Computation of exit times for potential landscapes in CRISPR binding

Pete Rigas, Abstract # C12.00006, Session C12: Physics of Proteins I

March 15, 2021

LAMBERT LAB.io
QUANTITATIVE SINGLE-CELL BIOPHYSICS

Cornell University
School of Applied and Engineering Physics

Binding thermodynamics

Fn Cas12a

- Target sequences are 20 base pairs long.
- The protein inspects target sequences upstream of a transcription factor, from which inspection for complementarity in the target sequence, and potential formation of a stable nuclease, occur.

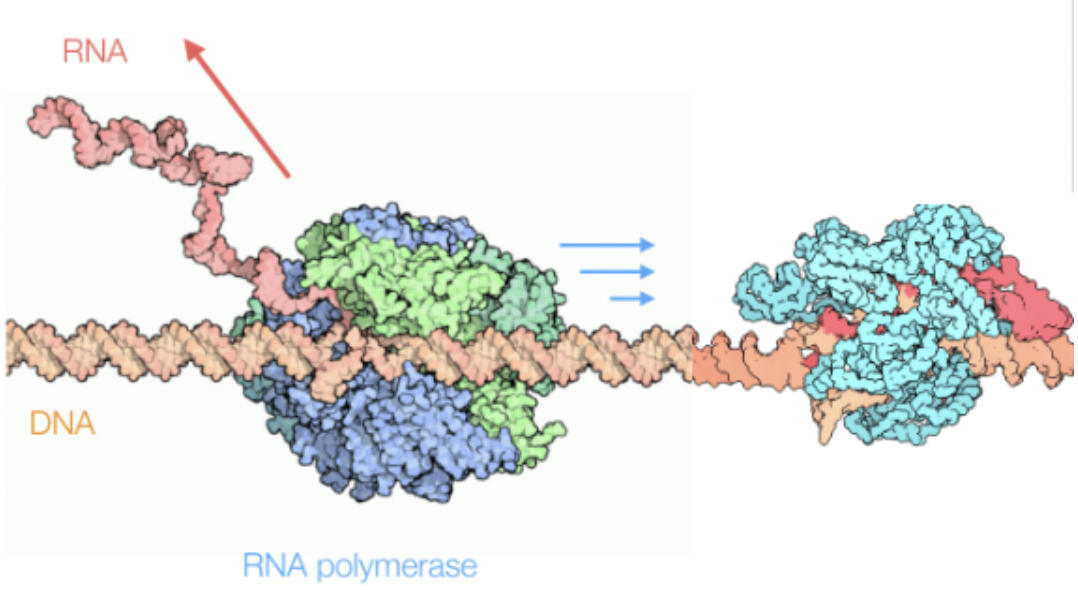


Figure 1: A CRISPR system diffusing in one dimension while inspecting the target sequence.

Introduction

- Exit/passage time distributions $\tau(x)$,
- Simulations of protein inspection process governed by,

$$\frac{\mathrm{d}^2\tau}{\mathrm{d}x^2} - \frac{\mathcal{U}'(x)}{k_b T}\frac{\mathrm{d}\tau}{\mathrm{d}x} = -\frac{1}{k_b T} \ ,$$

under landscape $\mathcal{U}'(x)$, $\tau(n) = 0$ and $\tau'(0) = 0$, $n \in \mathbf{Z}$.

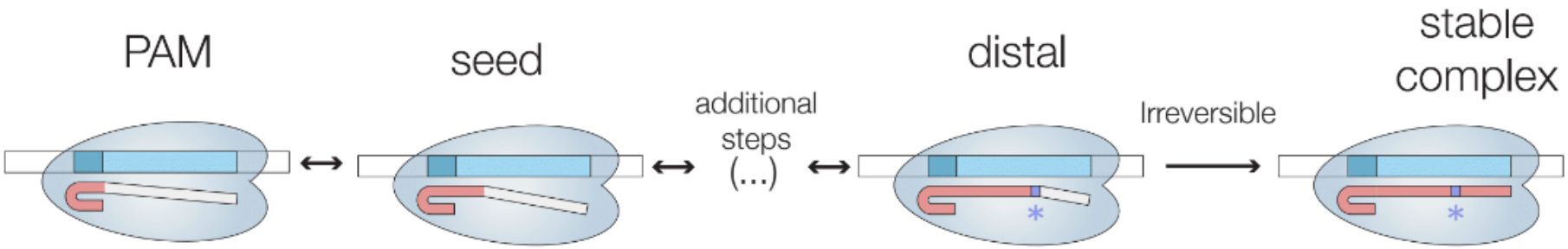


Figure 2: A first passage process, in which the $n = 0$ step of the landscape is determined by the PAM inspection of the target sequence, with subsequent $n = 1, 2, \cdots$ steps resulting in the formation of a DNA bubble in which a complex stabilizes.

