# draft

Pete Rigas, Lambert Lab

November 16, 2020

## 1 Introduction

Families of CRISPR proteins have attracted significant attention in recent studies, a few of which have devoted attention towards the regulation of microbial metabolic rate [19], genome engineering applications in Cas9 & Cas12 [6,11,13,15,17,27,35], dynamic imaging of telomeres [7], theoretically driven predictions of protein folding and atomistic simulation [22,24,25,34], and kinetic models detailing the process by which different Cas proteins identify target sequences and impact vital cellular functions [8-9,12,14,20,21,28]. In [29], a thermodynamic approach to model binding determinants is introduced, which is primarily rooted in analyses of the individual stages of Fn Cas12a binding which is comprised of PAM & crRNA inspection, followed by a reconfiguration stage. In the discussion, the authors reflect upon potential generalizations of the thermodynamics approach, particularly in making use of the formalism to interpret binding activity of catalytically active Cas12a nuclease.

Despite other studies which have implemented machine learning techniques to simulate trajectories in the energy landscape in efforts to provide analyses of binding for different proteins [10,16,30], as well as first principled models quantifying the expression of genes through transcription factors [5,33], generalizing the thermodynamic approach of [29] is advantageous in providing more interpretations of individual stages of binding for different Cas proteins which are known to variably depend on the random walk motion that the protein undergoes throughout the PAM inspection phase [29,32], in addition to blunt versus staggered cuts that are characteristic of Cas9 & Cas12, respectively [17,31,35]. To systematically quantify the rates at which particles diffuse across subsequent base pairs to an absorbing boundary as the Cas protein inspects a target sequence for complementarity, a dimensionless ODE from a well posed IVP is solved to obtain exit times. With numerical approximations to the solution, numerical approximations of the exit time of variable absorbing boundary length are obtained through studies of Fokker Plank type equations, whose IVPs can be placed into correspondence with those of the Langevin equation [21].

Computations of mean exit, or passage, times have been previously applied under diverse geometrical and biological constraints, with one study detailing a procedure for the reconstruction of drift terms through a change of variables transformation of the backward Kolmogorov equation to the Schrodinger equation, in addition to a mapping into the Euler Lagrange equations which recovers potentials [2,8]. Numerical manipulations of the solution to the ODE for numerical approximations to the first passage time can be readily adapted to obtain passage times across other base pairs in the target sequence by numerically adjusting the upper limit of integration in the solution, which in the case of simple classes of potentials can be approximated by Gaussians, yielding estimations for the first passage time from Kramer's Result, with other studies of similarly posed diffusion processes obtained from solutions of the Smoluchowski equation in [18].

For CRISPR-Cas binding, an IVP corresponding to the first passage problem can be formulated by enforcing initial conditions which stipulate that the position of the particle undergoing diffusion is centered at the origin when target inspection is initiated, and that the particle subject to a unit initial velocity. As the protein inspection continues for remaining base pairs in the sequence, passage times can be computed by making use of numerical relations from the closed form of solutions to the ODE, primarily based in obtaining three variational formulas involving participation from several terms. To satisfactorily generate realistic binding energy landscapes that proteins encounter throughout inspection, we reflect upon separate approaches to determine mean exit times, from one approach which is capable of obtaining the exit time through numerical approximations of solutions to a stochastically driven oscillator, while another method raises an inverse problem, similar to that studied in [2], in which potential energy landscapes can be uniquely reconstructed from distributions of exit times. The inverse problem formulation presented here is focused towards the construction of the binding landscape potential from collections of exit times up to an absorbing membrane of variable length. Additional comparisons between probability measures, in which probability measures with another Hamiltonian, against the probability measure $p_i = \exp(\nabla U_i)/Z$ with potential $U_i$, will also be established.

