determining mean exit times from solutions of the dimensionless oscillator

Pete Rigas, Lambert Lab

October 14, 2020

since last time...

to accommodate experimental measurements,

- desirable properties of potentials will be introduced,
- solutions to a damped oscillator will be obtained,
- we will physically interpret the duration of the mean time by examining 4 individual pieces from solutions to the second order ODE,
- detail the variation in the mean first exit times that are obtained from different binding sequences, which is reflective of particular solutions to the oscillator depending on the assumption of the potential,
- and finally, describing ways in which the mean exit time for one protein can be adapted to qualitatively describe diffusion processes in other similar proteins within the same family

objectives

- phrase solutions to the differential equation as an IVP over a subinterval of the genome to formulate thermodynamic interactions between neighboring nucleotides,
- introduce solutions to the differential equations with "over simplified" assumptions on the potential of the energy landscape, from which more complexity can be obtained with higher degree potentials that can be adapted to distinct proteins,
- delineate recognizable patterns in the solutions to the second order differential equation from a non constant potential energy term,
- explore polynomial interpolation techniques as a means to introduce more suitably defined potentials for proteins for which data does not exist, as a means to predict the different in protein kinetics

IVP, solutions of the first and second types for potentials with a few polynomial terms

Solutions to the dimensionless second order differential equation,

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

correspond to,

$$(n+1)^{\frac{1}{n+1}}\,\mathrm{i}\Gamma\left(\frac{1}{n+1}, \frac{x^{n+1}}{n+1}\right)\int_0^{t_{\mathrm{crit}}} \frac{\exp\left(\frac{u^{n+1}}{n+1}\right)}{n+1}\mathrm{d}u$$
$$-\int_0^{t_{\mathrm{crit}}} \frac{(n+1)^{\frac{1}{n+1}}\exp\left(\frac{u^{n+1}}{n+1}\right) i\Gamma\left(\frac{1}{n+1}, \frac{u^{n+1}}{n+1}\right)}{n+1}\mathrm{d}u + 1 \ ,$$

where n is the highest degree in the potential, or alternatively,...

continued...

for different choices of a potential, the solutions are of the form,

$$= \int_0^{t_{\rm crit}} \exp\Big(-\mathbf{1} \Big) \mathrm{d}u \int_0^{t_{\rm crit}} -\exp\Big(\mathbf{1} \Big) \mathrm{d}u$$
$$+ \int_0^{t_{\rm crit}} \int_0^{v} \exp\Big(\mathbf{1} \Big) \exp\Big(-\mathbf{1}' \Big) \mathrm{d}v \mathrm{d}u + 1 \ ,$$

where

$$\mathbf{1} := \frac{u^n(\text{higher order terms})}{\text{normalization}} \ ,$$
$$\mathbf{1}' := \frac{v^n(\text{higher order terms})}{\text{normalization}} \ ,$$
$$|\text{higher order terms}| = \max_n\{ n \text{ such that } \exists\ \alpha_i x^i \text{ in U } \} \ ,$$

which can be numerically evaluated to determine the first mean ...

... exit time corresponding to the expected average time for particles to reach an absorbing boundary, of unit distance, in which chemical reactants are consumed up to some time t_{crit}, n is the degree of the highest term in the potential, and with boundary conditions $F(0) = 0$, and $\frac{dF}{dx}(0) = 1$.

Rmk: to numerically obtain mean exit times, observe:

- that the 2 integrals in solutions of both the first and second types involve integrals in the second term of the solution which has one dummy variable which is a parameter to numerically determine the first mean exit time that does not have to be specified or tuned,
- qualitatively, we can interpret solutions of the first passage problem for binding by recognizing that in solutions of the first and second types, the numerical discrepancy between the first and second terms is dependent on $|v - t_{\mathrm{crit}}|$,
- can alternatively be expressed as $\int_0^{t_{\mathrm{crit}}} \exp(-\mathbf{1}) \left(\int_v^{t_{\mathrm{crit}}} \exp(\mathbf{1}) \mathrm{d}v \right) \mathrm{d}u$

how the potential motif impacts solutions

for 20 degree potentials for the nucleotide length of the binding sequence, potential terms in the power of the exponent give solutions with more terms,

$$\left(\int_0^x \exp\left(-\sum_{i=2}^{20}\frac{u^i}{i}\right)\mathrm{d}u\right)\left(\int_0^x -\exp\left(\sum_{i=2}^{20}\frac{u^i}{i}\right)\mathrm{d}u\right)$$
$$+\left(\int_0^x \exp\left(\sum_{i=2}^{20}\frac{v^i}{i}\right)\mathrm{d}v\right)\left(\int_0^v \exp\left(-\sum_{i=2}^{20}\frac{u^i}{i}\right)\mathrm{d}u\right)$$
$$+\int_0^x \exp\left(-\sum_{i=2}^{20}\frac{u^i}{i}\right)\mathrm{d}u\ ,$$

in contrast to solutions for less complicated potentials, an additional term $\neq 1$ appears

rearranging terms in the $v \neq x$ case

from the solution, 2 of the terms can be combined,

$$-\int_0^x \exp\left(\prod_{i=2}^{20} \frac{u^i}{i}\right)\left(\int_0^x \{\prod_{i=2}^{20} \exp\left(\frac{u^i}{i}\right) - 1\}\mathrm{d}u\right)\mathrm{d}u \ ,$$

(Product P)

yielding, for $v \neq x$,

$$\mathcal{S}_{v\neq x}(v,x) = (P) + \left(\int_0^x \exp\left(\sum_{i=2}^{20} \frac{v^i}{i}\right)\mathrm{d}v\right)\left(\int_0^v \exp\left(-\sum_{i=2}^{20} \frac{u^i}{i}\right)\mathrm{d}u\right)$$

