatively explains the dynamics of Cas9 observed in single-molecule experiments. We then formalize the mapping between facilitated diffusion and Anderson localization. This approach permits us to determine the localization length of Cas9 on typical long DNA strands and explain the discrepancy between the sliding length in [6] and the localization length in [14] in terms of a hopping mechanism. The mapping presented in this Letter can be used to study the dynamics of other DNA binding proteins, such as transcription factors.

We consider a Cas9 protein that binds on a DNA chain of length N and slides along it before detaching, see Fig. 1. Our aim is to quantify the distribution of duration of binding events depending on the arrangement of specific PAM sites along the DNA chain. We introduce the probability $p_n(t)$ that Cas9 is bound at site n at time t, given that it had attached on the DNA at time $t = 0$. Each site represents a nucleotide position $n = 1, \ldots, N$. We assume attachment to be nonspecific, so that $p_n(t = 0) = 1/N$.

We distinguish between two types of DNA sites. PAM sites are those at the beginning of a PAM sequence, where Cas9 can bind specifically. We consider every other site as nonspecific, including the two other base pairs constituting a PAM, see Fig. 1. We call E_n the binding energy of Cas9 at position n. We assume that all nonspecific sites have the same binding energy $E_n = 0$. If n is a PAM site, then $E_n = -\beta$, with $\beta > 0$. All energies are expressed in units of $k_B T$, where k_B is the Boltzmann constant and T the temperature. Our aim is to analyze single binding events and therefore we do not consider rebinding after detachment.

The probabilities $p_n(t)$ evolve according to the master equation

$$\frac{d}{dt}p_n(t) = D_{n,n+1}p_{n+1} + D_{n,n-1}p_{n-1} - (D_{n+1,n} + D_{n-1,n} + k_n)p_n, \quad (1)$$

in which $D_{n,m} = De^{E_m}$ and $k_n = ke^{E_n}$, where the diffusion rate D and the unbinding rate k are given parameters. We impose vanishing fluxes at the boundaries, $D_{0,1} = D_{1,0} = D_{N,N+1} = D_{N+1,N} = 0$. This choice of rates satisfies the detailed balance condition $D_{n,m}e^{-E_m} = D_{m,n}e^{-E_n}$.

We express the model in vector notation by defining $\mathbf{p}(t) = [p_1(t), p_2(t), \ldots, p_N(t)]$. We write Eq. (1) as $d\mathbf{p}/dt = \hat{A}\mathbf{p}$, where the elements $A_{m,n}$ of the matrix $\hat{A}$ are given by

$$A_{m,n} = \begin{cases} De^{E_n} & \text{if } |n-m| = 1 \\ -(k+2D)e^{E_m} & \text{if } n = m. \end{cases} \quad (2)$$

The formal solution to the master equation is $\mathbf{p}(t) = e^{\hat{A}t}\mathbf{p}(0)$, where $\mathbf{p}(0)$ is the uniform initial condition. The eigenvalue equation associated with the master equation is

$$\hat{A}\boldsymbol{\psi} = -\lambda\boldsymbol{\psi}. \quad (3)$$

Equation (3) is solved by a set of eigenvalues $\lambda = \lambda_1, \lambda_2, \ldots, \lambda_N$ and associated right eigenvectors $\boldsymbol{\psi} = \boldsymbol{\psi}^{(1)}, \boldsymbol{\psi}^{(2)}, \ldots, \boldsymbol{\psi}^{(N)}$, assumed to be normalized. The solution of the master equation can be decomposed into eigenvalues

$$\mathbf{p}(t) = \sum_{i=1}^{N} e^{-\lambda_i t} c_i \boldsymbol{\psi}^{(i)}, \quad (4)$$

where the coefficients c_i are determined by the initial condition. Because of detachment, one has $\lim_{t\to\infty} p_i(t) = 0$ for all i. This fact and the detailed balance condition imply that all eigenvalues must be real and positive. We sort the eigenvalues so that λ_1 is the smallest one.

