Temperature and cantilever dependent protein unfolding

A Thesis
Submitted to the Faculty
of
Drexel University
by
William Trevor King
in partial fulfillment of the
requirements for the degree
of
Doctor of Philosophy
July 2010

© Copyright 2010
William Trevor King.

This work is licensed under the terms of the Creative Commons Attribution-ShareAlike license Version 3.0. The license is available at http://creativecommons.org/licenses/by-sa/3.0/.
Dedications

Heaven Malaika King-Bailey
January 21st, 2006 – June 6th, 2010

What a bundle of joy.
We’ll miss you, love.
Acknowledgments

This work was produced using the following open source software projects:

\TeXlive/CTAN Typesetting.
PGF/Asymptote/Asyfig Graphical programming.
Python General purpose scripting.
SCons Build manager.
GNU Emacs Text editor.
Git Version control.
GNU/Linux/Gentoo Operating system and distribution.

and many, many more. I am deeply indebted to all of the smart, generous people who produce such wonderful tools.

\texttt{sawsim} work was supported by National Institutes of Health Grants R01-GM071793.
Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>iv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>iv</td>
</tr>
<tr>
<td>Abstract</td>
<td>vii</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Mechanical unfolding experiments</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Thesis Outline</td>
<td>3</td>
</tr>
<tr>
<td>2. Unfolding Theory</td>
<td>5</td>
</tr>
<tr>
<td>3. Chain Tension</td>
<td>6</td>
</tr>
<tr>
<td>3.1 Polymer Models</td>
<td>6</td>
</tr>
<tr>
<td>3.1.1 Wormlike chains</td>
<td>6</td>
</tr>
<tr>
<td>3.1.2 Freely-jointed chains</td>
<td>6</td>
</tr>
<tr>
<td>3.2 Folded domain tension</td>
<td>6</td>
</tr>
<tr>
<td>4. Theoretical Unfolding Force Distributions</td>
<td>7</td>
</tr>
<tr>
<td>4.1 Overview</td>
<td>7</td>
</tr>
<tr>
<td>4.2 Review of current research</td>
<td>7</td>
</tr>
<tr>
<td>4.2.1 Who’s who</td>
<td>7</td>
</tr>
<tr>
<td>4.2.2 Evolution of unfolding modeling</td>
<td>8</td>
</tr>
<tr>
<td>4.2.3 History of simulations</td>
<td>8</td>
</tr>
<tr>
<td>4.2.4 History of experimental AFM unfolding experiments</td>
<td>8</td>
</tr>
<tr>
<td>4.2.5 History of experimental laser tweezer unfolding experiments</td>
<td>8</td>
</tr>
<tr>
<td>4.3 Single-domain proteins under constant loading</td>
<td>9</td>
</tr>
<tr>
<td>4.3.1 Constant unfolding rate</td>
<td>9</td>
</tr>
<tr>
<td>4.3.2 Bell model</td>
<td>10</td>
</tr>
<tr>
<td>4.3.3 Saddle-point Kramers’ model</td>
<td>11</td>
</tr>
<tr>
<td>4.4 Double-integral Kramers’ theory</td>
<td>12</td>
</tr>
<tr>
<td>4.4.1 Cusp-like potentials</td>
<td>12</td>
</tr>
<tr>
<td>4.4.2 Quartic potentials</td>
<td>12</td>
</tr>
<tr>
<td>5. Monte Carlo Mechanical Unfolding Simulation</td>
<td>13</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>13</td>
</tr>
<tr>
<td>5.2 Methods</td>
<td>13</td>
</tr>
<tr>
<td>5.2.1 Generating force curves</td>
<td>13</td>
</tr>
<tr>
<td>5.2.2 Unfolding protein molecules by force</td>
<td>15</td>
</tr>
<tr>
<td>5.2.3 Choosing the simulation time steps</td>
<td>16</td>
</tr>
<tr>
<td>5.3 Results and Discussion</td>
<td>16</td>
</tr>
<tr>
<td>5.3.1 Force curves generated by simulation</td>
<td>16</td>
</tr>
<tr>
<td>5.3.2 Dependence of the unfolding force on the unfolding order and polymer length</td>
<td>17</td>
</tr>
<tr>
<td>5.3.3 The effect of cantilever force constant</td>
<td>20</td>
</tr>
<tr>
<td>5.3.4 Determination of $\Delta x_u$ and $k_{uo}$</td>
<td>20</td>
</tr>
<tr>
<td>5.4 Conclusions</td>
<td>22</td>
</tr>
<tr>
<td>6. Contour Length Space</td>
<td>25</td>
</tr>
<tr>
<td>7. Temperature Dependent Unfolding Theory</td>
<td>26</td>
</tr>
<tr>
<td>7.1 Energy landscape roughness</td>
<td>26</td>
</tr>
<tr>
<td>8. Apparatus</td>
<td>28</td>
</tr>
<tr>
<td>9. Cantilever Dependent Unfolding Experiments</td>
<td>29</td>
</tr>
</tbody>
</table>
List of Tables

List of Figures

1.1 Complex of biotin (red) and a streptavidin tetramer (green) (PDB ID: 1SWE). The correct streptavidin conformation creates the biotin-specific binding pockets. Biotin-streptavidin is a model ligand-receptor pair isolated from the bacterium *Streptomyces avidinii*. Streptavidin binds to cell surfaces, and bound biotin increases streptavidin’s cell-binding affinity. Figure generated with PyMol. 1
1.2 (a) A “double T” example of the pathway model of protein folding, in which the protein proceeds from the native state $N$ to the unfolded state $U$ via a series of metastable transition states $I_1$ and $I_2$ with two “dead end” states $I_{X1}$ and $I_{X2}$. Adapted from Bédard et al. $^4$. (b) The landscape model of protein folding, in which the protein diffuses through a multi-dimensional free energy landscape. Separate folding attempts may take many distinct routes through this landscape on the way to the folded state. Reproduced from Dill and Chan $^5$. 

1.3 Operating principle for an Atomic Force Microscope. A sharp tip integrated at the end of a cantilever interacts with the sample. Cantilever bending is measured by a laser reflected off the cantilever and incident on a position sensitive photodetector. 

1.4 (a) Schematic of the experimental setup for mechanical unfolding of proteins using an AFM (not to scale). An experiment starts with the tip in contact with the substrate surface, which is then moved away from the tip at a constant speed. $x_t$ is the distance traveled by the substrate, $x_c$ is the cantilever deflection, $x_u$ is the extension of the unfolded polymer, and $x_f = x_{f1} + x_{f2}$ is the extension of the folded polymer. (b) An experimental force curve from stretching a ubiquitin polymer with the rising parts of the peaks fitted to the WLC model (Section 3.1.1)$^6$. The pulling speed used was 1 $\mu$m/s. The irregular features at the beginning of the curve are due to nonspecific interactions between the tip and the substrate surface, and the last high force peak is caused by the detachment of the polymer from the tip or the substrate surface. Note that the abscissa is the extension of the protein chain $x_t - x_c$. 

5.1 Once the unfolding probability has been calculated, we need to determine whether or not a domain should unfold. We do this by generating a random number, and comparing that number to the unfolding probability $P$. The random number determines which of the possible paths we should follow for the current simulation. Such “statistical sampling” is the hallmark of the Monte Carlo approach$^7$. This cartoon translates the idea into the more familiar doors (possible paths) and dice (random numbers). 

5.2 (a) Three simulated force curves from pulling a polymer of eight identical protein molecules. The simulation was carried out using the parameters: pulling speed $v = 1$ $\mu$m/s, cantilever spring constant $\kappa_c = 50$ pN/nm, temperature $T = 300$ K, persistence length of unfolded proteins $p_u = 0.40$ nm, $\Delta x_u = 0.225$ nm, and $k_{u0} = 5 \cdot 10^{-5}$ s$^{-1}$. The contour length between the two linking points on a protein molecule is $L_{f1} = 3.7$ nm in the folded form and $L_{u1} = 28.1$ nm in the unfolded form. These parameters are those of ubiquitin molecules connected through the N-C termini$^6,8$. Detachment from the tip or substrate is assumed to occur at a force of 400 pN. In experiments, detachments have been observed to occur at a variety of forces. For clarity, the green and blue curves are offset by 200 and 400 pN respectively. (b) The distribution of the unfolding forces from 400 simulated force curves (3200 data points) such as that shown in (a). The frequency is normalized by the total number of points, i.e., the height of each bin is equal to the number of data points in that bin divided by the total number of data points (3200, for this histogram).
5.3 The dependence of the unfolding force on the temporal unfolding order for four polymers with 4, 8, 12, and 16 identical protein domains. Each point in the figure is the average of 400 data points. The first point in each curve represents the average of only the first peak in each of the 400 simulated force curves, the second point represents the average of only the second peak, and so on. The solid lines are fits of Eq. (5.13) to the simulated data, with best fit $\kappa_{WLC} = 203, 207, 161, \text{ and } 157 \, \text{pN/nm}$, respectively, for lengths 4 through 16. The insets show the force distributions of the first, fourth, and eighth peaks, left to right, for the polymer with eight protein domains. The parameters used for generating the data were the same as those used for Fig. 5.2a, except for the number of domains. The histogram insets were normalized in the same way as in Fig. 5.2b.

5.4 Simulated force curves obtained from pulling a polymer with eight protein molecules using cantilevers with different force constants $\kappa_c$. Parameters used in generating these curves are the same as those used in Fig. 5.2, except the cantilever force constant. Successive force curves are offset by 300 pN for clarity.

5.5 (a) The dependence of the unfolding forces on the pulling speed for three different model protein molecules characterized by the parameters $k_{u0}$ and $\Delta x_u$. The polymer length is eight molecules, and each symbol is the average of 3200 data points. (b) The dependence of standard deviation of the unfolding force distribution on the pulling speed for the simulation data shown in (a), using the same symbols. The insets show the force distribution histograms for the three proteins at the pulling speed of 1 $\mu$m/s. The left, middle, and right histograms are for the proteins represented by the top, middle, and bottom lines in (a), respectively.

5.6 Fit quality between an experimental data set and simulated data sets obtained using various values of unfolding rate parameters $k_{u0}$ and $\Delta x_u$. The experimental data are from octameric ubiquitin pulled at 1 $\mu$m/s, and the other model parameters are the same as those in Fig. 5.2. The best fit parameters are $\Delta x_u = 0.17$ nm and $k_{u0} = 1.2 \cdot 10^{-2} \, \text{s}^{-1}$. The simulation histograms were built from 400 pulls at for each parameter pair.

9.1 Energy landscape schematic.

9.2 EDC (TODO: full name)

9.3 Unfolding force velocity dependence for different cantilevers.

9.4 Unfolding force loading rate dependence simulations for different cantilevers.

9.5 Unfolding force peak index dependence simulations for different cantilevers.

9.6 I27, the immunoglobulin-like domain 27 from human Titin (PDB ID: 1TIT). Figure generated with PyMol.

9.7 Pulling speed dependence of I27 for different cantilever stiffnesses. The listed stiffnesses are averages across several individual cantilevers and calibrations. Each box is the average of some number of unfolding events, and the box area is proportional to that number. There are 82 unfolding events for the stiff cantilevers and 274 for the soft cantilevers.

9.8 Loading rate.
A.1 Integral contour $C$ enclosing the upper half of the complex plane. If the integrand $f(z)$ goes to zero “quickly enough” as the radius of $C$ approaches infinity, then the only contribution comes from integration along the real axis (see text for details). . . . . . 51
Abstract
Temperature and cantilever dependent protein unfolding
William Trevor King
Guoliang Yang, Ph.D.
Chapter 1: Introduction

In biological systems the most important molecules, such as proteins, nucleic acids, and polysaccharides, are all polymers. Understanding the properties and functions of these polymeric molecules is crucial in understanding the molecular mechanisms behind structures and processes in cells. The large size of these molecules imposes certain limitations on the information attainable from bulk measurements, because the macromolecules in a population can have diverse conformations and behaviors. Bulk measurements average over these differences, producing excellent statistics for the mean, but making it difficult to understand the variation. The individualized, and sometimes rare, behaviors of macromolecules can have important implications for their functions inside the cell. Single molecule techniques, in which the macromolecules are studied one at a time, allow direct access to the variation within the population without averaging. This provides important and complementary information about the functional mechanisms of several biological systems.

An organism’s genetic code is stored in DNA in the cell nucleus. DNA sequencing is a fairly well developed field, with fundamental work such as the Human Genome Project seeing major development in the early 2000s. It is estimated that human genetic information contains approximately 25,000 genes, each encoding a protein. Knowing the amino acid sequence for a particular protein, however, does not immediately shed light on the protein’s role in the body, or even the protein’s probable conformation. Indeed, a protein’s conformation is often vitally important in executing its biological tasks (Fig. 1.1). Unfortunately both predicting stable conformations of a given amino acid sequence and the inverse problem of finding sequences that form a given conformation have proven remarkably difficult.

Folding a protein via a brute force sampling of all possible conformations is impossibly inefficient, due to the exponential scaling of possible conformations with protein length, as outlined by Levinthal. This has lead to a succession of models explaining the folding mechanism. For a number of years, the “pathway” model of protein folding enjoyed popularity (Fig. 1.2a). More recently, the “landscape” or “funnel” model has come to the fore (Fig. 1.2b).