(1)

$v \equiv x$ case

enforcing this assumption gives solutions,

$$\boxed{\mathcal{S}_{v=x}(v,x) = -\int_0^x \exp\left(\prod_{i=2}^{20}\frac{u^i}{i}\right)\left(\int_0^x \{2\prod_{i=2}^{20}\exp\left(\frac{u^i}{i}\right) - 1\}\mathrm{d}u\right)\mathrm{d}u} \tag{2}$$

passage time between distinct base pairs

- **(1A)**

$$v \neq x_1, v \neq x_2 \Rightarrow -\int_{x_1}^{x_2} \exp\left(\prod_{i=2}^{20} \frac{u^i}{i}\right)\left(\int_{x_1}^{x_2} \{\prod_{i=2}^{20} \exp\left(\frac{u^i}{i}\right) - 1\}\mathrm{d}u\right)$$

$\Leftrightarrow \mathcal{S}_{v\neq x}(v, x_2) - \mathcal{S}_{v\neq x}(v, x_1)$

- **(1,2)** $v \neq x_1, v \equiv x_2 \Rightarrow (\mathbf{1})|_{x_1} - (\mathbf{2})|_{x_2}$

$\Leftrightarrow \mathcal{S}_{v\neq x}(v, x_1) - \mathcal{S}_{v=x}(v, x_2)$

- **(2)** $v \equiv x_1, v \neq x_2 \Rightarrow -\mathbf{(1,2)}$ $\Leftrightarrow \mathcal{S}_{v=x}(v, x_1) - \mathcal{S}_{v\neq x}(v, x_2)$

$v \equiv x_1, v \equiv x_2$ is not a possible case

previous works in the literature relating to CRISPR binding, energy landscapes, protein folding

- *Robert Brewster et al*, **The Transcription Factor Titration Effect Dictates Level of Gene Expression** (2014) [related to Miles' talks about plasmid copy number, transcription, metabolic function, ColE1 OR, in addition to my work on the thermodynamics end to construct relevant probability measures to measure the likelihood of melting of base pairs in DNA],
- *Behrouz Eslami-Mossallam et al*, **A kinetic model improves off-target predictions and reveals the physical basis of SpCas9 fidelity** [significant for quantifying the effect of off target effects in binding for Cas proteins, which can be examined through plots of the free energy landscape for different proteins which is completely dependent on the formulation of the partition function],

more works...

- *Francesco Mallamace et al*, **Energy landscape in protein folding and unfolding** (2015) [significant for quantifying, for thermodynamic reflection in our model, the role that base pairs, rather than the first 6 in 20 bp sequence, play in ensuring the ability of Cas12a and other Cas proteins to attain native state configurations],
- *Yongmoon Jeon et al*, **Direct observation of DNA target searching and cleavage by CRISPR-Cas12a** (2018) [commentary on different stages of the binding process for Cas proteins, related to the fact that cleavage steps in the binding process differ between Cas proteins, ie different types of cuts and varying lengths of sequences which are targeted]

for real solutions to the boundary problem,

- qualities of admissible candidate potentials will be introduced,
- a simple algorithm and conditions on the number of iterations of the algorithm will be provided,
- lastly, connections with the Hamiltonian in previous approaches, through mapping the thermodynamics of particles undergoing Brownian motion to an absorbing boundary with [as a novel direction not yet explored in the literature, instead of comparing the performance of a closed form Hamiltonian with other approaches, so as to establish connections between the expected time of first passage up to a boundary of length L with the expected number of visits of the random walk]

constant potential toy model

Under the assumption of a constant potential for all base pairs, the mean first exit time is obtained from solutions F to

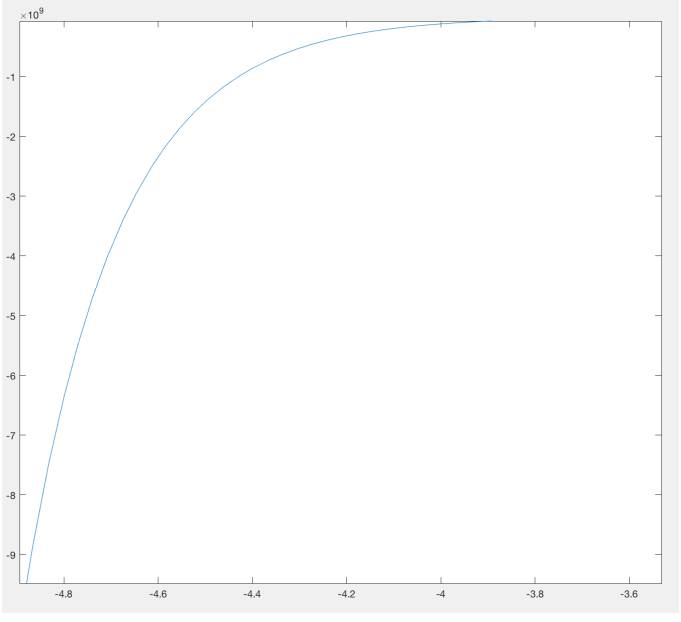
$$k_b T \frac{\mathrm{d}^2 F}{\mathrm{d}x^2} - U'(x) \frac{\mathrm{d}F}{\mathrm{d}x} = -\nu \ ,$$

where the Boltzmann factor in front of the second derivative term is constant, and drops out in the normalization to obtain a dimensionless ODE. Solutions of the first and second types are obtained from the dimensionless form after diving through by ν.

polynomial-exponential competition in solutions to the toy model, with the most simple potential for all base pairs

$U'(x) = 5\ \forall x$ gives a solution of the form,

$$\frac{6}{25} - \frac{6\exp(-5x)}{25} - \frac{x}{5}$$



structure of solutions for more complicated potentials

in this vein,

- family of first mean exit times, or mean exit times for any position along the binding sequence, with corresponding families of potentials to account for the probability of a particle passing through an energetic boundary which results from real solutions to the IVP,
- distributions of mean exit times for a fixed position,
- variability in the distribution of mean exit times ('conjecturally', would the variance of the distribution of mean exit times reflect the proportion of base pairs in the sequence, or can the mean and variance of these distributions be tuned for different binding processes?)

more complicated solutions

for higher degree potentials, the following non constant term in front of the second derivative in the differential equation,

$$\sum_{i \leq \text{degree}} \alpha_i x^i \ ,$$

which includes integrals of exponentials which is to be numerically approximated, for a given degree with arbitrary α_i, gives a solution whose form resembles that of the **second type**.