The total probability that Cas9 is still bound at a time t is given by $P(t) = \sum_n p_n(t)$. Since we are considering a single binding event, $P(t)$ is a decreasing function of t. We define the instantaneous detachment rate $g(t) = -d/dt P(t)$. Single-molecule experiments [6,14,15] observed that the temporal decay of $g(t)$, and therefore of $P(t)$, is characterized by two distinct exponential slopes at short and long times.

To understand these two regimes, we focus on $P(t)$ and define its instantaneous exponential slope $K(t) = -d/dt \ln P(t)$. We also define the total probability $P_{\mathrm{PAM}}(t) = [\sum_{n\in\mathrm{PAM}} p_n(t)]/P(t)$ of Cas9 being bound to a PAM site at time t, given that it had not detached yet. By summing Eq. (1) over n, we find that

$$K(t) = k[1 - P_{\mathrm{PAM}}(t)] + ke^{-\beta}P_{\mathrm{PAM}}(t). \quad (5)$$

Considering that $0 \le P_{\mathrm{PAM}}(t) \le 1$, the slope $K(t)$ is limited by the two unbinding rates:

$$ke^{-\beta} \le K(t) \le k. \quad (6)$$

The value of $K(t)$ in this range is determined by $P_{\mathrm{PAM}}(t)$. Since the initial distribution is uniform, at short times P_{PAM}

is equal to the fraction of PAM sites. Given that this fraction is usually small, Eq. (5) implies $K(t) \approx k$ at short times. In the long time limit, Eq. (4) leads to conclude that $K(t) = \lambda_1$.

Experiments in [6] measured the distribution of Cas9 binding events on DNA sequences containing from 0 to 5 PAM sites. We jointly fitted the parameters k, β, and D to these six experiments, see Fig. 2(a). Solutions of the master equation (1) with the best-fit parameters reproduce the double exponential behavior and fit well the experimental data, see Fig. 2(b). The fitted values of the parameters are $k = 1.94 \pm 0.10\ \mathrm{s}^{-1}$, $\beta = 3.34 \pm 0.07$, and $D = 52 \pm 9\ \mathrm{bp}^2\ \mathrm{s}^{-1}$. Experiments on a different variant of Cas9 find differences in binding energy between PAM and near-cognate sites that are comparable with our estimate of β [2]. A more detailed model where each non-PAM sequence is characterized by a different binding energy, leads to similar fitted values of the corresponding rates, see [18]. This evidence supports robustness of our results.

At an increasing number of PAM sites, the second slope in Fig. 2(a) becomes significantly less steep than the first. According to Eq. (5), this means that, at long times, Cas9 is much more localized on PAM sites compared with short times. Inspecting the eigenvectors $\boldsymbol{\psi}^{(1)}$ associated with the smallest eigenvalue λ_1 confirms this idea, see Fig. 2(c).

To gain further insight into the dynamics observed in Fig. 2(b), we analytically compute λ_1 and its eigenvector for an infinitely long chain with a single PAM site at $n = 0$. For $|n| > 1$, the eigenvector satisfies

