Molecular Dynamics Simulations of DNA Hybridization and Dynamic Force Spectroscopy

Raymond Jin

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Lutz Maibaum, Chair

Xiaosong Li

Paul A. Wiggins

Program Authorized to Offer Degree:
Chemistry

University of Washington

Abstract

Molecular Dynamics Simulations of DNA Hybridization and Dynamic Force Spectroscopy

Raymond Jin

Chair of the Supervisory Committee:
Assistant Professor Lutz Maibaum
Chemistry

A ubiquitous problem in chemistry is determining the energy change of moving from one state to another. Often these states are separated by a large energy barrier which can prevent the observation of the equilibrium distribution on a reasonable timescale. Techniques have been developed to accelerate the dynamics of these systems and determine the equilibrium energies. In this thesis, we examine two systems using non-equilibrium techniques.

We first studied DNA hybridization which involves two single-stranded DNA interacting through hydrogen bonds between their bases. We investigated DNA hybridization using coarse-grained molecular dynamics simulations and by creating a Markov State Model from the simulation data. We found that the mechanism of hybridization to be nucleation and zipping up. We also found the transition states contain one or two bonds. If the strands have three or more bonds, they are very likely to proceed to full hybridization.

We then studied dynamic force spectroscopy, a technique in which systems are biased out of equilibrium using external forces. We studied five models of dynamic force spectroscopy and found many differences between them. Overall we found the Hummer-Szabo intermediate model does the best. However these models tend to focus on the mean rupture force and largely do not account for the variance or distribution of forces. We found the variance to be less sensitive to the pulling protocol than the mean. We also found a relationship between the rupture force and the work performed on the system until rupture which requires further

study.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Number Page

GLOSSARY

k_0: the intrinsic rate of escape from the potential, without an external force

$\dot{F}$: the loading rate of a linearly increasing force

x_b: the distance of the transition barrier from the system minimum

$\Delta G^\ddagger$: the height of the barrier at x_b

k_sys: for a harmonic system, related to $\Delta G^\ddagger$ by $\Delta G^\ddagger = \frac{1}{2} k_\mathrm{sys} x_\mathrm{b}^2$

k_bias: the bias strength for a moving harmonic trap

v: the velocity of a moving harmonic trap

D: the diffusion coefficient

k_B: the Boltzmann constant

ACKNOWLEDGMENTS

I would like to thank Professor Maibaum and the rest of the Maibaum group for their support and knowledge. This work was facilitated though the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington.

Chapter 1

INTRODUCTION

In chemistry, the energy and changes of energies of systems are important for determining their behavior. However the calculation and measurement of these energies can be difficult because the energy barriers in the landscape of the system limit exploration of the state space. In order to accelerate the dynamics and exploration, techniques are used to modify the energy landscape which decrease the size of the barriers. In computational studies, these landscapes can be changed using methods such as Umbrella Sampling[61] or Metadynamics.[32] These methods change the energy of the systems based on collective variables (CVs). CVs are coordinates of interest chosen to span the range of configurations that are deemed important to the system. One challenge of using methods such as Umbrella Sampling and Metadynamics is choosing an appropriate CV which can adequately cover the entire range of configurations. If the CV is incorrect, then the calculated energy surface might not match reality. To alleviate this issue we will devise a method which does not require a prior selection of a CV.

Another way to change the energy landscape and accelerate the dynamics is to apply a force to the system. One particular experimental technique for applying a force is Dynamic Force Spectroscopy (DFS). In this method, an increasing force is applied to a system until the system responds and undergoes some conformational change which reduces the applied force. At different rates of increasing force, the systems can respond differently and respond at different forces. Models have been developed to analyze this relationship between the loading rate of force with the force at which it responds to extract information about the underlying system. However these models, which are based on the same data, often disagree with each other on the form of the underlying system.

1.1 Specific Aims

This thesis will be broken into two major parts. In the first part of the thesis, we investigated the mechanism of DNA hybridization through molecular dynamics simulations of two single-stranded DNA (ssDNA) into one double-stranded DNA (dsDNA). For the second part of the thesis, we tested the validity and accuracy of different pulling models by comparing them to data from Brownian dynamics simulations.

1.1.1 Studying the Mechanism of DNA Hybridization Through a Simulation Informed Markov State Model

We simulated DNA hybridization through coarse-grained molecular dynamics simulations, and created a Markov State Model from the raw molecular dynamics data. Using the Markov State Model (MSM), we were able to determine the pathways of hybridization with the highest probabilities, and showed that the observed mechanism matches the classic model of initial nucleation and zipping-up. We also used the MSM to identify the set of transition states of hybridization, of which the most important transition states had one or two hydrogen bonds. Another conclusion from the MSM analysis was that the hybridization process is generally not well-described by only the net number of base pairs between the strands.

1.1.2 Testing Popular Models of Dynamic Force Spectroscopy

Using Brownian dynamics simulations, we generated rupture force data with a variety of input variables. We investigated the effect of different force loading protocols including the differences between a force imposed on the system in a linearly increasing way as opposed to a force generated from a moving harmonic trap. In all our systems we performed simulations at many different loading rates to ensure we observed the system in many different regimes. Within the moving harmonic trap protocol, we explored the effect of trap stiffness on the rupture forces and found at high loading rates the effect of the trap stiffness on the rupture force is reduced. In addition, we probed to what extent the differences in rupture forces could be explained by different underlying free energy surfaces. We found that generally at high loading rates, there are neither differences in rupture forces between linearly increasing and moving harmonic trap protocols, nor differences between different underlying free energy surfaces. However at lower loading rates, the effect of the trap stiffness and the underlying free energy surface both affect the rupture forces.

We are also interested in determining the best model to fit the rupture force data. First we graphed each of the models using known values for the fitting parameters and observed the agreement between each model's prediction and our rupture force data. Using this technique we found the Bullerjahn-Sturm-Kroy and Hummer-Szabo models had the best match to our data. Then we used fitting techniques to find parameters that minimized the difference betwen the models' predictions and our rupture force data, and compared the resulting parameters to our known parameters. This again lead to Bullerjahn-Sturm-Kroy and Hummer-Szabo models with the best fits.

Chapter 2

STUDYING THE MECHANISM OF DNA HYBRIDIZATION THROUGH A SIMULATION-INFORMED MARKOV STATE MODEL

Reproduced from R. Jin and L. Maibaum. Mechanisms of DNA hybridization: Transition path analysis of a simulation-informed Markov model. *J. Chem. Phys.*, 150(10), 2019, with permission of AIP Publishing.

2.1 *Introduction*

Deoxyribonucleic acid (DNA) is a biopolymer that contains the information of life. Its crucial role in biology and its many uses in technological applications such as DNA origami[50,53] and nanoparticle linker[11,38] depend on DNA's ability to hybridize: two single-stranded DNA (ssDNA) molecules bind to form a double-stranded DNA (dsDNA) duplex. The duplex is stabilized by hydrogen bonds between the nucleobases adenine (A), thymine (T), guanine (G), and cytosine (C).

The mechanism of hybridization remains a topic of current research. Classically, hybridization is thought of as a two-step process in which the strands first form an initial nucleus of hydrogen-bonded base pairs, which is then followed by the strands "zipping up" to form the fully hybridized double helix.[66] It was later found that at least three base pairs were necessary for a thermodynamically stable binding nucleus.[46] Recent experiments on RNA, on the other hand, found a threshold of seven base pairs.[12] A more complex mechanism based on a three-step process has been proposed.[39] Here the two strands initially bind by non-specific interactions. The first stable base pair is then formed either through a one-dimensional sliding motion in which the strands search for an initial nucleation point, or through internal displacements in which the strands traverse each other in an inchworm-like motion. Once an initial nucleation point is found, the bases zip up to fully hybridize as in the classical mechanism.

It is challenging to observe hybridization in a base-by-base manner in experiments. Previous work has been done on detection of hybridization in bulk solution using a variety of techniques, including absorption spectroscopy,[44, 64, 66] surface-enhanced Raman spectroscopy,[4] and electrochemistry.[15] Recently, there have been advances in single molecule experiments to study hybridization using electrochemical techniques[56] and force probes.[33, 67] There have also been studies of DNA at the base level using fluorescence resonance energy transfer[9, 12, 14] and fluorescence correlation spectroscopy.[3, 10, 65]

Molecular Dynamics (MD) computer simulations can provide insight into the hybridization mechanism on a base-by-base level. While an atomistic description of the two strands, water, and counterions would provide the highest resolution, it is computationally too expensive to obtain a large dataset of hybridization trajectories. Coarse-grained models sacrifice detail for speed, and several such models of DNA can be found in the literature. In this work we use the recently developed oxDNA 2 model, in which each nucleotide is described by three interaction sites: a backbone site, a hydrogen bonding site, and a stacking site (Figure 2.1 top). It is parametrized to match the melting temperatures of the SantaLucia model[52] at various salt concentrations, as well as physical characteristics such as major-minor grooving and radius of gyration.[55] Interactions in the oxDNA2 model include stacking, hydrogen bonding and electrostatics. It has been used to study processes such as hybridization[42] and toehold-mediated strand displacement.[57] In their study on hybridization, Ouldridge and coworkers used Forward Flux Sampling to guide the system toward the duplexed state.[42] They found a mechanism consistent with a two step process of nucleation and zipping up similar to the classical mechanism.

Here we adopt a different approach to obtaining mechanistic information about the hybridization pathway from coarse-grained simulations. Starting with a large set of hybridization trajectories we construct a Markov State Model (MSM) that describes the stochastic

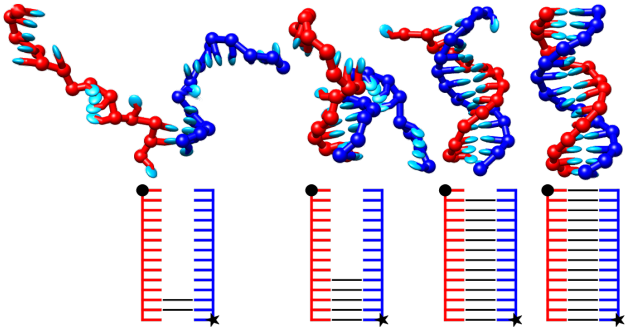


Figure 2.1: Top: Snapshots from a typical hybridization trajectory of a coarse-grained oxDNA2 simulation. Each strand has 14 bases, and each base is visualized by two particles: a sphere that represents the backbone and a disc that represents the base. The two strands make an initial contact near their ends, and then zip up to a fully hybridized state. Bottom: Schematic representation of the inter-strand hydrogen bonding pattern. Red and blue lines represent the two strands, and black lines denote hydrogen bonds. The circle and the star indicate the $5'$ end of strands 1 and 2, respectively.

dynamics of the system across a large set of discrete states that represent the two strands' hydrogen bonding pattern. The generation of MSMs from MD simulations to describe the dynamics of complex systems has been pioneered in the context of protein folding.[8,40,43,63] Having obtained an MSM, we use the framework of Transition Path Theory (TPT)[36] to obtain quantitative information about the ensemble of pathways by which DNA hybridizes.

2.2 Methods

2.2.1 Coarse-Grained Molecular Simulations

We use the coarse-grained oxDNA2 model to simulate DNA hybridization.[55] In this model, the solvent is treated implicitly, and each nucleotide is represented by three particles: one for the DNA backbone and two for the nucleobase. It is parametrized to reproduce the duplex melting temperatures of the SantaLucia model,[52] which it does accurately.[55] We used the sequence-averaged variant of oxDNA2, as in previous work,[42] which incorporates only a limited amount of sequence specificity. The strength of inter-base hydrogen bonds is considered to be the same for A-T and G-C base pairs, but mismatched bases cannot form hydrogen bonds. All other interactions between two bases have the same strength regardless of base identity.

We simulated two ($\alpha = 1, 2$) complementary strands of DNA with $N = 14$ bases each. The sequences are

`3’-GCTGTTCGGTCTAT-5’`

`5’-CGACAAGCCAGATA-3’`

and are designed to limit non-intended base-pairing.[42] At the beginning of a trajectory, the two strands are placed at random positions and relative orientations in a cubic simulation box with 10 nm side length. Both strands are initially helical and unbound. The dynamics of the system was propagated for up to 3 microseconds or until the strands were fully hybridized, i.e., each base formed a hydrogen bond with its complementary pair. Two bases were considered hydrogen bonded if their hydrogen bond contribution to the interaction energy was larger than 1.79 kcal/mol as done in previous work.[42] The simulation timestep was 15 femtoseconds. Temperature was held fixed at 300 Kelvin using an Anderson-like thermostat that stochastically resets a particle’s momentum, using a collision time of 103 steps and a collision probability of 0.02 as used in previous work.[42] Electrostatic interactions where computed for an effective salt concentration of 0.5 M.

2.2.2 Construction of the Markov State Model

In order to create an MSM, we partitioned the raw molecular dynamics data into individual states at each timestep. The states are defined based on the hydrogen bonding pattern of the two strands. The bases in each strand were labelled $i, j = 1...N$ from the $3'$ end to $5'$ end. At each timestep, we determined which bases are hydrogen bonded with each other. The hydrogen bonding pattern only includes bonding between the two strands; instances of intra-strand bonding were ignored. If a base has hydrogen bonding interactions with more than one other base, only the base with the strongest interaction was considered. From this list of hydrogen bonds, we labelled each state s with a number m that encodes the hydrogen bond sequence:

$$m(s) = \sum_{i=1}^{N} (N+1)^i \sum_{j=1}^{N} j\, \omega_{1ij}(s) \tag{2.1}$$

where

$$\omega_{\alpha ij}(s) = \begin{cases} 1 & \text{if base } i \text{ in strand } \alpha \text{ is bound to} \\ & \text{base } j \text{ in the other strand } 3{-}\alpha; \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

This scheme encodes the bonding pattern in a base-($N{+}1$) system. It ensures that each state label is unique, and that we can recover the bonding pattern from the number m alone. For example,

$$j = \left\lfloor \frac{m}{(N+1)^i} \right\rfloor \bmod (N+1) \tag{2.3}$$

is the index of the base in strand 2 to which base i in strand 1 is bound (or 0 if the latter is unbound).

Because each base can bind to at most one other base in the opposing strand, the total number of states encoded by (2.1) is

$$\sum_{k=0}^{N} \frac{(N!)^2}{k!((N-k)!)^2}, \tag{2.4}$$

where N is the number of bases in a single strand. For $N = 14$ we obtain approximately 1.6x10^{13} distinct states. By restricting hydrogen bonds to be between matched bases, this number is reduced to 1.9x10^{8} possible distinct states for our specific sequence. In practice, however, our simulations visit only $M = 8942$ of those. We therefore label these visited states sequentially from 1 to M. We adopt the convention that $s = 1$ refers to the unbound state, whereas $s = 2, \ldots, M$ indexes states with at least one hydrogen bond in arbitrary order.

Because each base can bind to at most one other base in the opposing strand, the total number of states encoded by (2.1) is

$$\sum_{k=0}^{N} \frac{(N!)^2}{k!((N-k)!)^2}, \tag{2.5}$$

where N is the number of bases in a single strand. For $N = 14$ we obtain approximately 1.6x10^{13} distinct states. Because the model considers pairing only between matched pairs, this number is reduced to 1.9x10^8 possible states for our specific sequence. In practice, however, our simulations visit only $M = 8942$ of those, which shows that the vast majority of states are not physically relevant.

The entire hydrogen bonding pattern between the two strands is a complex descriptor which makes it unwieldy to gain physical insight from the MSM. In principle, we can use any observable that is a function of the hydrogen bonding pattern, and we define two observables that provide limited but intuitive information about a state. The first is the net binding count $C(s)$, which is defined as the difference between the number of native base pairs (base pairs that also exist in the fully hybridized conformation) and the number of non-native base pairs:

$$C(s) = \sum_{i=1}^{N} 2\omega_{1i(N+1-i)}(s) - \sum_{j=1}^{N} \omega_{1ij}(s). \tag{2.6}$$

In principle the value of C ranges from $-N$ to N, with $C = N$ for the fully hybridized state. This observable measures the extent of matching for a specific hydrogen bonding pattern.

The second observable, center of hybridization $D(s)$, reports on the positions of bases in strand $\alpha = 1$ that are paired with bases in the other strand:

$$D(s) = \sum_{i=1}^{N} h_{1i}(s) \left(\frac{N+1}{2} - i \right), \tag{2.7}$$

where

$$h_{\alpha i}(s) = \sum_{j=1}^{N} \omega_{\alpha ij}(s) \tag{2.8}$$

is one if base i is hydrogen bonded to any base on the other strand, and zero otherwise. States with positive D have more bonding of strand 1 at the $3'$ end, while those with negative D have more bonding at the $5'$ end. As an example, the configuration shown in the first snapshot of Figure 2.1 has bonding pairs (i,j) of (2,13) and (3,12), which yields $m = 43425$, $C = 2$, and $D = 10$.

Having established the state space of the Markov model, we proceed by converting each oxDNA simulation trajectories into a sequence $m_1, m_2, \ldots$ of hydrogen bonding classifiers. The time Δ between steps in this sequence is determined by the output frequency of the coarse-grained simulations. We then use the MSMBuilder software package[5] to construct from these sequences a family of MSMs, parametrized by the lag time τ (which must be a multiple of Δ). This is done in two steps. First we determine the number of transitions between states that are a time τ apart using a sliding-window counting scheme. From this count matrix we obtain the transition probability matrix T using a Maximum Likelihood Estimator that enforces detailed balance.[47]

For each MSM we then solve the eigenvalue equation

$$\boldsymbol{\psi}_i T = \lambda_i \boldsymbol{\psi}_i, \tag{2.9}$$

i.e., $\boldsymbol{\psi}_i$ is i-th left eigenvector of T, and λ_i is its eigenvalue. If T is a regular stochastic matrix, then the Perron-Frobenius theorem guarantees that the largest eigenvalue (which we take to be λ_1) is equal to one, and the corresponding eigenvector $\boldsymbol{\psi}_1 \equiv \boldsymbol{\pi}$, when properly normalized, is the equilibrium probability distribution. All other eigenvalues are less than one in magnitude. In terms of these eigenvectors, the time evolution of the probability distribution $\boldsymbol{p}$ is

$$\boldsymbol{p}(t) = \boldsymbol{\psi}_1 + \sum_{i=2}^{M} b_i \mathrm{e}^{-t/t_i} \boldsymbol{\psi_i} \tag{2.10}$$

where $b_1, \ldots, b_M$ are constants determined by the initial condition, and the relaxation time scales t_i (also known as implied time scales) are

$$t_i = -\frac{\tau}{\ln \lambda_i}. \tag{2.11}$$

Equation (2.10) illustrates the meaning of each eigenvector $\boldsymbol{\psi}_i$ for $i \geq 2$: it describes how deviations from the equilibrium distribution decay. The sum of the elements of each eigenvector is zero, so that $\boldsymbol{\psi}_i$ corresponds to a redistribution of probability between states with positive elements and those with negative elements.

The parameter τ that enters the construction of an MSM must be carefully chosen. One typically desires a small lag time so that the MSM has the highest possible temporal resolution. However, if τ is too short then the MSM does not accurately describe the dynamics of the original system. To find the shortest acceptable lag time we use a basic property of MSMs:[60] the time scales (2.11) are independent of τ if the latter exceeds the time over which the system dynamics becomes Markovian. We therefore build MSMs for a range of lag times, and calculate their slowest relaxation time scales. We then choose the lag time as the time at which the implied time scales become independent of τ.

2.2.3 Transition Path Theory

Once an accurate MSM has been built, one can use different approaches to analyze its behavior in addition to the eigenvector decomposition (2.9). We use Transition Path Theory[18,36,40] to obtain statistical information about the DNA hybridization mechanism. A transition path connects a set of reactant states, A, to a set of product states, B, which we take to be the fully unbound and the fully hybridized state, respectively. All other states form the set of intermediates, I. Using the TPT functionality of the MSMBuilder software,[5] we calculate for each state the forward committor, q_i^+, which is the probability of reaching state A before reaching state B when starting from state i. Reactant and product states have committor values of 0 and 1, respectively, and the committor for the remaining states can be obtained by solving the set of equations[40]

$$-q_i^+ + \sum_{j \in I} T_{ij} q_j^+ = -\sum_{j \in B} T_{ij} \tag{2.12}$$

The forward committor is a measure of how far along a transition any individual state is, and is therefore the ideal reaction coordinate. An analogous quantity, the backward committor, is the reaction coordinate for the inverse path. For a transition matrix that obeys detailed balance, the backwards committor is $q_i^- = 1 - q_i^+$.

TPT allows us to calculate the most likely pathways of DNA hybridization. Likely paths are those that contain the most flux, which is the expected number of transitions between two states within one timestep. Limiting ourselves to reactive paths that connect A and B, the flux between two states i and j is[40]

$$f_{ij} = \pi_i q_i^- T_{ij} q_j^+ \tag{2.13}$$

where π_i is the equilibrium population of state i. Trajectories can go back and forth between two states, which builds up flux between those states but this built-up flux does not contribute towards reaching the product states B. The net flux[40]

$$f_{ij}^+ = \max(f_{ij} - f_{ji}, 0) \tag{2.14}$$

removes the effects of such recrossings. The capacity of an entire path is defined as the smallest net flux of each of its transitions. With these quantities at hand, one can calculate the highest capacity pathways that connect reactive and product states.

A related property is the "fraction visited" of a state, which is the fraction of reactive paths that visit the state along the way.[13] This quantity can also be calculated using the MSMBuilder software.

One way we calculate rate information is through the mean first passage time (MFPT). We compute the MFPT to reach the fully hybridized state for all other states by solving the equation[54]

$$(E - T')\boldsymbol{b} = \boldsymbol{c} \tag{2.15}$$

where T' is a matrix with elements

$$T'_{ij} = \begin{cases} 2, & \text{if } i \in B \text{ and } j = i \\ 0, & \text{if } i \in B \text{ and } j \neq i \\ T_{ij}, & \text{otherwise,} \end{cases} \tag{2.16}$$

E is the $M \times M$ identity matrix, $\boldsymbol{b}$ is the vector of mean first passage times in units of the lag time τ, and $\boldsymbol{c}$ is a column vector of length M that contains all ones except at the sink states, where it is 0.

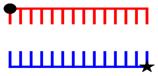
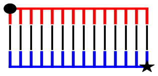
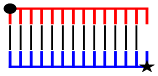
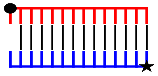
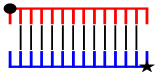
State	Contribution to Slowest Process	Equlibrium Population
	-1.00	3.7-5
	0.46	0.46
	0.16	0.16
	0.16	0.16
	0.05	0.06

Table 2.1: Five states with the largest (by magnitude) contributions to the eigenvector of the process with the longest relaxation time scale. The eigenvector is normalized so that the unbound state has a contribution of -1. This process mostly shifts probability from the unbound state to the completely hybridized and three mostly hybridized states.

As an alternative to the MFPT, we also calculate the time scale $t_{A\to B}$ of the hybridization process, defined as[40]

$$t_{A\to B} = \frac{\sum_{i=1}^{M} \pi_i q_i^-}{\sum_{j=2}^{M} \pi_1 T_{1j} q_j^+} \tau. \tag{2.17}$$

Here, the denominator corresponds to the expected number of transitions between A and B per time τ in either direction, and the numerator is the fraction of paths in the forward direction.

2.3 Results

We begin with a dataset of 1056 hybridization trajectories obtained from oxDNA simulations. Using the mapping (2.1), we convert them to trajectories in the discretized state space that describes the two DNA strands' hydrogen bonding pattern. As described in the

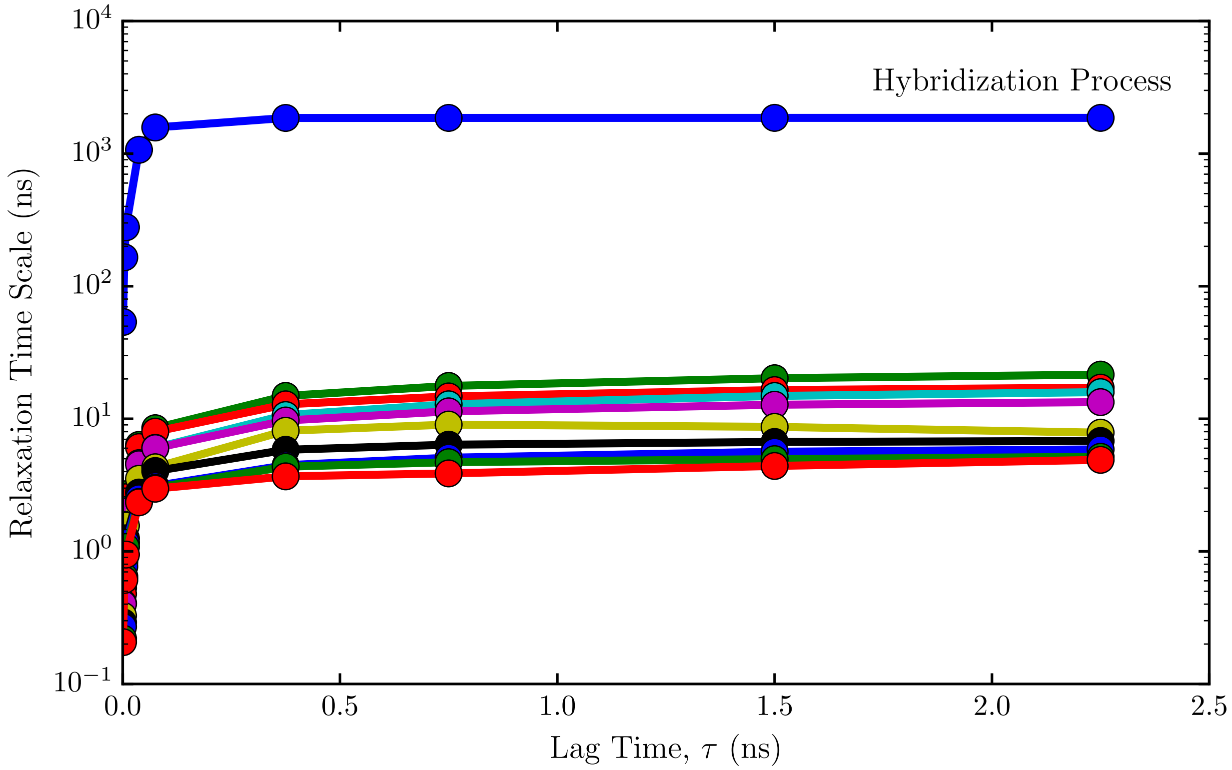


Figure 2.2: The ten longest relaxation time scales of MSMs built with increasing lag time τ. Only at sufficiently long lag times the relaxation time scales become independent of τ, which is a necessary condition for the system dynamics to be Markovian. The process with the longest relaxation time scale corresponds to the transition from the fully unbound to a hybridized state (Table 2.1).