'conjecture' spoiler: distribution of mean exit times for variable diffusivity in a Langevin equation

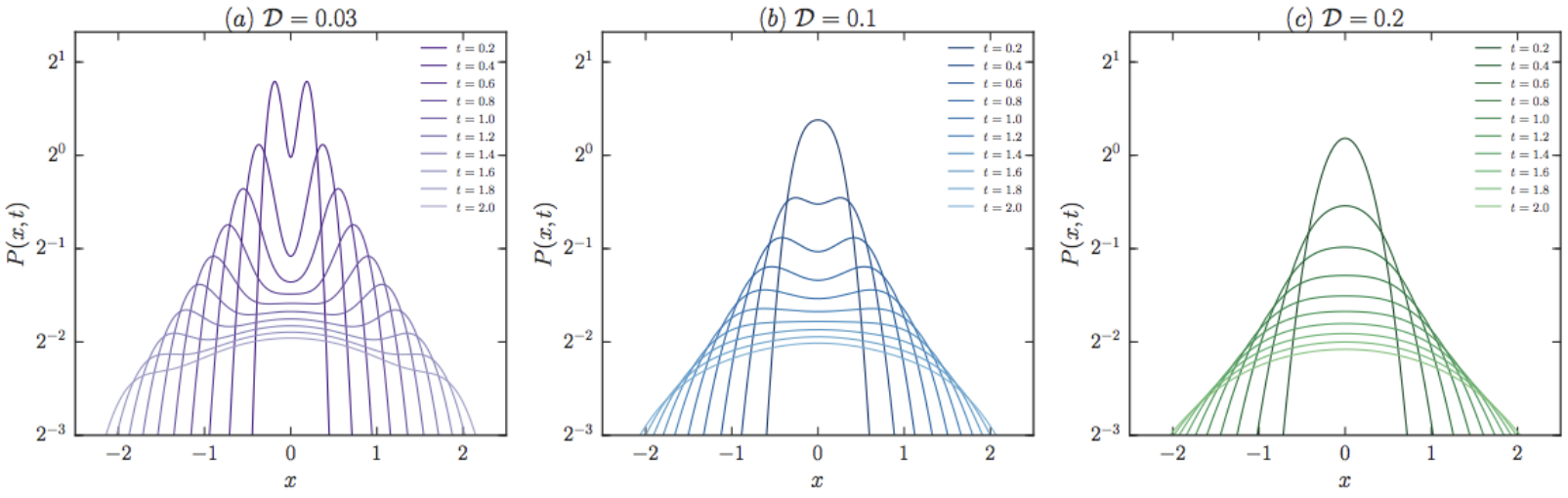


Figure 1. Plot of the probability density $P(x,t)$ in (10) for three different values of the diffusion constant $\mathcal{D}$.

Malakar, K. et al. Stead-state, relaxation and first passage properties of a run-and-tumble particle in one-dimension (2018)

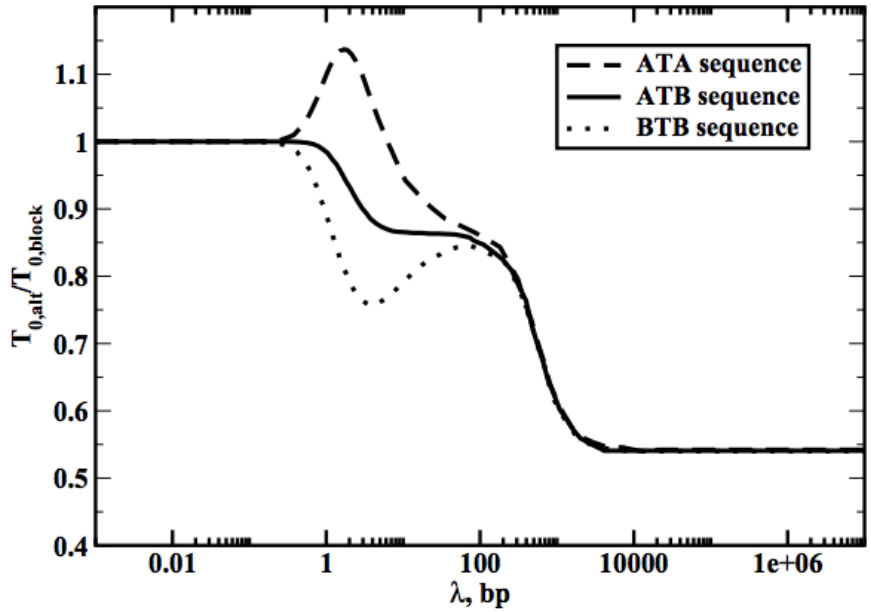


Fig. 7. The ratio of the mean search times for the alternating DNA sequences and for the block copolymer DNA sequences as a function of the scanning length $\lambda = \sqrt{u/k_{off}}$.