Single molecule techniques provide an opportunity to study protein folding and unfolding at the level of a single molecule, where the distinction between the pathway model and funnel model is clearer. These techniques include optical measurements, i.e., single molecule fluorescence microscopy and spectroscopy, and mechanical manipulations of individual macromolecules, i.e., force microscopy and spectroscopy using atomic force microscopes (AFMs), laser tweezers, magnetic

Figure 1.1: Complex of biotin (red) and a streptavidin tetramer (green) (PDB ID: 1SWE). The correct streptavidin conformation creates the biotin-specific binding pockets. Biotin-streptavidin is a model ligand-receptor pair isolated from the bacterium Streptomyces avidinii. Streptavidin binds to cell surfaces, and bound biotin increases streptavidin’s cell-binding affinity. Figure generated with PyMol.
Figure 1.2: (a) A “double T” example of the pathway model of protein folding, in which the protein proceeds from the native state $N$ to the unfolded state $U$ via a series of metastable transition states $I_1$ and $I_2$ with two “dead end” states $I_1^X$ and $I_2^X$. Adapted from Bédard et al.\textsuperscript{4}. (b) The landscape model of protein folding, in which the protein diffuses through a multi-dimensional free energy landscape. Separate folding attempts may take many distinct routes through this landscape on the way to the folded state. Reproduced from Dill and Chan\textsuperscript{5}.

Figure 1.3: Operating principle for an Atomic Force Microscope. A sharp tip integrated at the end of a cantilever interacts with the sample. Cantilever bending is measured by a laser reflected off the cantilever and incident on a position sensitive photodetector.

tweezers\textsuperscript{18}, biomembrane force probes\textsuperscript{19}, and centrifugal microscopes\textsuperscript{20}. Of these mechanical manipulation methods, AFM is the most widely used due to the availability of user-friendly commercial instruments. AFM has been employed on several types of biological macromolecules, mechanically unfolding proteins\textsuperscript{21} and forcing structural transitions in DNA\textsuperscript{22} and polysaccharides\textsuperscript{23}. An AFM uses a sharp tip integrated at the end of a cantilever to interact with the sample. Cantilever bending is measured by a laser reflected off the cantilever and incident on a position sensitive photodetector (Fig. 1.3). When the bending force constant of the cantilever is known\textsuperscript{24}, the force applied to the sample can be calculated.

The forces that can be applied and measured with an AFM range from tens of piconewtons to hundreds of nanonewtons. The investigation of the unfolding and refolding processes of individual protein molecules by the AFM is feasible because many globular proteins unfold under external forces in this range. Since elucidating the mechanism of protein folding is currently one of the most important problems in biological sciences, the potential of the AFM for revealing significant and unique information about protein folding has stimulated much effort in both experimental and theoretical research.
1.1 Mechanical unfolding experiments

A protein polymer is tethered between two surfaces: a flat substrate and an AFM tip. The polymer is stretched by increasing the separation between the two surfaces (Fig. 1.4a). The most common mode is the constant speed experiment in which the substrate surface is moved away from the tip at a uniform rate. The tethering surfaces, i.e., the AFM tip and the substrate, have much larger radii of curvature than the dimensions of single domain globular proteins that are normally used for folding studies. This causes difficulties in manipulating individual protein molecules because nonspecific interactions between the AFM tip and the substrate may be stronger than the forces required to unfold the protein when the surfaces are a few nanometers apart. To circumvent these difficulties, globular protein molecules are linked into polymers, which are then used in the AFM studies.

When such a polymer is pulled from its ends, each protein molecule feels the externally applied force, which increases the probability of unfolding by reducing the free energy barrier between the native and unfolded states. The unfolding of one molecule in the polymer causes a sudden lengthening of the polymer chain, which reduces the force on each protein molecule and prevents another unfolding event from occurring immediately. The force versus extension relationship, or force curve, shows a typical sawtooth pattern (Fig. 1.4b), where each peak corresponds to the unfolding of a single protein domain in the polymer. Therefore, the individual unfolding events are separated from each other in space and time, allowing single molecule resolution despite the use of multi-domain test proteins.

1.2 Thesis Outline

Chapter 2 of this thesis discusses the theory of protein unfolding for single domains. Chapter 3 discusses linker tension modeling. Chapter 4 pulls Chapters 2 and 3 together to discuss the theory of mechanical unfolding experiments. This theory makes straightforward analysis of unfolding results difficult, so Chapter 5 presents a Monte Carlo simulation approach to fitting unfolding parameters, and Chapter 6 presents the contour-length space analysis for converting force curves to unfolding pathway fingerprints. Chapter 7 wraps up the theory section by extending the analysis in Chapters 2 and 4 to multiple temperatures.

Chapter 8 describes our experimental apparatus and methods, as well as calibration procedures. With both the theory and procedure taken care of, Chapters 9 and 10 present and analyze AFM cantilever- and temperature-dependent unfolding behavior of the immunoglobulin-like domain 27 from human Titin (I27).

We close with Chapter 11, which presents our conclusions and discusses possible directions for future work.
Figure 1.4: (a) Schematic of the experimental setup for mechanical unfolding of proteins using an AFM (not to scale). An experiment starts with the tip in contact with the substrate surface, which is then moved away from the tip at a constant speed. \( x_t \) is the distance traveled by the substrate, \( x_c \) is the cantilever deflection, \( x_u \) is the extension of the unfolded polymer, and \( x_f = x_{f1} + x_{f2} \) is the extension of the folded polymer. (b) An experimental force curve from stretching a ubiquitin polymer with the rising parts of the peaks fitted to the WLC model (Section 3.1.1). The pulling speed used was 1 \( \mu \text{m/s} \). The irregular features at the beginning of the curve are due to nonspecific interactions between the tip and the substrate surface, and the last high force peak is caused by the detachment of the polymer from the tip or the substrate surface. Note that the abscissa is the extension of the protein chain \( x_t - x_c \).
Chapter 2: Unfolding Theory
Chapter 3: Chain Tension

3.1 Polymer Models

3.1.1 Wormlike chains

The unfolded forms of many domains can be modeled as Worm-Like Chains (WLCs)\textsuperscript{25,26}, which treats the unfolded polymer as an elastic rod of persistence length $p$ and contour length $L$. The relationship between tension $F$ and extension (end-to-end distance) $x$ is given to within XX% by Bustamante’s interpolation formula\textsuperscript{25,26}.

$$F_{\text{WLC}}(x, p, L) = \frac{kT}{p} \left[ \frac{1}{4} \left( \frac{1}{(1 - x/L)^2} - 1 \right) + \frac{x}{L} \right],$$  \hspace{1cm} (3.1)

where $p$ is the persistence length.

For a chain with $N_u$ unfolded domains sharing a persistence length $p_u$ and per-domain contour lengths $L_{u1}$, the tension of the WLC is determined by summing the contour lengths

$$F(x, p_u, L_{u1}, N_u) = F_{\text{WLC}}(x, p_u, N_u L_{u1})$$  \hspace{1cm} (3.2)

3.1.2 Freely-jointed chains

3.2 Folded domain tension

The unfolded polypeptide chain has been shown to follow the WLC model quite well (Section 3.1.1), though other polymer models, such as the Freely-Jointed Chain (FJC)\textsuperscript{27} (Section 3.1.2), can be used to fit the force-extension relationship\textsuperscript{28}. A short chain of folded proteins, however, cannot be described well by polymer models. Several studies have used WLC and FJC models to fit the elastic properties of the modular protein titin\textsuperscript{29,30}, but native titin contains hundreds of folded and unfolded domains. For the short protein polymers common in mechanical unfolding experiments, the cantilever dominates the elasticity of the polymer-cantilever system before any protein molecules unfold. After the first unfolding event occurs, the unfolded portion of the chain is already longer and softer than the sum of all the remaining folded domains, and dominates the elasticity of the whole chain. Therefore, the details of the tension model chosen for the folded domains has negligible effect on the unfolding forces, which was also suggested by Staple et al.\textsuperscript{31}. Force curves simulated using different models to describe the folded domains yielded almost identical unfolding force distributions (data not shown, TODO: show data).
Chapter 4: Theoretical unfolding force distributions

4.1 Overview

For testing the sawsim program, we need a few analytic solutions to unfolding distributions. We will start out discussing single-domain proteins under constant loading, and make some comments about multi-domain proteins and variable loading if we can make any progress in that direction. This note also functions as my mini-review article on unfolding theory, since I haven’t been able to find an official one.

4.2 Review of current research

Rief and Grubmüller\textsuperscript{32} provide a general review of force spectroscopy with a short section on protein unfolding. There’s not all that much information here, but it’s a good place to go to get a big-picture overview before diving into the more technical papers.

There are two main approaches to modeling protein domain unfolding under tension: Bell’s and Kramers\textsuperscript{33–35}. Bell introduced his model in the context of cell adhesion\textsuperscript{36}, but it has been widely used to model mechanical unfolding in proteins\textsuperscript{21,33,37} due to its simplicity and ease of use\textsuperscript{35}. Kramers introduced his theory in the context of thermally activated barrier crossings, which is how we use it here.

There is an excellent review of Kramers’ theory in Hänggi et al.\textsuperscript{38}. The bell model is generally considered too elementary to be worth a detailed review in this context, and yet I had trouble finding explicit probability densities that matched my own in Eqn. 4.20. Properties of the Bell model receive more coverage under the name of the older and equivalent Gompertz distribution\textsuperscript{39–41}. A warning about the “Gompertz” model is in order, because there seem to be at least two unfolding/dying rate formulas that go by that name. Compare, for example, Braverman and Mamdani\textsuperscript{42} Eqn. 5 and Juckett and Rosenberg\textsuperscript{43} Fig. 2.

4.2.1 Who’s who

The field of mechanical protein unfolding is developing along three main branches. Some groups are predominantly theoretical,

- Evans, University of British Columbia (Emeritus)
- Thirumalai, University of Maryland
  http://www.marylandbiophysics.umd.edu/
- Onuchic, University of California, San Diego
  http://guara.ucsd.edu/
- Hyeon, Chung-Ang University (Onuchic postdoc, Thirumalai postdoc?)
  http://physics.chem.cau.ac.kr/
- Dietz (Rief grad)
  http://www.hd-web.de/
- Hummer and Szabo, National Institute of Diabetes and Digestive and Kidney Diseases
and the experimentalists are usually either AFM based

- Rief, Technischen Universitt Mnchen
  [http://cell.e22.physik.tu-muenchen.de/gruppematthias/index.html](http://cell.e22.physik.tu-muenchen.de/gruppematthias/index.html)

- Fernandez, Columbia University
  [http://www.columbia.edu/cu/biology/faculty/fernandez/FernandezLabWebsite/](http://www.columbia.edu/cu/biology/faculty/fernandez/FernandezLabWebsite/)

- Oberhauser, University of Texas Medical Branch (Fernandez postdoc)
  [http://www.utmb.edu/ncb/Faculty/OberhauserAndres.html](http://www.utmb.edu/ncb/Faculty/OberhauserAndres.html)

- Marszalek, Duke University (Fernandez postdoc)
  [http://smfs.pratt.duke.edu/homepage/lab.htm](http://smfs.pratt.duke.edu/homepage/lab.htm)

- Guoliang Yang, Drexel University

- Wojcikiewicz, University of Miami
  [http://chroma.med.miami.edu/physiol/faculty-wojcikiewicz_e.htm](http://chroma.med.miami.edu/physiol/faculty-wojcikiewicz_e.htm)

or laser-tweezers based

- Bustamante, University of California, Berkley
  [http://alice.berkeley.edu/](http://alice.berkeley.edu/)

- Forde, Simon Fraser University

### 4.2.2 Evolution of unfolding modeling

Evans introduced the saddle-point Kramers’ approximation in a protein unfolding context 1997 (Evans and Ritchie, Eqn. 3). However, early work on mechanical unfolding focused on the simper Bell model. In the early ‘00’s, the saddle-point/steepest-descent approximation to Kramer’s model (Hänggi et al., Eqn. 4.56c) was introduced into our field. By the mid ‘00’s, the full-blown double-integral form of Kramer’s model (Hänggi et al., Eqn. 4.56b) was in use.

There have been some tangential attempts towards even fancier models. Dudko et al. attempted to reduce the restrictions of the single-unfolding-path model. Hyeon and Thirumalai attempted to measure the local roughness using temperature dependent unfolding.

### 4.2.3 History of simulations

Early molecular dynamics (MD) work on receptor-ligand breakage by Grubmüller 1996 and Izrailev 1997 (according to Evans 1997). Evans and Ritchie introduce a smart Monte Carlo (SMC) Kramers’ simulation.

### 4.2.4 History of experimental AFM unfolding experiments

- Rief et al.

### 4.2.5 History of experimental laser tweezer unfolding experiments

- Izrailev et al.
4.3 Single-domain proteins under constant loading

Let $x$ be the end to end distance of the protein, $t$ be the time since loading began, $F$ be tension applied to the protein, $P$ be the surviving population of folded proteins. Make the definitions

\[ v \equiv \frac{dx}{dt} \quad \text{the pulling velocity} \quad (4.1) \]

\[ k \equiv \frac{dF}{dx} \quad \text{the loading spring constant} \quad (4.2) \]

\[ P_0 \equiv P(t = 0) \quad \text{the initial number of folded proteins} \quad (4.3) \]

\[ D \equiv P_0 - P \quad \text{the number of dead (unfolded) proteins} \quad (4.4) \]

\[ \kappa \equiv -\frac{1}{P} \frac{dP}{dt} \quad \text{the unfolding rate} \quad (4.5) \]

The proteins are under constant loading because

\[ \frac{dF}{dt} = \frac{dF}{dx} \frac{dx}{dt} = kv, \quad (4.6) \]

a constant, since both $k$ and $v$ are constant (Evans and Ritchie\textsuperscript{44} in the text on the first page, Dudko et al.\textsuperscript{34} in the text just before Eqn. 4).