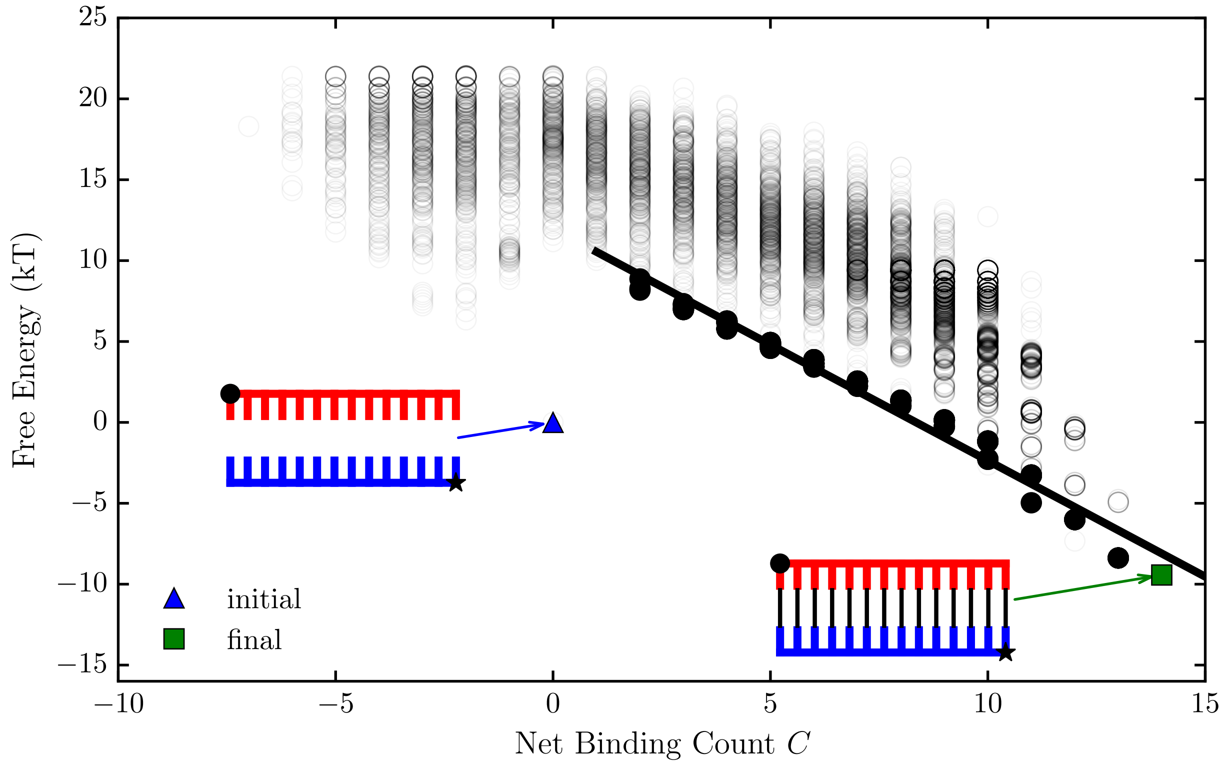


Figure 2.3: Free energy of each state, projected onto the Net Binding Count C. The free energy was calculated from the equilibrium probability distribution. Areas with more states appear darker. States with a large number of bound bases tend to have a low free energy. The filled points are the states that appear in the 20 most likely paths. The solid black line is the line of best fit of the filled points.

Methods section, we construct a series of MSMs from these discretized trajectories by counting transitions that occur over the lag time τ. If the Markovian dynamics described by an MSM is an accurate approximation of the underlying oxDNA dynamics, then the relaxation time scales (2.11) are independent of the lag time. To find the smallest value of τ for which this is the case, we plot in Figure 2.2 the implied time scales for the ten slowest relaxation processes of MSMs constructed using different lag times. We find that $\tau = 0.75$ns is the shortest acceptable lag time, and use the corresponding MSM in the following analysis. To verify the accuracy of this model, we repeat all further analysis for an MSM built using only one half of the coarse-grained trajectories, and find no significant variation of the results (see Supplemental Material for details).

The eigenvector of the transition matrix that has an eigenvalue of 1, when properly normalized, is the vector of equilibrium probabilities, π. From it, we compute the difference in free energy between the unbound and the i-th state as

$$\Delta G_i = -k_{\mathrm{B}} T \ln \left(\frac{\pi_i}{\pi_1} \right) . \tag{2.18}$$

Figure 2.3 shows these free energies projected onto the net binding count C, defined in equation (2.6). We find that the initial, unbound state has a significantly lower free energy than any other state with a net binding count close to zero. The reason for that is the loss of translational entropy when going from two independently diffusing strands to a single, partially bound duplex. Its magnitude depends on the size of the simulation box, and therefore on the concentration of DNA strands.[42] There is a large ensemble of duplex states that vary significantly in free energy, but there is a clear trend of decreasing free energy as the the net binding count increases. There is only a single state with $C = N$, the final and fully hybridized state, which is also the global free energy minimum of the system.

From this graph we estimate the average free energy gain of forming a base pair by calculating a linear fit to the free energies of those states that appear in the twenty most likely hybridization paths (see below). We obtain a negative slope of approximately $1.4 k_{\mathrm{B}} T$, or 0.83kcal/mol, per base pair. This is smaller than experimental measurements of base pair formation free energies[1,21,23,29,34,62] and oxDNA results obtained at the same temperature,[42] but matches oxDNA calculations at elevated temperatures.[41,42,55]

Figure 2.2 also shows that there is a large, 100-fold difference in the time scale between the slowest relaxation process and the second-slowest one. To illustrate the nature of this lengthy process, we show in Table 2.1 the five states that have the largest contributions to the corresponding eigenvector. The eigenvector is defined only up to a multiplicative constant, which we chose such that the unbound state has a contribution of -1. We see that the slowest process describes transformations between the unbound state on one side and completely or mostly hybridized states on the other. Among the latter, the perfect duplex has the largest weight, but states that show fraying at either or both ends also contribute significantly to this process. It is encouraging that the MSM identified the entirety of the hybridization process as the slowest dynamical process in the system. This result, which is physically intuitive, is based on a purely numerical analysis of the supplied simulation data.

For comparison, we include in Table 2.1 the equilibrium populations of the states that participate most in the slowest dynamical process. As expected from the results shown in Figure 2.3, the unbound state has negligible probability compared to the hybridized states. Interestingly, the equilibrium probabilities of the latter are essentially the same as their contributions to the slowest process. This also is intuitive: the slowest process corresponds to a transfer of probability from the unbound to the bound states. Because all other processes occur on shorter time scales, this transfer directly establishes the equilibrium probability distribution.

To obtain further kinetic information from the MSM, we calculate the typical time scale of hybridization. This can be done in multiple ways. First, we use the implied time scale (2.11) of the slowest relaxation process, which we have shown to correspond to DNA hybridization (Table 2.1). Second, we calculate the mean first passage time (2.15) from the the unbound to the fully bound state. Third, we compute the hybridization time scale from the total flux (2.17). All three methods yield a consistent estimate for the bimolecular rate constant of $k_{\text{hyb}} = ([\text{DNA}]\, t_{\text{hyb}})^{-1} = 1.6\text{x}10^{8}\ \text{M}^{-1}\text{s}^{-1}$. This is slower than the rate constant $7.7\text{x}10^{8}$ $\text{M}^{-1}\text{s}^{-1}$ obtained by Ouldridge and coworkers from oxDNA simulations, but still significantly faster than experimental measurements.[44]

A major advantage of describing the DNA dynamics by a Markov model is that we can obtain statistical information about the hybridization pathways using Transition Path Theory. We use MSMBuilder to calculate the paths with the highest net flux. The top eight paths, projected onto the net binding count C and the center of hybridization D, are shown in Figure 2.4 (an alternative illustration, in which the highest flux paths are superimposed onto the two-dimensional free surface, is included in the Supplemental Material as Figure S9). In this two-dimensional representation, the initial, unbound state is located at $(C, D) = (0, 0)$, and the final, fully hybridized state lies at $(N, 0)$. The eight highest-flux paths connect

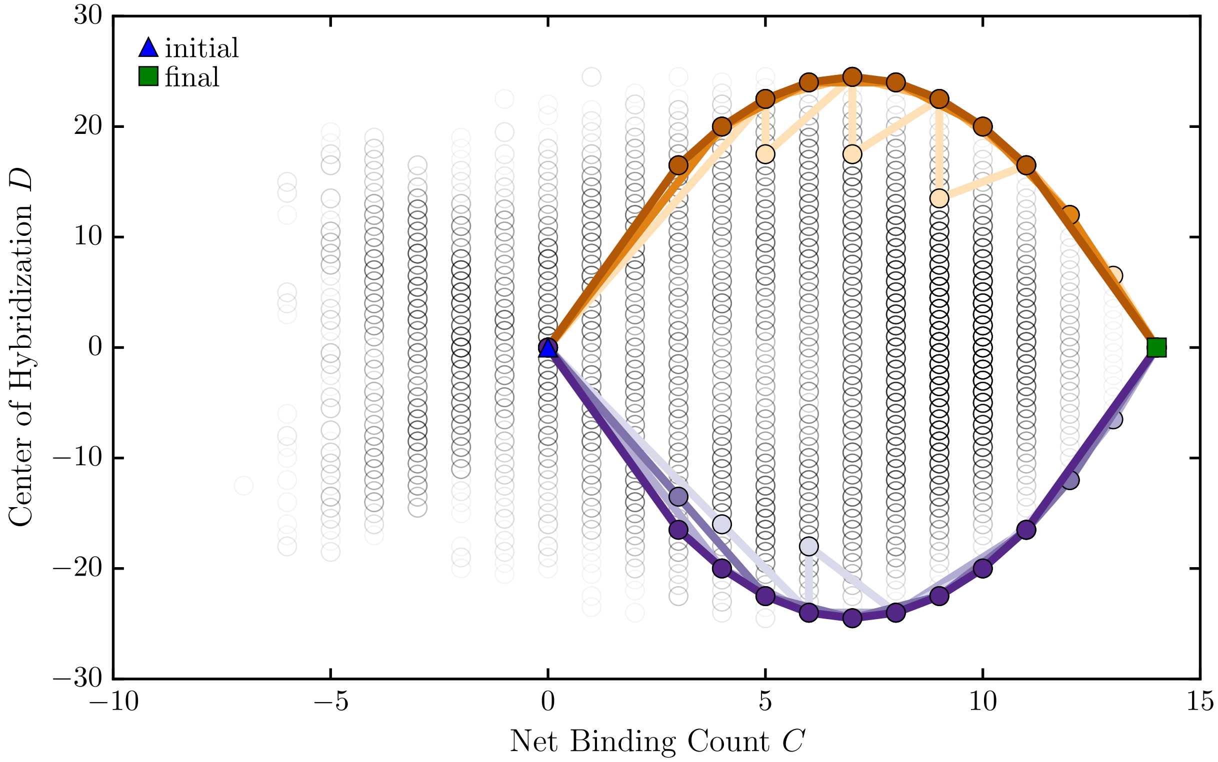


Figure 2.4: Eight most likely paths that connect the unbound state (triangle) to the completely hybridized state (square), shown on a two-dimensional projection of all states onto the net binding count C and the center of hybridization D.

these two states by a series of transitions that mainly occur along the upper or the lower boundary of the set of available states. Because each step in the Markov chain corresponds to a physical timestep of length τ, it is possible that hydrogen bonds between multiple bases form or break in a single step.

The location of these highest flux paths reveals the principal mechanism of hybridization: bases initially pair near either end of the strands, and the remaining bases then form hydrogen bonds with their opposing counterparts in a mostly sequential, zipper-like fashion. This process, however, is not entirely rigid: the pathways with the 6th and 7th highest flux contain transitions that appear as vertical lines in Figure 2.4. These transitions correspond to the formation of the next rung in the hydrogen bond ladder, while at the same time the base pair at the end of the strand breaks. Simple end fraying, i.e., the breaking and subsequent reforming of a terminal base pair, occurs frequently in the MSM trajectories but is not apparent in the path analysis, because it does not contribute to the net flux of a path.

To identify the states that contribute most to the hybridization process we compute the "fraction visited" of each state, defined as the fraction of reactive paths that pass through that state.[13] In the (C, D) representation, these states that are visited most frequently form a series of concentric layers, shown in Figure 2.5. States in the outer layer, which have the highest probability of being visited in a reactive trajectory, contain a single, consecutive segment of hybridized bases that begins on one end of the strands. The hybridization pattern of states in the second layer is similar, except that paired segment starts at the second-to-last base on either end of the strand. These states with a frayed end have a lower probability of being visited. States in even more central layers show more extensive end fraying. This figure illustrates that states with a single block of hybridized base pairs, which can be reached for example by the step-wise base pairing in a zipper-like mechanism, have the highest likelihood to be visited along a reactive trajectory.

Transition Path Theory is built around the committor q_i, which is defined as the probability that a trajectory starting from state i will reach the product state B before it reaches the reactant state A. It describes the position of a state along the $A \rightarrow B$ reaction, and

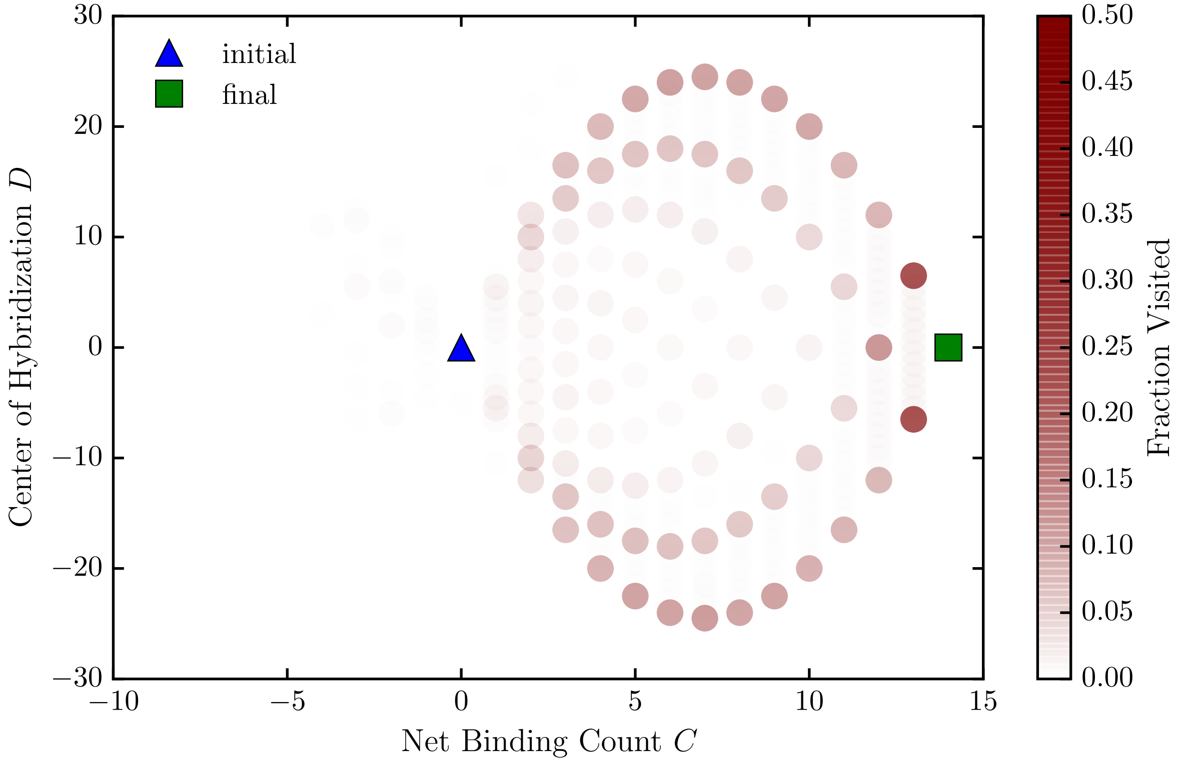


Figure 2.5: Illustration of the fraction visited of states along the hybridization pathways. States that are visited more often are shown in darker colors. Multiple nested pathways are discernible, each corresponding to a zipper mechanism. The outer layer contains states with adjacent bases bound, starting from either end of the strand. The next layer contains states in which base pairing starts at the second base from either end.

is considered the ideal reaction coordinate. It is completely determined by the choice of reactant and product states and the rules of the underlying dynamics, and does not require an arbitrary choice of a low-dimensional collective variable to measure the reaction progress. On the other hand, it can be difficult to obtain physical meaning from a set of committor values.

To measure the extent to which the committor values correlate with the physically intuitive net binding count, we show in Figure 2.6 a scatter plot of these two quantities. By construction, the unbound and the fully hybridized states have committor values of 0 and 1, respectively. We see that all states with $C \gtrsim 4$ have a committor value of nearly unity, which means that these states are almost certain to evolve towards a fully hybridized DNA duplex. States with $C \lesssim -4$, on the other hand, will almost always separate into two isolated strands. We find that states with a net binding count close to 0 show a large variability of their committor values, which span the entire range from 0 to 1. In this region, where states have a similar number of native and non-native bonds, it is not possible to infer the committor value from a measurement of the net binding count alone. In particular, we cannot select the subset of states with $q = 1/2$ by specifying C alone. The net binding count is therefore not a useful reaction coordinate, because one cannot describe the set of transition states with it.

The predictive power of the net binding count improves if we limit ourselves to only those states that have only native bonds between bases. This is the case for 4805 out of the 8942 states in the MSM. The observable C is then simply the number of such bonds, and its correlation with the committor is shown in Figure 2.7. We find that the range of committor values for a given choice of C is significantly shorter, and the majority of states with $q \approx 1/2$ have a binding count of 1. Nevertheless, there are also transition states with 2, 3, or 4 native base pairs.

Returning to the full state space shown in Figure 2.6, we find that 46 states have a committor between 0.45 and 0.55, which we consider transition states. To identify those that are most important to the hybridization process, we compute the fraction of reactive

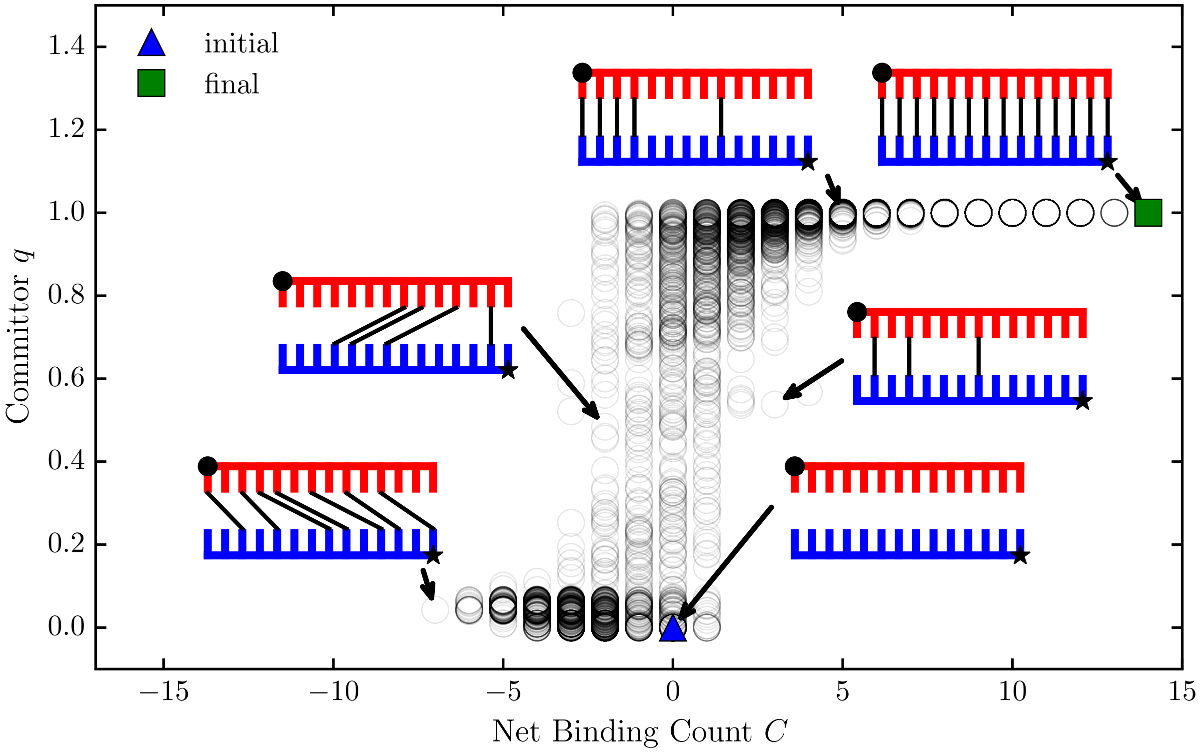


Figure 2.6: Committor values of each state, projected onto the net binding count. Higher opacity indicates a larger number of states. The committor, a measure of progress towards the hybridized product state, increases rapidly at small values of C. States with the same net binding count can have very different committor values.

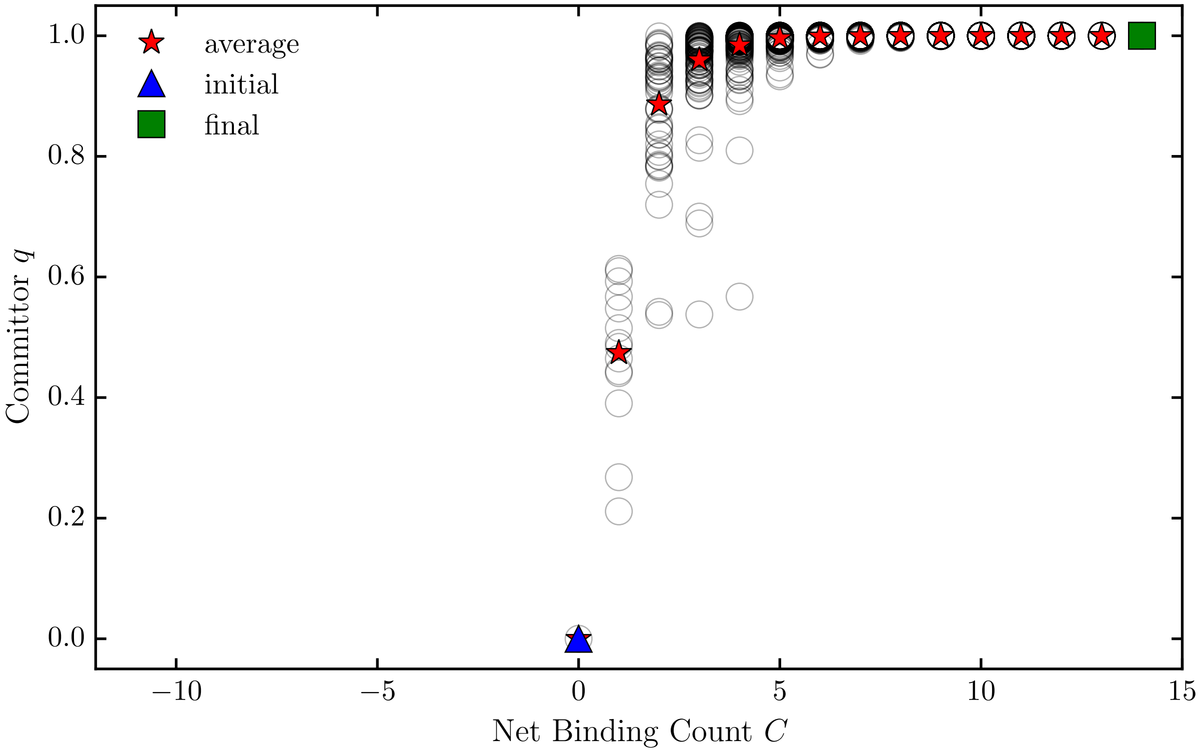


Figure 2.7: Committor values of the subset of states that have only native base pairs. Considering those states only, the average committor for states with a single, native base pair is 0.5.

trajectories that visit these states. Figure 2.8 shows the eight states with the highest fraction visited. The top two most visited transition states contain two native base pairs at either end, and together are visited in approximately 11% of hybridization trajectories. The next five transition states have only a single native base pair at different positions. Following with significantly lower probability, the eighth transition state is the first to contain nonnative in addition to native base pairs.