Goals

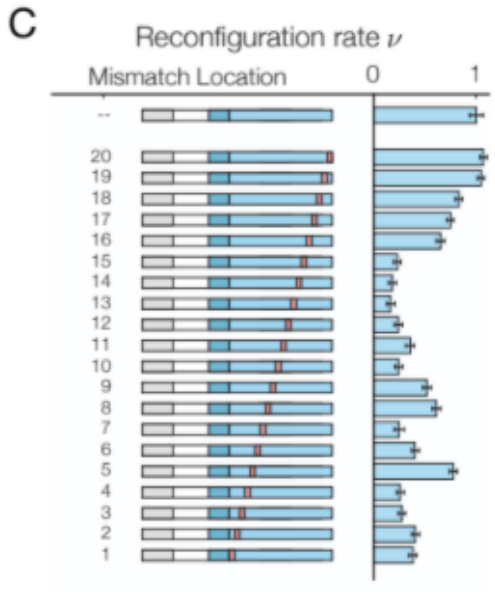
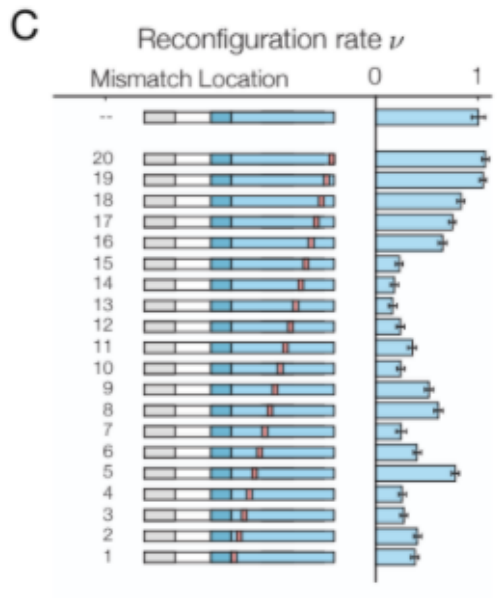


Figure 3: Landscape barrier costs for base pair mismatches.

- ☐ Simulate exit time dynamics for target sequences of variable length, in order to,
  - ☐ Verify predictions of Fn Cas 12a binding dynamics for 20 bp target sequences, dictating that,
    - (Region I) mismatches within the first 6 base pairs of the landscape pose the highest detriment to nuclease formation,



- ☐
  - (Region II) mismatches at base pairs that are either consecutive, or within a one to two base pair window, after the first 6 base pairs significantly impact nuclease formation,
  - (Region III) mismatches in the last several bases of the landscape pose minimal effect on the binding process (seed region from the last 6 rows).
- ☐ Introduce a temperature dependence on approximations.

δ landscape potential

Such a class of potentials

- resembles the energy barriers of binding, through fixed barrier heights in the landscape between neighboring base pairs throughout the inspection process,
- produces solutions τ dependent on the magnitude of the jump discontinuity of the landscape between neighboring base pairs,

$$\tau(x) = \frac{x}{|\mathcal{U}'(x_1)|} - \frac{\exp(|\mathcal{U}'(x_1)|\ x)}{|\mathcal{U}'(x_1)|^2} + \frac{1}{|\mathcal{U}'(x_1)|^2}\ ,$$

where the landscape $\mathcal{U}'(x)$ is evaluated at the barrier corresponding to inspected base pair x_1 in the target sequence (**similar to solutions obtained from preparing distributions over the landscape**)

Simulating energy barriers of variable height through the variance of the landscape distributions

With the following steps, we will control the height of the energy barrier for each step of protein inspection by introducing standard normal random variables as each base pair is inspected.

☑ Simulate exit time dynamics for target sequences of variable length.