Kinetics of Protein-DNA Interactions First Passage Analysis (2018)

related to our goal of designing sequences to observe fluctuations in the search time, which in this figure is expressed as a ratio of search times for different sequences

conditions to enforce on higher degree potentials

for more realistic potentials, additional fluctuations can be imposed per the requirements:

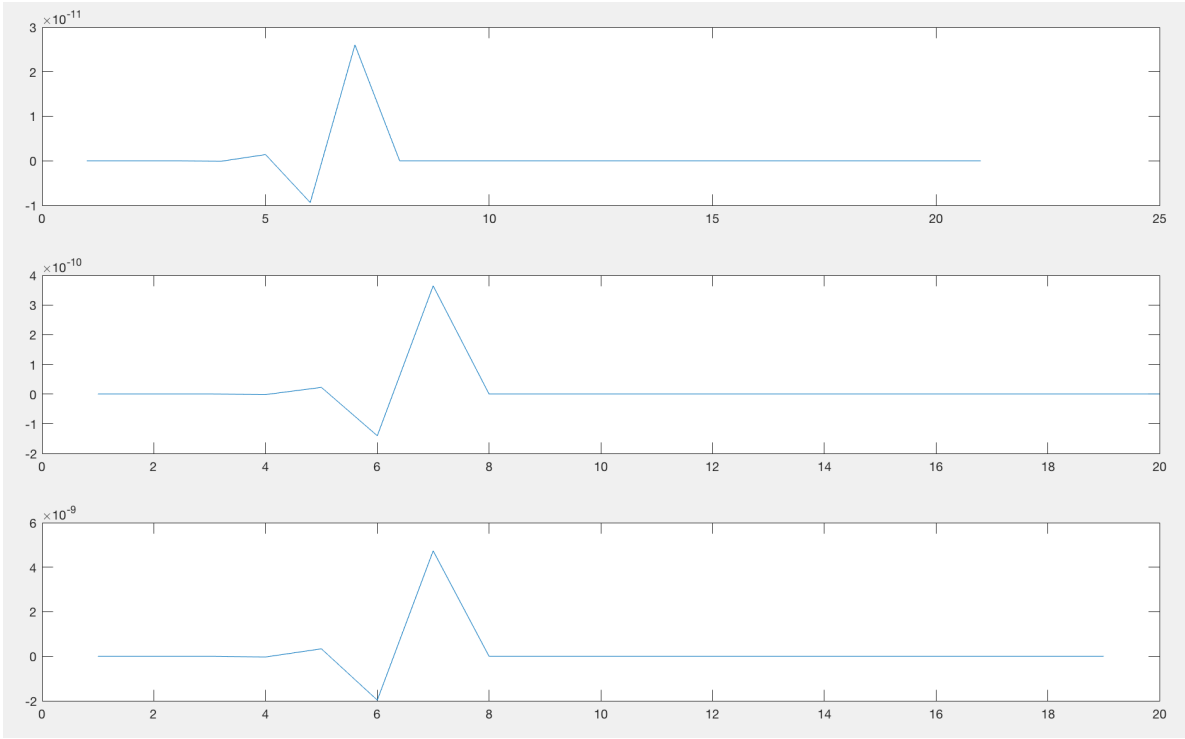
- polynomial interpolation requirements, which would make the polynomial have a minimum energy at each base pair,
- other requirements relating to the support of the potential about each base pair minima, which is qualitatively captured through the rate of decrease, or increase, of the potential about each base pair,
- final requirements to construct potentials for each binding sequence, stipulating that the potential which is the non constant term in front of the first derivative reflect less energetic cost for guide sequences which have no mismatches with the target sequence

candidate potentials ensuring that real solutions to the oscillator exist

fortunately, as motivation for constructing potentials that yield real solutions to the oscillator, we

- determine solutions for a potential that achieves local minima at each base pair of the sequence,
- from which additional constraints on the binding energy can be imposed by studying how the shape of the potential landscape varies amongst each base pair mismatch

example: typical fellows and their derivatives in the potential family



- *'kinks' could be smoothed, but exhibit sharpness of the phase transition in between base pairs*
- *obtaining information about the first and second derivatives of the potential relates to analytical approximations of exit times*

string initialization

to this end, we automate the desired procedure, for thousands of sequences, with the following easy steps:

- making use of routine string methods to isolate each integral term,
- from which we can declare function handles in Matlab/Python,
- finally which allows us to accommodate arbitrary solutions to the differential equation based on the potential that we enforce

generating potentials

given a starting potential, we can immediately implement the same procedure for other sequences in the database by

- shifting a new potential for other binding sequences, from the starting potential reflecting a successful conformational change for binding, numerically by ϵ if there exists a sequence for which a mismatch occurs at a position in the sequence, which perturbs the initial potential landscape for all remaining base pairs after the mismatch has occurred,
- obtaining solutions to the dimensionless ODE for appropriate boundary conditions,
- validating that there exists an achievable choice of parameters for which the potential landscape will prevent binding from occurring for a given number of mismatches between the guide and binding sequences

two approaches

- $\underline{\text{automate\_met}(\cdot)}$: generate potentials for other proteins outside of experimental data for Cas12a by permuting the concentration of nucleotides $\mathcal{N}$ relative to the total length of the binding sequence (details provided in the next few slides but parts of the code are still in progress, stemming from technicalities of the loops which is necessary for generating the space of all possible potentials),
- boxed relations ($\mathbf{1}$ & $\mathbf{2}$) from higher degree potentials: with the relations presented earlier based on solutions to 20 degree potentials, it is possible to determine the change in potentials from difference in exit times through a well posed *inverse problem*