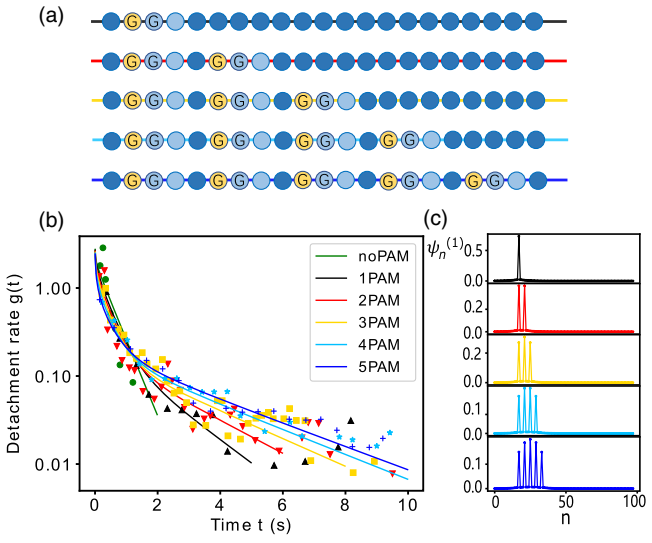

FIG. 2. (a) Arrangements of PAM sites used in the experiments in [6]. Line colors correspond to the different curves in panel (b). The figure shows only the portion of the DNA sequence of length $N = 98$ where the PAM sites are located. (b) Comparison of the prediction of our model (lines) with experiments [6] (points). Model parameters are determined by jointly fitting the experimental data for $j = 0, \ldots, 5$ PAM sites using maximum likelihood, see [18]. (c) Eigenvectors $\boldsymbol{\psi}^{(1)}$ for $j = 1, \ldots, 5$.

$$-\lambda_1 \psi_n^{(1)} = D(\psi_{n+1}^{(1)} + \psi_{n-1}^{(1)} - 2\psi_n^{(1)}) - k\psi_n^{(1)}. \quad (7)$$

We assume a solution of the form $\psi_n \propto e^{-|n|/\ell}$ where $|n| > 0$ and we define ℓ as the sliding length. Substituting into Eq. (7) we obtain

$$k - \lambda_1 = 2D[\cosh(1/\ell) - 1]. \quad (8)$$

By expanding the cosh at first order we find $\ell \approx \sqrt{D/(k-\lambda_1)}$. Note that $\lambda_1 \leq k$ due to Eq. (6). The three unknown λ_1, ℓ, and ψ_0 can be determined from Eq. (8) and the equivalents of Eq. (7) for $n = 0$ and $|n| = 1$. Substituting the fitted parameters of Fig. 2, we find $\ell \approx 6.2$ bp.

Both our model and experiments [6] show that the lifetime of long binding events increases at an increasing number of PAMs, see Fig. 2(b). In the model, this means that λ_1 is a decreasing function of the number of PAMs. This effect can be explained by interference among PAM sites, i.e., the fact that the eigenvector $\boldsymbol{\psi}^{(1)}$ for j PAM sites is not simply a superimposition of j single-PAM eigenvectors, unless the interval between the PAM sites is much larger than ℓ. Only in this limit binding events around each PAM site behave independently, and the long-time exponential slope becomes independent of the number of PAM sites, see Fig. 3. At shorter intervals, interference leads to an increase in target occupancy. This implies that, at large t, $P_{\mathrm{PAM}}(t)$, and therefore the typical lifetime of binding events $1/\lambda_1$, are decreasing functions of the interval between the PAM sites, see Eq. (5) and Fig. 3.

In summary, we found that the distribution of a Cas9 molecule in a region of DNA containing several PAM sites tends to be localized. We now study the behavior of Cas9 on a very long stretch of DNA including a disordered assortment of PAM sites. The theory of Anderson

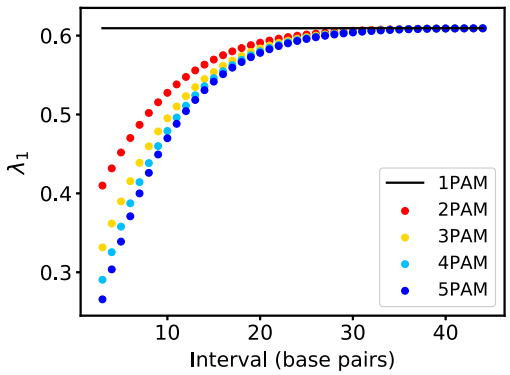

FIG. 3. Interference between $j = 1, \ldots, 5$ equally spaced PAM sites on an infinite DNA chain. Lowest eigenvalue λ_1 as a function of the interval between the PAM sites. Points are obtained by numerically diagonalizing the matrix $\hat{A}$ corresponding to each case, with $N = 220$. The horizontal line marks the value of λ_1 for a single PAM sequence, from the solution of Eq. (7).