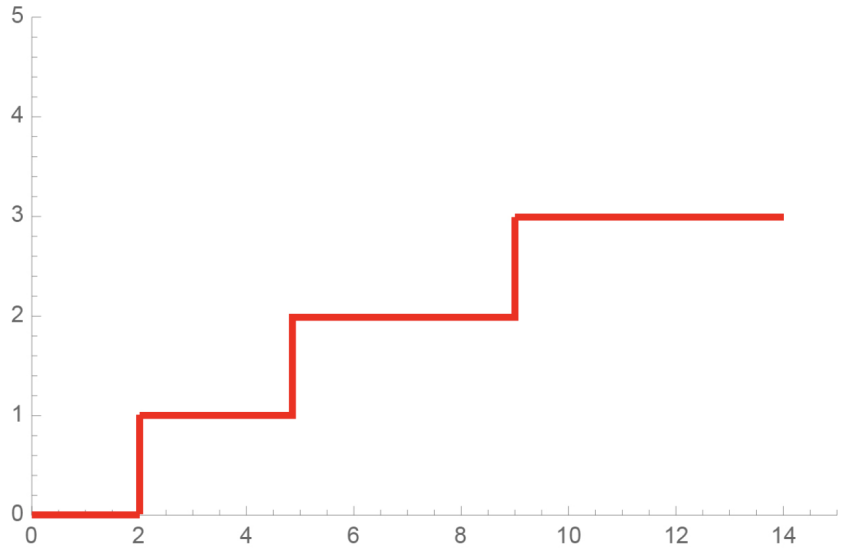


Figure 4: Typical δ landscape, over which normally distributed approximations simulate barrier heights.

Preparing Gaussian distributions over the landscape (0 & 1)

- Step 0: Set the mean of each Gaussian within some tolerance of the inspected base pair, from ongoing inspection of the target sequence.
- Step 1, *Gaussian translation invariance*:
  Consider the pmf horiztonally shifted upwards h units,

$$p(x) = \frac{\exp(-(\frac{x-\mu}{\sigma})^2)}{\sqrt{2\pi\sigma^2}} + h = \frac{\exp(-(\frac{x-\mu}{\sigma})^2) + \exp(\log\sqrt{2\pi\sigma h^2})}{\sqrt{2\pi\sigma^2}},$$

  equivalent to,

$$p(x) = \frac{\exp\{-(\frac{x-\mu}{\sigma})^2\}}{\sqrt{2\pi\sigma^2}}\left(1 + \exp\{\log\sqrt{2\pi\sigma h^2} + (\frac{x-\mu}{\sigma})^2\}\right),$$

  with $h = \mathcal{U}'(x_{i+1}) - \mathcal{U}'(x_i)$.

(2 & 3)

- Step 2, *numerically approximate exit times*: Solve the second order dimensionless equation with $\mathcal{U}'(x) = p(x)$, $\tau^i(0) = 0$ if $i = 1$ and $\tau^i(0) = \tau^{i-1}(1)$ for $i \geq 2$.

  Remark: Conveniently, $\boxed{\sup_x\{p(x)\} \approx \frac{0.4}{\sigma}}$

- Step 3, *Collected approximations from the simulations*: Characterize the dependence of the exit times τ in correspondence with variable energy barriers, constructed in Steps 0-2.

Analytical dependence of solutions τ for absorbing boundary of length n

With normally distributed approximations, the exit time solution takes the form,

$$\tau(x) = \frac{1}{k_b T}\Bigg(\underbrace{n^2}_{=\int_0^n \int_0^n f(u)f(-v)\mathrm{d}u\ \mathrm{d}v} - \underbrace{x^2}_{=\int_0^x \int_0^x f(u)\ f(-v)\ \mathrm{d}u\ \mathrm{d}v} - \underbrace{\int_x^n f(v) \int_0^v f(u)\ \mathrm{d}u\ \mathrm{d}v}_{\approx \int_0^x f(v) \int_0^v f(u)\ \mathrm{d}u\ \mathrm{d}v - \int_0^n f(v) \int_0^v f(u)\ \mathrm{d}u\ \mathrm{d}v} \Bigg),$$

where

$$f(u) = \exp\left(\frac{\sqrt{\pi\sigma^2}}{2k_b T} \left(\frac{2}{\sqrt{\pi}} \int_0^{\frac{\mu-u}{\sigma}} \exp(-t^2)\ \mathrm{d}t \right) \right)$$

Recursive modification

- Solution decomposition

$$\tau(0) \equiv \tau(n-1) + \sum_{j<n-1} \tau(j)$$

- Exit time fluctuation

$$\Delta_{n,n-1}\tau \equiv \frac{1}{k_b T}\left(n^2 - (n-1)^2 - \int_{n-1}^{n} f(v) \int_0^v f(u)\, \mathrm{d}u\, \mathrm{d}v \right) .$$

Initial configuration

- From $f(u)$, observe

$$\frac{\sqrt{\pi\sigma^2}}{2k_bT}\frac{2}{\sqrt{\pi}} \approx \frac{\sigma}{k_bT} \approx \frac{0.4}{h\ k_bT} \approx \boxed{\frac{0.4}{hT}}\ .$$