automate_met$(\cdot)$

in the construction of landscape potentials,

- $\cdot = (\mathrm{len}, \mathcal{N}, \mathcal{L})$ specifies the length of the nucleotide sequence that is to be targeted, the 'stretch' of base pairs over which the restriction $U_{\mathcal{N}}$ of the potential gives base pairs yielding subsequent local maxima satisfying $\mathrm{length} \leq \mathcal{N}$, and also the length $\mathcal{L}$ up to which the exit time is computed,
- mean exit times, of variable length and composition of nucleotides $\mathcal{N}$, from one protein can be realized for another protein by varying the extrema of the potential along binding sequences of length len,
- leading to thermodynamic interpretations of the binding energy not only through the expected passage time, but also through a correspondence relating the exit time, up to a fixed position, with the first-principled Hamiltonian $\mathcal{H}$ that has been studied in other approaches to qualitatively compare TFs

modular arithmetic conditions

the conditions below clarify the enumeration of candidate potentials attaining local maxima within an $\mathcal{N}$ base pair window, which is combinatorially equivalent to determining the number of ways in which a subsequence $\mathcal{N}$ bases long can be placed within the original sequence of length len

- $\underline{\mathrm{len} \bmod \mathcal{N} \equiv 0}$: to enumerate all possible number candidate potentials,
  - we position all base pairs of the window $\mathcal{N}$ in the first base pair of the sequence,
  - continuing the procedure until we reach the last position in the sequence for which the remaining base pairs of $\mathcal{N}$ conditional that all base pairs are included at the end of the sequence
    $\Rightarrow \mathrm{len} - \mathcal{N} + 1$

- $\underline{\text{len} \bmod \mathcal{N} \equiv 1}$: the possible number of candidate potentials is $\text{len} - 1$

⋮

(finite number of cases, in correspondence with the equivalence classes of the additive group $\mathbf{Z}_{\mathcal{N}}$, are looped over)

- $\underline{\text{len} \bmod \mathcal{N} \equiv q}$: the possible number of candidate potentials is $\text{len} - q$ for any $q \geq 0$

Obs: As $\mathcal{N} \uparrow^{-} \text{len}$, the number of candidate potentials decreases to 1 & the distribution of mean exit times for any position approaches the **uniform distribution**

algorithm complexity

from conditions on the space of candidate potentials, the runtime could be

- $\underline{\mathcal{O}(1)}$, if a single candidate potential is returned in the case that $\mathcal{N} \equiv \mathrm{len}$, resemblant of the time that it takes the algorithm to construct a single potential $[\mathrm{len} \bmod \mathcal{N} \equiv 0]$,

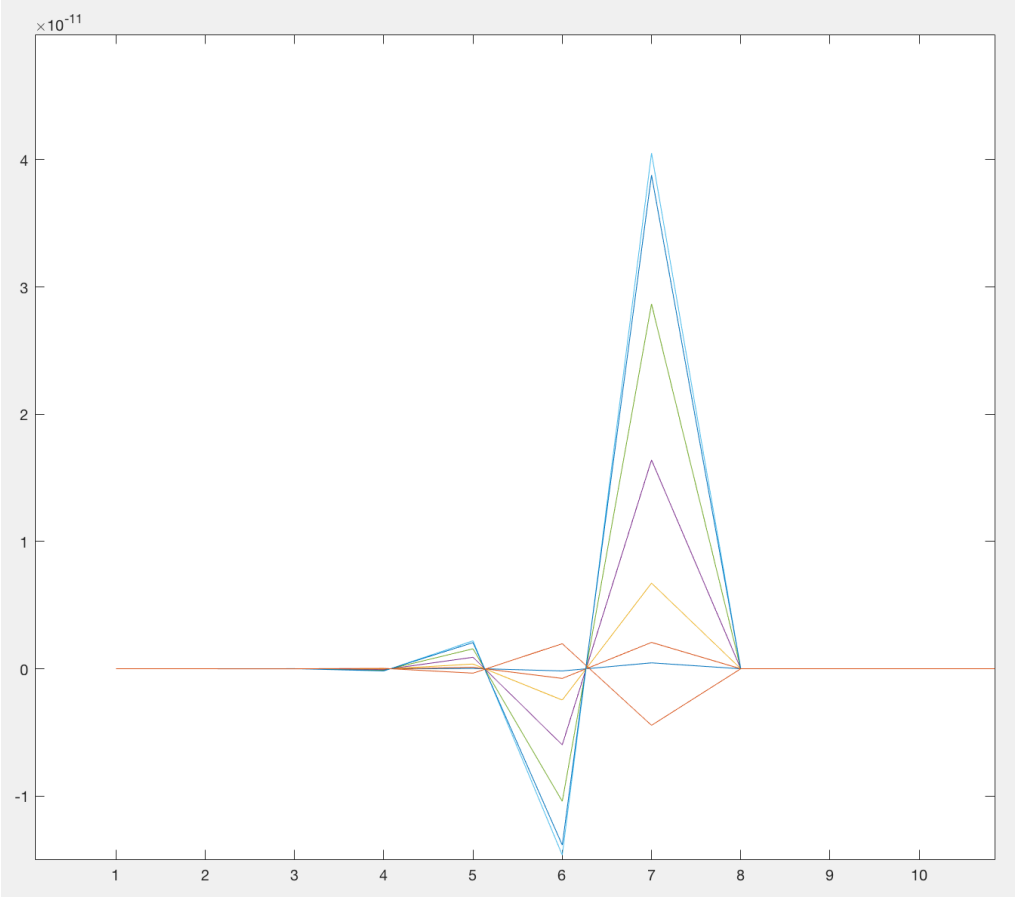
- $\underline{\mathcal{O}(n)}$, if a polynomial number of potentials are generated per iteration of the algorithm $[\mathrm{len} \bmod \mathcal{N} \not\equiv 0]$,

adjustment of the potential within neighborhoods of base pairs

one can completely identify the binding energy landscape with knowledge of the following free parameters,