The sum of these probabilities, i.e., the probability that a hybridization pathway visits any of these states, is significantly less than one. The reason for that is that due to relative large lag time of our model, multiple base pairs can form or break within a single step of the Markov chain. The system can therefore skip these states, and evolve from the unbound to the duplexed state without visiting any transition state along the way.

The results indicate that the initial formation of the first one or two base pairs is crucial to decide the fate of the newly formed complex. We therefore calculated which bases along the DNA strands are most likely to be involved in those initial contacts. To this end we calculated the probability that base i in strand α is bound to any other base in the ensemble of states that is weighted by their probability to be visited directly from the unbound state:

$$\frac{\displaystyle\sum_{s=2}^{M} T_{1s} h_{\alpha i}(s)}{1 - T_{11}}. \tag{2.19}$$

The results of this calculation are shown in Figure 2.9. We find that the bases at the third position from either end of the strand has the highest probability of being part of the initial base pairing. Bases at the ends or in the middle of the strands have a significantly lower probability. These observations are consistent across both strands.

Having established that bases close to, but not directly at, the ends of the DNA strands are more likely to participate in the initial base pairing, we now study the effect of these base pairs on the probability of successful hybridization. Figure 2.10 shows the committor values of the 14 states that have a single, native base pair. We find that if this base pair is located positions 3–6 or 12, then the probability that the DNA will fully hybridize is large.

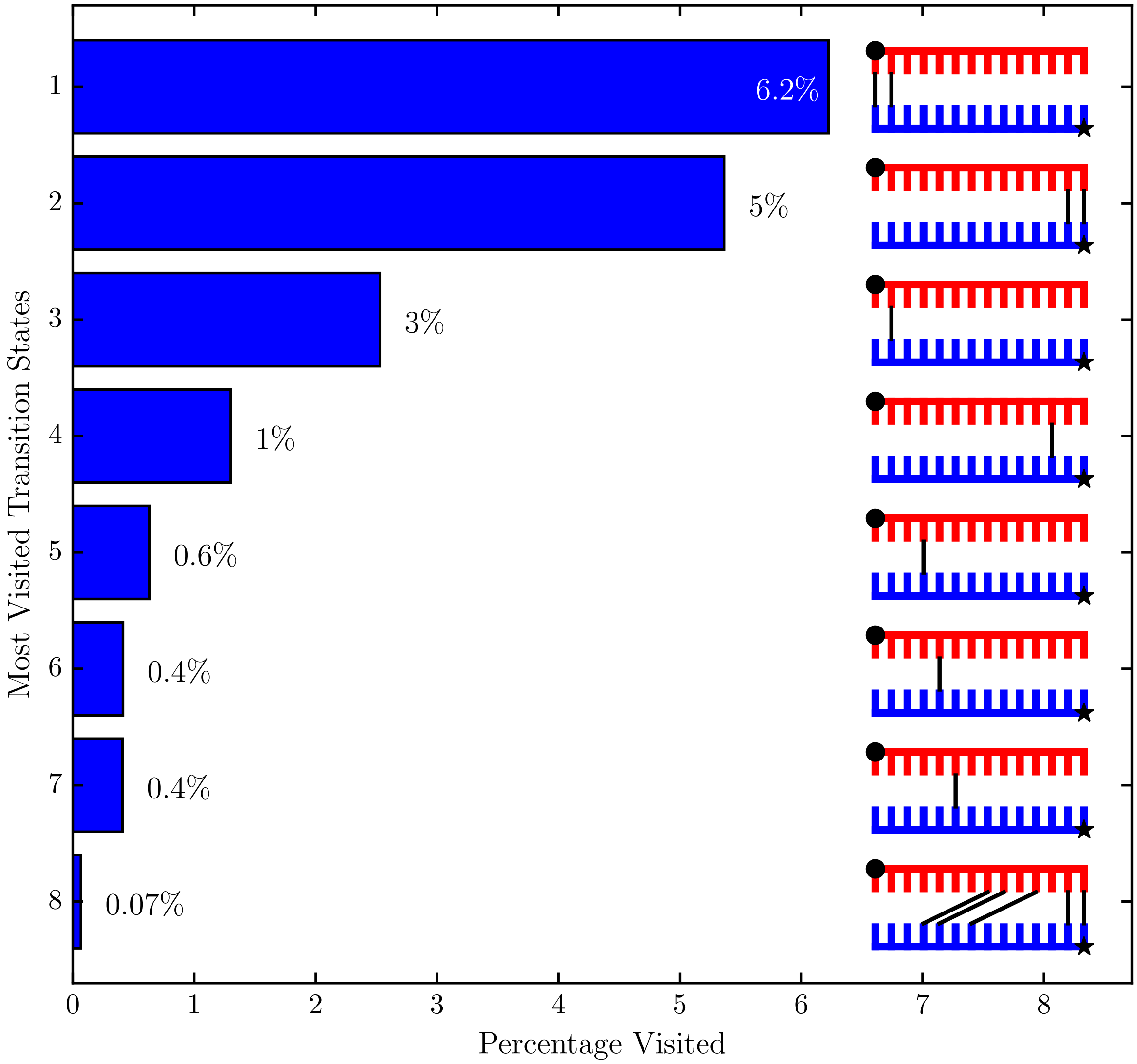


Figure 2.8: The fraction visited of transition states, i.e., states with committor values between 0.45 and 0.55. Transition states with native base pairs at the ends are often visited along the hybridization pathway.

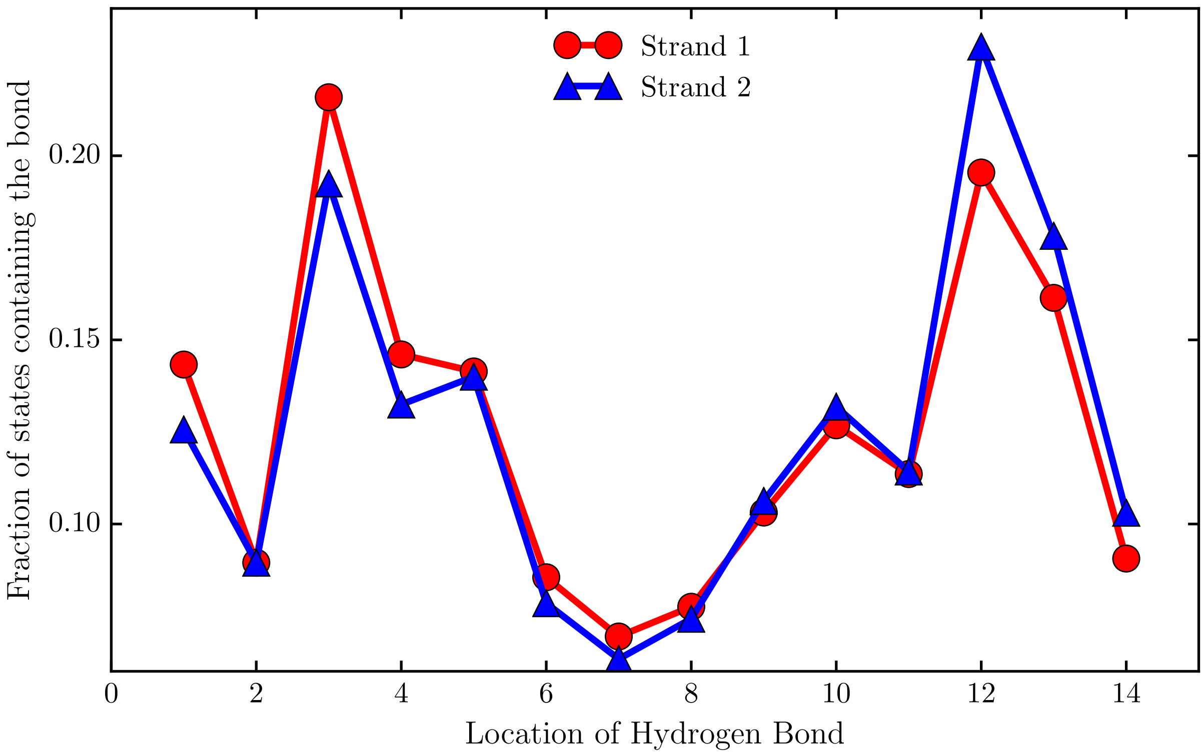


Figure 2.9: Probability that a base will appear in a first contact of the two strands. The third base from either end has a significantly higher probability to participate in the initial base pairing than a base in the middle or at the end of a strand.

If the base pair forms in the middle or at either end, then the nascent duplex is more likely to fall apart.

2.4 Discussion

We investigated how DNA hybridizes and identified the important states and pathways in going from ssDNA to dsDNA. We identified the important states and pathways by simulating coarse-grained DNA and creating an MSM from the simulation data. Our results suggest the most important transition states contain one or two native base pairs. To have a high likelihood of binding, we only required 3 native base pairs. Our most important pathways also show a zipping up process as suggested by previous work.[46,66] Our results suggest that once a few bases are bound, DNA is likely to finish binding fully. The DNA tended to

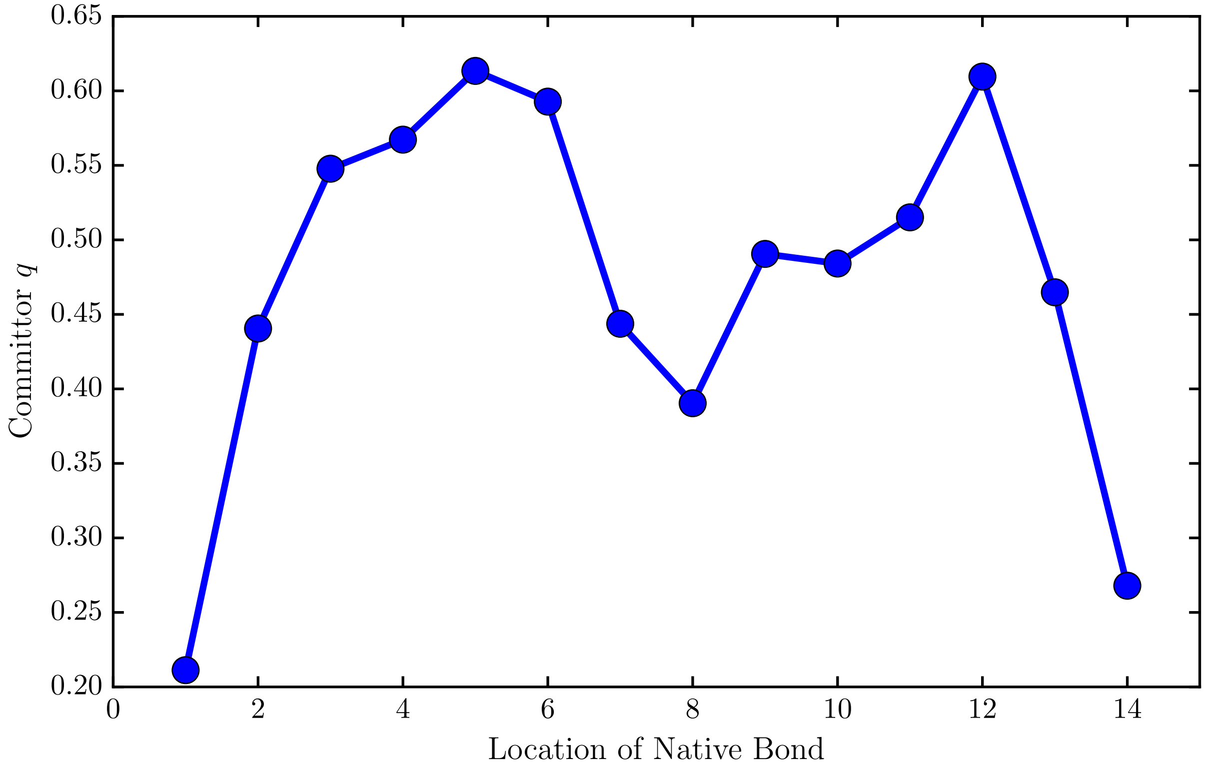


Figure 2.10: The committor of states that contain only a single native base pair and no non-native base pairs. If the single base pair is at the end or in the middle then the committor is low, indicating that this nascent duplex is more likely to fall apart than to proceed to full hybridization.

bind from one end instead of in the middle (Figure 2.5). That suggests the identity of the bases on the ends could control the kinetics more than those in the middle. The number of bases required before full hybridization in our model seems to be smaller than in other experiments,[12,46] but is consistent with results in previous oxDNA2 simulations.[42]

One strength of our approach was building up a model from simulation data. From this, we were able to get information on the large number of potential paths and the kinetics associated with those paths. By the nature of this problem, there is a large state space of theoretically 1.6x10^{13} states but we only access 8952 states. This shows the number of relevant states for hybridization is much smaller than the theoretical number of possible states.

Hybridization has been studied using oxDNA2 but using forward flux sampling, a technique for accelerating dynamics.[42] Our results are consistent with those found by Ouldridge and coworkers.[42] Qualitatively, they find a larger probability of initial binding near the ends of the strands. Additionally they find the probability of reaching the fully hybridized state is larger for base pairs near the middle of the strand in comparison towards the ends. They also observe a nucleation and zipping process.

Because the MSM is built from oxDNA simulations it can at best be as accurate as this coarse-grained force field. Any deficiency in the underlying force field will propagate into the MSM. Other coarse-grained models of DNA are available, for example the 3SPN force field .[25,26,51] Previously it was found that oxDNA and 3SPN exhibit different hybridization pathways.[42] Not surprisingly, our MSM is consistent with previous studies of the oxDNA force field.

Another factor that limits the analysis is the choice of variable of interest. We partitioned the states by hydrogen bonding and that does not seem to be an appropriate reaction coordinate. In a good reaction coordinate, each value on the reaction coordinate should have a single value of a committor.[45] This property is not seen in our committor plot (Figure 2.6). For instance, at a Net Binding Count of 0 we find committor values that range from 0 to 1. Therefore, while the extent of hydrogen bonding is a physically intuitive descriptor, it is not a good reaction coordinate. It does seem to be valid for states that only contain native base pairs (Figure 2.7).

Our study is limited to a single DNA sequence, which makes it difficult to determine whether our findings are specific to this particular sequence, or whether they generally hold for arbitrary sequences. It would therefore be valuable to perform similar analyses for a wider range of sequences. Of particular interest would be those that are uniform or highly repetitive. As described in the Methods section, such sequences can exhibit a vastly greater number of potential hydrogen bonding pattern. For example, a system of 14-base poly-A and poly-T strands has 1.6×10^{13} possible states, compared to the 1.9×10^{8} states of our specific sequences. The importance of this vastly enlarged state space has been confirmed in preliminary simulations of the former, in which 306562 states were visited; a large increase over the 8942 visited states in this work. This suggests that hybridization pathways may differ significantly across sequence, which warrants further investigation. Unfortunately, the much larger state space creates a significant computational burden on the accurate determination of the transition matrix and on the numerical operations required for the TPT analysis.

Creating an MSM has an important advantage over some other enhanced sampling techniques such as Umbrella Sampling,[61] Metadynamics,[32] or Forward Flux Sampling.[2] In those methods one must first define a coordinate of interest, often called collective variable (CV), before running the simulation. However, it is often not clear beforehand which CV provides a good description of the process of interest, and one would have to perform another set of simulations if one wanted to consider different CVs. In our approach there is no need to specify a proposed CV before collecting simulation data. The MSM itself does not describe the system dynamics in terms of a one-dimensional reaction coordinate. Instead it views the dynamics as a sequence of transition in a arbitrarily complex network of states. Where we did make a choice is in the construction of the MSM when we picked a state decomposition based on the pattern of inter-strand hydrogen bonds. Making a different choice would require the calculation of another MSM transition matrix, but it would not require a new set of molecular mechanics simulations which would be computationally expensive.

2.5 Conclusion

We have shown that the combination of coarse-grained Brownian dynamics simulations, Markov State Models, and Transition Path Theory forms a powerful tool to study the kinetic properties and dynamical pathways of DNA hybridization. Our results are consistent with the classical zipper model: once an initial nucleus of hybridized base pairs is formed, it grows and neighboring bases zip up sequentially. We find the size of the initial nucleus to lie between one and two base pairs, and that in general the net number of base pairs is not a sufficiently informative observable to characterize the transition state ensemble.

Chapter 3

DETERMINING THE BEST MODELS OF DYNAMIC FORCE SPECTORSCOPY

3.1 Experimental Pulling Motivation

An important aspect to further our knowledge of chemical systems is to learn about their underlying free energy surfaces. One way to investigate these surfaces is through dynamic force spectroscopy (DFS). Using DFS, expermentalists can physically manipulate (apply force to) systems and measure how the systems respond. The forces they apply increase over time and eventually the system ruptures. This rupture can come in many forms including protein unfolding,[48,49] DNA unbinding,[58,68] and protein-ligand unbinding.[35,59] Within these experiments, the rupture event of interest will be performed many times in order to get statistics on the strength of the system. One important aspect to these pulling experiments is that the rate at which the experimentalists pull is important for the statistics of rupture. For these pulling experiments, the faster you pull, the larger the force will be at the rupture point.

In order to determine the shape of the underlying system's potential, experimentalists perform many experiments on the same system, while using a large range of pulling velocities. Then they fit their data using one of many proposed models for the relationship between the rupture force as a function of pulling rate. These models include the Bell-Evans (Section 3.2.2), Hummer-Szabo (2003) (Section 3.2.3), Dudko-Hummer-Szabo (2006) (Section 3.2.4), Friddle-Noy-De Yoreo (Section 3.2.5), and Bullerjahn-Sturm-Kroy (2014) (Section 3.2.6). All of these models approach the pulling problem in slightly different ways, and all claim to give correct and reasonable answers. However, as we have investigated each of these models, we find that they all differ slightly and give varying results. We analyzed the performance of each of these models, both in how well they fit the simulated data given the correct fitting parameters, and how well they can output the correct fitting parameters given our simulated data.

3.2 Theoretical Models of Rupture

Many groups have proposed models that link the experimentally obtained rupture force data to the shape of the underlying potential. There have been two main approaches to modeling the rupture force. One approach is to determine how the rate responds to an external force in a phenomenlogical way. This approach is used in the Bell-Evans and the Friddle-Noy-De Yoreo models. The other approach is to model the response of the probability distribution $P(x,t)$ to force and then calculate a survival probability $S(t) = \int_{-\infty}^{x_\mathrm{b}} P(x,t)\mathrm{d}x$ where x_b is the location of the transition state. Assuming first-order kinetics holds, the survival probability of a system is

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \dot{S}(t) = -k(t)S(t) \tag{3.1}$$

and therefore $S(t)$ is

$$S(t) = \exp\Big(- \int_0^t k(t')\mathrm{d}t' \Big) \tag{3.2}$$

where $k(t)$ is the instantaneous dissociation rate at time t.

3.2.1 Kramers (1940)

Before incorporating the external pulling forces, it will be instructive to look at the rate of dissociation of a single underlying potential. This was done in 1940 with the work of Kramers.[31] Kramers solved the diffusion equation for Brownian dynamics with an assumption of a large barrier and found an important expression for the dissociation rate,

$$w = \frac{D|\sigma(x)\mathrm{e}^{U(x)/k_\mathrm{B}T}|_{x=B}^{A}}{\displaystyle\int_A^B \mathrm{e}^{U(x)/k_\mathrm{B}T}dx}, \tag{3.3}$$

where D is the diffusion constant, $\sigma(x)$ is the density distribution, $U(x)$ is the underlying potential,k_B is the Boltzmann constant, T is the temperature and A and B are two points on the x-coordinate. The derivation of this expression is shown below. One potential of interest is the harmonic cusp potential. This potential is described as

$$U(x) = \begin{cases} \frac{1}{2}k_{\mathrm{sys}}x^2 & \text{if } x < x_{\mathrm{b}} \\ -\infty & \text{otherwise} \end{cases} \tag{3.4}$$

For a harmonic cusp potential, the Kramers rate becomes

$$k_0 = \frac{2D\Delta G^{\ddagger}}{k_{\mathrm{B}}Tx_{\mathrm{b}}^2}\sqrt{\frac{\Delta G^{\ddagger}}{\pi k_{\mathrm{B}}T}}\mathrm{e}^{-\Delta G^{\ddagger}/k_{\mathrm{B}}T}, \tag{3.5}$$

where $\Delta G^{\ddagger}$ is the potential at the transition point x_{b}. The derivation of this expression from the generic expression (Equation 3.3) is shown below.

Steady State Equation

If the barrier to escape a well is large, we can assume the distribution of particles in the well are in a stationary state near the bottom of the well. If our particles are in a steady state, then the 1D Smoluchowski diffusion equation

$$\frac{\partial\sigma(x)}{\partial t} = -\frac{\partial}{\partial x}\Big(\frac{F(x)D}{k_{\mathrm{B}}T}\sigma(x) - D\frac{\partial\sigma(x)}{\partial x}\Big) \tag{3.6}$$

where $F(x)$ is the force acting on the particle due to the the underlying potential will be equal to 0. For this to occur, the inner piece

$$w = \frac{F(x)D}{k_{\mathrm{B}}T}\sigma(x) - D\frac{\partial\sigma(x)}{\partial x} \tag{3.7}$$

must be a constant. Here we have chosen the constant w as used in Kramers. This can be rewritten as

$$w = -D\mathrm{e}^{-U(x)/k_{\mathrm{B}}T}\frac{\partial}{\partial x}(\sigma(x)\mathrm{e}^{U(x)/k_{\mathrm{B}}T}). \tag{3.8}$$

Then take the integral from two points A and B on the x-coordinate to obtain

$$-\int_A^B \frac{\mathrm{e}^{U(x)/k_{\mathrm{B}}T}}{D}w\mathrm{d}x = \int_A^B \frac{\partial}{\partial x}\sigma(x)\mathrm{e}^{U(x)/k_{\mathrm{B}}T}\mathrm{d}x \tag{3.9}$$

By isolating w we find

$$w = \frac{D\Big|\sigma(x)\mathrm{e}^{U(x)/k_\mathrm{B}T}\Big|_{x=B}^{A}}{\int_A^B \mathrm{e}^{U(x)/k_\mathrm{B}T}\mathrm{d}x} \tag{3.10}$$

which is equation 13 in reference 31 and equation 3.3 of this work.

Derivation of the Escape Rate for the Harmonic Cusp Potential

Let us assume that we have a quasi-stationary state near A and essentially no particles have arrived at B yet. We can apply equation 3.10

$$w = D\sigma_A\Big(\int_A^B \mathrm{e}^{U(x)/k_\mathrm{B}T}\mathrm{d}x\Big)^{-1}, \quad \sigma_A = (\sigma(x)\mathrm{e}^{U(x)/k_\mathrm{B}T})_{\text{near A}} \tag{3.11}$$

Assuming $U(x)$ can be described near A as a harmonic oscillator by $\frac{1}{2}k_\mathrm{sys}x^2$, we can calculate the number n_A of particles near A:

$$n_a = \int_{-\infty}^{\infty} \sigma_A \mathrm{e}^{-k_\mathrm{sys}x^2/2k_\mathrm{B}T}\mathrm{d}x = \sigma_A\sqrt{\frac{2\pi k_\mathrm{B}T}{k_\mathrm{sys}}} \tag{3.12}$$

The rate k_0 is w/n_A. Overall k_0 is

$$k_0 = \frac{w}{n_A} = D\sqrt{\frac{k_\mathrm{sys}}{2\pi k_\mathrm{B}T}}\Big(\int_A^B \mathrm{e}^{U(x)/k_\mathrm{B}T}\mathrm{d}x\Big)^{-1} \tag{3.13}$$

To evaluate the integral, let us assume that the potential is exactly $U(x) = \frac{1}{2}k_\mathrm{sys}x^2$ with the transition state occuring at x_b. We Taylor expand around x_b and keep terms up to the linear portions. This gives us a new expression $U(x) = \frac{k_\mathrm{sys}x_\mathrm{b}^2}{2} + k_\mathrm{sys}x_\mathrm{b}(x - x_\mathrm{b})$. This is useful because we can now evaluate the integral.

$$\int_A^B \mathrm{e}^{U(x)/k_\mathrm{B}T}\mathrm{d}x = \int_{-\infty}^{x_\mathrm{b}} \mathrm{e}^{(k_\mathrm{sys}x_\mathrm{b}^2/2 + k_\mathrm{sys}x_\mathrm{b}(x-x_\mathrm{b}))/k_\mathrm{B}T}\mathrm{d}x = \frac{\mathrm{e}^{-k_\mathrm{sys}x_\mathrm{b}^2/2k_\mathrm{B}T}k_\mathrm{B}T}{k_\mathrm{sys}x_\mathrm{b}} \tag{3.14}$$

Combining equations 3.13 and 3.14, we now find k_0

$$k_0 = \frac{2D\Delta G^\ddagger}{x_\mathrm{b}^2 k_\mathrm{B}T}\sqrt{\frac{\Delta G^\ddagger}{\pi k_\mathrm{B}T}}\mathrm{e}^{-\Delta G^\ddagger/k_\mathrm{B}T}, \;\; \Delta G^\ddagger = k_\mathrm{sys}x_\mathrm{b}^2/2 \tag{3.15}$$

3.2.2 Bell-Evans (1997)

The Bell-Evans model of dynamic force spectroscopy was developed by Evans and colleagues in 1997 as an extension of the Bell model of off-rates modified by an external force and has become the most widley used model of dynamic force spectroscopy.[6,19] Bell first proposed a relationship between the applied force F and the modified off-rate as

$$k(F) = k_0 \mathrm{e}^{\frac{F x_\mathrm{b}}{k_\mathrm{B} T}} \tag{3.16}$$

They use this equation to calculate the rupture force probability density as well as to find the most probable rupture force. In the original 1997 paper, they perform Monte Carlo simulations of Brownian bond breakage under a force ramp, using a large range of loading rates. They then create a histogram at each loading rate and finds the most probable rupture force (F_r^*). They find three regions of differing proportionality between the loading rate and the rupture force. In the first region of slow loading rates, they find the rupture force is proportional to a weak power of loading rate ($F_\mathrm{r}^* \propto \dot{F}^n$). At intermediate rates, they find the rupture force is proportional to the logarithm of the loading rate ($F_\mathrm{r}^* \propto \ln \dot{F}$). And at an ultrafast regime the rupture force is proportional to the square root of the loading rate ($F_\mathrm{r}^* \propto \sqrt{\dot{F}}$).