localization predicts that, in such disordered one-dimensional systems, eigenvectors are exponentially localized:

$$\psi_n \sim e^{-|n-n^*|/\gamma(\lambda)}, \tag{9}$$

where n^* is the location of the eigenvector peak and $\gamma(\lambda)$ is the localization length associated with the eigenvalue λ. The localization length γ can be thought as the generalization of the sliding length ℓ: the former is defined for an arbitrary disordered DNA chain, whereas the latter is defined for a single target. Our hypothesis is that the localization length associated with the smallest eigenvalues of Cas9 dynamics can explain the results of DNA curtains experiments [14].

We sharpen the analogy between our problem and the Anderson localization by rescaling the components of our eigenvectors by the Boltzmann weight, $f_n = \psi_n \exp(E_n)$. With this transformation, Eq. (3) assumes the same form for PAM and non-PAM sites:

$$f_{n+1} + f_{n-1} - \left(2 + \frac{k - \lambda e^{-E_n}}{D}\right) f_n = 0. \tag{10}$$

This equation is formally similar to the discrete Schrödinger equation in Anderson's original work [16]. It can be solved by the transfer matrix method. We introduce the vector $\mathbf{f}_n = (f_n, f_{n-1})$ and the transfer matrix

$$\hat{T}_n = \begin{pmatrix} 2 + \frac{k-\lambda e^{-E_n}}{D} & -1 \\ 1 & 0 \end{pmatrix}. \tag{11}$$

With these definitions, we rewrite Eq. (10) as

$$\mathbf{f}_{n+1} = \hat{T}_n \mathbf{f}_n \tag{12}$$

and therefore

$$\mathbf{f}_N = \prod_{n=1}^{N-1} \hat{T}_n \mathbf{f}_1. \tag{13}$$

We assume that, in a typical long DNA sequence, each site n has a probability $1/16$ to be a PAM site, thereby affecting the value of E_n in the corresponding matrix T_n. In this view, Eq. (13) expresses the solution of the eigenvalue equation as a product of random matrices [17].

The localization length γ can be calculated from this product with an approach proposed by Herbert, Jones, and Thouless [21,22]. This approach rests on the idea that $f_N(\lambda)$, with appropriate boundary conditions, vanishes if λ is an eigenvalue and changes sign as a function of λ at every eigenvalue. This argument leads to the expression

$$\frac{1}{N}\ln f_N(\lambda) = \frac{1}{N}\sum_{n=1}^{N-1} \ln|\lambda_n - \lambda| + \frac{i\pi}{N}\sum_{n=1}^{N-1}\theta(\lambda - \lambda_n) + \frac{1}{N}\ln A, \tag{14}$$

where θ is the Heaviside step function and A is a finite constant. Taking the limit $N \to \infty$, we define

$$\Lambda(\lambda) = \lim_{N\to\infty}\frac{1}{N}\ln f_N(\lambda) = \lim_{N\to\infty}\frac{1}{N}\ln\left(\mathrm{Tr}\prod_{n=1}^{N}\hat{T}_n\right), \tag{15}$$

The Furstenberg theorem guarantees that $\Lambda(\lambda)$ is independent of the realization of the disorder and of the choice of $\mathbf{f}_1$ [23,24].

The inverse of the real part of $\Lambda(\lambda)$ can be identified with the localization length $\gamma(\lambda)$ thanks to a result known as the Borland conjecture [25]. The validity of this conjecture for our class of systems is supported by numerical and theoretical studies [23,26]. Further, Eq. (14) links the imaginary part of Λ with the cumulative density of states. Computing $\Lambda(\lambda)$ from the product of transfer matrices, we find that the localization length for the whole spectrum is always shorter than 11 base pairs, see Fig. 4(b).