- the initial energy in the landscape before protein inspection occurs,
- the final energy in the landscape to which the potential monotonically increases if a match or mismatch occurs,
- the change in binding landscape potential between the match or mismatch

return values for potential construction with pseudo random number generators

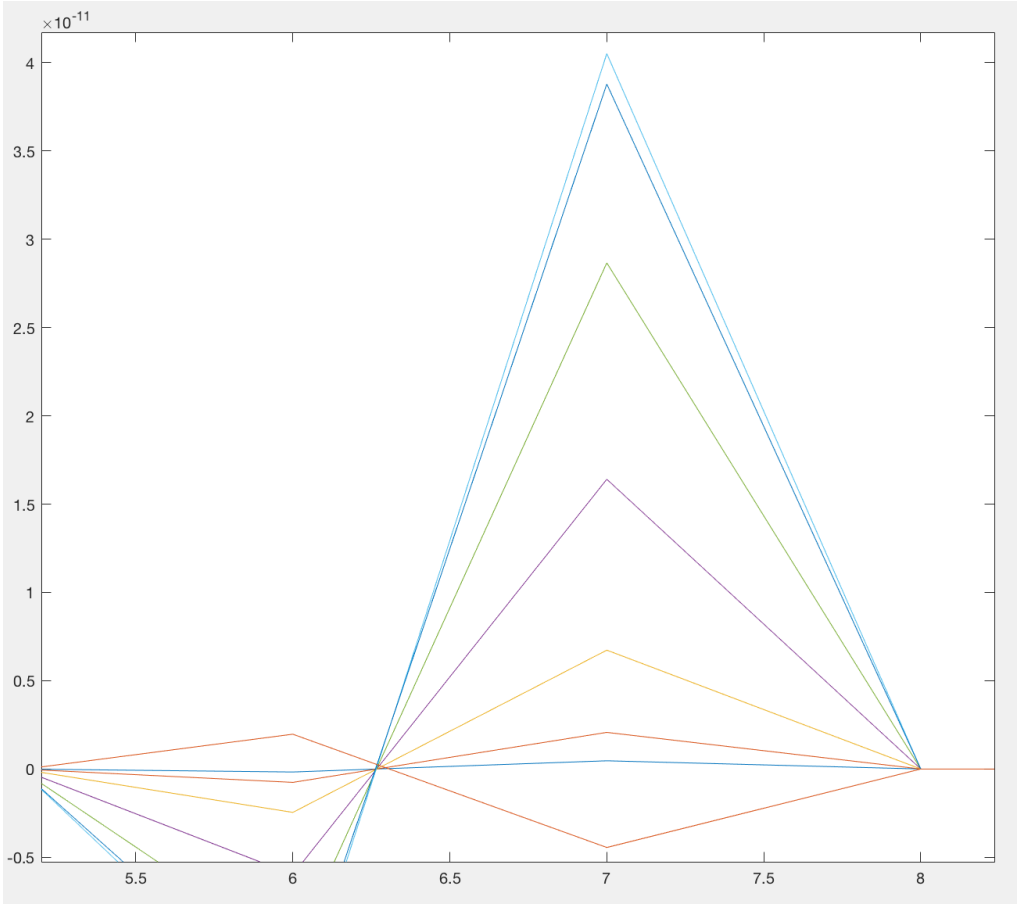


8 *potential function outputs*

qualities include:

- prominent, variable increases in ΔG at the first, second, or third, base pairs in the sequence throughout inspection,
- variable magnitude in ΔG for base pairs later in the sequence if complementarity is achieved for a sufficient number of base pairs,
- minimal curvature in the potential in the last few base pairs

neighboring base pair resolution



maxima of the potential influences protein kinetics, and makes passage time less probable

varying the $\mathcal{N}$ window is similarly possible

a note on guide sequence design

for our purposes,

- the correspondence between base pair mismatches and the accompanying potentials is assigned through the procedure of fixing a basis potential and then imposing fluctuations on the landscape given a fixed ordering,
- the problem which we have numerically posed can be implemented to compute the first and subsequent passage times,
- relies on the following steps,
  - polynomial interpolation to form the appropriate potential,
  - impose fluctuations on the potential by obtaining a new polynomial from a separate application of the polynomial interpolation amongst a new set of data points which can be randomly generated for samples within prescribed tolerance bounds on the base pair mismatch penalty

exit times τ from potentials that are polynomial, of degree $\mathrm{len} = 20$

from

$$\tau \approx -\int_0^x \exp\left(\prod_{i=2}^{20} \frac{u^i}{i}\right)\left(\int_0^x \{2\prod_{i=2}^{20} \exp\left(\frac{u^i}{i}\right) - 1\}\mathrm{d}u\right)\mathrm{d}u \ ,$$

set $v \equiv 0$, $t_{\mathrm{crit}} \equiv 1$

ongoing numerics

points of the potential $U(x)$	exit time τ
	$\approx$
	$\approx$
	$\approx$
	$\approx$
	$\approx$

varying the length of the absorbing boundary with t_{crit}

from

$$\tau_L \approx -\int_0^x \exp\Big(\prod_{i=2}^{20}\frac{u^i}{i}\Big)\Big(\int_0^x\{\prod_{i=2}^{20}\exp\Big(\frac{u^i}{i}\Big)-1\}\mathrm{d}u\Big)\mathrm{d}u+$$
$$\Big(\int_0^x \exp\Big(\sum_{i=2}^{20}\frac{v^i}{i}\Big)\mathrm{d}v\Big)\Big(\int_0^L \exp\Big(-\sum_{i=2}^{20}\frac{u^i}{i}\Big)\mathrm{d}u\Big)\mathrm{d}u\ ,$$