## 2 Methodology

### 2.1 Description

The inverse problem poised towards reconstruction of the binding potential from exit time distributions relies on the following framework. To study the rates at which particles diffuse across base pairs in the binding process throughout crRNA inspection, solutions $\tau$ to the dimensionless, second order ODE of the form,

$$-\mathcal{A}\frac{\mathrm{d}^2\tau}{\mathrm{d}x^2}+\mathcal{U}'(x)\frac{\mathrm{d}\tau}{\mathrm{d}x}=1\ ,$$

are determined where the normalization introduced to obtain the dimensionless equation is proportional to the product of the Boltzmann constant and ambient temperature of bond melting, $\mathcal{A}\equiv\frac{k_bT}{\nu}$, $\mathcal{U}'(x)\equiv\frac{U'(x)}{\nu}$, and $U'(x)$ is the potential landscape before normalization by the driving force $\nu$. To specify classes of binding potentials for which solutions are to be determined, we impose the criterion that candidate potentials from the admissible landscape space possess one degree of freedom for each base pair at which binding occurs. In numerical applications of potential landscape reconstruction for mean exit time distributions, enforcing straightforward conditions on the mean and variance of the exit distributions themselves can respectively be achieved through specifying the first sample that is drawn from the time distribution, in addition to the maximum and minimum sample that can be drawn afterwards to specify the variance of the distribution which is also related to the fatness of its tails. For such classes of potentials, solutions to the ODE take the form,

$$\tau\approx\int_0^x\int_0^x\prod_{i=2}^{20}\exp\Big(-\frac{u^i}{i}\Big)\prod_{i=2}^{20}\exp\Big(\frac{v^i}{i}\Big)\mathrm{d}u\ \mathrm{d}v+\int_0^x\int_0^v\{\prod_{i=2}^{20}\exp\Big(-\frac{u^i}{i}\Big)\mathrm{d}u\}\prod_{i=2}^{20}\exp\Big(\frac{v^i}{i}\Big)\mathrm{d}v+\int_0^x\prod_{i=2}^{20}\exp\Big(-\frac{u^i}{i}\Big)\mathrm{d}u$$

indeed reflective of contributions from all nonzero polynomial terms specified in the candidate potential. Under simple rearrangements, variatonal formulae for the difference of exit times $\Delta_\tau=\tau_{x_2}-\tau_{x_1}$ are obtained through analyzing, on a case by case basis, the space of possible combinations of absorbing membrane lengths at arbitrary positions $x_1<x_2$ of the target sequence. Once the formulae have been established, further discussion will be devoted towards the construction of admissible distributions from which potential landscapes can be reconstructed. Before derivations of the variational formulae, we characterize the dependence between solutions and the class of potentials that we have identified, with solutions for varying arrangements of absorbing membranes.

Within the potential space, the goal of numerically obtaining mean exit times is to generate small perturbations to the energy landscape corresponding to small perturbations in the exit time. Regardless of experimental constraints in experiments that have been carried through measurements of the rate at which reactants are consumed in Fn Cas12a binding [29], the inverse problem of recovering the landscape can be numerically realized readily in several ways. First, one method involves producing approximations of the mean exit time from a given potential through Gaussian approximations on the integral terms from $\tau$, while another second closely related numerical approach involves numerically approximating the exit time after rearranging terms from $\tau$ through possible values on the innermost variable $v$ of integration from the second term in $\tau$. Third, another approach entails that we rearrange terms from $\tau$ depending on the position of the exit time of interest $v$, in which it is possible to make use of linearity of the integral to obtain variational formulae below. Before proceeding to the computations, we denote solutions $\mathcal{S}$ for the variational formulae with,

$$\mathcal{S}_{v\neq x}(v,x)\equiv-\int_0^x\Big(\int_v^x\{\prod_{i=2}^{20}2\ \exp\Big(-\frac{u^i}{i}\Big)\}\Big)\mathrm{d}u\prod_{i=2}^{20}\exp\Big(\frac{v^i}{i}\Big)\mathrm{d}v+\int_0^x\prod_{i=2}^{20}\exp\Big(-\frac{u^i}{i}\Big)\mathrm{d}u\ ,$$