The instantaneous likelihood of a protein unfolding is given by $\frac{dD}{dF}$, and the unfolding histogram is merely this function discretized over a bin of width $W$ (This is similar to Dudko et al.\textsuperscript{34} Eqn. 2, remembering that $\ddot{F} = kv$, that their probability density is not a histogram ($W = 1$), and that their pdf is normalized to $N = 1$).

\[ h(F) = \frac{\frac{dF}{dbin}}{\frac{dD}{dF}} \cdot \frac{\frac{dF}{dbin}}{\frac{dD}{dF}} = W \frac{dD}{dF} = -W \frac{dP}{dF} = -W \frac{dP}{dt} \frac{dt}{dF} = \frac{W}{vk} P \kappa \quad (4.7) \]

Solving for theoretical histograms is merely a question of taking your chosen $\kappa$, solving for $P(f)$, and plugging into Eqn. 4.7. We can also make a bit of progress solving for $P$ in terms of $\kappa$ as follows:

\[ \kappa \equiv -\frac{1}{P} \frac{dP}{dt} \quad (4.8) \]

\[ -\kappa dt \frac{dF}{dt} = \frac{dP}{P} \quad (4.9) \]

\[ -\frac{1}{kv} \int \kappa dF = \ln(P) + c \quad (4.10) \]

\[ P = C \exp \left( -\frac{1}{kv} \int \kappa dF \right), \quad (4.11) \]

where $c \equiv \ln(C)$ is a constant of integration scaling $P$.

4.3.1 Constant unfolding rate

In the extremely weak tension regime, the proteins’ unfolding rate is independent of tension, we have

\[ P = C \exp \left( -\frac{1}{kv} \int \kappa dF \right) = C \exp \left( -\frac{1}{kv} \kappa F \right) = C \exp \left( -\frac{\kappa F}{kv} \right) \quad (4.12) \]

\[ P(0) \equiv P_0 = C \exp(0) = C \quad (4.13) \]

\[ h(F) = \frac{W}{vk} P \kappa = \frac{W \kappa P_0}{vk} \exp \left( -\frac{\kappa F}{kv} \right) \quad (4.14) \]
So, a constant unfolding-rate/hazard-function gives exponential decay. Not the most earth shattering result, but it’s a comforting first step, and it does show explicitly the dependence in terms of the various unfolding-specific parameters.

### 4.3.2 Bell model

Stepping up the intensity a bit, we come to Bell’s model for unfolding (Hummer and Szabo\(^35\) Eqn. 1 and the first paragraph of Dudko et al.\(^34\) and Dudko et al.\(^48\)).

\[
\kappa = \kappa_0 \cdot \exp \left( \frac{Fdx}{k_BT} \right) = \kappa_0 \cdot \exp(aF), \tag{4.15}
\]

where we’ve defined \(a \equiv dx/k_BT\) to bundle some constants together. The unfolding histogram is then given by

\[
P = C \exp \left( \frac{-1}{kv} \int \kappa dF \right) = C \exp \left[ \frac{-1}{kv} \frac{\kappa_0}{a} \exp(aF) \right] = C \exp \left[ \frac{-\kappa_0}{akv} \exp(aF) \right], \tag{4.16}
\]

\[
P(0) \equiv P_0 = C \exp \left( \frac{-\kappa_0}{akv} \right) \tag{4.17}
\]

\[
P = P_0 \exp \left( \frac{\kappa_0}{akv} \left[ 1 - \exp(aF) \right] \right) \tag{4.18}
\]

\[
h(F) = \frac{W}{vk} P = \frac{W}{vk} P_0 \exp \left\{ \frac{\kappa_0}{akv} \left[ 1 - \exp(aF) \right] \right\} \kappa_0 \exp(aF) = \frac{W \kappa_0 P_0}{vk} \exp \left\{ aF + \frac{\kappa_0}{akv} \left[ 1 - \exp(aF) \right] \right\}. \tag{4.19}
\]

The \(F\) dependent behavior reduces to

\[
h(F) \propto \exp \left[ aF - b \exp(aF) \right], \tag{4.20}
\]

where \(b \equiv \kappa_0/akv \equiv \kappa_0 k_BT/\nu kdz\) is another constant rephrasing.

This looks similar to the Gompertz / Gumbel / Fisher-Tippett distribution, where

\[
p(x) \propto z \exp(-z) \tag{4.22}
\]

\[
z \equiv \exp \left( -\frac{x - \mu}{\beta} \right), \tag{4.23}
\]

but we have

\[
p(x) \propto z \exp(-bz). \tag{4.24}
\]

Strangely, the Gumbel distribution is supposed to derive from an exponentially increasing hazard function, which is where we started for our derivation. I haven’t been able to find a good explanation of this discrepancy yet, but I have found a source that echos my result (Wu et al.\(^41\) Eqn. 1). TODO: compare Wu et al.\(^41\) with my successful derivation in Section 5.3.2.

Oh wait, we can do this:

\[
p(x) \propto z \exp(-bz) = \frac{1}{b} z' \exp(-z') \propto z' \exp(-z'), \tag{4.25}
\]

with \(z' \equiv bz\). I feel silly... From http://mathworld.wolfram.com/GumbelDistribution.html, the mean of the Gumbel probability density

\[
P(x) = \frac{1}{\beta} \exp \left[ \frac{x - \alpha}{\beta} - \exp \left( \frac{x - \alpha}{\beta} \right) \right] \tag{4.26}
\]

is given by \(\mu = \alpha - \gamma\beta\), and the variance is \(\sigma^2 = \frac{1}{6}\pi^2\beta^2\), where \(\gamma = 0.57721566 \ldots\) is the Euler-
Mascheroni constant. Selecting $\beta = 1/a = k_B T/dx$, $\alpha = -\beta \ln(\kappa \beta/\kappa v)$, and $F = x$ we have

$$P(F) = \frac{1}{\beta} \exp \left[ \frac{F + \beta \ln(\kappa \beta/\kappa v)}{\beta} - \exp \left( \frac{F + \beta \ln(\kappa \beta/\kappa v)}{\beta} \right) \right]$$

(4.27)

$$= \frac{1}{\beta} \exp(F/\beta) \exp[\ln(\kappa \beta/\kappa v)] \exp \{ - \exp(F/\beta) \exp[\ln(\kappa \beta/\kappa v)] \}$$

(4.28)

$$= \frac{1}{\beta} \frac{\kappa \beta}{\kappa v} \exp(F/\beta) \exp [ - \kappa \beta/\kappa v \exp(F/\beta) ]$$

(4.29)

$$= \frac{\kappa}{\kappa v} \exp(F/\beta - \kappa \beta/\kappa v \exp(F/\beta))$$

(4.30)

$$= \frac{\kappa}{\kappa v} \exp(aF - \kappa/akv \exp(aF))$$

(4.32)

$$= \frac{\kappa}{\kappa v} \exp(aF - b \exp(aF)) \propto h(F).$$

(4.33)

So our unfolding force histogram for a single Bell domain under constant loading does indeed follow the Gumbel distribution.

### 4.3.3 Saddle-point Kramers’ model

For the saddle-point approximation for Kramers’ model for unfolding (Evans and Ritchie Eqn. 3, Hänggi et al. Eqn. 4.56c, van Kampen Eqn. XIII.2.2).

$$\kappa = \frac{D}{l_b l_s} \exp \left( -\frac{E_b(F)}{k_B T} \right),$$

(4.34)

where $E_b(F)$ is the barrier height under an external force $F$, $D$ is the diffusion constant of the protein conformation along the reaction coordinate, $l_b$ is the characteristic length of the bound state $l_b \equiv 1/\rho_b$, $\rho_b$ is the density of states in the bound state, and $l_s$ is the characteristic length of the transition state $l_s = TODO$

(4.35)

Evans and Ritchie solved this unfolding rate for both inverse power law potentials and cusp potentials.

**Inverse power law potentials**

$$E(x) = \frac{-A}{x^n}$$

(4.36)

(e.g. $n = 6$ for a van der Waals interaction, see Evans and Ritchie in the text on page 1544, in the first paragraph of the section Dissociation under force from an inverse power law attraction). Evans then goes into diffusion constants that depend on the protein’s end to end distance, and I haven’t worked out the math yet. TODO: clean up.

**Cusp potentials**

$$E(x) = \frac{1}{2} \kappa_a \left( \frac{x}{x_0} \right)^2$$

(4.37)

(see Evans and Ritchie in the text on page 1545, in the first paragraph of the section Dissociation under force from a deep harmonic well).
4.4 Double-integral Kramers’ theory

The double-integral form of overdamped Kramers’ theory may be too complex for analytical predictions of unfolding-force histograms. Rather than testing the entire `sawin` simulation (Chapter 5), we will focus on demonstrating that the Kramers’ \( k(F) \) evaluations are working properly. If the Bell modeled histograms check out, that gives reasonable support for the \( k(F) \rightarrow \text{histogram} \) portion of the simulation.

Looking for analytic solutions to Kramers’ \( k(F) \), we find that there are not many available in a closed form. However, we do have analytic solutions for unforced \( k \) for cusp-like and quartic potentials.

4.4.1 Cusp-like potentials

4.4.2 Quartic potentials
Chapter 5: Monte Carlo mechanical unfolding simulation

5.1 Introduction

Much theoretical and computational work has been done in order to extract information about the structural, kinetic, and energetic properties of the protein molecules from the experimental data of force-induced protein unfolding measurements. Steered molecular dynamics simulations\(^\text{50}\), as well as calculations and simulations using lattice\(^\text{51}\) and off-lattice models\(^\text{52,53}\), have provided insights into structural and energetic changes during force-induced protein unfolding. However, these simulations often involve time scales that are orders of magnitude smaller than those of the experiments, and the parameters used in the calculations are often neither experimentally controllable nor measurable (TODO: example parameters of each type). As a result, a Monte Carlo simulation approach based on a simple two-state kinetic model for the protein is usually used to analyze data from mechanical unfolding experiments. A comparison of the force curves measured experimentally and those generated from simulations can yield the unfolding rate constant of the protein in the absence of force as well as the distance from the native state to the transition state along the pulling direction. The Monte Carlo simulation method has been used since the first report of mechanical unfolding experiments using the AFM\(^\text{21,23,37,54–57}\), but these previous implementations are neither fully described nor publicly available.

To fill this gap, I developed the \texttt{sawsim} simulation program\(^\text{58}\). In this chapter, I provide a detailed description of the simulation procedure, including the theories, approximations, and assumptions involved. I also explain the procedure for extracting kinetic properties of the protein from experimental data and introduce a quantitative measure of fit quality between simulation and experimental results. In addition, the effects of various experimental parameters on force curve appearance are demonstrated, and the errors associated with different methods of data pooling are discussed. These results should be useful in future experimental design, artifact identification, and data analysis for single molecule mechanical unfolding experiments.

5.2 Methods

In simulating the mechanical unfolding process, a force curve is generated by calculating the amount of cantilever bending as the substrate surface moves away from the tip. The cantilever bending is obtained by balancing the tension in the protein polymer and the Hookean force of the bent cantilever. The unfolding probability of the protein molecules in the polymer is then calculated for that tension, and whether an unfolding event occurs is determined according to a Monte Carlo method. The simulation was implemented in C\(^1\).

5.2.1 Generating force curves

The fundamental abstraction of the simulation is the “domain”, which represents a discrete chunk of the flexible chain between the substrate and the cantilever holder. Each of these domains is assigned a particular state; for example, the domain representing the cantilever is assigned to the “cantilever” state, and the domains representing protein molecules are assigned to either the “folded” or the “unfolded” state. When balancing the tension along the chain, we assume that the spatial order of domains along the chain is irrelevant\(^\text{59}\), and therefore, the domains can be rearranged and grouped by state. To determine the tension in the chain and the amount of cantilever bending when

\(^1\text{Source code available at http://www.physics.drexel.edu/~wking/sawsim/}\)
\( n \) states are populated, a system of \( n + 1 \) equations with \( n + 1 \) unknowns must be solved
\[
F_i(x_i) = F_t, \quad \sum_i x_i = x_t, \tag{5.1, 5.2}
\]
where \( F \) are tensions, \( x \) are extensions, and the subscripts \( i \) and \( t \) represent a particular state group and the total chain respectively (Fig. 1.4a). From this \( F(x_i) \) may be computed using any multi-dimensional root-finding algorithm.

Inside this framework, we chose a particular extension model \( F_i(x_i) \) for each domain state. Cantilever elasticity is described by Hooke’s law, which gives
\[
F = \kappa_c x_c, \tag{5.3}
\]
where \( \kappa_c \) is the bending spring constant and \( x_c \) is the deflection of the cantilever (Fig. 1.4a). Unfolded domains are modeled as WLCs (Section 3.1.1).

The chain of \( N_f \) folded domains is modeled as a string, free to assume any extension up to some fixed contour length \( L_f = N_f L_{f1} \)
\[
F = \begin{cases} 0 & \text{if } x_f < L_f, \\ \infty & \text{if } x_f > L_f, \end{cases} \tag{5.4}
\]
where \( L_{f1} \) is the separation of the two linking points of a folded domain, and \( x_f \) is the end-to-end length of the chain of folded domains. In this model, any non-zero tension will fully extend these folded domains. As discussed in Section 3.2, the contribution of the folded domains to the elastic behavior of the polymer-cantilever system is relatively insignificant.