We remark that the disordered arrangement of PAM sites is crucial for this result. In a long DNA chain containing a periodic arrangement of PAM sites, the eigenvectors are extended rather than localized, see [18].

The localization lengths in Fig. 4 are much shorter than those observed in DNA curtains experiments [14]. We assume that this discrepancy can be explained by the following idea. Measuring the distribution of Cas9 in an experiment amounts to performing an "ensemble average" which is potentially affected by search mechanisms other than sliding (such as hopping). In contrast, FRET

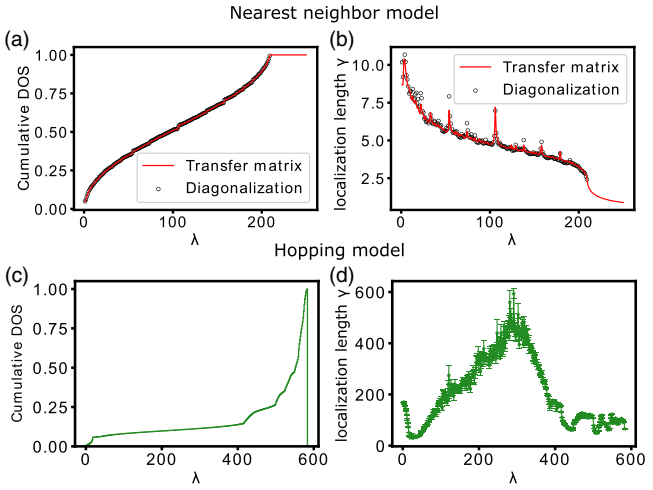

FIG. 4. (a) Cumulative density of states (DOS) and (b) localization length as function of λ for the nearest neighbor model, Eq. (1), computed using Eq. (14). Results obtained by the transfer matrix method agree with those obtained by direct diagonalization. The DNA chain length is $N = 10^6$ for the transfer matrix method and $N = 5000$ for the direct diagonalization. (c) Cumulative DOS and (d) localization length for the hopping model expressed by Eq. (16), computed using Eq. (17). In this case, the DNA chain length is $N = 2000$.

experiments focus on individual sliding events, which are unaffected by such mechanisms.

To test this idea, we generalize our model to include hopping. In a hopping event, Cas9 detaches and then reattaches to the DNA at a short distance. This amounts to include in our master equation diffusion among nonnearest neighboring sites:

$$D_{m,n} = De^{E_n} h(|n-m|), \tag{16}$$

where $h(n)$ is a positive decreasing function characterizing the probability of hopping events at a given distance n relative to sliding events. We impose $h(1) = 1$, so that nearest-neighbor sliding is consistent with Eq. (2). We determine the function $h(n)$ from the solution of a diffusion equation in cylindrical coordinates, see [13] and [18]. Unbinding rates in the hopping model are the same as in Eq. (2). For models with next to nearest neighbor interactions, such as our hopping model, the localization length can not be computed using Eqs. (14) and (15), see [27]. We therefore estimate the localization length by a more direct strategy, although computationally heavier. Assuming that a given eigenvector $\boldsymbol{\psi}^{(i)}$ associated with an eigenvalue λ_i is localized, we obtain from Eq. (9) that

$$\gamma(\lambda_i) \sim -\frac{(N-1)}{\ln\left[\psi_1^{(i)}\psi_N^{(i)}\right]}. \tag{17}$$

In this case, the localization length associated with the lowest eigenvalues is on the same order of the experimentally measured one [hundreds of base pairs, see Fig. 4(d)].

In conclusion, in this Letter we studied the search dynamics of Cas9 along the DNA. We have shown that the predictions of a facilitated diffusion model with a short sliding length are consistent with the result of single-molecule FRET experiments. By applying the theory of Anderson localization, we have argued that a hopping mechanism can explain how Cas9 is generically distributed along the DNA.