which are readily obtained from rearrangement with the intermediate variable of integration $v$, in addition to specification of the position up to which the mean exit time is to be computed. From $\mathcal{S}_{v\neq x}$, terms from the numerical approximation of exit times are implemented in the following cases. In the formulae, the passage time up to the first position $x_1$, interactions over the passage time to $x_2$, and the intermediate interactions between the first and second passage times numerically contribute, from which variations in one exit time parameter generate classes of potential landscapes. For the inverse problem, at onset we require specification of one basis element in the potential space, and its corresponding exit time, in addition to the deviation from the exit time through specification of the second exit time. The potential corresponding to the second exit time can be recovered through numerical approximation of the variational formulae. To distinguish between potential landscapes, we denote free variables of the potential associated with the second exit time $\tau_2$ at $x_2$ with

$u_{x_2} \equiv u_2$, and similarly, free variables of the potential associated with the remaining exit time at an earlier position $x_1$ with $u_{x_1} \equiv u_1$. Before numerically approximating the final expression, we must evaluate the inner order of integration with intermediate variables $v_1$ and $v_2$. Tables corresponding to each formula in *Section 2* illustrate approximations associated with the exit time for potential recovery up to $x_1$, besides the upper limit of the second order of integration that can be numerically adjusted to obtain exit times of varying base pair length. Over previous works that have been mentioned, advantages of this approach include the flexibility to determine fluctuations in the energy landscape up to a base pair position at which the exit time is determined. Sampling at random from exit time distributions enables potential recovery with the variational formulae.

In the formulae below, depending on the case it is necessary that we isolate intervals over which the exit time is computed to ease numerical approximation by combining integral terms together. Without loss of generality, we determine the exit time associated with $x_2$. We introduce the straightforward relations for linearity amongst the inner, and outermost, variables, respectively, where interchanging the order in the outermost variable in the second term yields,

$$-\int_0^{x_2}\Big(\int_{u_2}^{x_2}\{\prod_{i=2}^{20}\exp\Big(\frac{v_2^i}{i}\Big)-1\}\mathrm{d}v_2\Big)\prod_{i=2}^{20}\exp\Big(\frac{u_2^i}{i}\Big)\mathrm{d}u_2 = -\int_0^{x_1}\Big(\int_{u_2}^{x_2}\{\prod_{i=2}^{20}\exp\Big(\frac{v_2^i}{i}\Big)-1\}\mathrm{d}v_2\Big)\prod_{i=2}^{20}\exp\Big(\frac{u_2^i}{i}\Big)\mathrm{d}u_2 - \int_{x_1}^{x_2}\Big(\int_{u_2}^{x_2}\{\prod_{i=2}^{20}\exp\Big(\frac{v_2^i}{i}\Big)-1\}\mathrm{d}v_2\Big)\prod_{i=2}^{20}\exp\Big(\frac{u_2^i}{i}\Big)\mathrm{d}u_2 \ . \tag{\textbf{Lin}}$$

## 2.2 (Var) equality

In the realization of the variational formula, solutions can be fashioned towards recovering potential landscapes for the passage times through a decrement of $\tau_{x_1}$ with $\tau_{x_2}$, in which $\Delta_\tau$ takes the form,

$$\Delta_\tau \equiv \tau_{x_2} - \tau_{x_1} = \mathcal{S}_{v\neq x}(v, x_2) - \mathcal{S}_{v\neq x}(v, x_1) \ ,$$

after which substitution for solutions $\mathcal{S}$ gives,

$$-\int_0^{x_2}\Big(\int_{u_2}^{x_2}\{\prod_{i=2}^{20}2\,\exp\Big(-\frac{v_2^i}{i}\Big)\}\Big)\mathrm{d}v_2\prod_{i=2}^{20}\exp\Big(\frac{u_2^i}{i}\Big)\mathrm{d}u_2 + \int_0^{x_2}\prod_{i=2}^{20}\exp\Big(-\frac{u_2^i}{i}\Big)\mathrm{d}u_2 + \int_0^{x_1}\Big(\int_{u_1}^{x_1}\{\prod_{i=2}^{20}2\,\exp\Big(-\frac{v_1^i}{i}\Big)\}\Big)\mathrm{d}v_1\prod_{i=2}^{20}\exp\Big(\frac{u_1^i}{i}\Big)\mathrm{d}u_1 - \int_0^{x_1}\prod_{i=2}^{20}\exp\Big(-\frac{u_1^i}{i}\Big)\mathrm{d}u_1 \ .$$