In the simulation, the protein polymer is assumed to be stretched in the direction perpendicular to the substrate surface, which is a good approximation in most experimental situations, because the unfolded length of a protein molecule is much larger than that of the folded form. Therefore, after one molecule unfolds, the polymer becomes much longer and the angle between the polymer and the surface approaches 90 degrees\(^\text{60}\). The joints between domain groups are assumed to lie along a line between the surface tether point and the position of the tip (Eq. (5.2)). The effects of this assumption are also minimized due to greater length of the unfolded domain. Finally, the interactions between different parts of the polymer and between the chain and the surface (except at the tethering points) are not considered. This is reasonable since these interactions should not make substantial contributions to the force curve at the force levels of interest, where the polymer is in a relatively extended conformation.

Consider an experiment of pulling a polymer with \( N \) identical protein molecules at a constant speed. At the start of an experiment, the chain is unstretched \((x_t = 0)\), which means all the domains are unstretched, the cantilever is undeflected, and the tip is in contact with the surface. There is one domain in the cantilever state, \( N \) in the folded state, and none in the unfolded state. As the surface moves away from the tip at a constant speed \( v \), the chain becomes more extended (Fig. 1.4a), such that
\[
x_t = \sum_i x_i = vt. \tag{5.5}
\]
The simulation assumes that the pulling takes discrete steps in space and treats \( x_t \) as constant over the duration of one time step \( \Delta t \). Because of the adaptive time steps discussed in Section 5.2.3, the space steps \( \Delta x_t = v \Delta t \) may have different sizes, but each step will be “small”. At each step, the total extension is calculated using Eq. (5.5), and the tension \( F(x_t = vt) \) is determined by numerically solving Eqs. (5.1) and (5.2) using the models Eqs. (3.1), (5.3) and (5.4) for known values of the parameters in the various states \((N_u, N_f, v, \kappa_c, L_{f1}, L_{u1}, p_u)\). When one of the molecules in the polymer unfolds (Section 5.2.2), there will be one domain in the unfolded state and \( N - 1 \) in the folded state. In the next step, a newly balanced tension between the cantilever and the polymer is determined by solving for \( F(x_t) \) as discussed above, but with the total extension \( x_t \) incremented by

5.2 Methods
\( \Delta t \) and the new unfolded contour length \( L_u \) and folded contour length \((N - 1)L_f \). The sudden lengthening of the polymer chain results in a corresponding abrupt drop in the force, leading to the formation of one sawtooth in the force curve. As the pulling continues and more domains unfold, force curves with a series of sawteeth are generated (Fig. 5.2a).

The tension calculation assumes an equilibrated chain, so consideration must be given to the chain’s relaxation time, which should be short compared to the loading timescale. The relaxation time for a WLC is given by

\[
\tau \approx \eta \frac{k_B T p}{F^2},
\]

where \( \eta \) is the dynamic viscosity, \( F \) is the tension, and \( p \) is the persistence length. For forces greater than 1 pN, with \( \eta_{\text{water}}/k_B T = 2.45 \cdot 10^{-10} \text{ s/nm}^3 \), \( \tau \leq 2 \text{ ns} \) for the protein polymer used in the simulation. Therefore, the polymer chain is equilibrated almost instantaneously within a time step, which is on the order of tens of microseconds. The relaxation time of the cantilever can be determined by measuring the cantilever deflection induced by liquid motion and fitting the time dependence of the deflection to an exponential function. For a 200 \( \mu \)m rectangular cantilever with a bending spring constant of 20 pN/nm, the measured relaxation time in water is \( \sim 50 \mu \text{s} \) (data not shown). TODO: show data. This relatively large relaxation time constant makes the cantilever act as a low-pass filter and also causes a lag in the force measurement.

5.2.2 Unfolding protein molecules by force

According to the theory developed by Bell and extended by Evans and Ritchie, an external stretching force \( F \) increases the unfolding rate constant of a protein molecule

\[
k_u = k_{u0} \exp \left( \frac{F \Delta x_u}{k_B T} \right),
\]

where \( k_{u0} \) is the unfolding rate in the absence of an external force, and \( \Delta x_u \) is the distance between the native state and the transition state along the pulling direction. The probability for a protein molecule to unfold under an applied force is

\[
P_1 = k_u \Delta t,
\]

where \( \Delta t \) is the time duration for each pulling step, over which \( F \) is constant. This expression is accurate for \( P_1 \ll 1 \). From the binomial distribution, the probability of at least one of a group of \( N_f \) identical domains to unfold in a given time step is

\[
P = 1 - (1 - P_1)^{N_f} \approx N_f P_1,
\]

where the approximation is valid when \( N_f P_1 \ll 1 \).

To determine if an unfolding event occurs in a particular time step, the probability calculated using Eq. (5.9) is compared with a randomly generated number uniformly distributed between 0 and 1 (Fig. 5.1). If \( P \) is bigger than the random number, a domain unfolds, changing the population of each tension state, and a new balance between the polymer and the cantilever is determined. If no unfolding event occurs the pulling continues and the unfolding probability is calculated again in the next step at a higher force. When all the molecules in the polymer have unfolded, the pulling continues until a pre-determined force level is reached, where the polymer is assumed to detach from one of the tethering surfaces. The cantilever deflection becomes zero after this point.

Although the Bell model (Eq. (5.7)) is the most widely used unfolding model due to its simplicity and its applicability to various biopolymers, other theoretical models have been proposed to interpret mechanical unfolding data. For example, Schlierf and Rief used the mechanical unfolding data of the protein ddFLN4 to demonstrate that Kramers’ diffusion model fit the measured unfolding force data better than the Bell model for proteins with broad free energy barriers. For proteins with relatively narrow folded and transition states, the Bell model provides a good approximation.
Figure 5.1: Once the unfolding probability has been calculated, we need to determine whether or not a domain should unfold. We do this by generating a random number, and comparing that number to the unfolding probability $P$. The random number determines which of the possible paths we should follow for the current simulation. Such “statistical sampling” is the hallmark of the Monte Carlo approach. This cartoon translates the idea into the more familiar doors (possible paths) and dice (random numbers).

5.2.3 Choosing the simulation time steps

The demands on the time step vary throughout a simulated pull due to the non-linear elasticity of the polymer. Within a specified time duration (or pulling distance), the force change is small at low force levels and large at high force levels. To be efficient, the simulation algorithm adapts the time step to keep the time steps large where large time steps have little effect, while shrinking the time step where smaller steps are necessary.

Within each time step, the total chain extension $x_t$ is treated as a constant and a force balance is reached very quickly among the various domains (see Section 5.2.1 for equilibration timescales). This force is used to determine the unfolding probability (Eqs. (5.8) and (5.9)), which determines the domain state populations in the next time step. Therefore, the chain tension must not change appreciably over the course of the time step ($\Delta F < 1 \text{pN}$), and the unfolding probability is only calculated once for the entire step. The time step must also be short enough that the probability of unfolding in a single time step is low ($P < 10^{-3}$). Besides ensuring that the approximations made in Eqs. (5.8) and (5.9) are valid, this restriction makes time steps which should have multiple unfoldings in a single time step highly-unlikely. Experimentally measured unfolding are temporally separated, because the unfolding transition is characterized by multiple, Markovian attempts over a large energy barrier, where the probability of crossing the barrier in a single attempt is very low. A successful attempt quickly extends the chain contour length, reducing the tension, dramatically reducing the likelihood of a second escape in that time step. The time step used is recalculated for each step so that both of these criteria are satisfied.

5.3 Results and Discussion

5.3.1 Force curves generated by simulation

Figure 5.2a shows three simulated force curves from pulling a polymer composed of eight identical protein molecules using parameters from typical experimental settings. The order of the peaks in the force curves reflects the temporal sequence of the unfolding events instead of the positions of the protein molecules in the polymer. As observed experimentally (Fig. 1.4b), the forces at which identical protein molecules unfold fluctuate, revealing the stochastic nature of protein unfolding.
since no instrumental noise is included in the simulation. Figure 5.2b shows the distribution of the unfolding forces, i.e., the highest force in each peak (except the last peak in a force curve), from a total of 400 force curves (3200 force values). The unfolding forces have an average of 281 pN with a standard deviation of 25 pN.

5.3.2 Dependence of the unfolding force on the unfolding order and polymer length

Analysis of the mechanical unfolding data is complicated by the dependence of the average unfolding force on the unfolding order due to the serial linkage of the molecules. Under an external stretching force $F$, the probability of some domain unfolding in a polymer with $N_f$ folded domains is $N_f P_f$ (Eq. (5.9)), which is higher than the unfolding probability for a single molecule $P_f$. Consequently, the average unfolding force is lower for the earlier unfolding events when $N_f$ is larger, and the force should increase as more and more molecules become unfolded. However, there is a competing factor that opposes this trend. As the protein molecules unfold, the chain becomes softer and the force loading rate becomes lower when the pulling speed is constant, leading to a decrease in the unfolding force. The dependence of the average unfolding force on the unfolding order is the result of these two opposing effects. Figure 5.3 shows the dependence of the average unfolding force on the unfolding force peak order (the temporal order of unfolding events) for four polymers with 4, 8, 12, and 16 identical protein molecules. The effect of polymer chain softening dominates the initial unfolding events, and the average unfolding force decreases as more molecules unfold. After several molecules have unfolded, the softening for each additional unfolding event becomes less significant, the change in unfolding probability becomes dominant, and the unfolding force increases upon each subsequent unfolding event.

We validate this explanation by calculating the unfolding force probability distribution’s depending on the two competing factors. The rate of unfolding events with respect to force is

$$r_u = \frac{dN_f}{dF} = -\frac{dN_f/dt}{dF/dt} = \frac{N_f k_u}{\kappa v} \exp \left( \frac{F\Delta x_u}{k_B T} \right) = \frac{1}{\rho} \exp \left( \frac{F - \alpha}{\rho} \right),$$

(5.10)

where $N_f$ is the number of folded domain, $\kappa = (1/\kappa_c + N_u/\kappa_{WLC})^{-1}$ is the spring constant of the cantilever-polymer system ($\kappa_{WLC}$ is the effective spring constant of one unfolded domain, assumed constant for a particular polymer/cantilever combination), $\kappa v$ is the force loading rate, and $k_u$ is the unfolding rate constant (Eq. (5.7)). In the last expression, $\rho \equiv k_B T/\Delta x_u$, and $\alpha \equiv -\rho \ln(N_f k_u \rho / \kappa v)$. The event probability density for events with an exponentially increasing likelihood function follows the Gumbel (minimum) probability density, with $\rho$ and $\alpha$ being the scale and location parameters, respectively

$$P(F) = \frac{1}{\rho} \exp \left[ \frac{F - \alpha}{\rho} - \exp \left( \frac{F - \alpha}{\rho} \right) \right].$$

(5.12)

The distribution has a mean $\langle F \rangle = \alpha - \gamma_c \rho$ and a variance $\sigma^2 = \pi^2 \rho^2 / 6$, where $\gamma_c = 0.577\ldots$ is the Euler-Mascheroni constant. Therefore, the unfolding force distribution has a variance $\sigma^2 = (\pi k_B T/\Delta x_u)^2 / 6$, and the average

$$\langle F(i) \rangle = \frac{k_B T}{\Delta x_u} \left[ -\ln \left( \frac{N_f k_u}{\kappa v \Delta x_u} \right) - \gamma_c \right],$$

(5.13)

where $N_f$ and $\kappa$ depend on the domain index $i = N_u$. Curves based on this formula fit the simulated data remarkably well considering the effective WLC stiffness $\kappa_{WLC}$ is the only fitted parameter, and that the actual WLC stiffness is not constant, as we have assumed here, but a non-linear function of $F$.

From Fig. 5.3, it can be seen that the proper way to process data from mechanical unfolding.
Figure 5.2: (a) Three simulated force curves from pulling a polymer of eight identical protein molecules. The simulation was carried out using the parameters: pulling speed $v = 1 \, \mu m/s$, cantilever spring constant $\kappa_c = 50 \, pN/nm$, temperature $T = 300 \, K$, persistence length of unfolded proteins $p_u = 0.40 \, nm$, $\Delta x_u = 0.225 \, nm$, and $k_{u0} = 5 \cdot 10^{-5} \, s^{-1}$. The contour length between the two linking points on a protein molecule is $L_{f1} = 3.7 \, nm$ in the folded form and $L_{u1} = 28.1 \, nm$ in the unfolded form. These parameters are those of ubiquitin molecules connected through the N-C termini. Detachment from the tip or substrate is assumed to occur at a force of 400 pN. In experiments, detachments have been observed to occur at a variety of forces. For clarity, the green and blue curves are offset by 200 and 400 pN respectively. (b) The distribution of the unfolding forces from 400 simulated force curves (3200 data points) such as that shown in (a). The frequency is normalized by the total number of points, i.e., the height of each bin is equal to the number of data points in that bin divided by the total number of data points (3200, for this histogram).
Figure 5.3: The dependence of the unfolding force on the temporal unfolding order for four polymers with 4, 8, 12, and 16 identical protein domains. Each point in the figure is the average of 400 data points. The first point in each curve represents the average of only the first peak in each of the 400 simulated force curves, the second point represents the average of only the second peak, and so on. The solid lines are fits of Eq. (5.13) to the simulated data, with best fit $\kappa_{WLC} = 203, 207, 161,$ and 157 pN/nm, respectively, for lengths 4 through 16. The insets show the force distributions of the first, fourth, and eighth peaks, left to right, for the polymer with eight protein domains. The parameters used for generating the data were the same as those used for Fig. 5.2a, except for the number of domains. The histogram insets were normalized in the same way as in Fig. 5.2b.