While their simulations found these three regimes, the approach used to fit experimental models only focused on the logarithmic relationship. This lead to the equation to fit rupture force data (full derivation shown below):

$$F_\mathrm{r}^* = \frac{k_\mathrm{B} T}{x_\mathrm{b}} \ln \frac{\dot{F} x_\mathrm{b}}{k_\mathrm{B} T k_0} \tag{3.17}$$

When rupture force data is plotted on a F_r^* versus $\ln(\dot{F})$ graph, Equation 3.17 will be a straight line with slope $\frac{k_\mathrm{B} T}{x_\mathrm{b}}$. For this equation, there are two fit parameters, the position of the barrier, x_b and the intrinsic off-rate k_0. When rupture forces are measured in experiment, they often cannot be well fit with a single straight line. These rupture force graphs are interpreted as having two barriers and the rupture forces are then fit piecewise to give the heights and positions of these barriers.[7,30,35]

Derivation of Bell-Evans Equation

We start with expression 3.16 for the off-rate in the presence of a statice force F. Let the rate of escape as a function of the force, F be:

$$k(F) = k_0 \mathrm{e}^{\frac{F x_\mathrm{b}}{k_\mathrm{B} T}} \tag{3.18}$$

where k_0 is the intrinsic unbinding rate, and x_b is the location of the barrier. This description of the rate of escape comes from the model by Bell.[6] It requires the particle to be in quasi-equilibrium.

From Evans eq. (11):

$$P_t(F,t) = k(F) \mathrm{exp}\Big\{ - \int_0^t k[F(t')] \mathrm{d}t' \Big\} \tag{3.19}$$

where $P_t(F,t)$ is the probability that the particle escapes the well at time t and $F(t)$ is the force exerted on the particle by the external force at time t. To calculate the most likely rupture force, we compute the derivative

$$\frac{\partial P_t(F,t)}{\partial F} = \frac{\partial k(F)}{\partial F} \mathrm{exp}\Big\{ - \int_0^t k[F(t')] \mathrm{d}t' \Big\} + k(F) \frac{\partial}{\partial F} \mathrm{exp}\Big\{ - \int_0^t k[F(t')] \mathrm{d}t' \Big\} \tag{3.20}$$

Plug equation 3.16 into equation 3.20

$$\begin{aligned} \frac{\partial P_t(F,t)}{\partial F} =& \frac{\partial k_0 \mathrm{e}^{\frac{F x_\mathrm{b}}{k_\mathrm{B} T}}}{\partial F} \mathrm{exp}\Big\{ - \int_0^t k_0 \mathrm{e}^{\frac{F x_\mathrm{b}}{k_\mathrm{B} T}} \mathrm{d}t' \Big\} + \\ & k_0 \mathrm{e}^{\frac{F x_\mathrm{b}}{k_\mathrm{B} T}} \frac{\partial}{\partial F} \mathrm{exp}\Big\{ - \int_0^t k_0 \mathrm{e}^{\frac{F x_\mathrm{b}}{k_\mathrm{B} T}} \mathrm{d}t' \Big\} \\ =& k_0 \frac{x_\mathrm{b}}{k_\mathrm{B} T} \mathrm{e}^{\frac{F x_\mathrm{b}}{k_\mathrm{B} T}} \mathrm{exp}\Big\{ - \int_0^t k_0 \mathrm{e}^{\frac{F x_\mathrm{b}}{k_\mathrm{B} T}} \mathrm{d}t' \Big\} - \\ & k_0 \mathrm{e}^{\frac{F x_\mathrm{b}}{k_\mathrm{B} T}} \mathrm{exp}\Big\{ - \int_0^t k_0 \mathrm{e}^{\frac{F x_\mathrm{b}}{k_\mathrm{B} T}} \mathrm{d}t' \Big\} \frac{\partial}{\partial F} \int_0^t k_0 \mathrm{e}^{\frac{F x_\mathrm{b}}{k_\mathrm{B} T}} \mathrm{d}t' \end{aligned} \tag{3.21}$$

We know the following to be true about partial derivatives

$$\frac{\partial}{\partial t} = \frac{\partial F}{\partial t} \frac{\partial}{\partial F} = \dot{F} \frac{\partial}{\partial F} \tag{3.22}$$

So that

$$\begin{aligned}\frac{\partial P_t(F,t)}{\partial F} =& k_0 \frac{x_\mathrm{b}}{k_\mathrm{B}T} \mathrm{e}^{\frac{Fx_\mathrm{b}}{k_\mathrm{B}T}} \exp\Big\{ -\int_0^t k_0 \mathrm{e}^{\frac{Fx_\mathrm{b}}{k_\mathrm{B}T}} \mathrm{d}t' \Big\} - \\ & k_0 \mathrm{e}^{\frac{Fx_\mathrm{b}}{k_\mathrm{B}T}} \exp\Big\{ -\int_0^t k_0 \mathrm{e}^{\frac{Fx_\mathrm{b}}{k_\mathrm{B}T}} \mathrm{d}t' \Big\} \frac{1}{\dot{F}} \frac{\partial}{\partial t} \int_0^t k_0 \mathrm{e}^{\frac{Fx_\mathrm{b}}{k_\mathrm{B}T}} \mathrm{d}t' \\ =& k_0 \frac{x_\mathrm{b}}{k_\mathrm{B}T} \mathrm{e}^{\frac{Fx_\mathrm{b}}{k_\mathrm{B}T}} \exp\Big\{ -\int_0^t k_0 \mathrm{e}^{\frac{Fx_\mathrm{b}}{k_\mathrm{B}T}} \mathrm{d}t' \Big\} - \\ & k_0 \mathrm{e}^{\frac{Fx_\mathrm{b}}{k_\mathrm{B}T}} \exp\Big\{ -\int_0^t k_0 \mathrm{e}^{\frac{Fx_\mathrm{b}}{k_\mathrm{B}T}} \mathrm{d}t' \Big\} \frac{1}{\dot{F}} k_0 \mathrm{e}^{\frac{Fx_\mathrm{b}}{k_\mathrm{B}T}}\end{aligned} \tag{3.23}$$

This derivative is zero at the maximum value

$$\frac{\partial P_t(F,t)}{\partial F} = k_0 \mathrm{e}^{\frac{F_\mathrm{r}^* x_\mathrm{b}}{k_\mathrm{B}T}} \exp\Big\{ -\int_0^t k_0 \mathrm{e}^{\frac{Fx_\mathrm{b}}{k_\mathrm{B}T}} \mathrm{d}t' \Big\} \Big(\frac{x_\mathrm{b}}{k_\mathrm{B}T} - k_0 \mathrm{e}^{\frac{F_\mathrm{r}^* x_\mathrm{b}}{k_\mathrm{B}T}} \Big) = 0 \tag{3.24}$$

$$\frac{x_\mathrm{b}}{k_\mathrm{B}T} = \frac{1}{\dot{F}} k_0 \mathrm{e}^{\frac{F_\mathrm{r}^* x_\mathrm{b}}{k_\mathrm{B}T}} \tag{3.25}$$

Solving for F_r^* we find

$$F_\mathrm{r}^* = \frac{k_\mathrm{B}T}{x_\mathrm{b}} \ln \frac{\dot{F} x_\mathrm{b}}{k_\mathrm{B}T k_0} \tag{3.26}$$

which is equation 3.17.

3.2.3 Hummer-Szabo (2003)

Hummer-Szabo created a model based on a harmonic trap pulling out of a quadratic well with a cusp at the top of the barrier. This is the same potential used in the Kramers harmonic cusp rate (equation 3.4). They model a system which follows Brownian dynamics and can be solved through the stochastic differential equation,

$$\dot{x}(t) = -\frac{Dk_\mathrm{sys}}{k_\mathrm{B}T} x - \frac{Dk_\mathrm{bias}(x - vt)}{k_\mathrm{B}T} + R(t), \tag{3.27}$$

where k_bias is the spring constant of the moving biasing trap, v is the velocity of the moving biasing trap, and $R(t)$ is a Gaussian random force with zero mean, $\langle R(t) \rangle = 0$, and variance $\langle R(t)R(t') \rangle = 2D\delta(t - t')$.[27] This stochastic differential equation can be solved for the average position x at time t, with the initial position $x(0) = 0$.

$$\langle x \rangle(t) = \frac{v_e k_\mathrm{B}T}{Dk_\mathrm{tot}} \Big(\frac{Dk_\mathrm{tot}}{k_\mathrm{B}T} t + \mathrm{e}^{-Dk_\mathrm{tot}/k_\mathrm{B}Tt} - 1 \Big); \quad k_\mathrm{tot} = k_\mathrm{sys} + k_\mathrm{bias}, \tag{3.28}$$

where $v_e = vk_{\text{bias}}/k_{\text{tot}}$.

Hummer-Szabo created a model in a different way from the Bell-Evans model. Instead of using the Bell relationship of a purely exponential effect of force on the off-rate, they start with the Kramers off-rate and then model the pulling as a moving, shrinking barrier. For a slow pulling rate, one can make an adiabatic approximation which means the system has enough time for the particle to equilibriate in the modified potential. In the traditional Kramers formulation of the off-rate, there is no external pulling forces and it is only the rate of escape past the barrier by itself. Using the adiabatic approximation, the barrier becomes a time-dependant barrier with a shrinking height which moves closer to the minimum of the potential. This modifies the Kramers rate (equation 3.5) by replacing x_{b} with $x_{\text{b}} - \langle x\rangle(t)$ and k_{sys} with $k_{\text{tot}} = k_{\text{sys}} + k_{\text{bias}}$. We can obtain the new biased escape rate in relation to the unbiased escape rate,

$$k_0(x_{\text{b}}-\langle x\rangle(t);k_{\text{tot}}) = k_0\Big(\frac{k_{\text{tot}}}{k_{\text{sys}}}\Big)^{3/2}\frac{x_{\text{b}}-\langle x\rangle(t)}{x_{\text{b}}}\mathrm{e}^{\dfrac{k_{\text{sys}}x_{\text{b}}^2}{2k_{\text{B}}T}\Big(1-\dfrac{k_{\text{tot}}}{k_{\text{sys}}}\Big(\dfrac{x_{\text{b}}-\langle x\rangle(t)}{x_{\text{b}}}\Big)^2\Big)};\quad k_{\text{tot}} = k_{\text{sys}}+k_{\text{bias}} \tag{3.29}$$

In the slow pulling limit, we can rewrite $\langle x\rangle(t)$ as $vk_{\text{bias}}t/k_{\text{tot}}$ by Taylor expanding equation 3.28 around $t \to \infty$ and the equation becomes,

$$k_0(x_{\text{b}}-\langle x\rangle(t);k_{\text{tot}}) = k_0\Big(\frac{k_{\text{tot}}}{k_{\text{sys}}}\Big)^{3/2}\frac{x_{\text{b}}-vk_{\text{bias}}t/k_{\text{tot}}}{x_{\text{b}}}\mathrm{e}^{\dfrac{k_{\text{sys}}x_{\text{b}}^2}{2k_{\text{B}}T}\Big(1-\dfrac{k_{\text{tot}}}{k_{\text{sys}}}\Big(\dfrac{x_{\text{b}}-vk_{\text{bias}}t/k_{\text{tot}}}{x_{\text{b}}}\Big)^2\Big)}. \tag{3.30}$$

The force $F(t)$ acting on the particle from the biasing potential is $vk_{\text{bias}}t$ if we assume the particle is near 0. If $k_{\text{sys}} \gg k_{\text{bias}}$ then we can replace k_{tot} with k_{sys} and find the following equation in terms of the applied force

$$k_0(x_{\text{b}}-\langle x\rangle(t);k_{\text{tot}}) = k_0(1)^{3/2}\frac{x_{\text{b}}-F(t)/k_{\text{sys}}}{x_{\text{b}}}\mathrm{e}^{\dfrac{k_{\text{sys}}x_{\text{b}}^2}{2k_{\text{B}}T}\Big(1-1\Big(\dfrac{x_{\text{b}}-F(t)/k_{\text{sys}}}{x_{\text{b}}}\Big)^2\Big)}. \tag{3.31}$$

Simplifying and consolidating terms, we see that instead of an exponential relationship, we have something more complicated as a function of force.

$$k_0(x_\mathrm{b} - \langle x\rangle(t); k_\mathrm{tot}) = k_0\Big(1 - \frac{F(t)}{x_\mathrm{b}k_\mathrm{sys}}\Big)\mathrm{e}^{\dfrac{k_\mathrm{sys}x_\mathrm{b}^2}{2k_\mathrm{B}T}\Big(1 - \Big(1 - \dfrac{F(t)}{x_\mathrm{b}k_\mathrm{sys}}\Big)^2\Big)}. \tag{3.32}$$

In the case where $k_\mathrm{sys} \to \infty$, the Hummer-Szabo expression becomes the same as the Bell expression,

$$k_0(x_\mathrm{b} - \langle x\rangle(t); k_\mathrm{tot}) = k_0\mathrm{e}^{Fx_\mathrm{b}/k_\mathrm{B}T}. \tag{3.33}$$

Fast Pulling Velocities

Start with equation 12 from reference 27.

$$F_\mathrm{r} = -k_\mathrm{bias}(x_\mathrm{b} - vt^*) \tag{3.34}$$

where t^* is the time at rupture, and equation 13,

$$\langle x\rangle(t) = \frac{v_e k_\mathrm{B}T}{Dk_\mathrm{tot}}(\frac{Dk_\mathrm{tot}}{k_\mathrm{B}T}t + \mathrm{e}^{-Dk_\mathrm{tot}/k_\mathrm{B}Tt} - 1) \tag{3.35}$$

$$k_\mathrm{tot} = k_\mathrm{sys} + k_\mathrm{bias}, \quad v_e = vk_\mathrm{bias}/k_\mathrm{tot}. \tag{3.36}$$

For small times,

$$\langle x\rangle(t) \approx \frac{Dvk_\mathrm{bias}t^2}{2k_\mathrm{B}T}. \tag{3.37}$$

Solving for t^* in $\langle x\rangle(t^*) = x_\mathrm{b}$ and plugging in t^* for t in the exponent in eq. 3.35 we find

$$\langle x\rangle(t) \approx \frac{v_e k_\mathrm{B}T}{Dk_\mathrm{tot}}(\frac{Dk_\mathrm{tot}}{k_\mathrm{B}T}t + \mathrm{e}^{-Dk_\mathrm{tot}/k_\mathrm{B}T\sqrt{2x_\mathrm{b}k_\mathrm{B}T/Dvk_\mathrm{bias}}} - 1). \tag{3.38}$$

Set $\langle x\rangle(t) = x_\mathrm{b}$ in equation 3.38 and solve for t. Plug that value for t into 3.34

$$t^* = \frac{k_\mathrm{B}T\Big(\dfrac{Dk_\mathrm{tot}x_\mathrm{b}}{v_e k_\mathrm{B}T} + 1 - \mathrm{e}^{-Dk_\mathrm{tot}/k_\mathrm{B}T\sqrt{2x_\mathrm{b}k_\mathrm{B}T/Dvk_\mathrm{bias}}}\Big)}{Dk_\mathrm{tot}} \tag{3.39}$$

$$\langle F_\mathrm{r}\rangle = -k_\mathrm{bias}\Big(x_\mathrm{b} - \frac{vk_\mathrm{B}T}{Dk_\mathrm{tot}}\Big(\frac{Dk_\mathrm{tot}x_\mathrm{b}}{v_e k_\mathrm{B}T} + 1 - \mathrm{e}^{-Dk_\mathrm{tot}/k_\mathrm{B}T\sqrt{2x_\mathrm{b}k_\mathrm{B}T/Dvk_\mathrm{bias}}}\Big)\Big). \tag{3.40}$$

Plug-in definitions from equation 3.36 and simplify.

$$\langle F_\mathrm{r}\rangle = k_\mathrm{sys}x_\mathrm{b} + \frac{vk_\mathrm{bias}k_\mathrm{B}T}{Dk_\mathrm{tot}}\big(1 - \mathrm{e}^{-Dk_\mathrm{tot}/k_\mathrm{B}T\sqrt{2x_\mathrm{b}k_\mathrm{B}T/Dvk_\mathrm{bias}}}\big), \tag{3.41}$$

which is equation 14 from reference 27. We can perform similar calculations in the $\dot{F}$ ensemble and find the equation of motion for the average position of the particle

$$\langle x\rangle(t) = \frac{\dot{F}k_{\mathrm{B}}T}{Dk_{\mathrm{sys}}^2}\Big(\frac{Dk_{\mathrm{sys}}}{k_{\mathrm{B}}T}t + \mathrm{e}^{-Dtk_{\mathrm{sys}}/k_{\mathrm{B}}T} - 1\Big) \tag{3.42}$$

and for the average rupture force in the small time (large $\dot{F}$) limit

$$\langle F_{\mathrm{r}}\rangle = k_{\mathrm{sys}}x_{\mathrm{b}} + \frac{\dot{F}k_{\mathrm{B}}T}{Dk_{\mathrm{sys}}}\big(1 - \mathrm{e}^{-Dk_{\mathrm{sys}}/k_{\mathrm{B}}T\sqrt{2x_{\mathrm{b}}k_{\mathrm{B}}T/D\dot{F}}}\big). \tag{3.43}$$

Small and Intermediate Pulling Velocities

For small and intermediate velocities, where the barrier is still large and Kramers' theory still holds, they treat the barrier as a moving, shrinking barrier. This allows them to treat it as a first order reaction with a survival probability,

$$S(t) = \exp\Big(-\int_0^t k_0(x_b - \langle x\rangle(t); k_{\mathrm{tot}})\mathrm{d}t'\Big), \tag{3.44}$$

where $k_0(x_b - \langle x\rangle(t); k_{\mathrm{tot}})$ is the same as equation 3.32. One can then use this expression to determine the average rupture time and the average rupture force by plugging it into equation 3.34.

$$\langle F_{\mathrm{r}}\rangle = -k_{\mathrm{bias}}\Big(x_{\mathrm{b}} - v\int_0^\tau S(t)\mathrm{d}t\Big), \tag{3.45}$$

where τ is the time the average trajectory reaches the barrier. For very small velocities $v \ll k_0x_{\mathrm{b}}$, this leads to

$$\langle F_{\mathrm{r}}\rangle \approx \frac{k_{\mathrm{bias}}v}{k_0(k_{\mathrm{tot}}/k_{\mathrm{sys}})^{3/2}} - k_{\mathrm{bias}}x_{\mathrm{b}}, \tag{3.46}$$

which is equation 20 in reference 27. For intermediate velocities, $k_0x_{\mathrm{b}} < v < \frac{k_0}{x_{\mathrm{b}}k_{\mathrm{bias}}/k_{\mathrm{B}}T}\exp(\Delta G^{\ddagger}/k_{\mathrm{B}}T)$ they find

$$\langle F_{\mathrm{r}}\rangle = k_{\mathrm{bias}}x_{\mathrm{b}} - \Big(2k_{\mathrm{tot}}k_{\mathrm{B}}T\log\frac{k_0k_{\mathrm{B}}T\mathrm{e}^{\gamma + k_{\mathrm{sys}}x_{\mathrm{b}}^2/2k_{\mathrm{B}}T}}{k_{\mathrm{bias}}vx_{\mathrm{b}}(k_{\mathrm{sys}}/k_{\mathrm{tot}})^{3/2}}\Big)^{1/2} \tag{3.47}$$

where γ is the Euler-Mascheroni constant (≈ 0.577).

3.2.4 Dudko-Hummer-Szabo (2006)

Dudko-Hummer-Szabo is an extension and combination of the work done by Hummer-Szabo (2003)[27] and Dudko *et al.* (2003).[16] In Hummer-Szabo, the intermediate speed relationship was $\langle F_{\mathrm{r}} \rangle \propto (\log \dot{F})^{1/2}$, while in Dudko *et al.* the intermediate speed relationship was $\langle F_{\mathrm{r}} \rangle \propto (\log \dot{F})^{2/3}$ In order to account for this discrepancy, they found they could combine both expressions into a unified form. The difference between the two models, came from a difference in the underlying potential. For the Hummer-Szabo model, it is a quadratic, cusp potential, and the Dudko *et al.* model is a linear-cubic potential. One other difference between the HS (2003) model and DHS (2006) model is that in the DHS (2006) model there is an assumption that $k_{\mathrm{bias}} << k_{\mathrm{sys}}$ This leads to an equation that is similar to the weak spring outcome from HS. (Equation 3.32) This lead to the revised expression for the force-dependant rate constant

$$k(F) = k_0 \Big(1 - \frac{\nu F x_{\mathrm{b}}}{\Delta G^{\ddagger}}\Big)^{1/\nu - 1} \mathrm{e}^{\frac{\Delta G^{\ddagger}}{k_{\mathrm{B}}T}(1-(1-\nu F x_{\mathrm{b}}/\Delta G^{\ddagger})^{1/\nu})}, \tag{3.48}$$

where $\nu = 2/3$ for the linear cubic potential and $\nu = 1/2$ for the quadratic potential (Equation 3 from[17]). If $\nu = 1$ the expression becomes identical to the Bell expression.

For this new expression of the off-rate, the new expression for $\langle F_{\mathrm{r}} \rangle$ is

$$\langle F_{\mathrm{r}} \rangle \approx \frac{\Delta G^{\ddagger}}{\nu x_{\mathrm{b}}} \Big(1 - \Big(\frac{k_{\mathrm{B}}T}{\Delta G^{\ddagger}} \log \frac{k_0 k_{\mathrm{B}} T \mathrm{e}^{\Delta G^{\ddagger}/k_{\mathrm{B}}T + \gamma}}{x_{\mathrm{b}} k_{\mathrm{bias}} v}\Big)^{\nu}\Big) \tag{3.49}$$

where γ is again the Euler-Mascheroni constant. Equation 3.49 reduces to the HS (2003) equation for intermediate velocities (equation 3.47) when $k_{\mathrm{sys}} >> k_{\mathrm{bias}}$ and if $\nu = 1/2$ in equation 3.49.

3.2.5 Friddle-Noy-De Yoreo (2012)

Friddle-Noy-De Yoreo approach dynamic force spectroscopy through a master equation of a two-state Markov process.[22] This approach to DFS modeling is novel because it includes an explicit rebinding term. Their master equation is written as

$$\frac{\mathrm{d}P_b}{\mathrm{d}t} = -k_u(t) P_b + k_b(t) P_u, \tag{3.50}$$

where P_b is the probability the system is in the bound state and P_u is the probability the system is in the unbound state. k_u and k_b are given by

$$k_u(F) = k_0 \exp((F x_\mathrm{b} - \frac{1}{2} k_\mathrm{bias} x_\mathrm{b}^2)/k_\mathrm{B}T)) \tag{3.51}$$

$$\begin{aligned} k_b(F) &= k_b^0 \exp(-\frac{k_\mathrm{bias}}{2k_\mathrm{B}T}(\frac{F}{k_\mathrm{bias}} - x_\mathrm{b})^2) \\ &= k_u(F)\exp((\Delta G_{bu} - \frac{F^2}{2k_\mathrm{bias}})/k_\mathrm{B}T), \end{aligned} \tag{3.52}$$

where $\Delta G_{bu} = k_\mathrm{B}T \log(k_0/k_b^0)$ and k_b^0 is the natural frequency of the transducer.[22] The expression for $k_u(F)$ comes from the difference between a unbiased potential and a biased potential (derivation shown below). This expression is similar to the Bell off-rate, in that it is an exponential relationship with respect to the force. However there is an extra $\exp(-\frac{1}{2k_\mathrm{B}T} k_\mathrm{bias} x_\mathrm{b}^2)$ in comparison to the Bell off-rate (equation 3.16). The rebinding rate $k_b(F)$ is related to the energy of a spring extended between the barrier and the location of the biasing potential's minimum. There is a unique force F_eq at which the two rates cross $k_u(F_\mathrm{eq}) = k_b(F_\mathrm{eq})$. F_eq is related to ΔG_{bu} by

$$F_\mathrm{eq} = \sqrt{2 k_\mathrm{bias} \Delta G_{bu}}. \tag{3.53}$$

In previous models including HS and DHS, the nature of their underlying cusp potential means no rebinding is possible. One consequence of adding a rebinding term is that there is a minimum, positive force required to break the bond even at the slowest loading rates. At forces below this minimum force, the rebinding rate dominates the kinetics and In contrast the HS and DHS models allow for so-called negative rupture forces to occur at small loading rates.