For exit times in which the length of the absorbing boundaries at exit times $\tau_{x_1}$ and $\tau_{x_2}$ for $x_1, x_2$ along the genome, the variational formula **(Var)** permits for solutions to the inverse exit time problem through specification of the exit time parameters and distribution. We further rearrange terms to obtain the desired relation, through relevant applications of **(Lin)** to collect like terms,

$$\int_0^{x_2}\{\int_{u_2}^{x_2}\prod_{i=2}^{20}2\exp\Big(-\frac{v_2^i}{i}\Big)\mathrm{d}v_2+1\}\prod_{i=2}^{20}\exp\Big(-\frac{u_2^i}{i}\Big)\mathrm{d}u_2 - \int_0^{x_1}\{\int_{x_1}^{u_1}\{\prod_{i=2}^{20}2\,\exp\Big(-\frac{v_1^i}{i}\Big)\mathrm{d}v_1-1\}\}\prod_{i=2}^{20}\exp\Big(-\frac{u_1^i}{i}\Big)\mathrm{d}u_1 \ . \tag{\textbf{Var}}$$

## 2.3 Numerical test cases

We review the composition of each variational relation for potential recovery.

- **(Var), (Var1)**: The first two variational formulas encapsulates straightforward numerical behaviors of the fluctuations in the potential landscape from corresponding fluctuations in the exit time. From the expression, fluctuations up to first order are captured from terms in the potential expansion.

| Formula | Recovery formulation $\mathcal{R}$ of $u_1$ terms of potential |
|---|---|
| **Var** | $\Delta_\tau - \int_0^{x_2}\{\int_{u_2}^{x_2}\{\prod_{i=2}^{20} 2\,\exp\big(-\frac{v_2^i}{i}\big)\}\mathrm{d}v_2 + 1\}\prod_{i=2}^{20}\exp\big(-\frac{u_2^i}{i}\big)\mathrm{d}u_2$ |

Table 1: *General formula for potential recovery from* **(Var)**. The relation is obtained through straightforward rearrangements of $\Delta_\tau$, returning an expression for the potential energy landscape up to $x_1$ . To readily apply the formula, we introduce approximations of $\exp\big(\sum_{i=2}^{20} -\frac{u_1^i}{i}\big) \sim 1 + \frac{\big(\sum_i -\frac{u_1^i}{i}\big)^2}{2} + \cdots + \frac{\big(\sum_i -\frac{u_1^i}{i}\big)^n}{n} + \cdots$. We make use of this guiding approximation to obtain recovered modes in *Table 2* below. Fluctuations of the energy landscape are obtained by incorporating terms from the mode approximation of the landscape potential as the position of exit time along the target sequence increases.

| Formula | $x_1, x_2$ | Approximation of recovered modes |
|---|---|---|
| **Var1**, | $x_2 > 1, x_1 \equiv 1$ | $\mathcal{R} \approx \sum_{i=2}^{20}\left(-\frac{x_1^{i+2}}{(i+1)(i+2)} + \frac{x_1^{i+1}}{i+1}\right)$ |
| **Var2** | $x_2 \equiv 1, x_1 < 1$ | $\mathcal{R} \approx \mathbf{Var1} + \frac{1}{2}\sum_{i=2}^{20}\frac{x_1^{3i+2}}{i^2(i+1)}\left(\frac{1}{3i+2} - \frac{1}{2i+1}\right)$ |
| **Var3** | $x_2 > x_1, x_1 > 1$ | $\mathcal{R} \approx \mathbf{Var2} + \frac{1}{3}\sum_{i=2}^{20}\frac{x_1^{i+5}}{3i+3}\left(\frac{1}{i+5} - \frac{1}{4}\right)$ |
| general **Var** | $x_2 > x_1, x_1 > 1$ | $\mathcal{R} \approx \mathbf{Var3} + \sum_{j\in\mathcal{M}_{-3}}\sum_{i=2}^{20}\frac{x_1^{i+j+2}}{j(i+1)}\left(\frac{1}{i+j+2} - \frac{1}{j+1}\right)$ |