Experiments is to group the curves according to the length of the polymer and to perform statistical analysis separately for peaks with the same unfolding order. However, in most experiments, the tethering of the polymer to the AFM tip is by nonspecific adsorption; as a result, the polymers being stretched between the tip and the substrate have various lengths. In addition, the interactions between the tip and the surface often cause irregular features in the beginning of the force curve (Fig. 1.4b), making the identification of the first peak uncertain. Furthermore, it is often difficult to acquire a large amount of data in single molecule experiments. These difficulties make the aforementioned data analysis approach unfeasible for many mechanical unfolding experiments. As a result, the values of all force peaks from polymers of different lengths are often pooled together for statistical analysis. To assess the errors caused by such pooling, simulation data were analyzed using different pooling methods and the results were compared. Figure 5.2b shows that, for a polymer with eight protein molecules, the average unfolding force is 281 pN with a standard deviation of 25 pN when all data is pooled. If only the first peaks in the force curves are analyzed, the average force is 279 pN with a standard deviation of 22 pN. While for the fourth and eighth peaks, the average force are 275 pN and 300 pN, respectively, and the standard deviations are 23 pN and 25 pN, respectively. As expected from the Gumbel distribution, the width of the unfolding force distribution (insets in Fig. 5.3) is only weakly effected by unfolding order, but the average unfolding force can be quite different for the same protein because of the differences in unfolding order and polymer length.
5.3.3 The effect of cantilever force constant

In mechanical unfolding experiments, the ability to observe the unfolding of a single protein molecule depends on the tension drop after an unfolding event such that another molecule does not unfold immediately. The magnitude of this drop is determined by many factors, including the magnitude of the unfolding force, the contour and persistence lengths of the protein polymer, the contour length increase from unfolding, and the stiffness (force constant) of the cantilever. Among these, the effect of the cantilever force constant is particularly interesting because cantilevers with a wide range of force constants are available. In addition, different single molecule manipulation techniques, such as the AFM and laser tweezers, differ mainly in the range of the spring constants of their force transducers. Figure 5.4 shows the simulated force curves from pulling an octamer of protein molecules using cantilevers with different force constants, while other parameters are identical. For this model protein, the appearance of the force curve does not change much until the force constant of the cantilever reaches a certain value ($\kappa_c = 50$ pN/nm). When $\kappa_c$ is lower than this value, the individual unfolding events become less identifiable. In order to observe individual unfolding events, the cantilever needs to have a force constant high enough so that the bending at the maximum force is small in comparison with the contour length increment from the unfolding of a single molecule. Figure 5.4 also shows that the back side of the force peaks becomes more tilted as the cantilever becomes softer. This is due to the fact that the extension (end-to-end distance) of the protein polymer has a large sudden increase as the tension rebalances after an unfolding event.

It should also be mentioned that the contour length increment from each unfolding event is not equal to the distance between adjacent peaks in the force curve because the chain is never fully stretched. This contour length increase can only be obtained by fitting the curve to WLC or other polymer models (Fig. 1.4b).

5.3.4 Determination of $\Delta x_u$ and $k_{u0}$

The zero-force unfolding rate $k_{u0}$ and the distance $\Delta x_u$ from the native state to the transition state are the two kinetic parameters obtainable for mechanical unfolding experiments by matching the simulated data with measured results. Fig. 5.5a shows the dependence of the unfolding force on the pulling speed for different values of $k_{u0}$ and $\Delta x_u$. As expected, the unfolding force increases linearly with the pulling speed in the linear-log plot. While the magnitude of the unfolding forces is affected by both $k_{u0}$ and $\Delta x_u$, the slope of speed dependence is primarily determined by $\Delta x_u$. Figure 5.5b shows that the width of the unfolding force distribution is very sensitive to $\Delta x_u$, as expected from the Gumbel distribution discussed in Section 5.3.2. To obtain the values of $k_{u0}$ and $\Delta x_u$ for the protein, the pulling speed dependence and the distribution of the unfolding forces from simulation, such as those shown in Fig. 5.5a and the insets of Fig. 5.5b, are compared with the experimentally measured results. The values of $k_{u0}$ and $\Delta x_u$ that provide the best match are designated as the parameters describing the protein under study. Since $k_{u0}$ and $\Delta x_u$ affect the unfolding forces differently, the values of both parameters can be determined simultaneously. The data used in plotting Fig. 5.5 includes all force peaks from the simulated force curves because most experimental data is analyzed that way.

In most published literature, determination of the values of $k_{u0}$ and $\Delta x_u$ was mostly done by carrying out simulations using a handful of possible unfolding parameters and selected the best fit by eye. This approach does not allow estimation of uncertainties in the fitting parameters, as shown by Best et al. A more rigorous approach involves quantifying the quality of fit between the experimental and simulated force distributions, allowing the use of a numerical minimization algorithm to pick the best fit parameters. We use the Jensen-Shannon divergence, a measure of the similarity between two probability distributions.

$$D_{JS}(p_e,p_s) = D_{KL}(p_e,p_m) + D_{KL}(p_s,p_m),$$

where $p_e(i)$ and $p_s(i)$ are the the values of the $i^{th}$ bin in the experimental and simulated unfolding
Figure 5.4: Simulated force curves obtained from pulling a polymer with eight protein molecules using cantilevers with different force constants $\kappa_c$. Parameters used in generating these curves are the same as those used in Fig. 5.2, except the cantilever force constant. Successive force curves are offset by 300 pN for clarity.
force histograms, respectively. $D_{KL}$ is the Kullback-Leibler divergence

$$D_{KL}(p, q) = \sum_i p(i) \log_2 \left( \frac{p(i)}{q(i)} \right),$$

(5.15)

where the sum is over all unfolding force histogram bins. $p_m$ is the symmetrized probability distribution

$$p_m(i) \equiv \frac{[p_e(i) + p_s(i)]}{2}.$$  

(5.16)

Figure 5.6 shows the Jensen-Shannon divergence calculated using Eq. (5.14) between an experimental data set and simulation results obtained using a range of values of $k_u 0$ and $\Delta x_u$. There is an order of magnitude range of $k_u 0$ that produce reasonable fits to experimental data (Fig. 5.6), which is consistent with the results Best et al. obtained using a chi-square test. The values of $k_u 0$ and $\Delta x_u$ can be determined to higher precision by using both the pulling speed dependent data and the unfolding force distribution, as well as any relevant information about the protein from other sources.

5.4 Conclusions

We have described the method of performing Monte Carlo simulations based on a simple two-state model for the mechanical unfolding of protein molecules and discussed the complications involved in the simulation procedure. In addition to the extraction of kinetic properties of the protein from mechanical unfolding data, such simulations can help to elucidate the effects of various experimental parameters on the appearance of force curves and to estimate the errors associated with data pooling. To date, the force-induced unfolding approach has been used to investigate several different types of proteins. As the technique is used to study a wider range of proteins, this simple simulation method will be useful for data analysis, experimental design, and artifact identification.

5.4 Conclusions
Figure 5.5: (a) The dependence of the unfolding forces on the pulling speed for three different model protein molecules characterized by the parameters $k_{u0}$ and $\Delta x_u$. The polymer length is eight molecules, and each symbol is the average of 3200 data points. (b) The dependence of standard deviation of the unfolding force distribution on the pulling speed for the simulation data shown in (a), using the same symbols. The insets show the force distribution histograms for the three proteins at the pulling speed of 1 $\mu$m/s. The left, middle and right histograms are for the proteins represented by the top, middle, and bottom lines in (a), respectively.

5.4 Conclusions
Figure 5.6: Fit quality between an experimental data set and simulated data sets obtained using various values of unfolding rate parameters $k_{u0}$ and $\Delta x_u$. The experimental data are from octameric ubiquitin pulled at $1 \mu$m/s, and the other model parameters are the same as those in Fig. 5.2. The best fit parameters are $\Delta x_u = 0.17$ nm and $k_{u0} = 1.2 \cdot 10^{-2}$ s$^{-1}$. The simulation histograms were built from 400 pulls at each parameter pair.
Chapter 6: Contour length space

TODOPuchner et al.\textsuperscript{66}.
Chapter 7: Temperature dependent unfolding theory

7.1 Energy landscape roughness

I’m skeptical about H&T eq. 8 to H&T eq. 9, so I’ll rework as much of their math as I am capable of...

\[ f^* = \frac{k_B T}{\Delta x(f^*)} \left[ \log \left( \frac{r f \Delta x(f^*)}{\nu_D(f^*) e^{-\beta \Delta F_0^i(f^*) k_B T}} \right) + \log \left( 1 + f^* \frac{\Delta x(f^*)'}{\Delta x(f^*)} - \frac{\Delta F_0^i(f^*)'}{\Delta x(f^*)} \right) + \nu_D(f^*)' \cdot \frac{k_B T}{\Delta x(f^*)} \right] + \log (\langle e^{\beta F_1} \rangle)^2 \]  

(H&T eq. 8)

We simplify by dropping the 2nd term (“In obtaining Eq. 9, we have assumed that the second term in Eq. 8 is small.”), and defining \( \alpha \equiv k_B T \), \( \rho \equiv \log \left( \frac{r f \Delta x(f^*)}{\nu_D(f^*) e^{-\beta \Delta F_0^i(f^*) k_B T}} \right) \), and \( e^{\beta \rho} \equiv \langle e^{\beta F_1} \rangle \), yielding

\[ f^* = \frac{\alpha}{\Delta x(f^*)} \left( \rho + \frac{\epsilon^2}{\alpha^2} \right) \]  

(7.1)

We obtain our version of H&T eq. 9 by taking two measurements of equal mode force

\[ 0 = f_1^* - f_2^* \]  

(7.2)

\[ = \frac{1}{\Delta x(f^*)} \left( \alpha_1 \rho_1 + \frac{\epsilon^2}{\alpha_1} - \alpha_2 \rho_2 - \frac{\epsilon^2}{\alpha_2} \right) \]  

(7.3)

\[ \epsilon^2 \left( \frac{1}{\alpha_2} - \frac{1}{\alpha_1} \right) = \alpha_1 \rho_1 - \alpha_2 \rho_2 \]  

(7.4)

\[ \epsilon^2 \cdot \frac{\alpha_1 - \alpha_2}{\alpha_1 \alpha_2} = \text{TODO} \]  

(7.5)

\[ \epsilon^2 = \frac{\alpha_1 \alpha_2}{\alpha_1 - \alpha_2} (\alpha_1 \rho_1 - \alpha_2 \rho_2) \]  

(7.6)

\[ \epsilon^2 = \frac{k_B T_1 k_B T_2}{k_B T_1 - k_B T_2} \left[ k_B T_1 \log \left( \frac{r f_1 \Delta x_1(f^*)}{\nu_D(f^*) e^{-\beta_1 \Delta F_0^i(f^*) k_B T_1}} \right) - k_B T_2 \log \left( \frac{r f_2 \Delta x_2(f^*)}{\nu_D(f^*) e^{-\beta_2 \Delta F_0^i(f^*) k_B T_2}} \right) \right] \]  

(7.7)

Which is different from H&T eq. 9 by the sign in the prefactor, and the replacement \( \nu_D(f^*) \rightarrow k(f^*) \).

\[ \epsilon^2 = \frac{k_B T_1 k_B T_2}{k_B T_2 - k_B T_1} \left[ k_B T_1 \log \left( \frac{r f_1 \Delta x_1(f^*)}{\nu_D(f^*) k_B T_1} \right) - k_B T_2 \log \left( \frac{r f_2 \Delta x_2(f^*)}{\nu_D(f^*) k_B T_2} \right) \right] \]  

(H&T eq. 9)

Alternatively, noting that \( \Delta x(f^*) \) can vary as a function of temperature, we follow Nevo et al.\textsuperscript{67}
in keeping it in. Using $\delta \equiv \Delta x(f^*)$

\[ 0 = f^*_1 - f^*_2 = \alpha_1 \rho_1 + \frac{\epsilon^2}{\delta_1} - \alpha_2 \rho_2 - \frac{\epsilon^2}{\delta_2} \]  \hspace{1cm} (7.8)

\[ \epsilon^2 \left( \frac{1}{\delta_2 \alpha_2} - \frac{1}{\delta_1 \alpha_1} \right) = \frac{\alpha_1 \rho_1}{\delta_1} - \frac{\alpha_2 \rho_2}{\delta_2} \]  \hspace{1cm} (7.9)

\[ \frac{\Delta_1 \alpha_1 - \delta_2 \alpha_2}{\delta_1 \delta_2 \alpha_1 \alpha_2} = \frac{\delta_2 \alpha_1 \rho_1 - \delta_1 \alpha_2 \rho_2}{\delta_1 \delta_2} \]  \hspace{1cm} (7.10)

\[ \epsilon^2 = \frac{\alpha_2 \rho_2}{\delta_1 \alpha_1 - \delta_2 \alpha_2} (\delta_2 \alpha_1 \rho_1 - \delta_1 \alpha_2 \rho_2) \]  \hspace{1cm} (7.11)

\[ \epsilon^2 = \frac{k_B T_1 k_B T_2}{\Delta x_1(f^*) k_B T_1 - \Delta x_2(f^*) k_B T_2} \left[ \Delta x_2(f^*) k_B T_1 \log \left( \frac{r_{f_1} \Delta x_1(f^*)}{k_1(f^*) k_B T_1} \right) - \Delta x_1(f^*) k_B T_2 \log \left( \frac{r_{f_2} \Delta x_2(f^*)}{k_2(f^*) k_B T_2} \right) \right] \]  \hspace{1cm} (7.12)
Chapter 8: Apparatus
Chapter 9: Cantilever dependent unfolding experiments

Understanding a protein’s free energy landscape is important to effectively model protein folding and unfolding behavior. Force spectroscopy has been a useful technique for exploring these free energy landscapes and those of the related field of ligand-receptor kinetics. In force spectroscopy with the atomic force microscope (AFM), it is common practice to use spring constants in the range of 50 pN/nm, but the effect of the cantilever itself on the free energy landscape is generally ignored. However, in AFM biotin-streptavidin unbinding experiments last year, Walton et al. demonstrated a surprisingly strong effect on unbinding force due to cantilever stiffness. The unbinding force approximately doubled due to a change from a 35 pN/nm cantilever to a 58 pN/nm cantilever. Alarmed by the magnitude of the shift, we repeated their experiment on octomeric I27 to determine the magnitude for our mechanical protein unfolding experiments.