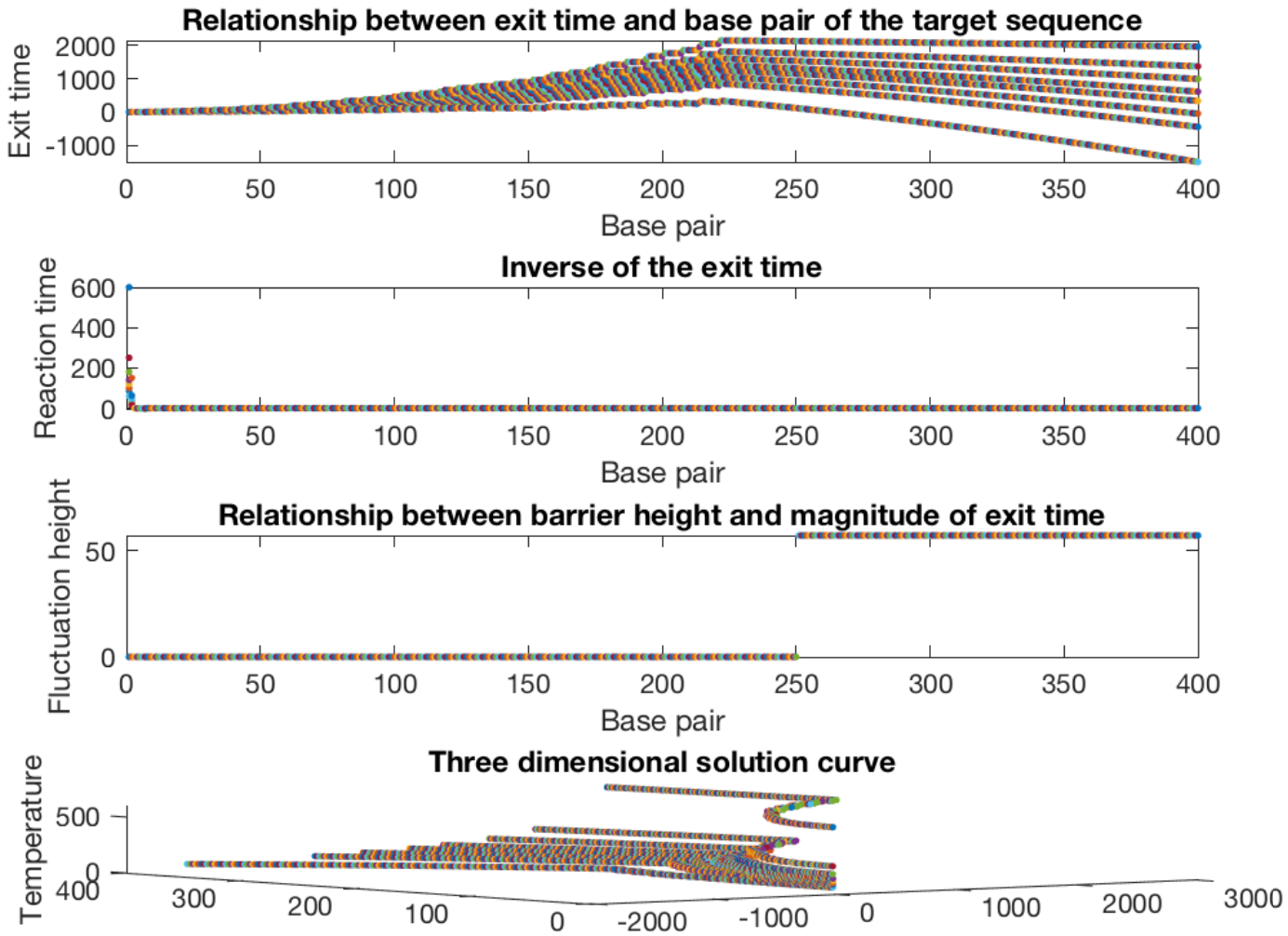
- For the orange term on the previous slide, introduce the trapezoidal approximation,

$$\int_0^v f(u)\ \mathrm{d}u \approx \frac{u}{2}\left(f(0) + f(\frac{u}{2}) + f(u)\right)\ .$$

- Set $h_i = 0.004$, $h_f = 25.004$, $\Delta x = 0.1$, $T \in \{60, 80, 100, 120, 600\}$.