Table 2: *Numerical instances of* **(Var)** *with membranes of absorbing unit or varying boundary length.* In each instance, Variational formulae from **(Var)** are obtained from enforcing numerical approximations with the corresponding values of $x_1$ and $x_2$. From the specification of these free parameters, corresponding numerics can be performed to recover all corresponding potential modes, in light of the formulation from *Table 1.* Fluctuations of the potential landscape from the exit time corresponding to $x_1$ can be approximated from the deviation between exit times $\tau_{x_1}$ & $\tau_{x_2}$. The approximation returns a polynomial in $x_1$. As a generalization of the approximation for recovered modes, in the final row an expression for additional terms in the expansion is provided. The $j$ summation is taken over the collection of modes $\mathcal{M}$ indexed by the naturals, and $\mathcal{M}_{-3} = \mathcal{M}\backslash\{1,2,3\}$.

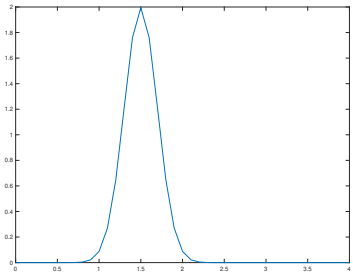

Figure 1: Normal distribution from which time samples can be drawn for approximation of time increments $\Delta_\tau$.

- **(Var2)**: For exit times at $x_1$ & $x_2$, the second variational formula encapsulates higher order fluctuations in the landscape with an additional term in the series expansion. For convenience, we rearrange terms by collecting powers of the common $x_1$ term, while leaving the remaining fractional terms separately.
- **(Var3)**: For exit times at $x_1$ & $x_2$, cubic order terms from the expansion in **(Var3)** capture fluctuations farther along the target sequence.
- general **(Var)**: For exit times at $x_1$ & $x_2$, the final variational formula is comprised of a mixture of contributions similar to the previous ones, in which contributions from the series are indexed by $j$ which runs along $\mathcal{M}$ which could contain an arbitrary number of modes. All previous terms in **(Var1), (Var2), (Var3)** can be determined by respectively taking the $j = 1, 2, 3$ mode from the series.

### 2.4 Generating exit time distributions

We readily generate exit time distributions from which potential landscapes will be recovered. In a suitable free parameter generalization for different Cas proteins, the exit time formalism must be capable of determining the perturbation in the landscape given an initial potential choice with degrees of freedom at each base pair. The three variational formulae that have been presented are capable of exectuting the potential reconstruction, in addition to other quantities pertaining to the drift terms associated with well characterized properties of protein kinetics [2].

This final subsection is devoted towards not only a description of how the visit distributions for the exit times can be generated, but also how apriori choicess of the exit times at both positions, in addition to the potential corresponding to the first exit time, can be enforced. The implementation of our approach is capable of recovering potentials associated with the free exit time, permitting for numerical simulations of energy landscapes for different Cas proteins. To numerically implement the variational formulae for landscape reconstruction, we make use of individual instances of the relation described extensively above. Primarily, we are interested in determining the order of difference in the exit times across subsequences of the target sequence that are at least one base pair long, despite being able to compute orders of magnitude of difference between exit times across shorter nucleotide lengths than 1. To identify the order of fluctuations within the parameter space that are valuable for physical interpretation, we devote the most attention towards cases of the formulae which are applicable in experiments for analyzing curvature of the energy landscape.

### 2.5 Procedure

We provide numerical results and interpretations in light of the relations given in the previous section. The ingredients are listed below.