The equation derived for the mean rupture force from the master equation is

$$\langle F_\mathrm{r} \rangle = F_\mathrm{eq} + \frac{k_\mathrm{B}T}{x_\mathrm{b}} \mathrm{e}^{1/R(F_\mathrm{eq})} E_1\Big(\frac{1}{R(F_\mathrm{eq})}\Big), \quad R(F_\mathrm{eq}) = \frac{\dot{F} x_\mathrm{b}}{k_u(F_\mathrm{eq}) k_\mathrm{B}T} \tag{3.54}$$

where $E_1(z) = \int_z^\infty \frac{\exp -s}{s} \mathrm{d}s$ is the expontential integral. An additional approximation is $\mathrm{e}^z E_1(z) \approx \log(1 + \mathrm{e}^{-\gamma}/z)$, where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. This allows us to make two approximations at slow and fast loading rates. At slow loading rates, Equation 3.54 reduces to

$$\langle F_{\mathrm{r}} \rangle = F_{\mathrm{eq}} + \frac{\dot{F}}{k_u(F_{\mathrm{eq}})} \tag{3.55}$$

and at large loading rates it reduces to

$$\langle F_{\mathrm{r}} \rangle = \frac{k_{\mathrm{B}}T}{x_{\mathrm{b}}} \log\left(\mathrm{e}^{-\gamma}\frac{\dot{F}x_{\mathrm{b}}}{k_u(F_{\mathrm{eq}})k_{\mathrm{B}}T}\right) \tag{3.56}$$

Derivation of $k_u(F)$

Let us start with a quadratic potential of the form

$$U(x) = \frac{1}{2}k_{\mathrm{sys}}x^2 \tag{3.57}$$

where $U(x)$ is the potential energy of the system and there is a transition state at x_{b}. Our intrinsic off-rate can be described using an Arrhenius equation.

$$k_0 = A\mathrm{e}^{-E_a/k_{\mathrm{B}}T} \tag{3.58}$$

where $E_a = 1/2k_{\mathrm{sys}}x_{\mathrm{b}}^2$ is the difference between the minimum and the transition state, and A is a pre-exponential factor is constant but depends on the system. We can then apply a force using a harmonic trap and now our overall potential is

$$U'(x) = \frac{1}{2}k_{\mathrm{sys}}x^2 + \frac{1}{2}k_{\mathrm{bias}}(x - x_{\mathrm{bias}})^2 \tag{3.59}$$

where x_{bias} is the minimum of our harmonic trap. For this biased potential, we can also use the Arrhenius equation, however our E_a will be different.

$$\begin{aligned} E'_a &= U'(x_{\mathrm{b}}) - U'(0) \\ &= \frac{1}{2}k_{\mathrm{sys}}x_{\mathrm{b}}^2 + \frac{1}{2}k_{\mathrm{bias}}(x_{\mathrm{b}} - x_{\mathrm{bias}})^2 - \frac{1}{2}k_{\mathrm{sys}}0^2 - \frac{1}{2}k_{\mathrm{bias}}(0 - x_{\mathrm{bias}})^2 \\ &= \frac{1}{2}k_{\mathrm{sys}}x_{\mathrm{b}}^2 + \frac{1}{2}k_{\mathrm{bias}}(x_{\mathrm{b}}^2 - 2x_{\mathrm{b}}x_{\mathrm{bias}}) \end{aligned} \tag{3.60}$$

We calculate our force induced off-rate with respect to k_0,

$$\begin{aligned} k(F)/k_0 &= \mathrm{e}^{-(E'_a - E_a)/k_{\mathrm{B}}T} \\ &= \mathrm{e}^{-(\frac{1}{2}k_{\mathrm{bias}}(x_{\mathrm{b}}^2 - 2x_{\mathrm{b}}x_{\mathrm{bias}}))/k_{\mathrm{B}}T} \\ &= \mathrm{e}^{(k_{\mathrm{bias}}x_{\mathrm{bias}}x_{\mathrm{b}} - \frac{1}{2}k_{\mathrm{bias}}x_{\mathrm{b}}^2)/k_{\mathrm{B}}T} \end{aligned} \tag{3.61}$$

which is Equation 3.51 when we replace $k_{\text{bias}}x_{\text{bias}}$ with F which is only strictly true if the particle starts at $x = 0$.

Derivation of equation 3.54

The expectation value of an observable is the integral of that observable multiplied by its probability distribution

$$\langle F_{\text{r}} \rangle = \int_{F_{\text{eq}}}^{\infty} F P_r(F) \mathrm{d}F. \tag{3.62}$$

Defining probability of rupture at a certain force $P_r(F)$ as

$$P_r(F) = -\frac{\mathrm{d}P_b(F)}{\mathrm{d}F}, \tag{3.63}$$

where $P_b(F)$ is the probability the system is still bound. Let us assume that before the force F_{eq}, the rebinding rate dominates the kinetics and after F_{eq} the unbinding rate is the dominant rate. This allows us to simplify our equation by looking at only forces greater than F_{eq} and neglecting the effect of the rebinding rate. The master equation (equation 3.50)can then be rewritten as

$$\frac{\mathrm{d}P_b}{\mathrm{d}t} = -k_u(t)P_b. \tag{3.64}$$

Solving for an expression of P_b as a function of force, we set up the following integral which is Friddle eq. 5,

$$\int_1^{P_b} \frac{\mathrm{d}P_b'}{P_b'} \approx \frac{1}{\dot{F}} \int_{F_{\text{eq}}}^{F} k_u(F')\mathrm{d}F' \tag{3.65}$$

and we have used the relationship

$$\frac{\mathrm{d}P_b}{\mathrm{d}t} = \frac{\mathrm{d}F}{\mathrm{d}t}\frac{\mathrm{d}P_b}{\mathrm{d}F} = \dot{F}\frac{\mathrm{d}P_b}{\mathrm{d}F} \tag{3.66}$$

to solve in terms of the force.

Taking the integral on both sides, we find

$$P_b(F) = \mathrm{e}^{-\frac{F_\beta}{\dot{F}}(k_u(F)-k_u(F_{\text{eq}}))} \tag{3.67}$$

$$\langle F_{\text{r}} \rangle = -\int_{F_{\text{eq}}}^{\infty} F \frac{\mathrm{d}P_b(F)}{\mathrm{d}F} \mathrm{d}F, \tag{3.68}$$

where F_β is $k_\mathrm{B}T/x_\mathrm{b}$. Using integration by parts, we split up this integral

$$\langle F_\mathrm{r}\rangle = -FP_b(F)\Big|_{F_\mathrm{eq}}^{\infty} + \int_{F_\mathrm{eq}}^{\infty} P_b(F)\mathrm{d}F. \tag{3.69}$$

$P_b(F)$ goes to 0 as $f \to \infty$ and $P_b(F_\mathrm{eq}) = 1$, therefore

$$\langle F_\mathrm{r}\rangle = F_\mathrm{eq} + e^{\frac{F_\beta k_u(F_\mathrm{eq})}{\dot{F}}} \int_{F_\mathrm{eq}}^{\infty} \mathrm{e}^{\frac{F_\beta k_u(F)}{\dot{F}}}\mathrm{d}F. \tag{3.70}$$

Using the exponential integral $\mathrm{Ei}(x) = -\int_{-x}^{\infty} \frac{\mathrm{e}^{-t}\mathrm{d}t}{t}$ we find the antiderivative of the integral is $F_\beta\mathrm{Ei}(-\frac{F_\beta k_u(F)}{\dot{F}})$ and the expression becomes,

$$\langle F_\mathrm{r}\rangle = F_\mathrm{eq} + e^{\frac{F_\beta k_u(F_\mathrm{eq})}{\dot{F}}} F_\beta \mathrm{Ei}\Big(-\frac{F_\beta k_u(F)}{\dot{F}}\Big)\Big|_{F_\mathrm{eq}}^{\infty}. \tag{3.71}$$

$k_u(\infty) \to \infty$ so $\mathrm{Ei}(-F_\beta k_u(\infty)/\dot{F}) \to \mathrm{Ei}(\infty)$ and $\mathrm{Ei}(\infty) \to 0$. Evaluating $\mathrm{Ei}\Big(-\frac{F_\beta k_u(F)}{\dot{F}}\Big)\Big|_{F_\mathrm{eq}}^{\infty}$ then leads to

$$\langle F_\mathrm{r}\rangle = F_\mathrm{eq} - F_\beta \mathrm{e}^{\frac{F_\beta k_u(F_\mathrm{eq})}{\dot{F}}} \mathrm{Ei}\Big(-\frac{F_\beta k_u(F_\mathrm{eq})}{\dot{F}}\Big). \tag{3.72}$$

$E_1(x) = \int_x^{\infty} \frac{\mathrm{e}^{-t}\mathrm{d}t}{t}$ so by definition $E_1(x) = -\mathrm{Ei}(-x)$ and

$$\langle F_\mathrm{r}\rangle = F_\mathrm{eq} + F_\beta \mathrm{e}^{\frac{F_\beta k_u(F_\mathrm{eq})}{\dot{F}}} E_1\Big(\frac{F_\beta k_u(F_\mathrm{eq})}{\dot{F}}\Big). \tag{3.73}$$

Make $\frac{1}{R(F)} = \frac{F_\beta k_u(F_\mathrm{eq})}{\dot{F}}$ substitution to match the notation used in reference 22

$$\langle F_\mathrm{r}\rangle = F_\mathrm{eq} + F_\beta e^{\frac{1}{R(F)}} E_1\Big(\frac{1}{R(F)}\Big), \tag{3.74}$$

which is equation 6 from Friddle-Noy-De Yoreo.[22]

3.2.6 Bullerjahn-Sturm-Kroy (2014)

Bullerjahn-Sturm-Kroy approach the problem of DFS through the Fokker-Planck equation of a particle in a harmonic potential being pulled by an external force

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}[A(x,t)P(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[B(x,t)P(x,t)] \tag{3.75}$$

where

$$A(x,t) = \frac{D}{k_{\mathrm{B}}T}F(x,t) \tag{3.76}$$

$$B(x,t) = 2D \tag{3.77}$$

They use this framework to solve for the flux accross the barrier $j(x_{\mathrm{b}}, t)$ and the rate $k(t) \approx j(x_{\mathrm{b}}, t)/S(t)$, where $S(t)$ is the the survival probability up to time t. Then they find the rupture force distribution

$$P_{\mathrm{r}}(F) \approx \frac{1}{\dot{F}(F)}\frac{\mathrm{d}}{\mathrm{d}t}\exp\Big(-\int_0^t \frac{j(x_{\mathrm{b}}, t')}{S(t')}\mathrm{d}t'\Big) \tag{3.78}$$

which has a similar form to equation 17 in reference 27. For $k_{\mathrm{bias}} \ll k_{\mathrm{sys}}$, they find an analytical solution to the rupture force distribution

$$P_{\mathrm{r}}(F) \approx \frac{k_0}{\dot{F}}\Big(1+\frac{F}{F_c}-\frac{2\langle x\rangle(F)}{x_{\mathrm{b}}}\Big)\mathrm{e}^{\Delta G^{\ddagger}(1-(1-\langle x\rangle(F)/x_{\mathrm{b}})^2)/k_{\mathrm{B}}T}\exp\Big(-\frac{k_{\mathrm{B}}Tk_0}{\dot{F}x_{\mathrm{b}}}(\mathrm{e}^{\Delta G^{\ddagger}(1-(1-\langle x\rangle(F)/x_{\mathrm{b}})^2)/k_{\mathrm{B}}T}-1)\Big) \tag{3.79}$$

where $\langle x\rangle(F)$ is the maximum of a distribution,

$$\langle x\rangle(F) = x_{\mathrm{b}}\Big(\frac{F}{F_c} - \frac{k_{\mathrm{B}}T\dot{F}x_{\mathrm{b}}}{DF_c^2}(1-\mathrm{e}^{-2D\Delta G^{\ddagger}F/k_{\mathrm{B}}T\dot{F}x_{\mathrm{b}}^2})\Big) \tag{3.80}$$

and k_0 is the Kramers off-rate (equation 3.5)

They suggest using Bayesian methods to fit to the enetire set of distributions, but also provide expressions for $\langle F_{\mathrm{r}}\rangle$. For small loading rates, they obtain a previously observed result for the Bell-Evans model[24]

$$\langle F_{\mathrm{r}}\rangle \approx \frac{k_{\mathrm{B}}T\mathrm{e}^{qX}E_1(qX)}{\chi x_{\mathrm{b}}} \tag{3.81}$$

where

$$\chi = 1 + \frac{k_{\mathrm{bias}}x_{\mathrm{b}}^2}{2\Delta G^{\ddagger}},\ q = \mathrm{e}^{\Delta G^{\ddagger}(1-\chi)/k_{\mathrm{B}}T},\ X = \chi^{1/2}k_0\frac{k_{\mathrm{B}}T}{\dot{F}x_{\mathrm{b}}} \tag{3.82}$$

For fast loading rates, they obtain the result

$$\langle F_{\mathrm{r}}\rangle \approx \sqrt{\frac{2\dot{F}x_{\mathrm{b}}k_{\mathrm{B}}T}{D}}. \tag{3.83}$$

To interpolate between the fast and the slow regimes, they simply add the two equations together.

$$\langle F_{\mathrm{r}} \rangle \approx \frac{k_{\mathrm{B}} T \mathrm{e}^{qX} E_1(qX)}{\chi x_{\mathrm{b}}} + \sqrt{\frac{2\dot{F} x_{\mathrm{b}} k_{\mathrm{B}} T}{D}}. \tag{3.84}$$

3.2.7 Relationship Between Rupture Force and Rupture Work

Another observable we can measure is the work done by the external force. We show how we can relate the rupture work and the rupture force in this section.

Pulling without an Underlying Potential

Assume that we have no underlying potential, and only a moving harmonic trap. Our potential and force are

$$U(x,t) = \frac{1}{2} k_{\mathrm{bias}} (x - vt)^2 \tag{3.85}$$

$$F(x,t) = -k_{\mathrm{bias}} (x - vt) \tag{3.86}$$

The Fokker-Planck equation (FPE) for this system is

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}[A(x,t)P(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[B(x,t)P(x,t)] \tag{3.87}$$

where

$$A(x,t) = \frac{D}{k_{\mathrm{B}} T} F(x,t) \tag{3.88}$$

$$B(x,t) = 2D \tag{3.89}$$

where D is the diffusion constant, and $P(x,t)$ is the probability distribution function of the system. Assume that the trap moves to a new position $vt + v\Delta t$, and the new potential becomes

$$U(x, t+\Delta t) = \frac{1}{2} k_{\mathrm{bias}} (x - (vt + v\Delta t))^2. \tag{3.90}$$

The change in the potential is

$$\Delta U = k_{\mathrm{sys}} (vt - x)(v\Delta t) + \frac{k_{\mathrm{bias}}}{2} (v\Delta t)^2. \tag{3.91}$$

Keeping only the linear term of Δt, we have

$$\begin{aligned}\Delta U &= k_{\text{bias}}(vt - x)v\Delta t \\ &= F(x,t)v\Delta t.\end{aligned} \tag{3.92}$$

If $v = 0$, then $F(x,t) = -k_{\text{sys}}x$, and the FPE becomes

$$\frac{\partial P(x,t)}{\partial t} = -\frac{D}{k_{\text{B}}T}\Big(- k_{\text{sys}}P(x,t) - k_{\text{sys}}x\frac{\partial}{\partial x}P(x,t)\Big) + \frac{B(x,t)}{2}\frac{\partial^2}{\partial x^2}P(x,t). \tag{3.93}$$

This can be solved through a differential equation solver and has two stationary solutions where $\frac{\partial P(x,t)}{\partial t} = 0$:

$$P(x) = \mathrm{e}^{-\frac{k_{\text{sys}}x^2}{2k_{\text{B}}T}} \tag{3.94}$$

and

$$P(x) = \sqrt{\frac{2k_{\text{B}}T}{k_{\text{sys}}}}\text{DawsonF}\Big(\frac{\sqrt{k_{\text{sys}}x}}{\sqrt{2k_{\text{B}}T}}\Big) \tag{3.95}$$

where $\text{DawsonF}(x) = \exp(-x^2)\int_0^x \exp(y^2)\mathrm{d}y$. Because P(x) must be non-negative everywhere only the first solution is physical.

Following the solutions from reference 20, where their $\tilde{\gamma}(t) = k_{\text{bias}}D/k_{\text{B}}T$ and $\tilde{\beta}(t) = (k_{\text{bias}}vD/k_{\text{B}}T)t$, equation 18 of that paper becomes

$$H(t,t_1) = (k_{\text{bias}}/\gamma)(t - t_1) \tag{3.96}$$

For the intial condition

$$P(x, t = 0) = \delta(x - x_0) \tag{3.97}$$

which is the Dirac delta function centered at x_0, the distribution $P(x,t|x_0,0)$ at time t is Gaussian with mean (equation 23 of reference 20)

$$\langle x\rangle(t) = x_0\mathrm{e}^{-k_{\text{bias}}Dt/k_{\text{B}}T} + vt + \frac{vk_{\text{B}}T}{k_{\text{bias}}D}(\mathrm{e}^{-k_{\text{bias}}Dt/k_{\text{B}}T} - 1) \tag{3.98}$$

and variance (equation 24 of reference 20)

$$\sigma_x^2(t) = \frac{k_{\text{B}}T}{k_{\text{bias}}}(1 - \mathrm{e}^{-2k_{\text{bias}}Dt/k_{\text{B}}T}) \tag{3.99}$$

If the intial distribution at $t = 0$ is the equilibrium distribution (equation 3.94 with a normalization such that $\int_{-\infty}^{\infty} P(x)\mathrm{d}x = 1$)

$$P_{eq}(x) = \frac{1}{\sqrt{2\pi k_\mathrm{B}T/k_\mathrm{bias}}}\mathrm{e}^{-\frac{k_\mathrm{bias}x^2}{2k_\mathrm{B}T}} \tag{3.100}$$

then the distribution at time t,

$$P(x,t) = \int_{-\infty}^{\infty} P(x,t|x_0,0)P_{eq}(x_0)\mathrm{d}x_0 \tag{3.101}$$

is Gaussian with mean

$$\langle x\rangle(t) = vt - \frac{vk_\mathrm{B}T}{k_\mathrm{bias}D}(1-\mathrm{e}^{-k_\mathrm{bias}Dt/k_\mathrm{B}T}) \tag{3.102}$$

and time-independent variance

$$\sigma_x^2 = \frac{k_\mathrm{B}T}{k_\mathrm{bias}}. \tag{3.103}$$

At long times t,the system reaches a steady state and we can calculate the lag in position between the particle and the moving trap as

$$vt - \langle x\rangle(t) = \frac{vk_\mathrm{B}T}{k_\mathrm{bias}D}, \tag{3.104}$$

which means the steady state force the particle feels from the moving trap is

$$k_\mathrm{bias}(vt - \langle x\rangle(t)) = \frac{vk_\mathrm{B}T}{D} \tag{3.105}$$

which now only depends on the velocity of the moving trap and not on the strength.

Pulling Out of a Harmonic Well

Now we investigate the effect of including an underlying harmonic potential. Our total potential is

$$\begin{aligned} U_{tot}(x) &= \frac{1}{2}k_\mathrm{sys}x^2 + \frac{1}{2}k_\mathrm{bias}(x-vt)^2 \\ &= \frac{1}{2}k_\mathrm{tot}(x-x_\mathrm{tot})^2 + C \end{aligned} \tag{3.106}$$

where

$$\begin{aligned} k_{\text{tot}} &= k_{\text{sys}} + k_{\text{bias}} \\ x_{\text{tot}} &= \frac{k_{\text{bias}} vt}{k_{\text{sys}} + k_{\text{bias}}} = v_{\text{tot}} t \\ C &= \frac{k_{\text{sys}} k_{\text{bias}} (vt)^2}{2(k_{\text{sys}} + k_{\text{bias}})} \end{aligned} \tag{3.107}$$

with $v_{\text{tot}} = k_{\text{bias}} v/(k_{\text{sys}} + k_{\text{bias}})$.

Because the total potential is quadratic with curavture k_{tot} and a minimum moves to the right with velocity v_{tot} the equations from the previous section apply. For a particle starting at position x_0 the average position at time t is (from equation 3.98)

$$\langle x \rangle_{x_0}(t) = x_0 \mathrm{e}^{-k_{\text{tot}} Dt/k_{\mathrm{B}}T} + v_{\text{tot}} t + \frac{v_{\text{tot}} k_{\mathrm{B}} T}{k_{\text{tot}} D} (\mathrm{e}^{-k_{\text{tot}} Dt/k_{\mathrm{B}}T} - 1) \tag{3.108}$$

and variance (from equation 3.99)

$$\sigma_x^2 = \frac{k_{\mathrm{B}} T}{k_{\text{tot}}} (1 - \mathrm{e}^{-2k_{\text{tot}} Dt/k_{\mathrm{B}}T}) \tag{3.109}$$

but if the intial condition comes from the equilibrium distribution

$$P_{eq}(x) = \frac{1}{\sqrt{2\pi k_{\mathrm{B}} T/k_{\text{tot}}}} \mathrm{e}^{-\frac{k_{\text{tot}} x^2}{2k_{\mathrm{B}}T}} \tag{3.110}$$

the mean and variance follow from equations 3.102 and 3.103

$$\langle x \rangle(t) = v_{\text{tot}} t - \frac{v_{tot} k_{\mathrm{B}} T}{k_{\text{tot}} D} (1 - \mathrm{e}^{-k_{\text{tot}} Dt/k_{\mathrm{B}}T}) \tag{3.111}$$

$$\sigma_x^2 = \frac{k_{\mathrm{B}} T}{k_{\text{tot}}} \tag{3.112}$$

To compute the work, we first note that

$$\begin{aligned} x(t) &= x_0 + \int_0^t \dot{x}(t') \mathrm{d}t' \\ &= x_0 + \frac{D}{k_{\mathrm{B}} T} \int_0^t -k_{\text{sys}} x(t') - k_{\text{bias}} (x(t') - vt') + \xi(t') \mathrm{d}t' \\ &= x_0 - \frac{k_{\text{tot}} D}{k_{\mathrm{B}} T} \int_0^t x(t') \mathrm{d}t' + \frac{k_{\text{bias}} v D t^2}{2 k_{\mathrm{B}} T} + \frac{D}{k_{\mathrm{B}} T} \int_0^t \xi(t') \mathrm{d}t' \end{aligned} \tag{3.113}$$

so that

$$\frac{k_{\text{tot}}D}{k_{\text{B}}T}\int_0^t x(t')\mathrm{d}t' = x_0 - x(t) + \frac{k_{\text{bias}}vt^2D}{2k_{\text{B}}T} + \frac{D}{k_{\text{B}}T}\int_0^t \xi(t')\mathrm{d}t'. \tag{3.114}$$

The work performed by the trapping potential for a single trajectory is

$$\begin{aligned} W(t) &= v\int_0^t -k_{\text{bias}}(x(t') - vt')\mathrm{d}t' \\ &= -k_{\text{bias}}v\int_0^t x(t')\mathrm{d}t' + \frac{k_{\text{bias}}v^2t^2}{2} \\ &= \frac{k_{\text{bias}}k_{\text{B}}T}{k_{\text{tot}}D}v(x(t) - x_0) + \frac{k_{\text{bias}}k_{\text{sys}}v^2t^2}{2k_{\text{tot}}} - \frac{k_{\text{bias}}v}{k_{\text{tot}}}\int_0^t \xi(t')\mathrm{d}t' \end{aligned} \tag{3.115}$$

We can average this result over all trajectories that start at x_0 to give

$$\begin{aligned} \langle W\rangle_{x_0}(t) &= \frac{k_{\text{bias}}k_{\text{B}}T}{k_{\text{tot}}D}v(\langle x\rangle_{x_0} - x_0) + \frac{k_{\text{bias}}k_{\text{sys}}v^2t^2}{2k_{\text{tot}}} \\ &= \frac{k_{\text{bias}}k_{\text{B}}T}{k_{\text{tot}}D}v\Big(x_0 + \frac{v_{tot}k_{\text{B}}T}{k_{\text{tot}}D}\Big)(\mathrm{e}^{-k_{\text{tot}}Dt/k_{\text{B}}T} - 1) + \frac{k_{\text{bias}}^2k_{\text{B}}T}{k_{\text{tot}}^2D}v^2t + \frac{k_{\text{bias}}k_{\text{sys}}v^2t^2}{2k_{\text{tot}}} \end{aligned} \tag{3.116}$$

where the expression for $\langle x\rangle_{x_0}(t)$ came from equation 3.108.