set $v \equiv 0$, $t_{\mathrm{crit}} \equiv L$

ongoing numerics

potential $U(x)$	L	exit time τ_L
		$\approx$
		$\approx$
		$\approx$
		$\approx$

difficulties with $\mathrm{automate\_met}$

- tricky to implement for different arithmetic conditions, and for arbitrary sequence lengths len and window length $\mathcal{N}$,
- adjustment of individual terms in the potential does not provide an explicit relationship between the variation of mean exit time with the terms in the potential that are changed,

the boxed relations given earlier in the presentation can more easily allow us to generate potentials for the landscape by specifying the distribution of exit times from a mean exit time (**not** previously explored in the literature)

background on solving the inverse problem determining a potential from the distributions of first mean passage times [Bal & Chou], **2004**

to systematically generate potentials from expected visit distributions, the paper provides guarantees on uniqueness of reconstructed stochastic equations from visit distributions, in addition to routines for

- reconstruction of the drift or diffusion terms associated with the kinetics,
- the transformation of the backward Kolmogorov equation in the Schrodinger equation with a suitable change of variables,
- solving SDEs via a mapping of the Euler Lagrange equation

we proceed in the spirit of a similarly posed inverse problem

making use of relations **1** & **2**

- **(1 & 2)**: with $\tau_2 > \tau_1$, set $\Delta_\tau \equiv \tau_2 - \tau_1 > 0$, from which subtracting τ_2 from τ_1 gives,

$$\Rightarrow \boxed{\tau_2 - \tau_1} = \mathcal{S}_{v \neq x}(v, x_2) - \mathcal{S}_{v=x}(v, x_1)$$

$$= (P)|_{x_2} + \left(\int_0^{x_2} \exp\left(\sum_{i=2}^{20} \frac{v^i}{i} \right) \mathrm{d}v \right) \left(\int_0^{v} \exp\left(- \sum_{i=2}^{20} \frac{u^i}{i} \right) \mathrm{d}u \right) -$$

$$\int_0^{x_1} \exp\left(\prod_{i=2}^{20} \frac{u^i}{i} \right) \left(\int_0^{x_1} \{ 2 \prod_{i=2}^{20} \exp\left(\frac{u^i}{i} \right) - 1 \} \mathrm{d}u \right) \mathrm{d}u$$

$\Updownarrow$

combining (P) and the integral associated with exit time τ_1 yields

$$(P)|_{x_2} - \int_0^{x_1} \exp\left(\prod_{i=2}^{20} \frac{u^i}{i}\right)\left(\int_0^{x_1} \{2\prod_{i=2}^{20} \exp\left(\frac{u^i}{i}\right) - 1\}\mathrm{d}u\right)\mathrm{d}u \cdots$$

$$= \boxed{\int_0^{x_1}\left(\int_0^{x_1}\prod_{i=2}^{20}\exp\left(\frac{u^i}{i}\right)\mathrm{d}u\right)\mathrm{d}u - \int_0^{x_1}\int_{x_1}^{x_2}\cdots}$$

$$\boxed{\cdots\{2\prod_{i=2}^{20}\exp\left(\frac{u^i}{i}\right) - 1\}\mathrm{d}u\ \mathrm{d}u - \int_{x_1}^{x_2}\int_0^{x_2}\exp\left(\frac{u^i}{i}\right)\{\prod_{i=2}^{20}\exp\left(\frac{u^i}{i}\right) - 1\}}$$

$$\boxed{\mathrm{d}u\ \mathrm{d}u}\ ,$$

(Variational One)

because ...

$$(P)|_{x_2} = -\left(\{\int_0^{x_1} \exp\left(\prod_{i=2}^{20} \frac{u^i}{i}\right)\left(\int_0^{x_2} \{\prod_{i=2}^{20} \exp\left(\frac{u^i}{i}\right) - 1\}\mathrm{d}u\right)\mathrm{d}u\}+ \right.$$
$$\left.\{\int_{x_1}^{x_2} \exp\left(\prod_{i=2}^{20} \frac{u^i}{i}\right)\left(\int_0^{x_2} \{\prod_{i=2}^{20} \exp\left(\frac{u^i}{i}\right) - 1\}\mathrm{d}u\right)\mathrm{d}u\}\right) ,$$

"integral linearity applied to the outermost variable while holding lower and upper bounds for the innermost variable constant"

- other case with $v \equiv x_2, v \neq x_1$: with the same quantities as in the previous case, subtracting τ_2 from τ_1 gives a different relation, in the case in which $x_2 \equiv 1$ & $x_1 < x_2$,

$$\Rightarrow \boxed{\tau_2 - \tau_1} = \mathcal{S}_{v=x}(v, x_2) - \mathcal{S}_{v\neq x}(v, x_1)$$

$$= -\int_0^{x_2} \exp\left(\prod_{i=2}^{20} \frac{u^i}{i}\right)\left(\int_0^{x_2} \{2\prod_{i=2}^{20} \exp\left(\frac{u^i}{i}\right) - 1\}\mathrm{d}u\right)\mathrm{d}u +$$

$$\int_0^{x_1} \exp\left(\prod_{i=2}^{20} \frac{u^i}{i}\right)\left(\int_0^{x_1} \{\prod_{i=2}^{20} \exp\left(\frac{u^i}{i}\right) - 1\}\mathrm{d}u\right)\mathrm{d}u$$

$$-\left(\int_0^{x_1} \exp\left(\sum_{i=2}^{20} \frac{v^i}{i}\right)\mathrm{d}v\right)\left(\int_0^{x_1} \exp\left(-\sum_{i=2}^{20} \frac{u^i}{i}\right)\mathrm{d}u\right)$$

$$\cdots ,$$

from which rearrangements give ...