#### 2.5.1 Representative distributions

We execute numerical experiments to recover potential landscapes from candidate distributions. Normal distributions as shown in *Figure 1* are one suitable distribution class given that any exit time can be sampled from the distribution within two standard deviations to obtain fluctuations in the energy landscape. To recover corresponding potential landscapes, we evaluate terms of the series given in **(Var1), (Var2), (Var3),** general **(Var)**, in turn obtaining $x_1$ variables for the potential which are normalized by terms dependent on $i$. In the vanishing limit of the variance of the distribution, the concentration of exit times about the mean of the distribution provides more fine grain numerical measures of the perturbation to the energy landscape at given positions of exit $x_1$ and $x_2$.

### 2.6 Computations for potential recovery at $x_2$

To recover approximations of landscape modes, we further illustrate how sampling from typical distributions, as shown above, enable for efficient landscape recovery. First, it is necessary that we specify the initial exit time up to $x_1$, from

which the corresponding potential landscape can be obtained. Recall that from approximations of $\tau$ for solutions of the dimensionless second order ODE, specifying $\tau_{x_1}$ readily yields the following landscape associated with the exit time, through rearrangements for the potential modes given a polynomial locus in $u_1$. Term by term we determine the landscape associated with $x \equiv x_1$, from the approximations of $\tau$. We analyze contributions from well posed solutions through numerical investigation of the relations provided in *Table 2*.

## 3 Numerical experiments

To further study recovered modes of landscapes, we implement the procedure to estimate $\mathcal{R}$ from *Table 1*, to then apply higher modes of the landscape given in *Table 2*. Specifically, we establish the following criterion to generate theoretical predictions with regards to fluctuations of the landscape up to positions $x_2$ given initial data of the exit time magnitude and potential landscape at $x_1$.

## 4 References


[1] Allison, D. & Wang, G. R-loops: formation, function, and relevance to cell stress. *Cell Stress* **3**(2) (2019).

[2] Bal, G. & Chou, T. On the reconstruction of diffusions from first-exit time distributions. *Inverse Problems* **20**(4) (2004).

[3] Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., & Phillips, R. Transcriptional regulation by the numbers: applications. *Current opinion in genetics & development*, **15**(2), 125–135 (2015).

[4] Borys, P. & Grzywna, Z. The Fokker-Planck Equation for Chaotic Maps. *Acta Physica Polonica B* **37**(2) (2006).

[5] Brewester, R., Weinert, F., Garcia, H., Song, D., Rydenfelt, M. & Phillips, R. The Transcription Factor Titration Effect Dictates Level of Gene Expression. *Cell* **6**, 1312-1323 (2014).

[6] Chen, J., Dagdas, Y., Kleinstiver, B., Welch, M., Sousa, A., Harrington, L., Sternberg, S., Joung, J., Yildiz, A. & Doudna, J. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*, **550**, 407-410 (2017).

[7] Chen, B., Gilbert, L., et al. Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell*, **155**(7) 1479-1491 (2013).

[8] Chou, T. & D'Orsogna, M. First Passage Problems in Biology. *Arxiv* (2014).

[9] D'Orsogna, M., Lakatos, G. & Chou, T. Stochastic self-assembly of incommensurate clusters. *J Chem Phys* **136**(8) (2012).

[10] Eitzinger, S., Asif, A., Watters, K. E., Iavarone, A. T., Knott, G. J., Doudna, J. A., & Minhas, F. Machine learning predicts new anti-CRISPR proteins. *Nucleic acids research*, **48**(9), 4698–4708 (2020).

[11] Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Yaung, S. J., & Church, G. M. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nature methods, **10**(11), 1116–1121 (2013).

[12] Eslami-Mossallam, B., Klein, M., Smagt, C., Sanden, K., Jones Jr, S., Hawkins, J., A kinetic model improves off-target predictions and reveals the physical basis of SpCas9 fidelity. *Preprint.*

[13] Garneau, J. et al. the CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468** 67-71 (2010).