The bias force is

$$F_{\text{b}}(x(t),t) = -k_{\text{bias}}(x(t) - vt) \tag{3.117}$$

which can be rewritten as

$$t = \frac{F_{\text{b}}(x(t),t)}{vk_{\text{bias}}} + \frac{x(t)}{v}. \tag{3.118}$$

Plugging this into equation 3.115, we obtain the work-force relationship

$$W(t) = \frac{k_{\text{sys}}(F_{\text{b}}(x(t),t) + k_{\text{bias}}x(t))^2 + 2k_{\text{bias}}^2vk_{\text{B}}T(x(t) - x_0)/D}{2k_{\text{bias}}k_{\text{tot}}} - \frac{k_{\text{bias}}v}{k_{\text{tot}}}\int_0^t \xi(t')\mathrm{d}t' \tag{3.119}$$

If we consider trajectories until they reach a distance x_{b}, which happens at time t_{r}, at this moment the trapping force will be $F_{\text{r}} = -k_{\text{bias}}(x_{\text{b}} - vt_{\text{r}})$. The work performed up to that point, W_{r} is

$$W_{\text{r}} = \frac{k_{\text{sys}}(F_{\text{r}} + k_{\text{bias}}x_{\text{b}})^2 + 2k_{\text{bias}}^2vk_{\text{B}}T(x_{\text{b}} - x_0)/k_{\text{B}}T}{2k_{\text{bias}}k_{\text{tot}}} - \frac{k_{\text{bias}}v}{k_{\text{tot}}}\int_0^t \xi(t')\mathrm{d}t' \tag{3.120}$$

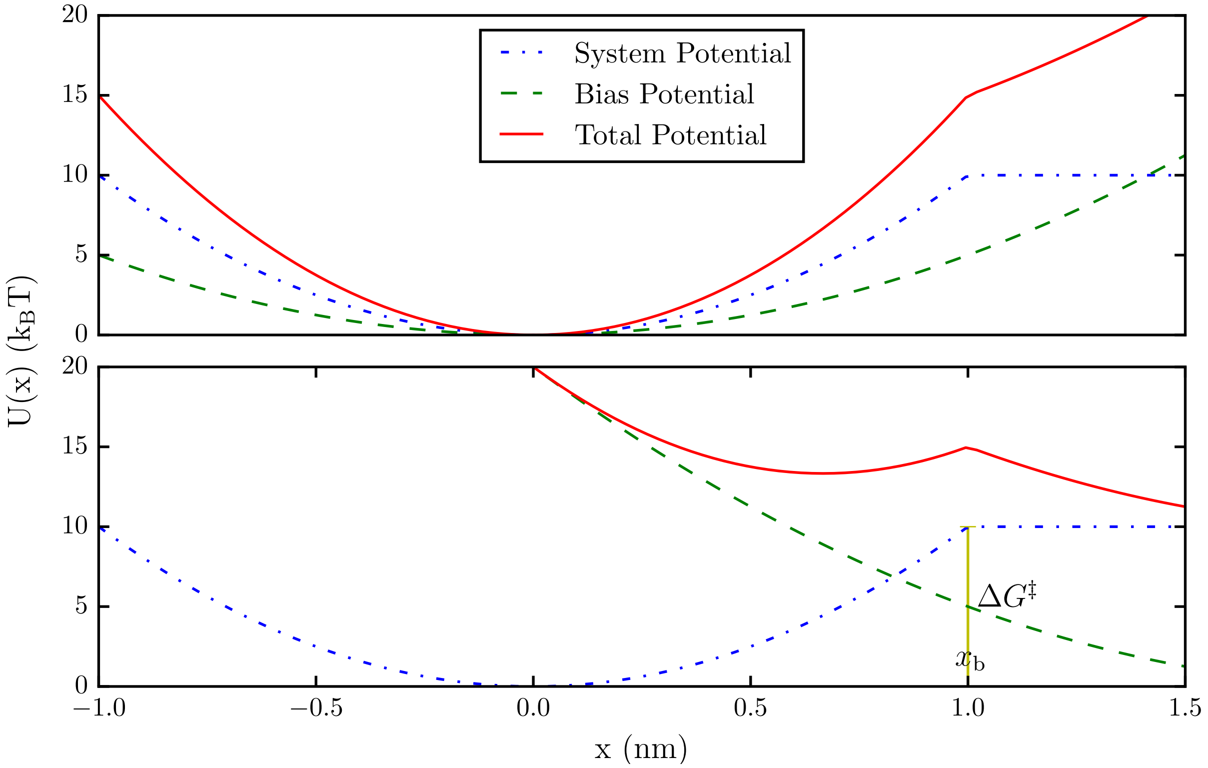


Figure 3.1: Schematic of the quadratic system with a moving harmonic trap. The top figure shows the system at $t = 0$ and the bottom picture is when the moving trap minimum has moved to 2 nm. We see the minimum of the total potential the particle feels has shifted to the right and the difference in energy between the potential minimum and the transition state at x_b has decreased.

3.3 Brownian Dynamics Simulations of Pulling

In order to test these models for their ability to accurately predict off-rates, we performed 1D Langevin dynamics simulations in which we can precisely control the underlying free energy surface and the dynamics of the pulling process. We changed the underlying potential by changing its functional form, as well as testing different barrier heights to determine the differences between them. For the pulling protocol, we used a linearly increasing force, and harmonic traps with differing strengths.

3.3.1 Simulation Methods

We wrote a Brownian dynamics code to perform these 1D pulling simulations. These particles followed the equation of motion

$$x(t+\Delta t) = D(F_{\mathrm{sys}}(t) - F_{\mathrm{b}}(t))\Delta t + \xi(t)\sqrt{2D\Delta t} \tag{3.121}$$

where $F(t)$ for the linearly increasing force was described as

$$F_{\mathrm{b}}(t) = \dot{F}t \tag{3.122}$$

where $\dot{F} = \frac{\Delta F}{\Delta t}$ is the loading rate. For the moving harmonic trap $F(t)$ was described as

$$F_{\mathrm{b}}(t) = -k_{\mathrm{bias}}(x(t) - vt). \tag{3.123}$$

where k_{bias} is the strength of the moving trap, and v is the velocity the trap moves to the right. The force from the system F_{sys} is described in a piecewise way. For the quadratic underlying potential, the force the particle feels from the system is

$$F_{\mathrm{sys}}(t) = \begin{cases} -k_{\mathrm{sys}}x(t) & \text{if } x(t) < x_{\mathrm{b}} \\ 0 & \text{otherwise} \end{cases}. \tag{3.124}$$

$\xi(t)$ is a Gaussian random fluctuation with mean $\langle\xi\rangle(t) = 0$ and variance $\sigma_\xi^2 = 1$. The potential surface the particle is moving on is illustrated in Figure 3.1.

At the beginning of the simulation, we allowed the particle to equilibrate in the underlying potential and later applied the biasing force through either the linearly increasing method or moving trap method. The simulation ran until the particle reached $1.5x_{\mathrm{b}}$, while recording each time the particle crossed x_{b}. For our system, $k_{\mathrm{sys}} = \frac{20k_{\mathrm{B}}T}{\mathrm{nm}^2}$, $x_{\mathrm{b}} = 1$ nm, and $D = 1000\frac{\mathrm{nm}^2}{\mathrm{s}}$. We tested four different k_{bias}: $\frac{0.01k_{\mathrm{B}}T}{\mathrm{nm}^2}$, $\frac{0.1k_{\mathrm{B}}T}{\mathrm{nm}^2}$, $\frac{1k_{\mathrm{B}}T}{\mathrm{nm}^2}$, $\frac{10k_{\mathrm{B}}T}{\mathrm{nm}^2}$ which can be converted into units of (N/m) at room temperature by multiplying by 4.14x10^{-3}. In the moving trap ensemble, the loading rate is calculated as $k_{\mathrm{bias}}v$.

3.3.2 Results and Discussion

When we use a lineraly increasing force, we see from Figure 3.2 there are three regimes of pulling as determined by the relationship between the rupture force and the loading rate. For slow loading rates, thermal fluctuations dominate and the off-rates are essentially the same as the unbiased potential. This means the rupture time is constant, but the average rupture force is increasing linearly because of the constant loading rate. At high loading rates, we see there is a $\sqrt{\dot{F}}$ relationship to $\langle F_{\mathrm{r}} \rangle$. At these high loading rates, the particle does not have enough time for thermal equilibrium in the potential, and it is actively being driven along to the rupture point. In the intermediate region, we see a logarithmic relationship similar to the expected result from Bell-Evans (equation 3.17).

When we compare the rupture forces of linearly increasing force to the rupture forces of the moving harmonic trap, we see that for very weak trap stiffnesses, the harmonic trap acts exactly like the linearly increasing trap. (Figure 3.3) As we increase the pulling trap stiffness, we get increasing deviation at lower pulling speeds. However, at larger pulling speeds, all of the rupture forces using different pulling strengths collapse onto the same curve, indicating that at large pulling speeds, the form of the bias force does not matter. For the harmonic trap at slow pulling speeds, we also observe the unphysical negative rupture force. For real systems and experiments, this might be interpreted as non-adhesion of the system.

To determine the ability of each model to fit to our data, we first plotted the analytical result of each model using known parameters of our underlying system. (Figures 3.2 and 3.3) The FNDY, BSK, and HS models need a k_{bias} as an input parameter, but our linearly increasing force has no k_{bias}. Another way to imagine the linearly increasing force is a harmonic trap with $k_{\text{bias}} \to 0$ and $v \to \infty$ that satisfy $k_{\text{bias}} v = \dot{F}$. So for FNDY, BSK, and HS we set $k_{\text{bias}} = 0$ for plotting the analytical results. For our linearly increasing biasing force, the BSK model had good agreement at slow and fast loading rates but had some deviation from the data at intermediate rates. Each piece of the HS model also did well in matching the observed data in the slow, intermediate and fast regimes. The DHS model did

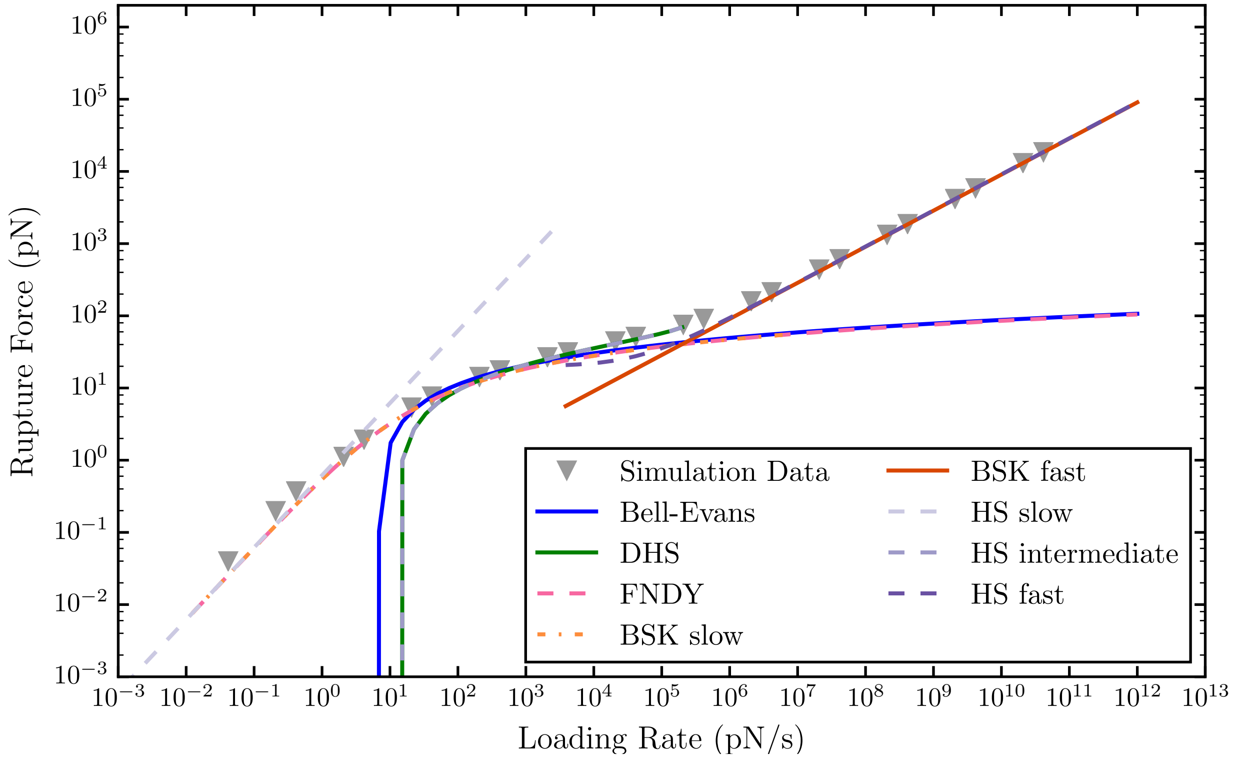


Figure 3.2: Mean rupture force of the linearly increasing biasing force and the expected fits of the models using known values of $\Delta G^{\ddagger}$, x_{b}, k_{bias}, and D. At the high loading rate and low loading rate limits, different models can predict the same behavior. At low loading rates, the BSK slow equation and the HS slow equation are identical. At high loading rates, the BSK fast equation and HS fast equation are identical. Also at high loading rates, the BSK slow equation, the FNDY equation, and the Bell-Evans equation all converge. For the linearly increasing biasing force, the DHS equation and HS intermediate equation are identical as expected.

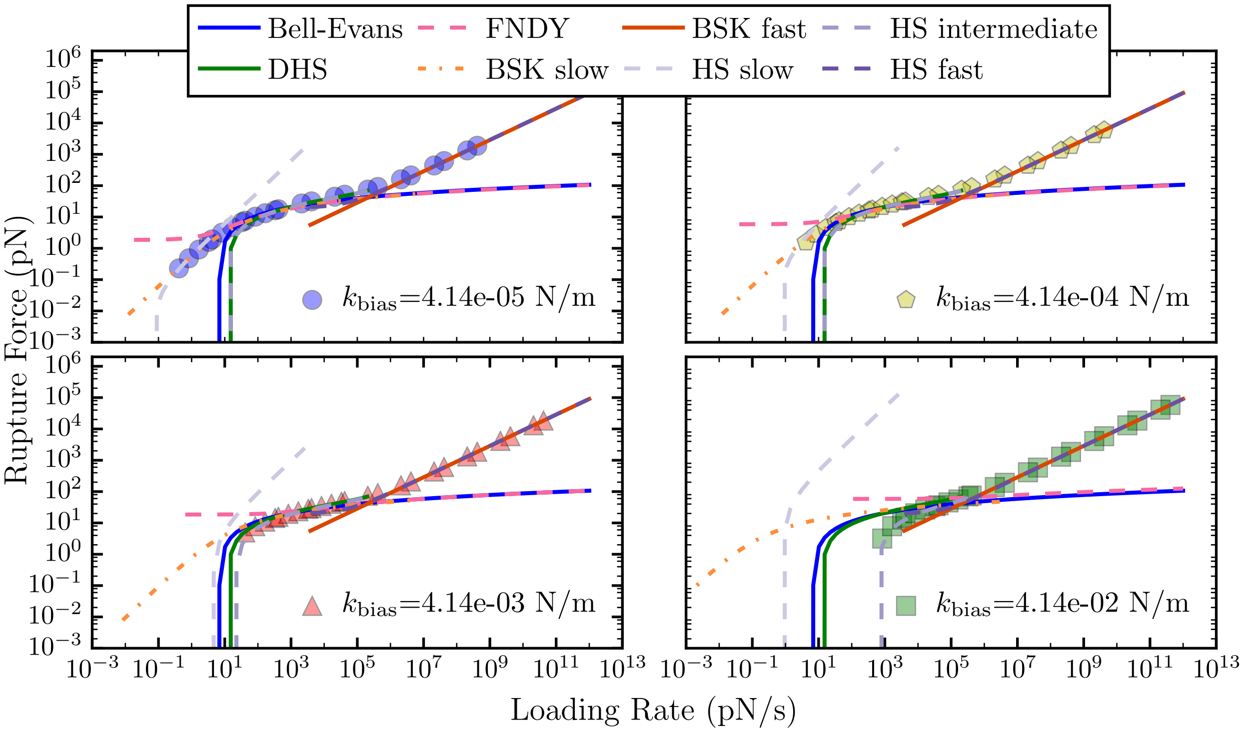


Figure 3.3: Expected fits of models using known values of $\Delta G^{\ddagger}$, x_{b}, k_{bias}, and D. Each subplot shows the effect of using different spring stiffnesses. The top left plot is the least stiff and most like a linearly increasing force, and the bottom right plot is the most stiff with the most deviation from the linearly increasing force. As seen here, each model predicts different behavior for the rupture force as a function of the loading rate.

well for only a small portion of the data in the intermediate region from ∼10pN/s to 10^5 pN/s. The Bell-Evans and FNDY models both performed similarly to the DHS model. Only a small portion of the intermediate regime was matched well, while the slow and fast regimes both deviate from the $\ln(\dot{F})$ relationship that these two models predict.

The models' performance for the moving harmonic trap is similar to their performance for the linearly increasing bias as seen in Figure 3.3. The HS models seem to do the best still in their respective regimes of slow, intermediate, and fast. However as the k_{bias} gets larger, the HS-slow model (equation 3.46) seems to be less accurate. The intermediate and fast HS models do well for all values of k_{bias} we used. The BSK-fast model continued to do well for the fast loading rates. The BSK-slow model matched the data for the two smaller

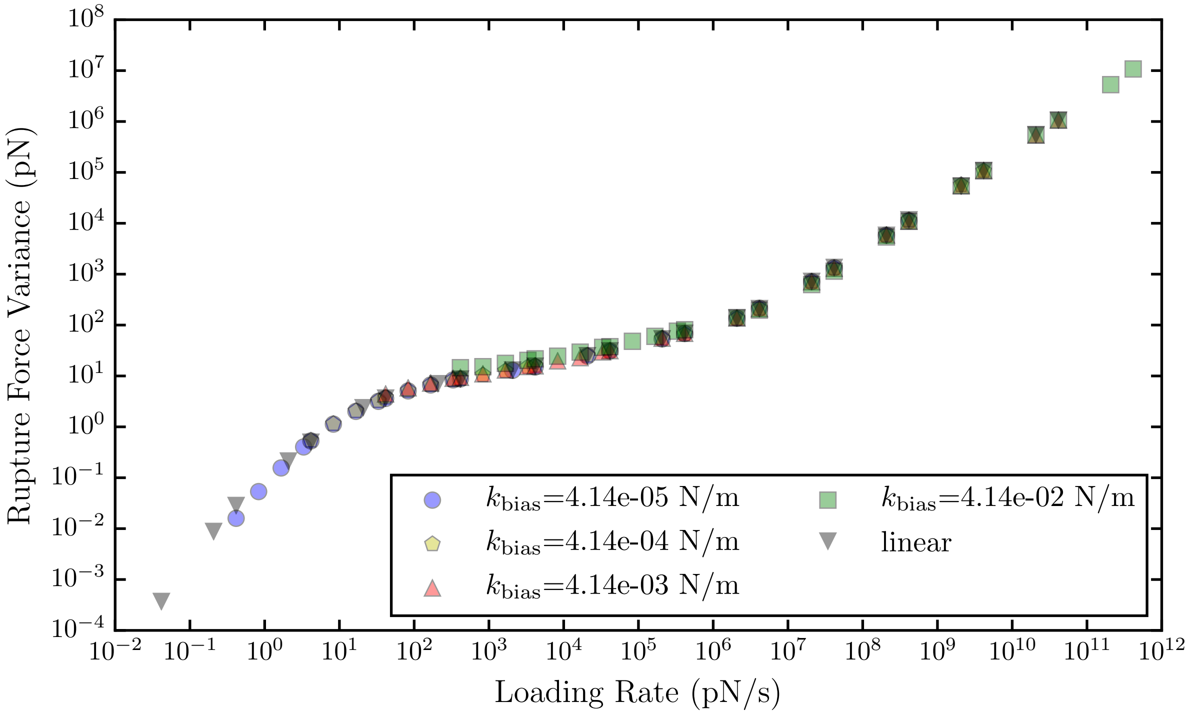


Figure 3.4: The variance of rupture forces at varying loading rates. The variance at each loading rate is independent of the value of k_{bias}.

k_{bias}, but performed badly for the two larger k_{bias}. The Bell-Evans and DHS models do not depend on the k_{bias} of the trap so these models predict the same behavior regardless of k_{bias}. As k_{bias} gets larger, we see that the simulation data of the moving trap deviates from that of the linearly increasing force. An increasing k_{bias} has the effect of increasing the F_{eq} in the FNDY model. The F_{eq} determines the value at which the FNDY model plateaus at low loading rates.

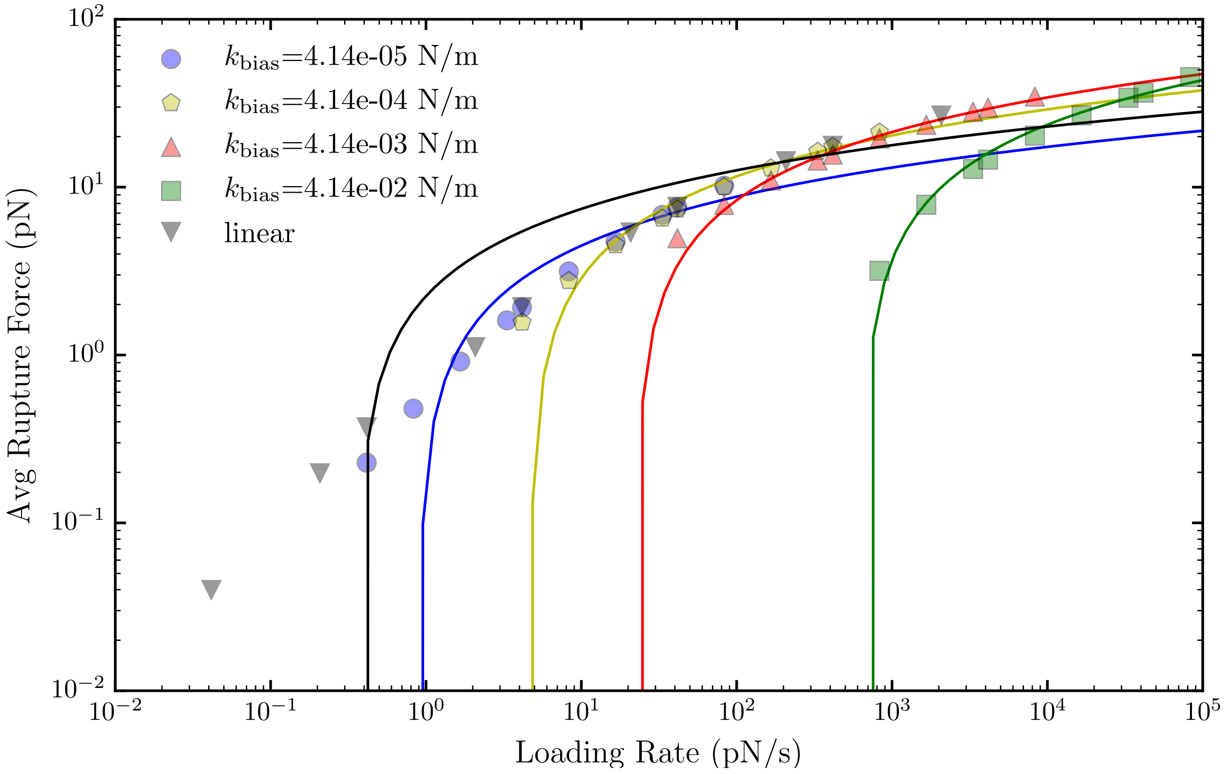
The simulation data was performed through individual trials, which means we not only have the mean rupture force $\langle F_{\mathrm{r}} \rangle$, we also have the variance of the rupture force $\sigma^2_{F_{\mathrm{r}}}$. We have plotted the variance of our rupture forces and find there is no large dependence on the k_{bias} as there was for the mean rupture forces (Figure 3.4). This suggests the variance of rupture forces could be a more robust observable than the mean rupture force to extract parameters from.

To test each model's ability to obtain accurate information about the system parameters if those were unknown, we fit each set of data with different k_{bias} using each proposed model. We used a least squares fitting method with a smooth l1 loss function where the error is given by $2(\sqrt{(}1 + z) - 1)$, and z is the absolute difference between the function and our data. To ensure we were not falling into a local minumum, we performed fits using a variety of initial guesses and then selected the fit with the smallest loss function. Another factor to consider when fitting the data is whether to fit against the $\langle F_{\text{r}} \rangle$ or $\log\langle F_{\text{r}} \rangle$. Since the differences between the largest and smallest $\langle F_{\text{r}} \rangle$ are many orders of magnitude, fitting it using only $\langle F_{\text{r}} \rangle$ will bias the fitting towards accurately fitting the largest $\langle F_{\text{r}} \rangle$.