The mapping to Anderson localization introduced in this Letter is a powerful tool that can be applied to any protein performing facilitated diffusion, such as transcription factors. Modern immunoprecipitation techniques permit us to measure binding profiles of transcription factors along the DNA at the base pair resolution [28]. However, the interpretation of these binding profiles is still under debate [29]. Our approach can be combined with sequence-dependent models of facilitated diffusion by transcription factors [10–12] to shed light on this crucial problem in biophysics.

We are grateful to Chirlmin Joo and Viktorija Globyte for sharing experimental data.

*simone.pigolotti@oist.jp

[1] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, Science **337**, 816 (2012).
[2] I. Farasat and H. M. Salis, PLoS Comput. Biol. **12**, e1004724 (2016).
[3] A. A. Shvets and A. B. Kolomeisky, Biophys. J. **113**, 1416 (2017).
[4] M. Klein, B. Eslami-Mossallam, D. G. Arroyo, and M. Depken, Cell Rep. **22**, 1413 (2018).
[5] D. L. Jones, P. Leroy, C. Unoson, D. Fange, V. Ćurić, M. J. Lawson, and J. Elf, Science **357**, 1420 (2017).
[6] V. Globyte, S. H. Lee, T. Bae, J.-S. Kim, and C. Joo, EMBO J. **38**, e99466 (2019).
[7] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf, Science **336**, 1595 (2012).
[8] O. G. Berg, R. B. Winter, and P. H. von Hippel, Biochemistry **20**, 6929 (1981).
[9] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj, J. Phys. A **42**, 434013 (2009).
[10] M. Slutsky and L. A. Mirny, Biophys. J. **87**, 4021 (2004).
[11] M. Bauer, E. S. Rasmussen, M. A. Lomholt, and R. Metzler, Sci. Rep. **5**, 10072 (2015).
[12] M. Cencini and S. Pigolotti, Nucleic Acids Res. **46**, 558 (2018).
[13] M. A. Lomholt, B. van den Broek, S.-M. J. Kalisch, G. J. Wuite, and R. Metzler, Proc. Natl. Acad. Sci. U.S.A. **106**, 8204 (2009).
[14] S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, and J. A. Doudna, Nature (London) **507**, 62 (2014).
[15] D. Singh, S. H. Sternberg, J. Fei, J. A. Doudna, and T. Ha, Nat. Commun. **7**, 12778 (2016).
[16] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[17] A. Crisanti, G. Paladin, and A. Vulpiani, *Products of Random Matrices in Statistical Physics* (Springer Science & Business Media, New York, 2012), Vol. 104.
[18] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.127.208102 for details on the maximum likelihood method; results for a sequence-dependent model and for regularly spaced PAMs; and details on the hopping model, which includes Refs. [19,20].
[19] M. E. Bonomo and M. W. Deem, Phys. Biol. **15**, 041002 (2018).
[20] E. A. Boyle, J. O. Andreasson, L. M. Chircus, S. H. Sternberg, M. J. Wu, C. K. Guegler, J. A. Doudna, and W. J. Greenleaf, Proc. Natl. Acad. Sci. U.S.A. **114**, 5461 (2017).
[21] D. Herbert and R. Jones, J. Phys. C **4**, 1145 (1971).
[22] D. Thouless, J. Phys. C **5**, 77 (1972).
[23] K. Ishii, Prog. Theor. Phys. Suppl. **53**, 77 (1973).
[24] H. Furstenberg, Trans. Am. Math. Soc. **108**, 377 (1963).
[25] R. Borland, Proc. R. Soc. A **274**, 529 (1963).
[26] H. Matsuda and K. Ishii, Prog. Theor. Phys. Suppl. **45**, 56 (1970).
[27] J. Biddle, D. J. Priour, Jr., B. Wang, and S. Das Sarma, Phys. Rev. B **83**, 075105 (2011).
[28] H. S. Rhee and B. F. Pugh, Cell **147**, 1408 (2011).
[29] K. L. MacQuarrie, A. P. Fong, R. H. Morse, and S. J. Tapscott, Trends Genet. **27**, 141 (2011).