9.1 Theory

The presence of attached linkers and cantilevers alters the free energy landscape. Tension in the linkers favors domain unfolding, but that tension is not necessarily independent of the unfolding reaction coordinate. For sufficiently stiff cantilevers and linkers, even the small extension of the domain as it shifts from its bound to transition state noticeably reduces the effective tension. Assuming the bound and transition state extensions are relatively independent of the applied tension, the energy of the transition state will be

\[ E_b(f) = E_b(f = 0) - \int_0^{\Delta x} f(x) \, dx \]

\[ = E_b(f = 0) - \int_0^{\Delta x} [f(x = 0) - \kappa x] \, dx \]

\[ = E_b(f = 0) - f(x = 0) \Delta x + \frac{1}{2} \kappa \Delta x^2 , \]

where \( \kappa \) is the effective linker spring constant for that tension. The Bell-model unfolding rate is thus

\[ k(f) = k_0 \exp \left( \frac{f \Delta x - \frac{1}{2} \kappa \Delta x^2}{k_B T} \right) , \]

and stiffer linkers will increase the mean unfolding force.

Unfolded I27 domains can be well-modeled as wormlike chains (WLCs, Section 3.1.1)\(^{21} \text{, where } \]
\[ p \approx 4 \, r_f A \text{ is the persistence length, and } L \approx 28 \text{ nm is the contour length of the unfolded domain.} \]

Figure 9.1: Energy landscape schematic.
9.2 Simulations

We can model the I27 multimers as an array of Bell-model unfolders in series with a cantilever. Any unfolded domains also contribute to the tension according to the WLC tension formula. Completing 1000 simulated pulls for each cantilever/pulling-speed/multimer-number combination with our 
\textit{sawsim} Monte Carlo simulator yielded the following results (Figs. 9.3 to 9.5).

9.3 Methods

The experiments were carried out on octomers of I27 (Fig. 9.6). I27 is a model protein that has been used in mechanical unfolding experiments since the first use of synthetic chains\textsuperscript{217}. It was
Figure 9.4: Unfolding force loading rate dependence simulations for different cantilevers.

Figure 9.5: Unfolding force peak index dependence simulations for different cantilevers.

9.3 Methods
Figure 9.6: I27, the immunoglobulin-like domain 27 from human Titin (PDB ID: 1TIT). Figure generated with PyMol.

I27’s unfolding mechanism seems to involve stretching into a metastable intermediate state followed by Bell-model escape to the unfolded state, although there is not yet a consensus of the presence of the proposed intermediate.

The I27 octamers were stored in a TODO buffer solution. Mechanical unfolding experiments were carried out on I27 octomers (AthenaES) in PBS on gold-coated coverslips. We used both cantilevers on Olympus’s OMCL-TR400-PSA-1, which are nominally 80 and 20 pN/nm. Promising sawtooth curves were selected by eye and fit to WLCs to identify I27 unfolding events. The results were sorted into two bins according to cantilever stiffness, and then averaged across each cantilever-stiffness/pulling-speed group to produce Fig. 9.7.

Unfortunately, the data are not of high enough quality to extract the unfolding parameters $k$ or $\Delta x$. Note that the increase in mean unfolding force is not entirely due to the increased loading rate of the stiffer cantilever, because the difference is still present in the loading rate dependence (Fig. 9.8). The loading rates were extracted from the data by taking the slope of the fit WLC at unfolding.

9.4 Discussion

Walton et al. demonstrated the large effect of cantilever spring constant on ligand-receptor binding. Our theoretical, computational, and experimental results show that this effect, though reduced, is still important for multi-domain protein unfolding experiments. Studies attempting precision unfolding force measurements should account for this effect, but an analytical correction is not immediately clear. We suggest the unfolding parameters $k_0$ and $\Delta x$ be extracted by fitting the per-pulling-speed unfolding force histograms in parallel with Monte Carlo simulations that take the effect of cantilever spring constant into effect (such as our GPLed sawsim simulator, available at http://www.physics.drexel.edu/~wking/rsrch/simulation/).
Figure 9.7: Pulling speed dependence of I27 for different cantilever stiffnesses. The listed stiffnesses are averages across several individual cantilevers and calibrations. Each box is the average of some number of unfolding events, and the box area is proportional to that number. There are 82 unfolding events for the stiff cantilevers and 274 for the soft cantilevers.
Figure 9.8: Loading rate.
Chapter 10: Temperature dependent unfolding experiments

I27 (see page 30), ...
Chapter 11: Conclusions and future work


itsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman,
T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott,
http://www.sciencemag.org/cgi/content/short/291/5507/1304.

editors, *Mossbauer Spectroscopy in Biological Systems*, pages 22–24, Allerton House, Monticello,
levinthal/levinthal.html.

Using mechanical force to probe the mechanism of pausing and arrest during continuous
elongation by *Escherichia coli* RNA polymerase. *Proceedings of the National Academy of Sciences
URL http://www.pnas.org/content/99/18/11682.

[18] S. B. Smith, L. Finzi, and Carlos Bustamante. Direct mechanical measurements of the elasticity
of single DNA molecules by using magnetic beads. *Science*, 258(5085):1122–1126, November
content/abstract/258/5085/1122.

ISSN 0028-0836. doi: 10.1038/16219. URL http://www.nature.com/nature/journal/v397/n6714/
full/397050a0.html.


E. Broedel, Jane Clarke, and Julio M. Fernandez. Mechanical and chemical unfolding of
a single protein: A comparison. *Proceedings of the National Academy of Sciences USA*, 96
content/abstract/96/7/3694.

html.

sciencemag.org/cgi/content/abstract/275/5304/1295.

[24] R Lévy and Mounir Maaloum. Measuring the spring constant of atomic force microscope
10.1088/0957-4484/13/1/307. URL http://stacks.iop.org/0957-4484/13/33. Good review of
temporal calibration to 2002, but not much on the derivation of the Lorentzian fit.

ISSN 0024-9297. URL http://pubs3.acs.org/acs/journals/doilookup?in_doi=10.1021/
ma00130a008. Derivation of the Worm-like Chain interpolation function.

[26] Carlos Bustamante, John F. Marko, Eric D. Siggia, and S. B. Smith. Entropic elasticity of
lambda-phage DNA. *Science*, 265(5178):1599–1600, September 1994. ISSN 0036-8075. doi:
10.1126/science.8079175. URL http://www.sciencemag.org/cgi/content/abstract/265/5178/1599. WLC interpolation formula.


[33] Michael Schlierf and Matthias Rief. Single-molecule unfolding force distributions reveal a funnel-shaped energy landscape. Biophys. J., 90(4):L33–L35, February 2006. ISSN 0006-3495. doi: 10.1529/biophysj.105.077982. URL http://www.biophysj.org/cgi/content/abstract/90/4/L33. The inspiration behind my sawtooth simulation. Bell model fit to $f_{\text{unfold}}(v)$, but Kramers model fit to unfolding distribution for a given $v$. Eqn. 3 in the supplement is Evans and Ritchie’s 1999’s Eqn. 2, but it is just “[dying percent] * [surviving population] = [deaths]” (TODO, check). $\nu \equiv k$ is the force/time-dependent off rate... (TODO) The Kramers’ rate equation (second equation in the paper) is Hänggi et al. 38 Eq. 4.56b (page 275). It is important to extract $k_0$ and $\Delta x$ using every available method.


1109–1112, May 1997. doi: 10.1126/science.276.5315.1109. URL http://www.sciencemag.org/cgi/content/abstract/276/5315/1109. Seminal paper for force spectroscopy on Titin. Cited by Dietz and Rief\(^7\) (ref 9) as an example of how unfolding large proteins is easily interpreted (vs. confusing unfolding in bulk), but Titin is a rather simple example of that, because of its globular-chain structure.


[46] Changbong Hyeon and Devarajan Thirumalai. Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? Proceedings of the National Academy of Sciences USA, 100(18):10249–10253, September 2003. ISSN 0027-8424. doi: 10.1073/pnas.1833310100. URL http://www.pnas.org/cgi/content/abstract/100/18/10249. Derives the major theory behind my thesis. The Kramers rate equation is Hänggi et al.\(^38\) Eq. 4.56c (page 275).


Unfolding order not from protein-surface interactions. Mechanical unfolding of a chain of interleaved domains \(ABAB\ldots\) yielded a run of \(A\) unfoldings followed by a run of \(B\) unfoldings.


[70] Hendrik Dietz and Matthias Rief. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proceedings of the National Academy of Sciences USA*, 101(46):
DRAFT
February 22, 2011

44


[72] Tsvi Tlusty, Amit Meller, and Roy Bar-Ziv. Optical gradient forces of strongly localized fields. *Phys Rev Lett*, 81(8):1738–1741, August 1998. doi: 10.1103/PhysRevLett.81.1738. also at http://nanoscience.bu.edu/papers/p1738_1_Meller.pdf. Cited by Grossman and Stout for derivation of thermal response fn. However, I only see a referenced thermal energy when they list the likelihood of a small particle (radius $R_c$) escaping due to thermal energy, where $R_c$ is roughly $R_c \sim (k_B T/\alpha I_0)^{1/3}$, $\alpha$ is a dielectric scaling term, and $I_0$ is the maximum beam energy density. I imagine Grossman and Stout mixed up this reference.


[88] C. Grossman and A. Stout. Optical tweezers advanced lab. 2005. Fairly complete overdamped PSD derivation in section 4.3., cites Tlusty et al. and Bechhoefer and Wilson for further details. However, Tlusty (listed as reference 8) doesn’t contain the thermal response fn. derivation it was cited for. Also, the single sided PSD definition credited to reference 9 (listed as Bechhoefer) looks more like Press (listed as reference 10). I imagine Grossman and Stout mixed up their references, and meant to refer to Bechhoefer and Wilson and Press et al. respectively instead.
Appendix A: Cantilever Calibration

A.1 Overview

In order to measure forces accurately with an Atomic Force Microscope (AFM), it is important to measure the cantilever spring constant. The force exerted on the cantilever can then be deduced from its deflection via Hooke’s law $F = -kx$.

The basic idea is to use the equipartition theorem

$$\frac{1}{2} k \langle x^2 \rangle = \frac{1}{2} k_B T,$$

(A.1)

where $k_B$ is Boltzmann’s constant, $T$ is the absolute temperature, and $\langle x^2 \rangle$ denotes the expectation value of $x^2$ as measured over a very long interval $t_T$,

$$\langle A \rangle = \lim_{t_T \to \infty} \frac{1}{t_T} \int_{-t_T/2}^{t_T/2} A \, dt.$$

(A.2)

Solving the equipartition theorem for $k$ yields

$$k = \frac{k_B T}{\langle x^2 \rangle},$$

(A.3)

so we need to measure (or estimate) the temperature $T$ and variance of the cantilever position $\langle x^2 \rangle$ in order to estimate $k$.

A.1.1 Related papers

Various corrections taking into account higher order modes, and cantilever tilt have been proposed and reviewed, but we will focus here on the derivation of Lorentzian noise in damped simple harmonic oscillators that underlies all frequency-space methods for improving the basic $k \langle x^2 \rangle = k_B T$ method.

Roters and Johannsmann describe a similar approach to deriving the Lorentizian power spectral density.

WARNING: It is popular to refer to the power spectral density as a “Lorentzian” even though Eq. (A.50) differs from the classic Lorentzian.

$$L(x) = \frac{1}{\pi} \frac{1}{2\Gamma} \frac{1}{(x-x_0)^2 + \left(\frac{1}{2\Gamma}\right)^2},$$

(A.4)

where $x_0$ sets the center and $\Gamma$ sets the width of the curve. It is unclear whether the references are due to uncertainty about the definition of the Lorentzian or to the fact that Eq. (A.50) is also peaked. In order to avoid any uncertainty, we will leave Eq. (A.50) unnamed.