Based on Figures 3.2 and 3.3 we fit only a subset of the data where it appeared the models fit best. We fit the $\langle F_{\text{r}} \rangle$ of the ten slowest loading rates in the linearly increasing ensemble and for each k_{bias} in the harmonic trap ensemble. Starting with the Bell-Evans model (Figure 3.5) we see the fitting parameters are largely on the same order of magnitude, but still not very accurate. For the highest trap strength, we see a much larger k_0 than expected from the Kramers off-rate. The FNDY fits (Figure 3.6) also have parameters which have the correct order of magnitude except for the fits to the largest k_{bias} which again gives parameters that are very different from the actual values. For the largest k_{bias} the F_{eq} is negative which is an unphysical value and the k_0 is much higher.

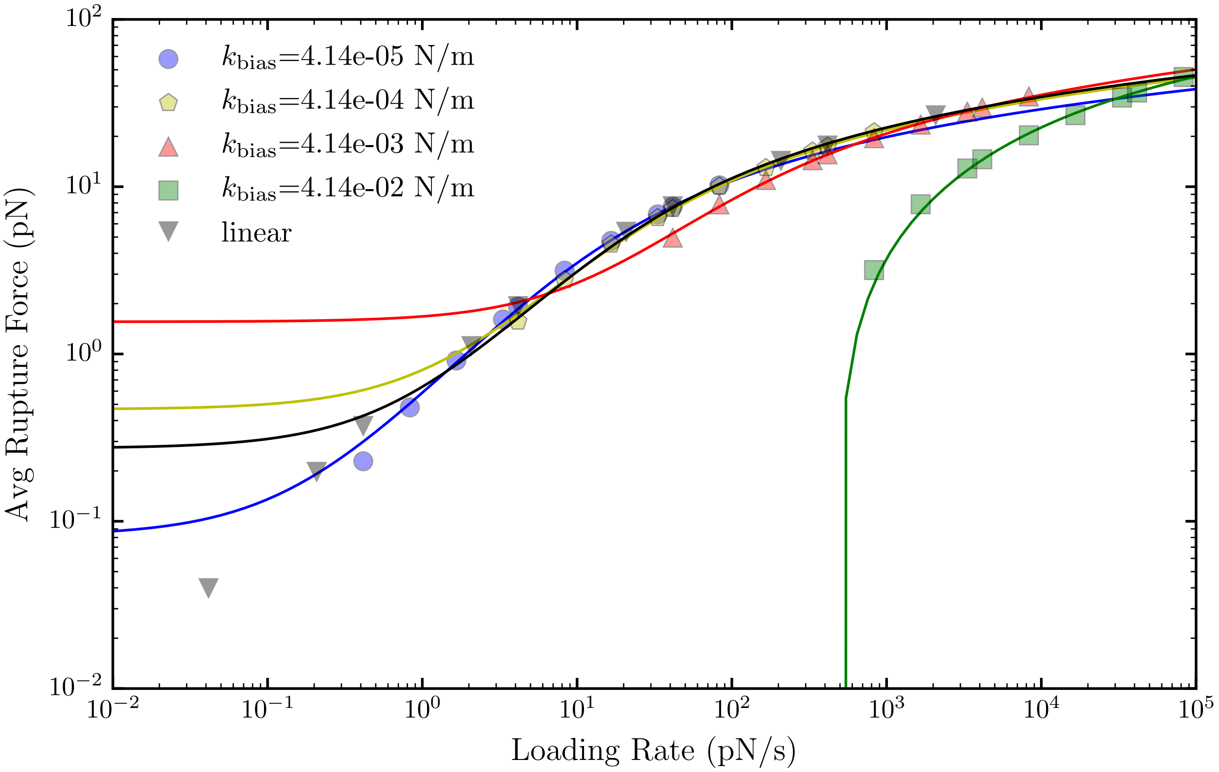
When we fit the system using the HS and DHS models, they produce the same curves. This is somewhat surprising because the HS model explicitly contains a term for k_{bias} while the DHS model does not. However their parameters are different and for the large k_{bias}, the HS model does a much better job at predicting the correct parameters. The BSK model performs the best for low k_{bias}. At higher k_{bias} the predicted D and x_{b} are quite far off.

An important feature we have neglected so far is the shape of the underlying potential. Up to this point we have solely focused on the quadratic cusp potential, which may not fully describe all systems. To determine the effect of the shape of the potential on the $\langle F_{\text{r}} \rangle$ we performed simulations on two other systems: a quartic potential with one barrier and a quartic potential with two barriers. For the single barrier quartic cusp potential, we see no



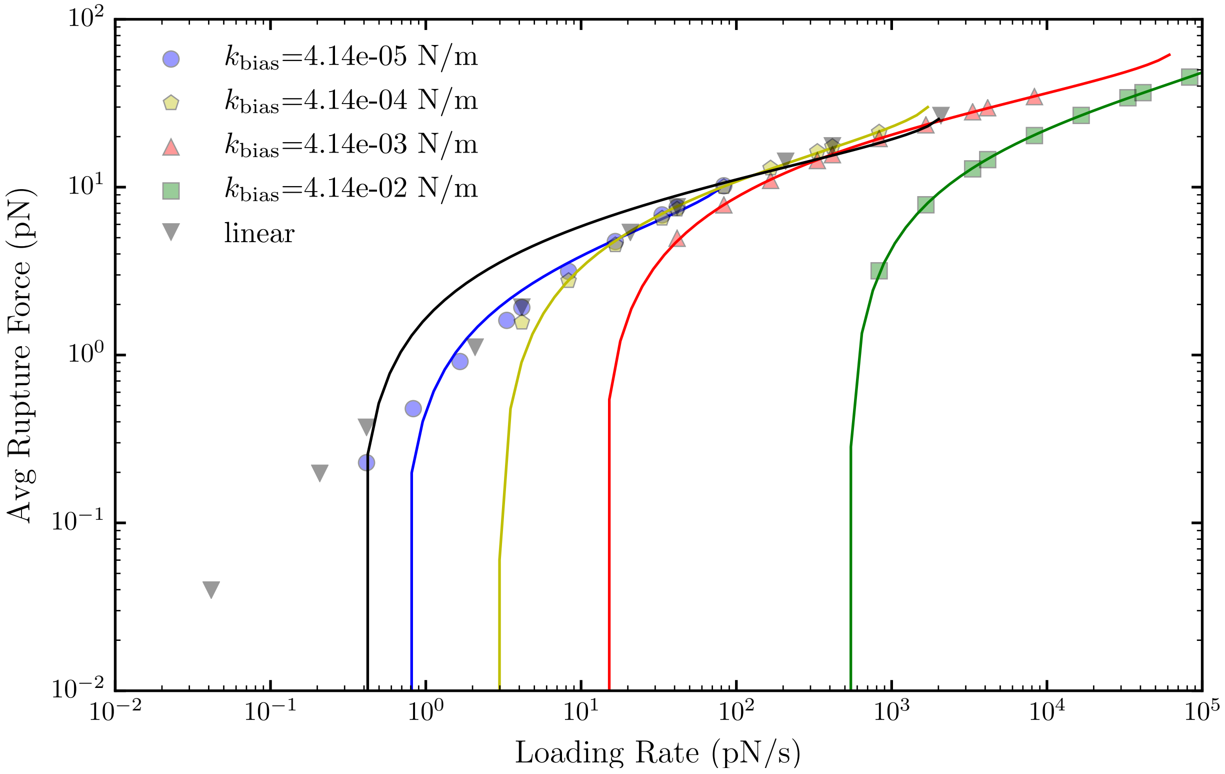
k_{bias} (N/m)	x_{b} (nm)	k_0 (s^{-1})
4.14x10^{-5}	2.22	0.48
4.14x10^{-4}	1.09	1.24
4.14x10^{-3}	0.74	4.02
4.14x10^{-2}	0.48	76
linear	1.85	0.16
expected	1	1.62

Figure 3.5: The Bell-Evans fitting parameters determined from the least squares fit of the ten slowest loading rates of each k_{bias} and $\dot{F}$. The expected k_0 comes from solving the Kramers off-rate (equation 3.5). The solid lines are equation 3.17 with each the fitted parameters of each line in the table.



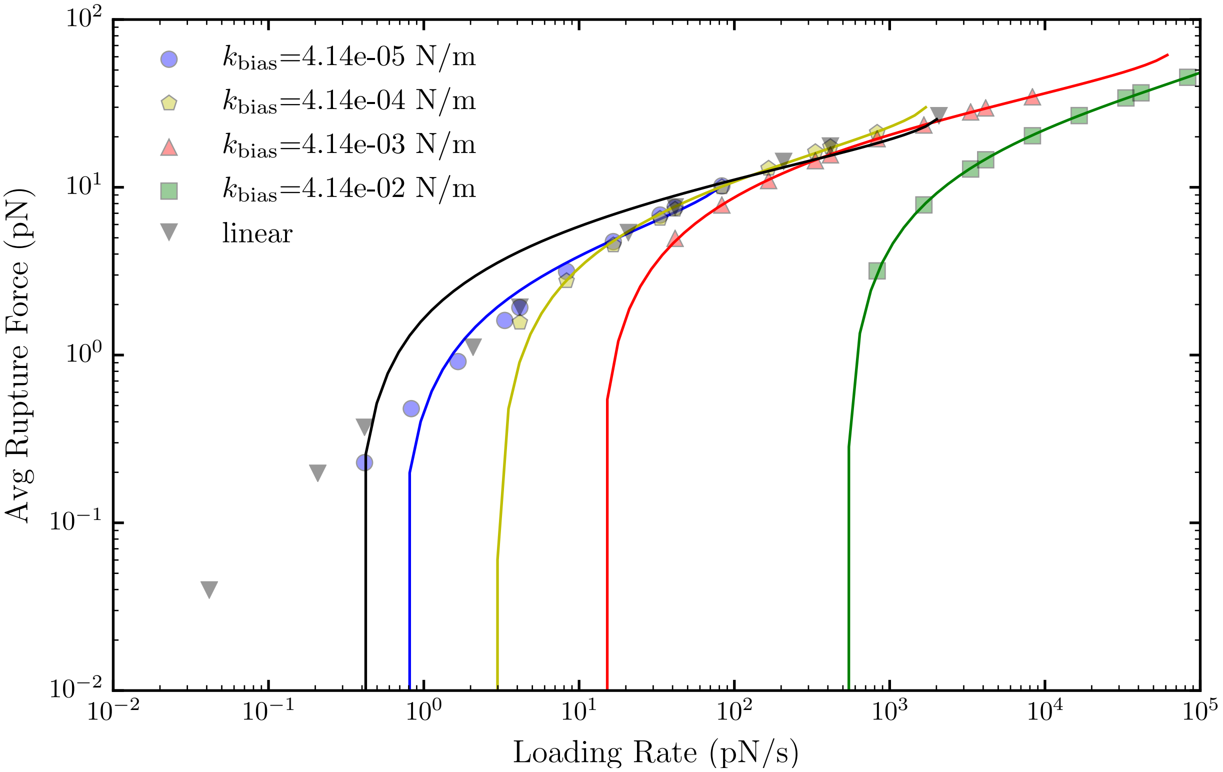
k_{bias} (N/m)	x_{b} (nm)	F_{eq} (pN)	k_0 (s^{-1})
4.14x10^{-5}	1.03	0.34	1.02
4.14x10^{-4}	0.82	1.94	1.49
4.14x10^{-3}	0.64	6.46	3.86
4.14x10^{-2}	0.40	-22.0	79.2
linear	0.80	1.12	1.41
expected	1	$\sqrt{8.28\text{x}10^4 k_{\mathrm{bias}}}$	1.62

Figure 3.6: The Friddle-Noy-De Yoreo fitting parameters determined from the least squares fit of the ten slowest loading rates of each k_{bias} and $\dot{F}$. The expected k_0 comes from solving the Kramers off-rate (equation 3.5). The expected F_{eq} is dependent on k_{bias} so would be expected to be different for each pulling protocol. The solid lines are equation 3.54 with the fitted parameters of each line in the table.



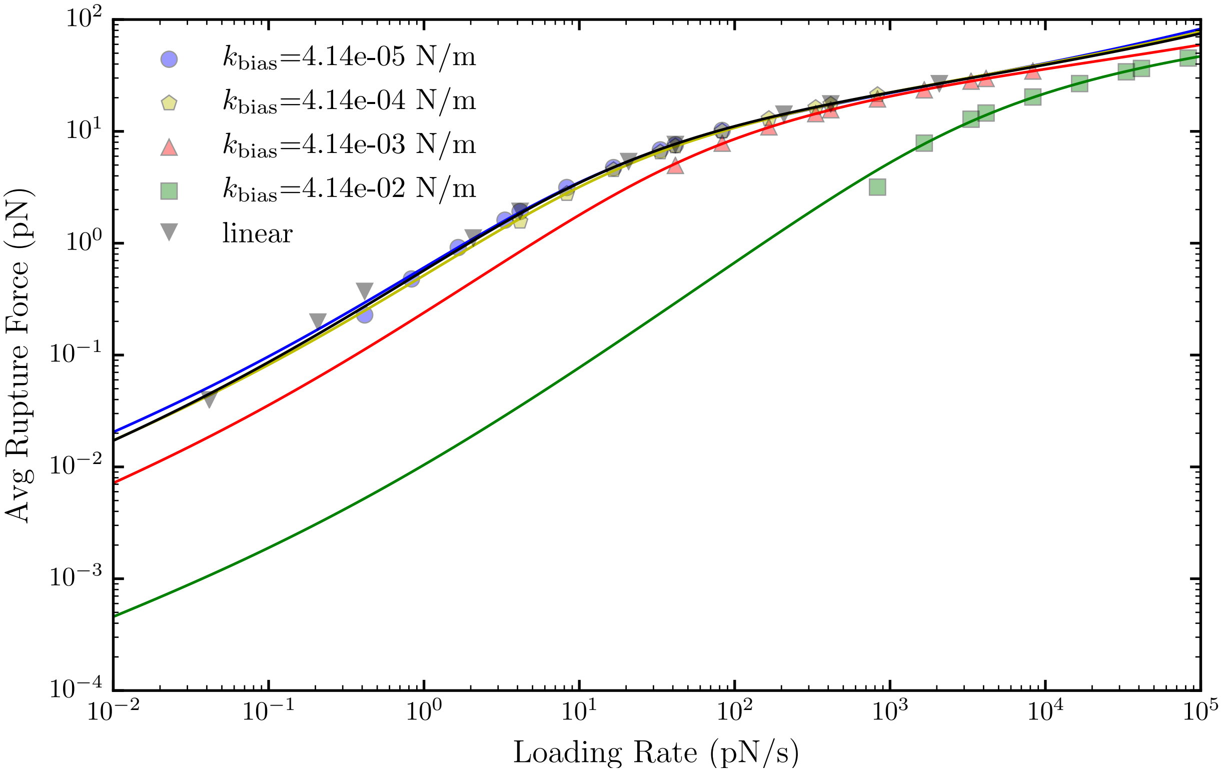
k_{bias} (N/m)	x_{b} (nm)	k_{sys} (N/m)	D (nm^2/s)	$\Delta G^{\ddagger}$ ($k_{\text{B}}T$)
4.14x10^{-5}	3.46	3.40x10^{-3}	40.3	4.93
4.14x10^{-4}	1.66	1.96x10^{-2}	56.8	6.55
4.14x10^{-3}	1.09	6.29x10^{-2}	354	9.10
4.14x10^{-2}	0.95	8.53x10^{-2}	574	9.36
linear	2.65	1.02x10^{-2}	181	8.66
expected	1	8.28x10^{-2}	1000	10

Figure 3.7: The Hummer-Szabo fitting parameters determined from the least squares fit of the ten slowest loading rates of each k_{bias} and $\dot{F}$. The solid lines are the intermediate Hummer-Szabo formula (equation 3.47) for each trap strength. The k_0 in the formula was calculated using the Kramers off-rate for a harmonic cusp potential (equation 3.5).



k_{bias} (N/m)	x_{b} (nm)	$\Delta G^{\ddagger}$ ($k_{\mathrm{B}}T$)	D (nm^2/s)
4.14x10^{-5}	3.42	4.87	40
4.14x10^{-4}	1.63	6.42	57
4.14x10^{-3}	1.03	8.54	354
4.14x10^{-2}	0.64	6.31	574
linear	2.65	8.66	181
expected	1	10	1000

Figure 3.8: The Dudko-Hummer-Szabo model fitting parameters determined from the least squares fit of the ten slowest loading rates of each k_{bias} and $\dot{F}$. The solid lines are equation 3.49 with the parameters given in the table.



k_{bias} (N/m)	x_{b} (nm)	$\Delta G^{\ddagger}$ ($k_{\mathrm{B}}T$)	D (nm^2/s)
4.14x10^{-5}	1.20	8.87	404
4.14x10^{-4}	1.06	9.16	501
4.14x10^{-3}	0.77	10.6	2310
4.14x10^{-2}	0.35	13.2	19100
linear	1.10	9.53	588
expected	1	10	1000

Figure 3.9: The Bullerjahn-Sturm-Kroy model fitting parameters determined from the least squares fit of the ten slowest loading rates of each k_{bias} and $\dot{F}$. The solid lines are equation 3.84.

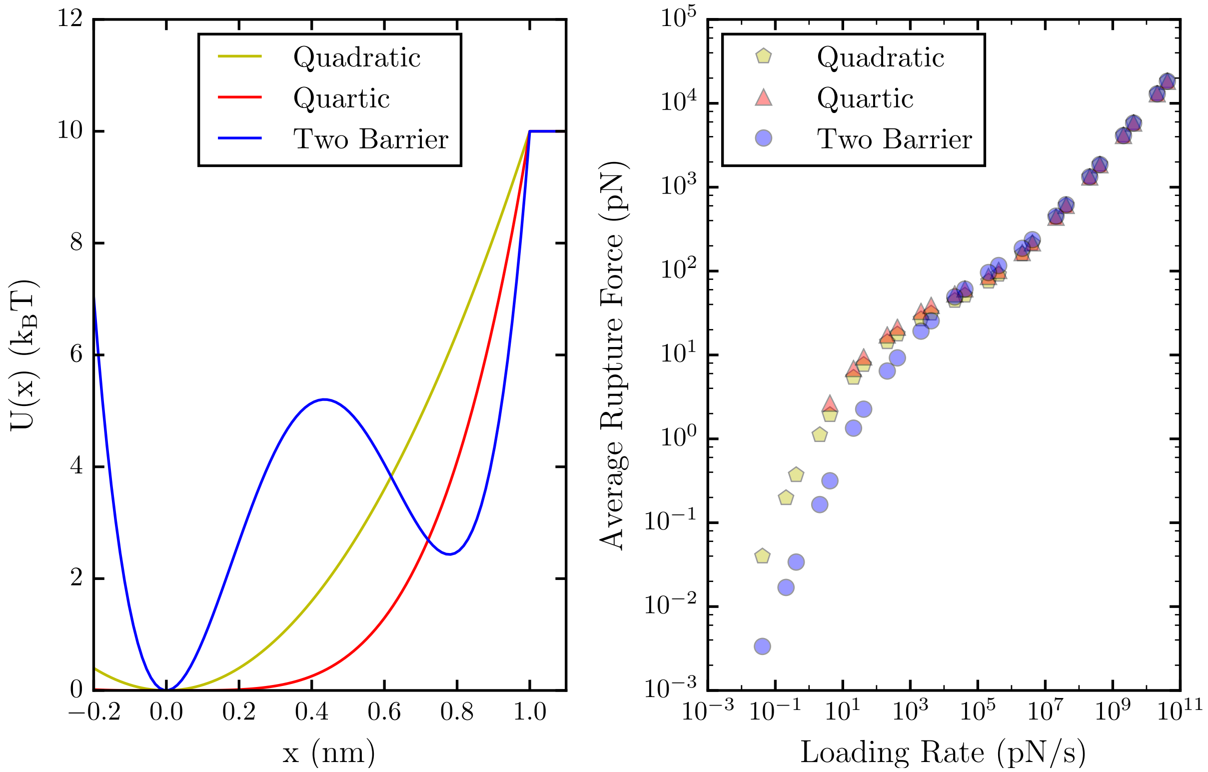


Figure 3.10: The average rupture force of linearly increasing pulling out of a quadratic, quartic, and two barrier potential.

major differences between the average rupture forces, especially at high loading rates (Figure 3.10). For small loading rates, we only see very slight differences. We also explored the effect of multiple barriers present in the underlying potential because many experimental results have been built on the Bell-Evans model, which expects a simple logarithmic relationship between $\dot{F}$ and $\langle F_{\mathrm{r}} \rangle$. When there is a deviation from the simple logarithmic behavior it has been interpreted as the existence of two barriers and two slopes are fit to the graph.[7,30,35] For potentials with two barriers, we qualitatively see the same relationship between $\langle F_{\mathrm{r}} \rangle$ and $\dot{F}$ for the two-barrier potential as the one-barrier quadratic and quartic potentials. However at lower loading rates there is a large deviation which is likely due the relative height of the second barrier being smaller than the height of the barrier in the quadratic and quartic potentials. These results suggest the existence of a nonlogarithmic relationship can be equally well-described by a simple one barrier quadratic potential.

Another avenue in single molecule pulling experiments is using the work as our observable of interest instead of the force In these types of simulations, the particle is moved in both directions across the barrier and a free energy surface is recreated in the dimension of the pulling.[37] This can be hard to translate to experiments because the reverse direction is not guaranteed to follow the same pathway across the barrier. Instead there would likely be alternative paths due to rearrangements in the higher order structure such as a protein unfolding, or a DNA strand unbinding. In our moving trap simulations, we measured the work done by the biasing trap. Surprisingly, we find there to be a very strong correlation between the rupture work, and the rupture force (3.11). However it does not match our expected rupture work (equation 3.120). To determine why, we used our rupture force and rupture work data to calculate the initial position of the particle, x_0. We expected this to be a Gaussian distribution with mean at 0, but this was only true for the higher velocity traps (Figure 3.13). For our slower moving traps, our initial positions calculated in this way seem to be skewed heavily towards negative values (Figure 3.12).

Because we use a simulation, we can easily traverse many more magnitudes of loading rate than are achievable in experiment. This allows us to see all three regions of slow, linear relationship, intermediate, logarithmic relationship, and fast, square root relationship. Experimentalists do not have the benefit of knowing the details of the underlying potential and so have to think about the best way to obtain data. Based on what we have seen, it is better to get information at lower loading rates because that will give you more information about the shape of the potential. One benefit of high loading rates is that the strength of the trap does not have an effect any more. Overall though, it seems it is important to take measurements at a wide range of loading rates in order to identify the slow, intermediate, and fast regions.

Another difference between experiment and our method of simulation, is how to determine when the bond has ruptured. In our simulations, since we know the position of our particle, we know when it crosses the predetermined barrier position. In dynamic force specroscopy, one has to look for a drop in the force in a force-time plot to determine what the rupture

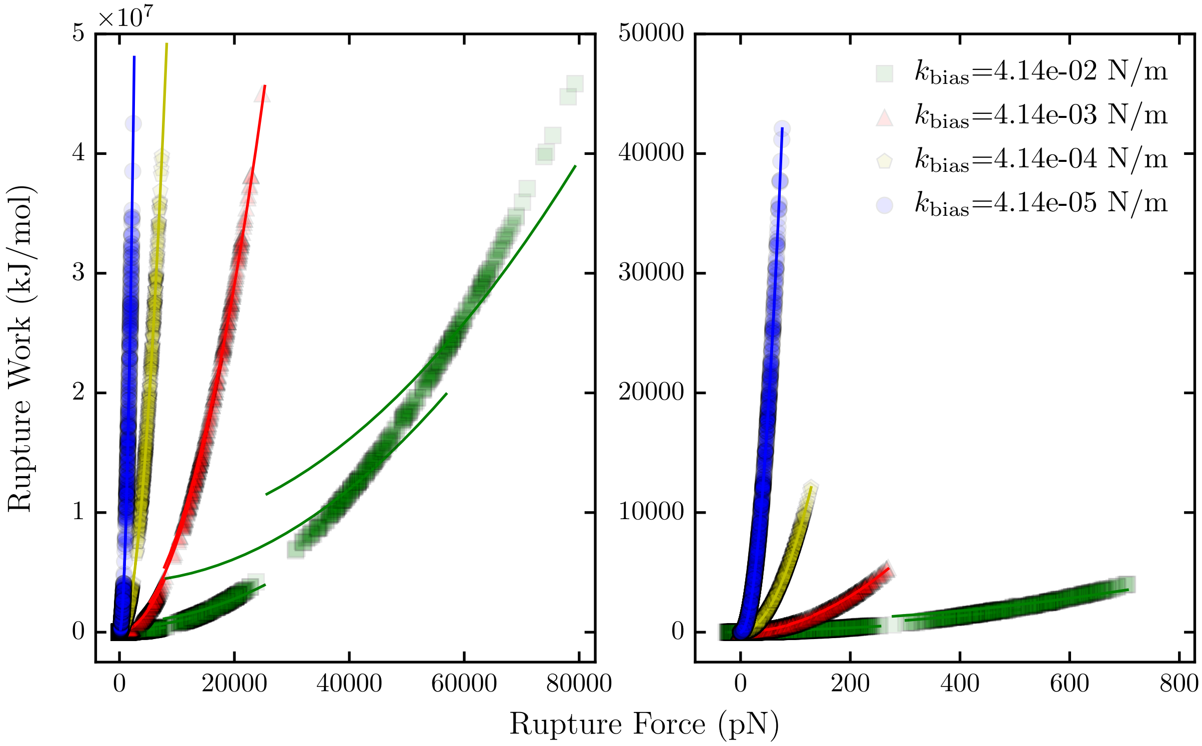


Figure 3.11: Rupture Work vs Rupture Force for the Harmonic Trap. Left is all trap velocities, while right is the 17 slowest. The solid lines are the expected work for a particle with an initial position of 0 (equation 3.120) We see that for lower trap strengths, the expected work matches well with the observed work. However for the large trap strength, this does not appear to match well. This is partially due to assuming 0 is the initial starting position for all points, even though we know that is not the case.