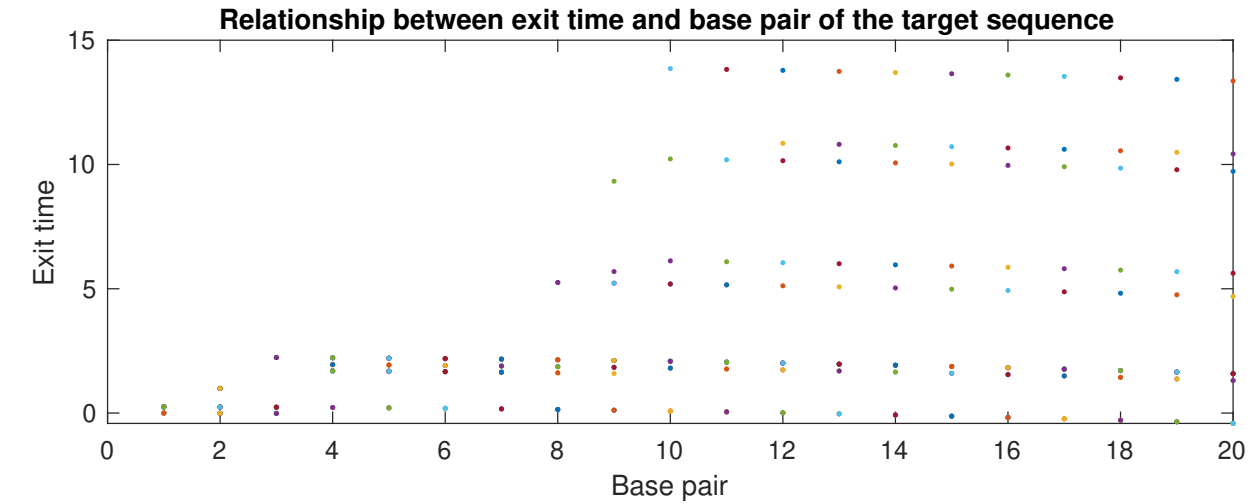
Simulation approximations for $\approx$ 400 bp target sequences (8 temperature level sets)

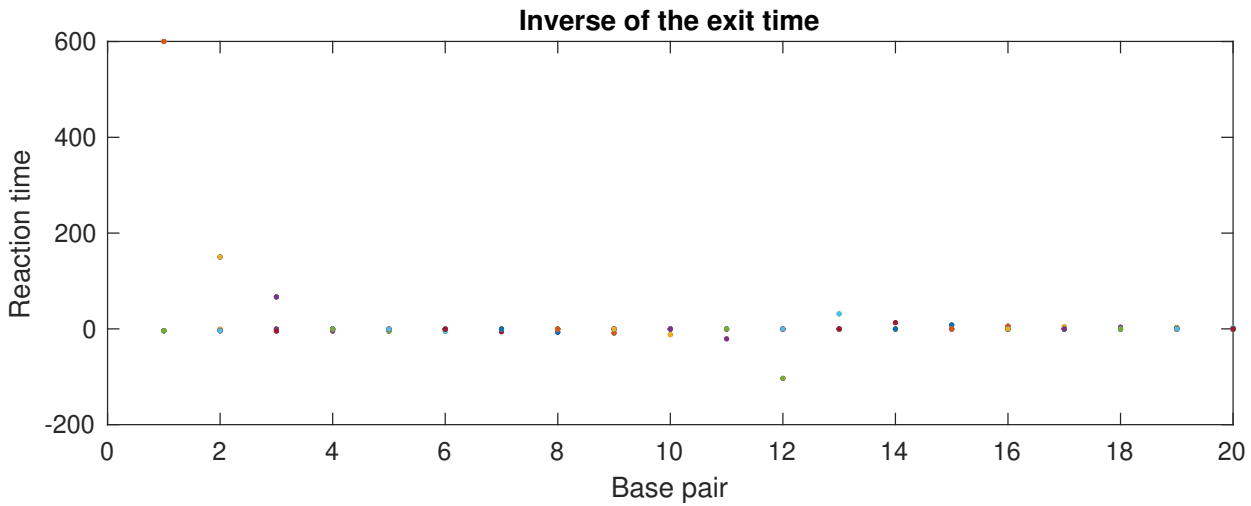
☑ Introduce a temperature dependence on approximations.



Passing to a height subsequence of the landscape for Fn Cas12a specific predictions

☑ Base pair mismatch-reconfiguration rate associated with Regions I, II, III





Concluding remarks

- Binding dynamics (*Figures 1-4*),
- analytical solutions to first passage problems with variable absorbing boundary length,
- simulations of Cas12a binding, with aspirations to apply numerical properties of similar analytical solutions to sequencing for other CRISPR Cas members from my colleague David.

Acknowledgements

Thank you to all members of the Lambert Lab.

Thank you for you listening. Are there any questions or remarks from the audience?