a similar relationship, composed of

- interactions over the passage time to x_1,
- interactions over the passage time to x_2,
- intermediate interactions for the regime between the first and second passage times at x_1 and x_2, respectively

$$\int_0^{x_1} \exp\left(\prod_{i=2}^{20} \frac{u^i}{i}\right) \int_0^{x_1} \{3 \prod_{i=2}^{20} \exp\frac{u^i}{i} - 2 \} + \int_0^{x_1} \int_{x_1}^{x_2} \left(\prod_{i=2}^{20} \exp\left(\frac{u^i}{i}\right)\right)$$

$$\ldots - \ldots$$

(Variational Two)

advantages (!!) of modifying **1** & **2**

instead of having to specify the difference of the potential amongst base pairs, the inverse problem that we have posed, in which we specify the magnitude of Δ_τ is convenient for

- solving the "reverse" problem of obtaining suitable potential families from distributions of exit times which are much easier to specify than individual potentials themselves,
- numerically adjusting for the variance between exit times within a distribution from which corresponding changes in potentials can be observed

this can be applied to several proteins at once with apriori knowledge of visit times, which for Cas12a & Cas9 seems well known

adapting the relations for potentials that are 'concentrated' amongst different base pairs of the binding sequence

for CRISPR proteins longer or shorter than dCas12 or Fn Cas12a,

- adjustments to the target gene of interest, in addition to variable $\mathcal{N}$ can be implemented through devoting attention towards
  - varying the local maxima and minima of the potential in order to numerically control the output exit time (in the spirit of Method 1),
  - directly varying the expected exit time (in the spirit of Method 2)

Fn Cas12a predictions. comparisons

from the reads in this data, predictions for additional reads can be achieved by

- generating classes of exit distributions,
- determining corresponding potentials,

while comparisons can be achieved by

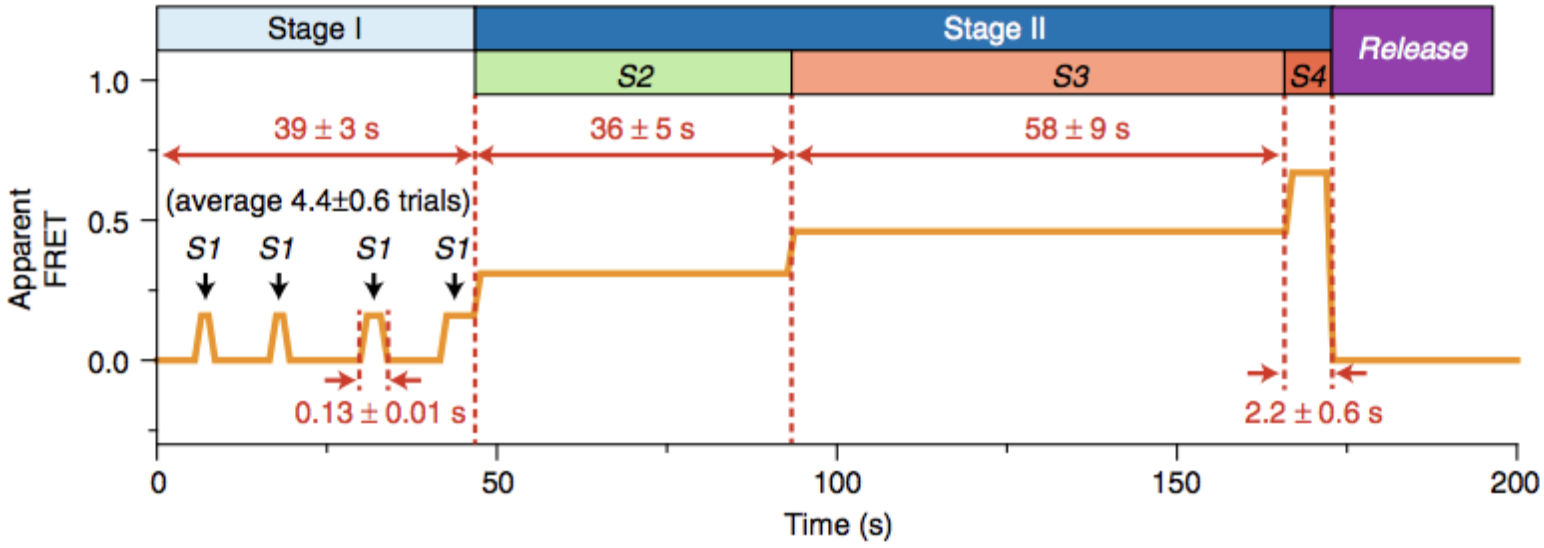
- varying the base pairs along which extrema are achieved,
- varying the strength of the couplings $\mathcal{J}_{ij}$ (aim of future work in presenting a unification with previous constructions of μ)

recap, automation steps: solving hundreds of differential equations in parallel with a specified potential

to this end, we automate our search for solutions to the equation, in order to understand variation in mean exit times, by

- constructing potential functions, in light of the desirable requirements,
- forming the corresponding second order ODE with each candidate potential that is constructed from the number of matching base pairs between the target and guide sequences,
- plotting the distribution of mean exit times, to determine the extent to which changes to separate polynomial terms in the potential impact the Brownian motion that particles undergo under the influence of stochastic force,
- making use of solutions of the ODE for the mean exit time as a proxy to study kinetics on macroscopic and microscopic scales of the binding process

PAM recogniton and DNA unzipping in a 2018 work relating to the intermediate stages of the ternary complex in Cas12a activation



More references for the time scales of dynamics before and after cleavage can be used for numerical tuning of the potential. *Conformational Activation Promotes CRISPR-Cas12a ..., 2018*