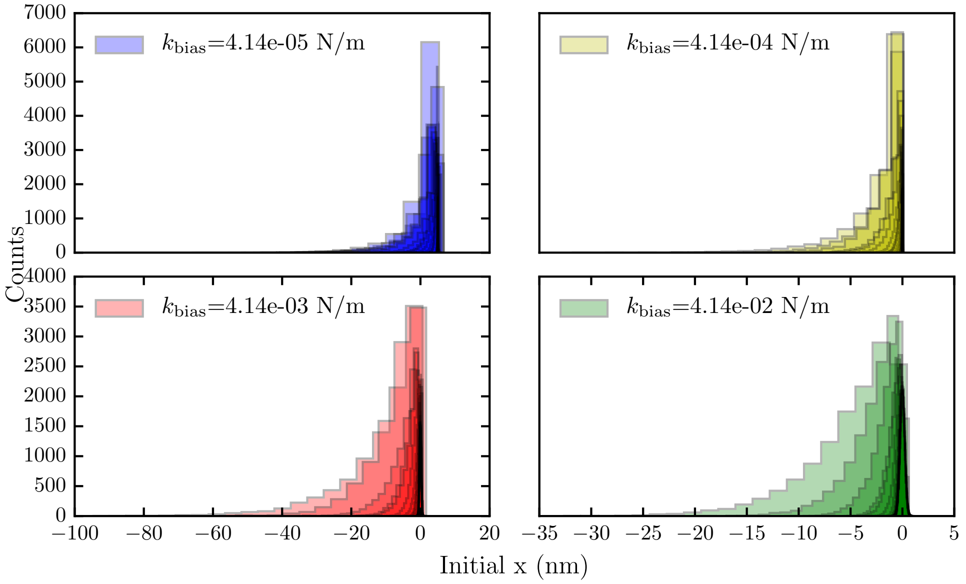


Figure 3.12: Histograms of the initial position of the particle calculated from the rupture work and rupture force for all pulling velocities. Each histogram is overlayed on each other, with darker shades indicating more histograms. We see that when we try to calculate initial position from the rupture work and the rupture force, we get a large skew to negative initial positions. We expect the initial positions to be a Gaussian distribution centered around 0.

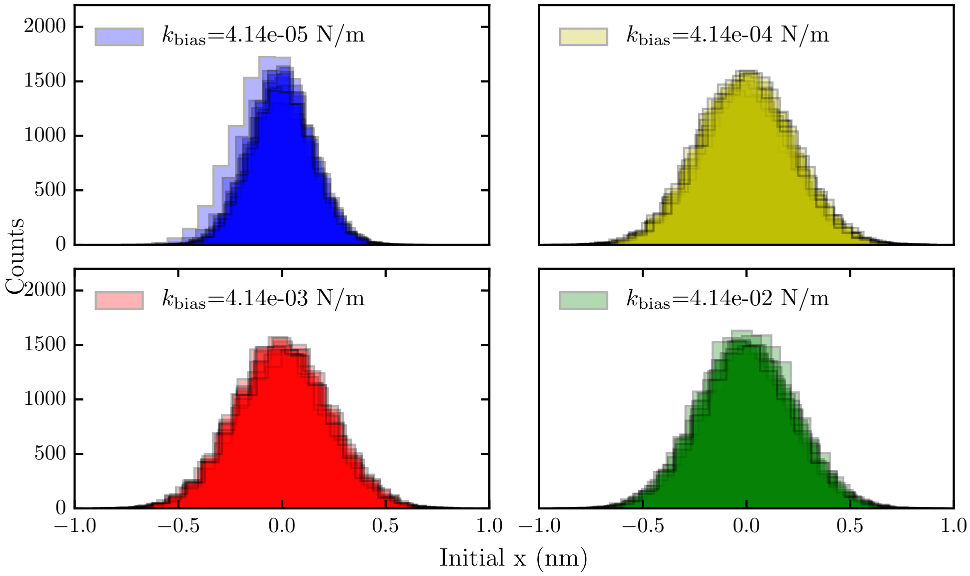


Figure 3.13: Initial position of the particle calculated from the rupture work and rupture force for only the 10 fastest moving traps. For these fast moving traps, we see the initial positions behave as we would expect unlike when we look at all pulling velocities.

force should be. When we plot force vs. time for the linearly increasing force protocol, we do not observe this drop in applied force because the applied force is constant at all points in space. When we do the same plot for the harmonic bias, we do see this drop in applied force for only the slow moving traps (Figure 3.14). For high velocity traps, we do not see any changes or obvious drop at the barrier position. We do not see this drop in force because our particle still feels some friction. This causes the particle to always lag behind the position of the trap and feel a force (equation 3.105). This force only depends on the velocity of the trap and not the strength of the trap. If the force on the particle due to lagging behind the trap is larger than the force required to pull the particle over the barrier, there will not be a drop in the applied force. This suggests that even if we can simulate a high pulling velocity, it may not be a good analog to experimentally derived results because in an experiment the rupture force is determined from the drop in measured force.

3.4 Conclusion

Using Brownian dynamics simulations, we were able to compare and contrast five different models of dynamic force spectroscopy. First we found the rupture force is most sensitive to the pulling protocol and the underlying potential at slow loading rates. This suggests the best regime to do experiments is at low loading rates, however this presents its own challenges because the rupture forces are so small. For the five models we tested, we found most performed the best for small trap strengths k_{bias}. At larger k_{bias} assumptions underlying the models broke down which lead to large deviations between the fitted parameters and the actual parameters. An interesting effect we found is that the rupture force variance seems to be much less sensitive to the trap strengths k_{bias}. This could lead into further investigation into extracting information about the potential surface from the variance as well as the mean.

Our simulations are low cost which allows us to perform many simulations over a wide range of loading rates. Experimentally it can be difficult to obtain data over such a wide range so the model does not need to fit over the slow, intermediate, and fast regimes at the same time. Instead the data could be fit using only the model that works well in that

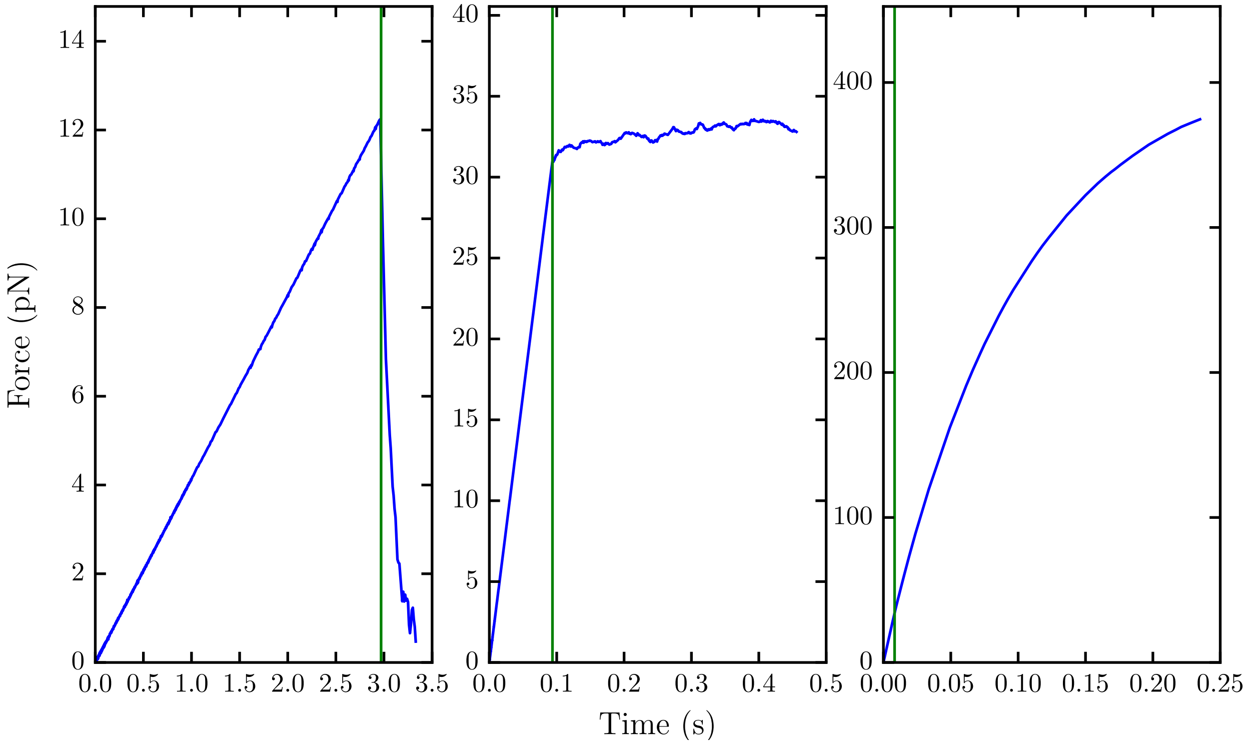


Figure 3.14: Force vs. time for a slow, intermediate, and fast moving harmonic trap. The left plot was pulled at a rate of 4.14 pN/s, the center plot was pulled at a rate of 33pN/s, and the right plot was pulled at a rate of 4140 pN/s. All k_{bias} are 4.14x10^{-5} N/m. The green vertical line represents the time when the particle left the well. For the slow moving trap, we see the time at which the particle leaves the well corresponds to a steep drop in the force applied on the particle by the harmonic trap. For the fast moving trap, we do not see a large change of force at the time of rupture. Instead we see the force applied to the particle continue to increase well past the point of rupture. This is due to friction such that even if the trap were to pull the particle over a flat potential, there would still be a lag between the average particle position and the trap minimum. This lag in particle position and trap minimum means that the particle will still feel a force as it is being dragged (equation 3.105).

regime. At slow regions it seems the BSK slow (equation 3.81), and HS slow (equation 3.46) work equally well except when the trap strength is large. For intermediate regimes HS intermediate (equation 3.47) works the best regardless of trap strength. And for the fast loading rates both the BSK fast (equation 3.83) and HS fast models (equation 3.41) work best.

Chapter 4

SUMMARY

In this thesis we have investigated how the properties of equilibrium systems can be obtained from non-equilibrium studies. Our first project focused on the simulation of DNA hybridization from two single-stranded DNA. We were able to create a Markov State Model (MSM) from raw simulation data and use it to find the free energy of states, and the major paths of hybridization. We found that DNA hybridizes through a nucleation and zipping up process. We also found that the major transition states are when the strands had two bases bound at either end of the strand. In order for two strands to be likely to continue to the fully hybridized state, they needed two or three correct bonds. However, the net number of bonds is not sufficient to characterize the transition state ensemble. Another advantage of this technique is in determining the free energy of this system without preselecting a particular collective variable (CV). Our simulations explored the state space in an unbiased way and we reconstructed the equilibrium properties of the system.

Our second project studied models of Dynamic Force Spectroscopy (DFS). We generated rupture force data using Brownian dynamics simulations and compared this data to five different models. We found all of the models worked, but only for certain regimes of pullling. The Bell-Evans, Friddle-Noy-De Yoreo, and Dudko-Hummer-Szabo models worked only at intermediate pulling rates, and only for weaker trap stiffnesses. The Hummer-Szabo and Bullerjahn-Sturm-Kroy models have different equations for different regimes of pulling. Their fast loading rate models both work well at fast rates. The Bullerjahn-Sturm-Kroy slow model and Hummer-Szabo slow model only worked for smaller trap strengths. The Hummer-Szabo intermediate model worked well for intermediate pulling rates regardless of the trap strength. We also found the rupture force variance to be less sensitive to the pulling protocol than the mean rupture force, which suggests the variance should be used instead.

Overall we have contributed to the field in two different ways through non-equilibrium techniques. The method of creating an MSM to determine the energetics and mechanism of DNA hybridization can be applied to additional sequences to study effects of base repetition and sequence dependence of hybridization. We have also elucidated the differences between different models of DFS to determine which model is most accurate. This validation of the accuracy of DFS models provides guidance to future experiments on how to extract meaningful data of the underlying potential.

BIBLIOGRAPHY

[1] F. Aboul-ela, D. Koh, I. Tinoco, and F. H. Martin. Base-base mismatches. thermodynamics of double helix formation for $dCA_3XA_3G + dCT_3YT_3G$ (X, Y = A,C,G,T). *Nucleic Acids Res.*, 13(13):4811–4824, 1985.

[2] R. J. Allen, C. Valeriani, and P. Rein ten Wolde. Forward flux sampling for rare event simulations. *J. Phys.: Condens. Matter*, 21(46):463102, 2009.

[3] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky. Bubble dynamics in double-stranded DNA. *Phys. Rev. Lett.*, 90(13):138101, 2003.

[4] A. Barhoumi and N. J. Halas. Label-free detection of DNA hybridization using Surface Enhanced Raman Spectroscopy. *J. Am. Chem. Soc.*, 132(37):12792–12793, 2010.

[5] K. A. Beauchamp, G. R. Bowman, T. J. Lane, L. Maibaum, I. S. Haque, and V. S. Pande. MSMBuilder2: Modeling conformational dynamics at the picosecond to millisecond scale. *J. Chem. Theory Comput.*, 7(10):3412–3419, 2011.

[6] G. I. Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–627, 1978.

[7] C. A. Bippes, A. Zeltina, F. Casagrande, M. Ratera, M. Palacín, D. J. Müller, and D. Fotiadis. Substrate binding tunes conformational flexibility and kinetic stability of an amino acid antiporter. *J. Biol. Chem.*, 284(28):18651–18663, 2009.

[8] G. R. Bowman, K. A. Beauchamp, G. Boxer, and V. S. Pande. Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.*, 131:124101, 2009.

[9] C. Chen, W. Wang, Z. Wang, F. Wei, and X. S. Zhao. Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization. *Nucleic Acids Res.*, 35(9):2875–2884, 2007.

[10] X. Chen, Y. Zhou, P. Qu, and X. S. Zhao. Base-by-base dynamics in DNA hybridization probed by fluorescence correlation spectroscopy. *J. Am. Chem. Soc.*, 130(50):16947–16952, 2008.

[11] L. Y. T. Chou, K. Zagorovsky, and W. C. W. Chan. DNA assembly of nanoparticle superstructures for controlled biological delivery and elimination. *Nat. Nanotechnol.*, 9(2):148–55, 2014.

[12] I. I. Cissé, H. Kim, and T. Ha. A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat. Struct. Mol. Biol.*, 19(6):623–7, 2012.

[13] A. Dickson and C. L. Brooks. Quantifying hub-like behavior in protein folding networks. *J. Chem. Theory Comput.*, 8(9):3044–3052, 2012.

[14] V. V. Didenko. DNA probes using fluorescence resonance energy transfer (FRET): designs and applications. *BioTechniques*, 31(5):1106–16, 1118, 1120–1, 2001.

[15] T. G. Drummond, M. G. Hill, and J. K. Barton. Electrochemical DNA sensors. *Nat. Biotechnol.*, 21(10):1192–1199, 2003.

[16] O. K. Dudko, A. E. Filippov, J. Klafter, and M. Urbakh. Beyond the conventional description of dynamic force spectroscopy of adhesion bonds. *Proc. Natl. Acad. Sci. U. S. A.*, 100(20):11378–11381, 2003.

[17] O. K. Dudko, G. Hummer, and A. Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, 96(10):108101, 2006.

[18] W. E and E. Vanden-Eijnden. Towards a theory of transition paths. *J. Stat. Phys.*, 123(3):503–523, 2006.

[19] E. Evans and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophys. J.*, 72(4):1541–55, 1997.

[20] K. S. Fa. Fokker-Planck equation with linear and time dependent load forces. *Eur. J. Phys.*, 37(6), 2016.

[21] S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner. Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. U. S. A.*, 83(24):9373–7, 1986.

[22] R. W. Friddle, A. Noy, and J. J. De Yoreo. Interpreting the widespread nonlinear force spectra of intermolecular bonds. *Proc. Natl. Acad. Sci. U. S. A.*, 109(34):13573–13578, 2012.

[23] B. L. Gaffney, L. A. Marky, and R. A. Jones. The influence of the purine 2-amino group on DNA conformation and stability-II: Synthesis and physical characterization of d[CGT(2-NH_2)ACG], d[CGU(2-NH_2)ACG], and d[CGT(2-NH_2)AT(2-NH_2)ACG]. *Tetrahedron*, 40(1):3–13, 1984.

[24] C. Gergely, J. Voegel, P. Schaaf, B. Senger, M. Maaloum, J. K. Hörber, and J. Hemmerlé. Unbinding process of adsorbed proteins under external stress studied by atomic force microscopy spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.*, 97(20):10802–7, 2000.

[25] D. M. Hinckley, G. S. Freeman, J. K. Whitmer, and J. J. de Pablo. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.*, 139(14):144903, 2013.

[26] D. M. Hinckley, J. P. Lequieu, and J. J. de Pablo. Coarse-grained modeling of DNA oligomer hybridization: length, sequence, and salt effects. *J. Chem. Phys.*, 141(3):035102, 2014.

[27] G. Hummer and A. Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophys. J.*, 85(1):5–15, 2003.

[28] R. Jin and L. Maibaum. Mechanisms of DNA hybridization: Transition path analysis of a simulation-informed Markov model. *J. Chem. Phys.*, 150(10), 2019.

[29] Y. Kawase, S. Iwai, H. Inoue, K. Miura, and E. Ohtsuka. Studies on nucleic acid interactions I. Stabilities of mini-duplexes ($dG_2A_4XA_4G_2 \cdot dC_2T_4YT_4C_2$) and self-complementary d(GGGAAXYTTCCC) containing deoxyinosine and other mismatched bases. *Nucleic Acids Res.*, 14(19):7727–7736, 1986.

[30] B.-H. Kim, N. Y. Palermo, S. Lovas, T. Zaikova, J. F. W. Keana, and Y. L. Lyubchenko. Single-molecule atomic force microscopy force spectroscopy study of Aβ-40 interactions. *Biochemistry*, 50(23):5154–5162, jun 2011.

[31] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, Apr 1940.

[32] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.*, 99(20):12562–6, 2002.

[33] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco, and C. Bustamante. Reversible unfolding of single RNA molecules by mechanical force. *Science*, 292(5517):733–7, 2001.

[34] F. H. Martin, M. M. Castro, F. Aboul-ela, and I. Tinoco. Base pairing involving deoxyinosine: implications for probe design. *Nucleic Acids Res.*, 13(24):8927–8938, 1985.

[35] R. Merkel, P. Nassoy, A. Leung, K. Ritchie, and E. Evans. Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. *Nature*, 397(6714):50–53, Jan 1999.

[36] P. Metzner, C. Schütte, and E. Vanden-Eijnden. Transition Path Theory for Markov Jump Processes. *Multiscale Model. Simul.*, 7(3):1192–1219, 2009.

[37] D. D. L. Minh and J. D. Chodera. Estimating equilibrium ensemble averages using multiple time slices from driven nonequilibrium processes: Theory and application to free energies, moments, and thermodynamic length in single-molecule pulling experiments. *J. Chem. Phys.*, 134(2):024111, jan 2011.

[38] C. A. Mirkin, R. L. Letsinger, R. C. Mucic, and J. J. Storhoff. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature*, 382(6592):607–609, 1996.

[39] G. Niranjani and R. Murugan. Theory on the mechanism of DNA renaturation: Stochastic nucleation and zipping. *PLoS ONE*, 11(4):e0153172, 2016.

[40] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. U. S. A.*, 106(45):19011–6, 2009.

[41] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.*, 134(8):085101, 2011.

[42] T. E. Ouldridge, P. Šulc, F. Romano, J. P. K. Doye, and A. A. Louis. DNA hybridization kinetics: Zippering, internal displacement and sequence dependence. *Nucleic Acids Res.*, 41(19):8886–8895, 2013.

[43] V. S. Pande, K. Beauchamp, and G. R. Bowman. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods*, 52(1):99–105, 2010.

[44] K. M. Parkhurst and L. J. Parkhurst. Kinetic studies by fluorescence resonance energy transfer employing a double-labeled oligonucleotide: Hybridization to the oligonucleotide complement and to single-stranded DNA. *Biochemistry*, 34(1):285–292, 1995.

[45] B. Peters and B. L. Trout. Obtaining reaction coordinates by likelihood maximization. *J. Chem. Phys.*, 125(5):054108, 2006.

[46] D. Pörschke and M. Eigen. Co-operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic–oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.*, 62(2):361–381, 1971.

[47] J. H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schtte, and F. Noé. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.*, 134(17):174105, 2011.

[48] F. Rico, L. González, I. Casuso, M. Puig-Vidal, and S. Scheuring. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. *Science*, 342(6159):741–743, 2013.

[49] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276(5315):1109–1112, 1997.

[50] P. W. K. Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, 2006.

[51] E. J. Sambriski, D. C. Schwartz, and J. J. de Pablo. Uncovering pathways in DNA oligonucleotide hybridization via transition state analysis. *Proc. Natl. Acad. Sci. U. S. A.*, 106(43):18125–30, 2009.

[52] J. SantaLucia and D. Hicks. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33:415–440, 2004.

[53] N. C. Seeman. Nucleic acid junctions and lattices. *J. Theor. Biol.*, 99(2):237–247, 1982.

[54] N. Singhal, C. D. Snow, and V. S. Pande. Using path sampling to build better Markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.*, 121:415, 2004.

[55] B. E. K. Snodin, F. Randisi, M. Mosayebi, P. Šulc, J. S. Schreck, F. Romano, T. E. Ouldridge, R. Tsukanov, E. Nir, A. A. Louis, and J. P. K. Doye. Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *J. Chem. Phys.*, 142(23):234901, 2015.

[56] S. Sorgenfrei, C.-y. Chiu, R. L. Gonzalez, Y.-J. Yu, P. Kim, C. Nuckolls, and K. L. Shepard. Label-free single-molecule detection of DNA-hybridization kinetics with a carbon nanotube field-effect transistor. *Nat. Nanotechnol.*, 6(2):126–132, 2011.

[57] N. Srinivas, T. E. Ouldridge, P. Šulc, J. M. Schaeffer, B. Yurke, A. A. Louis, J. P. K. Doye, and E. Winfree. On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.*, 41(22):10641–10658, 2013.

[58] T. Strunz, K. Oroszlan, R. Schäfer, and H. J. Güntherodt. Dynamic force spectroscopy of single DNA molecules. *Proc. Natl. Acad. Sci. U. S. A.*, 96(20):11277–82, Sep 1999.

[59] T. A. Sulchek, R. W. Friddle, K. Langry, E. Y. Lau, H. Albrecht, T. V. Ratto, S. J. DeNardo, M. E. Colvin, and A. Noy. Dynamic force spectroscopy of parallel individual Mucin1-antibody bonds. *Proc. Natl. Acad. Sci. U. S. A.*, 102(46):16638–16643, 2005.

[60] W. C. Swope, J. W. Pitera, and F. Suits. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B*, 108(21):6571—6581, 2004.

[61] G. Torrie and J. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23(2):187–199, 1977.

[62] D. H. Turner, N. Sugimoto, R. Kierzek, and S. D. Dreiker. Free energy increments for hydrogen bonds in nucleic acid base pairs. *J. Am. Chem. Soc.*, 109(12):3783–3785, 1987.

[63] V. A. Voelz, G. R. Bowman, K. Beauchamp, and V. S. Pande. Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). *J. Am. Chem. Soc.*, 132(5):1526–8, 2010.

[64] X. Wang, H. J. Lim, and A. Son. Characterization of denaturation and renaturation of DNA for DNA hybridization. *Environ. Health Toxicol.*, 29:e2014007, 2014.

[65] S. Weiss. Fluorescence spectroscopy of single biomolecules. *Science*, 283(5408):1676–83, 1999.

[66] J. G. Wetmur and N. Davidson. Kinetics of renaturation of DNA. *J. Mol. Biol.*, 31(3):349–370, 1968.

[67] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314(5801):1001–4, 2006.

[68] J. Zhang, Y. Yan, S. Samai, and D. S. Ginger. Dynamic melting properties of photoswitch-modified DNA: Shearing versus unzipping. *J. Phys. Chem. B*, 120(41):10706–10713, Oct 2016.