[14] Horlbeck, M. A., Witkowsky, L. B., Guglielmi, B., Replogle, J. M., Gilbert, L. A., Villalta, J. E., Torigoe, S. E., Tjian, R., & Weissman, J. S. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. eLife, **5**, e12677 (2016).

[15] Hultquist, J. F., Hiatt, J., Schumann, K., McGregor, M. J., Roth, T. L., Haas, P., Doudna, J. A., Marson, A., & Krogan, N. J. CRISPR-Cas9 genome engineering of primary CD4+ T cells for the interrogation of HIV-host factor interactions. *Nature protocols*, **14**(1), 1–27 (2019).

[16] Jackson, S., Suma, A. & Micheletti, C. How to fold intricately: using theory and experiments to unravel the properties of knotted proteins. *Current Opinion in Structural Biology* **42**, 6-14 (2017).

[17] Jeon, Y. et al. Direct observation of DNA target searching and cleavage by CRISPR-Cas12a. *Nature Communications*, **9**, 2777 (2018).

[18] Keener, J. Mathematical Biology Course Lecture Notes.

[19] Kim, B., Kim, H. & Lee, S. Regulation of Microbial Metabolic Rates Using CRISPR Interference with Expanded PAM Sequences. *Frontiers in Microbiology* **11**(282), (2020).

[20] Kinney, J., Tkacik, G. & Callan, G. Precise physical models of protein-DNA interaction from high-throughput data. *PNAS*, **104**(2), 501-506 (2007).

[21] Krapivsky, P., Redner, S. & Ben-Naim, E. A Kinetic View of Statistical Physics. *Cambridge University Press* 978-0-521-85103-9 (2010).

[22] Mallamace, F. et al. Energy landscape in protein folding and unfolding. *PNAS*, **113**(12), 3159-3163 (2016).

[23] Mekler, V., Minakhin, L., Semenova, E., Kuznedelov, K., & Severinov, K. Kinetics of the CRISPR-Cas9 effector complex assembly and the role of 3'-terminal segment of guide RNA. *Nucleic acids research*, **44**(6), 2837–2845 (2016).

[24] Mirny, L. & Shakhnovich, E. Protein Folding Theory: From Lattice to All-Atom Models. *Annual Reviews Biophysics*, **30** 261-96 (2001).

[25] Morra, G., Genoni, A. & Colombo, G. Protein Dynamics and Drug Design: The Role of Molecular Simulations. *Protein-Protein Complexes*, 340-385 (2010).

[26] Peled, R. Topics in Statistical Physics and Probability Theory. *Lecture Notes*.

[27] Ran, F., Hsu, P., Wright, J. Agarwala, V., Scott, D. & Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8** 2281-2308 (2013).

[28] Satori, P. & Leibler, S. Lessons from equilibrium statistical physics regarding the assembly of protein complexes. *PNAS* **117**(1) 114-120 (2020).

[29] Spetcht, D., Xu, Y. & Lambert, G. Massively parallel CRISPRi assays reveal concealed thermodynamic determinants of dCas12a binding. *PNAS* **117**(21) 11274-11282 (2020).

[30] Slutsky, M. & Mirny, L. Kinetics of Protein-DNA Interaction: Facilitated Target Location in Sequence-Dependent Potential. *Biophysical Journal*, **87** 4021-4035 (2004).

[31] Stella, S. et al. Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell* **175** 1856-1871 (2018).

[32] Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**(7490), 62–67 (2014).

[33] Tkacik, G., Callan, C. & Bialek, W. Information flow and optimization in transcriptional regulation. *PNAS*, **105**(34), 12265-12270 (2008).

[34] Wolf, S., Lickert, B., Bray, S. & Stock, G. Multisecond ligand dissociation dynamics from atomistic simulations. *Nature Communications*, **11**, 2918 (2020).

[35] Xu, X., Duan, D. & Chen, S. CRISPR-Cas9 cleavage efficiency correlates strongly with target-sgRNA folding stability: from physical mechanism to off-target assessment. *Sci Rep* **7**, 143 (2017).