A.2 Methods

To find $\langle x^2 \rangle$, the raw photodiode voltages $V_p(t)$ are converted to distances $x(t)$ using the photodiode sensitivity $\sigma_p$ (the slope of the voltage vs. distance curve of data taken while the tip is in contact with the surface) via

$$x(t) = \frac{V_p(t)}{\sigma_p}.$$

(A.5)
Rather than computing the variance of $x(t)$ directly, we attempt to filter out noise by fitting the power spectral density (PSD) of $x(t)$ to the theoretically predicted PSD for a damped harmonic oscillator (Eq. (A.50))

$$\ddot{x} + \beta \dot{x} + \omega_0^2 x = \frac{F_{\text{thermal}}}{m}$$

(A.6)

$$\text{PSD}(x, \omega) = \frac{G_1}{(\omega_0^2 - \omega^2)^2 + \beta^2 \omega^2}$$

(A.7)

where $G_1 \equiv G_0/m^2$, $\omega_0$, and $\beta$ are used as the fitting parameters (see Eq. (A.50)). The variance of $x(t)$ is then given by Eq. (A.53)

$$\langle x(t)^2 \rangle = \frac{\pi G_1}{2 \beta \omega_0^2}$$

(A.8)

which we can plug into the equipartition theorem (Eq. (A.1)) yielding

$$k = \frac{2 \beta \omega_0^2 k_B T}{\pi G_1}.$$  

(A.9)

From Eq. (A.55), we find the expected value of $G_1$ to be

$$G_1 \equiv G_0/m^2 = \frac{2 \pi m}{\pi G_1} k_B T \beta.$$  

(A.10)

A.2.1 Fitting deflection voltage directly

In order to keep our errors in measuring $\sigma_p$ separate from other errors in measuring $\langle x(t)^2 \rangle$, we can fit the voltage spectrum before converting to distance.

$$\ddot{V}_p/\sigma_p + \beta \dot{V}_p/\sigma_p + \omega_0^2 V_p/\sigma_p = F_{\text{thermal}}$$

(A.11)

$$\dot{V}_p/\sigma_p + \beta V_p/\sigma_p + \omega_0^2 V_p = \frac{F_{\text{thermal}}}{m}$$

(A.12)

$$\ddot{V}_p + \beta \dot{V}_p + \omega_0^2 V_p = \frac{F_{\text{thermal}}}{m_p}$$

(A.13)

$$\text{PSD}(V_p, \omega) = \frac{G_{1p}}{(\omega_0^2 - \omega^2)^2 + \beta^2 \omega^2}$$

(A.14)

$$\langle V_p(t)^2 \rangle = \frac{\pi G_{1p}}{2 \beta \omega_0^2} = \frac{\pi \sigma_p^2 G_1}{2 \beta \omega_0^2} = \sigma_p^2 \langle x(t)^2 \rangle,$$

(A.15)

where $m_p \equiv m/\sigma_p$, $G_{1p} \equiv G_0/m_p^2 = \sigma_p^2 G_1$. Plugging into the equipartition theorem yeilds

$$k = \frac{\sigma_p^2 k_B T}{\langle V_p(t)^2 \rangle} = \frac{2 \beta \omega_0^2 \sigma_p^2 k_B T}{\pi G_{1p}}.$$  

(A.16)

From Eq. (A.10), we find the expected value of $G_{1p}$ to be

$$G_{1p} \equiv \sigma_p^2 G_1 = \frac{2 \pi m}{\pi m} k_B T \beta.$$  

(A.17)

A.2.2 Fitting deflection voltage in frequency space

Note: the math in this section depends on some definitions from section Appendix A.3.

As yet another alternative, you could fit in frequency $f \equiv \omega/2\pi$ instead of angular frequency $\omega$. But we must be careful with normalization. Comparing the angular frequency and normal frequency
unitary Fourier transforms

\[ \mathcal{F}\{x(t)\}(\omega) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt \]  
(A.18)

\[ \mathcal{F}_f\{x(t)\}(f) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t)e^{-2\pi ift}dt = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt = \sqrt{2\pi} \cdot \mathcal{F}\{x(t)\}(\omega = 2\pi f), \]  
(A.19)

from which we can translate the PSD

\[ \text{PSD}(x, \omega) \equiv \lim_{t_\beta \to \infty} \frac{1}{2|\mathcal{F}\{x(t)\}(\omega)|^2} \]  
(A.20)

\[ \text{PSD}_f(x, f) \equiv \lim_{t_\beta \to \infty} \frac{1}{2|\mathcal{F}_f\{x(t)\}(f)|^2} = 2\pi \cdot \lim_{t_\beta \to \infty} \frac{1}{2|\mathcal{F}\{x(t)\}(\omega = 2\pi f)|^2} = 2\pi \text{PSD}(x, \omega = 2\pi f). \]  
(A.21)

The variance of the function \(x(t)\) is then given by plugging into Eq. (A.33) (our corollary to Parseval’s theorem)

\[ \langle x(t)^2 \rangle = \int_0^\infty \text{PSD}(x, \omega)d\omega = \int_0^\infty \frac{1}{2\pi} \text{PSD}_f(x, f)2\pi df = \int_0^\infty \text{PSD}_f(x, f)df. \]  
(A.22)

Therefore

\[ \text{PSD}_f(V_p, f) = 2\pi \text{PSD}(V_p, \omega) = \frac{2\pi G_{1p}}{(4\pi f_0^2 - 4\pi^2 f^2)^2 + \beta^2 24\pi^2 f^2} = \frac{2\pi G_{1f}}{16\pi^4 (f_0^2 - f^2)^2 + \beta^2 24\pi^2 f^2} \]  
(A.23)

\[ \langle V_p(t)^2 \rangle = \frac{\pi G_{1f}}{2\beta f_0^2} \cdot \frac{8\pi^3}{16\pi^3} \]  
(A.24)

where \(f_0 \equiv \omega_0/2\pi, \beta_f \equiv \beta/2\pi, \) and \(G_{1f} \equiv G_{1p}/8\pi^3.\) Finally

\[ k = \frac{\sigma_p^2 k_B T}{\langle V_p(t)^2 \rangle} = \frac{2\beta f_0^2 \sigma_p^2 k_B T}{2\pi G_{1f}}. \]  
(A.25)

From Eq. (A.10), we expect \(G_{1f}\) to be

\[ G_{1f} = \frac{G_{1p}}{8\pi^3} = \frac{\sigma_p^2 G_1}{8\pi^3} = \frac{2\pi \sigma_p^2 k_B T \beta}{8\pi^3} = \frac{\sigma_p^2 k_B T \beta}{4\pi^4 m}. \]  
(A.26)

### A.3 Theoretical power spectral density for a damped harmonic oscillator

Our cantilever can be approximated as a damped harmonic oscillator

\[ m\ddot{x} + \gamma \dot{x} + kx = F(t), \]  
(A.27)

where \(x\) is the displacement from equilibrium, \(m\) is the effective mass, \(\gamma\) is the effective drag coefficient, \(k\) is the spring constant, and \(F(t)\) is the external driving force. During the non-contact phase of calibration, \(F(t)\) comes from random thermal noise.

In the following analysis, we use the unitary, angular frequency Fourier transform normalization

\[ \mathcal{F}\{x(t)\} \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt. \]  
(A.28)
We also use the following theorems (proved elsewhere):

\[
\cos\left(\frac{\theta}{2}\right) = \pm \sqrt{\frac{1}{2}[1 + \cos(\theta)]}, \tag{A.29}
\]

\[
\mathcal{F}\left\{ \frac{d^n x(t)}{dt^n}\right\} = (i\omega)^n x(\omega), \tag{A.30}
\]

\[
\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |x(w)|^2 d\omega. \tag{Parseval’s} \tag{A.31}
\]

As a corollary to Parseval’s theorem, we note that the one sided power spectral density per unit time (PSD) defined by

\[
PSD(x, \omega) \equiv \lim_{t_T \to \infty} \frac{1}{t_T} \int_{t_T/2}^{t_T/2} |x(\omega)|^2 \, dt = \lim_{t_T \to \infty} \frac{1}{t_T} \int_{-\infty}^{\infty} |x(\omega)|^2 d\omega = \int_{0}^{\infty} PSD(x, \omega) \, d\omega, \tag{A.32}
\]

where \( t_T \) is the total time over which data has been acquired.

We also use the Wiener-Khinchin theorem, which relates the two sided power spectral density \( S_{xx}(\omega) \) to the autocorrelation function \( r_{xx}(t) \) via

\[
S_{xx}(\omega) = \mathcal{F}\{ r_{xx}(t) \}, \tag{Wiener-Khinchin} \tag{A.34}
\]

where \( r_{xx}(t) \) is defined in terms of the expectation value

\[
r_{xx}(t) \equiv \langle x(\tau)\bar{x}(\tau-t) \rangle, \tag{A.35}
\]

and \( \bar{x} \) represents the complex conjugate of \( x \).

### A.3.1 Highly damped case

For highly damped systems, the inertial term becomes insignificant \((m \to 0)\). This model is commonly used for optically trapped beads. Because it is simpler and solutions are more easily available, it will serve to outline the general approach before we dive into the general case.

Fourier transforming Eq. (A.27) with \( m = 0 \) and applying Eq. (A.30) we have

\[
(i\gamma + k)x(\omega) = F(\omega)
\]

\[
|x(\omega)|^2 = \frac{|F(\omega)|^2}{k^2 + \gamma^2 \omega^2}. \tag{A.37}
\]

We compute the PSD by plugging Eq. (A.37) into Eq. (A.32)

\[
PSD(x, \omega) = \lim_{t_T \to \infty} \frac{1}{t_T} \frac{2|F(\omega)|^2}{k^2 + \gamma^2 \omega^2}. \tag{A.38}
\]

Because thermal noise is white (not autocorrelated + Wiener-Khinchin Theorem), we can denote the one sided thermal power spectral density per unit time by

\[
PSD(F, \omega) = G_0 = \lim_{t_T \to \infty} \frac{1}{t_T} \frac{2|F(\omega)|^2}{k^2 + \gamma^2 \omega^2}. \tag{A.39}
\]
Plugging Eq. (A.39) into Eq. (A.45) we have

\[ \text{PSD}(x, \omega) = \frac{G_0}{k^2 + \gamma^2 \omega^2}. \]  \tag{A.40}

This is the formula we would use to fit our measured PSD, but let us go a bit farther to find the expected PSD and thermal noise given \( m, \gamma \) and \( k \).

Integrating over positive \( \omega \) to find the total power per unit time yields

\[ \int_0^\infty \text{PSD}(x, \omega) d\omega = \int_0^\infty \frac{G_0}{k^2 + \gamma^2 \omega^2} d\omega = \frac{G_0 \pi}{2 \gamma k}, \]  \tag{A.41}

where the integral is solved in Appendix A.5.

Plugging into our corollary to Parseval’s theorem (Eq. (A.33)),

\[ \langle x(t)^2 \rangle = \frac{G_0 \pi}{2 \gamma k}. \]  \tag{A.42}

Plugging Eq. (A.42) into Eq. (A.1) we have

\[ k \frac{G_0 \pi}{2 \gamma k} = k_B T \]  \tag{A.43}
\[ G_0 = \frac{2 \gamma k_B T}{\pi}. \]  \tag{A.44}

So we expect \( x(t) \) to have a power spectral density per unit time given by

\[ \text{PSD}(x, \omega) = \frac{2 \pi}{k^2 + \gamma^2 \omega^2} \cdot \frac{\gamma k_B T}{\omega_0^2 - \omega^2}. \]  \tag{A.45}

### A.3.2 General form

The procedure here is exactly the same as the previous section. The integral normalizing \( G_0 \), however, becomes a little more complicated.

Fourier transforming Eq. (A.27) and applying Eq. (A.30) we have

\[ (-m \omega^2 + i \gamma \omega + k) x(\omega) = F(\omega) \]  \tag{A.46}
\[ (\omega_0^2 - \omega^2 + i \beta \omega) x(\omega) = \frac{F(\omega)}{m} \]  \tag{A.47}
\[ |x(\omega)|^2 = \frac{|F(\omega)|^2 / m^2}{(\omega_0^2 - \omega^2)^2 + \beta^2 \omega^2}, \]  \tag{A.48}

where \( \omega_0 \equiv \sqrt{k/m} \) is the resonant angular frequency and \( \beta \equiv \gamma/m \) is the drag-acceleration coefficient.

We compute the PSD by plugging Eq. (A.48) into Eq. (A.32)

\[ \text{PSD}(x, \omega) = \lim_{T \to \infty} \frac{1}{2T} \frac{2|F(\omega)|^2 / m^2}{(\omega_0^2 - \omega^2)^2 + \beta^2 \omega^2}. \]  \tag{A.49}

Plugging Eq. (A.39) into Eq. (A.49) we have

\[ \text{PSD}(x, \omega) = \frac{G_0 / m^2}{(\omega_0^2 - \omega^2)^2 + \beta^2 \omega^2}. \]  \tag{A.50}

A.3 Theoretical power spectral density for a damped harmonic oscillator
Figure A.1: Integral contour $C$ enclosing the upper half of the complex plane. If the integrand $f(z)$ goes to zero “quickly enough” as the radius of $C$ approaches infinity, then the only contribution comes from integration along the real axis (see text for details).

Integrating over positive $\omega$ to find the total power per unit time yields

$$\int_0^\infty \text{PSD}(x, \omega)d\omega = \frac{G_0}{2m^2} \int_{-\infty}^{\infty} \frac{1}{(\omega_0^2 - \omega^2)^2 + \beta^2 \omega^2} d\omega = \frac{G_0 \pi}{2m^2 \beta \omega_0^2} = \frac{G_0 \pi}{2m^2 \beta \frac{k}{m}} \quad (A.51)$$

$$= \frac{G_0 \pi}{2m \beta k} \quad (A.52)$$

The integration is detailed in Appendix A.5. By the corollary to Parseval’s theorem (Eq. (A.33)), we have

$$\langle x(t)^2 \rangle = \frac{G_0 \pi}{2m^2 \beta \omega_0^2} \quad (A.53)$$

Plugging Eq. (A.53) into the equipartition theorem (Eq. (A.1)) we have

$$\frac{kG_0 \pi}{2m \beta k} = k_B T \quad (A.54)$$

$$G_0 = \frac{2}{\pi} k_B T m \beta \quad (A.55)$$

So we expect $x(t)$ to have a power spectral density per unit time given by

$$\text{PSD}(x, \omega) = \frac{2k_B T \beta}{\pi m \left(\frac{k}{m} - \omega^2\right)^2 + \gamma^2 \omega^2} \quad (A.56)$$

As expected, the general form Eq. (A.56) reduces to the extremely overdamped form Eq. (A.45). Plugging in for $\beta \equiv \gamma/m$ and $\omega_0 \equiv \sqrt{k/m}$,

$$\lim_{m \to 0} \text{PSD}(x, \omega) = \lim_{m \to 0} \frac{2k_B T \gamma}{\pi m \left[(k/m - \omega^2)^2 + \gamma^2/m^2 \omega^2\right]} = \lim_{m \to 0} \frac{2k_B T \gamma}{\pi \left[(k - m \omega^2)^2 + \gamma^2 \omega^2\right]}$$

$$= \frac{2 \gamma k_B T}{\pi \left(k^2 + \gamma^2 \omega^2\right)} \quad (A.57)$$

$$= \frac{2 \gamma k_B T}{\pi \left(k^2 + \gamma^2 \omega^2\right)} \quad (A.58)$$

A.4 Contour integration

As a brief review, some definite integrals from $-\infty$ to $\infty$ can be evaluated by integrating along the contour $C$ shown in Fig. A.1.

A sufficient condition on the function $f(z)$ to be integrated, is that $\lim_{|z| \to \infty} |f(z)|$ falls off at least as fast as $\frac{1}{|z|^2}$. When this is the case, the integral around the outer semicircle of $C$ is 0, so the

$$\int_C f(z)dz = \int_{-\infty}^{\infty} f(z)dz.$$
We can evaluate the integral using the residue theorem,

$$\int_C f(x)\,dz = \sum_{z_p \in \text{poles in } C} 2\pi i \text{Res} \left( z = z_p, f(z) \right), \quad (A.59)$$

where for simple poles (single roots)

$$\text{Res} \left( z = z_p, f(z) \right) = \lim_{z \to z_p} \frac{1}{z - z_p} \frac{d}{dz} \left[ (z - z_p)^n \cdot f(z) \right], \quad (A.60)$$

and in general for a pole of order \(n\)

$$\text{Res} \left( z = z_p, f(z) \right) = \frac{1}{(n-1)!} \lim_{z \to z_p} \frac{d^{n-1}}{dz^{n-1}} \left[ (z - z_p)^n \cdot f(z) \right]. \quad (A.61)$$

### A.5 Integrals

#### A.5.1 Highly damped integral

$$I = \int_0^\infty \frac{1}{k^2 + z^2} \,dz = \frac{1}{2} \int_0^\infty \frac{1}{k^2 + z^2} \,dz = \frac{1}{2k} \int_{-\infty}^\infty \frac{1}{u^2 + 1} \,du, \quad (A.62)$$

where \(u \equiv z/k\) and \(du = dz/k\). There are simple poles at \(u = \pm i\).

$$I = \frac{1}{2k} \cdot 2\pi i \text{Res} \left( z = i, f(u) \right) = \frac{1}{2k} \cdot \frac{2\pi i}{i + i} = \frac{\pi}{2k}. \quad (A.63)$$

#### A.5.2 General case integral

We will show that, for any \((a, b > 0) \in \mathbb{R}\),

$$I = \int_C \frac{1}{(a^2 - z^2)^2 + b^2 z^2} \,dz = \frac{\pi}{b a^2}. \quad (A.64)$$

First we note that \(|f(z)| \to 0\) like \(|z^{-4}|\) for \(|z| \gg 1\), and that \(f(z)\) is even, so

$$I = \int_C \frac{1}{(a^2 - z^2)^2 + b^2 z^2} \,dz, \quad (A.65)$$

where \(C\) is the contour shown in Fig. A.1.

Because the denominator is of the form \(A^2 + B^2\), we can factor it into \((A + iB)(A - iB)\).

$$\left( a^2 - z^2 \right)^2 + b^2 z^2 = (a^2 - z^2 + ibz)(a^2 - z^2 - ibz) \quad (A.66)$$

And the roots of \(z^2 + ibz - a^2\)

$$z_{r, \pm} = \pm \frac{ib}{2} \left( 1 \pm \sqrt{1 - 4 \frac{-a^2}{(ib)^2}} \right) = \pm \frac{ib}{2} \left( 1 \pm \sqrt{1 - 4 \frac{a^2}{b^2}} \right) = \pm \frac{ib}{2} \left( 1 \pm S \right), \quad (A.67)$$

where \(S \equiv \sqrt{1 - 4 \frac{a^2}{b^2}}\).

To determine the nature and locations of the roots, consider the following cases

- \(a < b/2\), overdamped.
- \(a = b/2\), critically damped.
- \(a > b/2\), underdamped.
Our factored function $f$ is

$$f(z) = \frac{1}{(z - z_{r+})(z + z_{r+})(z + z_{r-})(z - z_{r-})}. \quad (A.68)$$

Applying Eqs. (A.59) and (A.60) we have

$$I = 2\pi i \left( \text{Res} \left( z = z_{r+}, f(z) \right) + \text{Res} \left( z = z_{r-}, f(z) \right) \right)$$

$$= 2\pi i \left( \frac{1}{(z_{r+} + z_{r+})(z_{r+} + z_{r-})(z_{r+} - z_{r-})} + \frac{1}{(z_{r-} - z_{r+})(z_{r-} + z_{r+})(z_{r-} + z_{r-})} \right)$$

$$= \frac{\pi i}{z_{r+}^2 - z_{r-}^2} \left( \frac{1}{z_{r+} - z_{r-}} + \frac{1}{z_{r+} + z_{r-}} \right)$$

$$= \frac{\pi i}{z_{r+}^2 - z_{r-}^2} \left( \frac{1}{z_{r+} - z_{r-}} \right) = \frac{\pi i}{z_{r+} - z_{r-}}$$

$$= 4\pi \left( \frac{1}{b^3(1 - S^2)} \right) - \frac{4\pi}{b^3[1 - (1 - 4a^2)^2]} = \frac{4\pi}{b^3 \cdot 4a^2} = \frac{\pi}{ba^2}. \quad (A.73)$$

Critically damped

Our factored function $f(z)$ is

$$f(z) = \frac{1}{(z - z_{r+})^2(z - z_{r-})^2}. \quad (A.74)$$

Applying Eqs. (A.59) and (A.61) we have

$$I = 2\pi i \text{Res} \left( z = z_{r+}, f(z) \right) = 2\pi i \left( \frac{1}{2z} \lim_{z \to z_{r+}} \frac{d}{dz} \left( \frac{1}{z + z_{r+}} \right) \right)$$

$$= \pi i \lim_{z \to z_{r+}} -2 \cdot \frac{1}{(z_{r+} + z_{r+})^3}$$

$$= -2\pi i \frac{1}{z_{r+}^3} = -2\pi i \frac{1}{6 \cdot \left( \frac{b}{2} \right)^3} = \frac{\pi}{b^3 \cdot 2^3} = \frac{\pi}{ba^2}, \quad (A.76)$$

which matches Eq. (A.73).
Appendix B: Hydrodynamic effects in fast AFM single-molecule force measurements

Müller notes

We had some trouble with their notation, so I’ll try and clear some things up...

<table>
<thead>
<tr>
<th>Müller</th>
<th>Trevor</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z_{\text{surface}}$</td>
<td>$z_{\text{surface}}$</td>
<td>Distance from the surface to the equilibrium cantilever position (increases on pulling)</td>
</tr>
<tr>
<td>$z_{\text{cantilever}}$</td>
<td>$z_{\text{cantilever}}$</td>
<td>Cantilever deflection from its equilibrium position (downward deflection positive)</td>
</tr>
<tr>
<td>$h$</td>
<td>$h$</td>
<td>$h = z_{\text{surface}} - z_{\text{cantilever}}$ the distance between the tip and surface</td>
</tr>
<tr>
<td>$v_{\text{tip}}$</td>
<td>$v_{\text{tip,surface}}$</td>
<td>$v_{\text{tip,surface}} = dh/dt$, tip velocity relative to the surface</td>
</tr>
<tr>
<td>$v_{\text{eq,surface}}$</td>
<td>$v_{\text{eq,surface}}$</td>
<td>$v_{\text{eq,surface}} = dz_{\text{surface}}/dt$, pulling speed</td>
</tr>
<tr>
<td>$v_{\text{tip,eq}}$</td>
<td>$v_{\text{tip,eq}}$</td>
<td>$v_{\text{tip,eq}} = dz_{\text{cantilever}}/dt$, tip velocity relative to its equilibrium position</td>
</tr>
<tr>
<td>$F_{\text{measured}}$</td>
<td>$F_{\text{measured}}$</td>
<td>Measured force deflecting the cantilever</td>
</tr>
<tr>
<td>$F_{\text{net}}$</td>
<td>$F_{\text{protein}}$</td>
<td>Force applied to stretch the protein</td>
</tr>
<tr>
<td>$F_d$</td>
<td>$F_d$</td>
<td>Drag force acting on the cantilever</td>
</tr>
<tr>
<td>$\Delta F$</td>
<td>$F_{d:\text{tip,eq}}$</td>
<td>Drag force due to only to $v_{\text{tip,eq}}$</td>
</tr>
<tr>
<td>$F_{d:\text{eq,surface}}$</td>
<td>$F_{d:\text{eq,surface}}$</td>
<td>Drag force due to only to $v_{\text{eq,surface}}$, the drag on an untethered cantilever</td>
</tr>
<tr>
<td>$F_{\text{meas,zeroed}}$</td>
<td>$F_{\text{meas,zeroed}}$</td>
<td>$F_{\text{meas,zeroed}} = F_{\text{measured}} - F_{d:\text{eq,surface}}$, defined for zero force in the detached region</td>
</tr>
</tbody>
</table>

Müller equations:

\[ F_d = \frac{6 \pi \eta a_{\text{eff}}^2}{h + d_{\text{eff}}} \cdot v_{\text{tip}} \]  
\[ h = z_{\text{surface}} - z_{\text{cantilever}} \]  
\[ v_{\text{tip}} = \frac{dh}{dt} \]  
\[ \Delta F = F_d(v, h) - F_d(v_{\text{tip}}, h) \]  
\[ F_{\text{net}} = F_{\text{measured}} + \Delta F \]

Trevor derivations:

For Eqn. B.4, we assume that all the fluid in the cell moves with the surface (i.e., fluid flow does not depend on height above the surface). So the drag force is proportional to the speed of the tip relative to the surface.

\[ F_d = D(h)v_{\text{tip,surface}} \]

Where $D(h)$ is some constant that can depend on $h$ (like $6 \pi \eta a_{\text{eff}}^2/(h + d_{\text{eff}})$). This is Müller Eqn B.1. Substituting in $v_{\text{tip,surface}} = v_{\text{eq,surface}} - v_{\text{tip,eq}}$ we have

\[ F_d = D(h)v_{\text{eq,surface}} - D(h)v_{\text{tip,eq}} = F_{d:\text{eq,surface}} - F_{d:\text{tip,eq}} \]  
\[ F_{d:\text{tip,eq}} = F_{d:\text{eq,surface}} - F_d \]

This is Müller Eqn B.4. The measured force deflecting the cantilever is then

\[ F_{\text{measured}} = F_{\text{protein}} + F_d \]  
\[ F_{\text{protein}} = F_{\text{measured}} - F_d = F_{\text{measured}} - (F_{d:\text{eq,surface}} - F_{d:\text{tip,eq}}) \]  
\[ = F'_{\text{meas,zeroed}} + F_{d:\text{tip,eq}} = F'_{\text{meas,zeroed}} + D(h)v_{\text{tip,eq}} \]

This is Müller Eqn B.5.
The treatment assumes the drag force on a detached cantilever doesn’t depend on distance (see dashed line in Figure 4b,c), which doesn’t make sense because

$$F_{d,eq,\text{surface}} = D(h)v_{eq,\text{surface}}$$  \hspace{1cm} (B.12)

And $D(h)$ depends on $h$. Therefore, this treatment uses $F_{\text{meas,zeroed}}'$, not $F_{\text{meas,zeroed}}$, where

$$F_{\text{meas,zeroed}}' = F_{\text{measured}} - F_{d,eq,\text{surface},h \approx 300 \text{ nm}} = F_{\text{meas,zeroed}} + [D(h) - D(300 \text{ nm})]v_{eq,\text{surface}}$$  \hspace{1cm} (B.13)

What can we do about this?

The correction from $F_{\text{meas,zeroed}}'$ (solid line in Figure 3a) to $F_{\text{protein}}$ (dashed line) comes from adding $F_{d,\text{tip,eq}}$, which is why $F_{\text{protein}} = F_{\text{meas,zeroed}}'$ when

$$0 = F_{d,\text{tip,eq}} \propto v_{\text{tip,eq}} = \frac{dz_{\text{cantilever}}}{dt} \propto \frac{dF_{\text{meas,zeroed}}}{dt},$$  \hspace{1cm} (B.14)

why $F_{\text{protein}} < F_{\text{meas,zeroed}}'$ when the cantilever is rebounding ($v_{\text{tip,eq}} < 0$), and why $F_{\text{protein}} > F_{\text{meas,zeroed}}'$ when the cantilever is loading the protein ($v_{\text{tip,eq}} > 0$).

This last ($F_{\text{protein}} > F_{\text{meas,zeroed}}'$ on loading) is why the raw measurement underestimates the unfolding force.
Nomenclature

\[ \langle s(t) \rangle \] Mean (expectation value) of a time-series \( s(t) \)

\[ \beta \] Damped harmonic oscillator drag-acceleration coefficient \( \beta \equiv \gamma / m \)

\[ z \] Complex conjugate of \( z \)

\[ \ddot{s} \] Second derivative of the time-series \( s(t) \) with respect to time. \( \ddot{s} = \frac{d^2 s}{dt^2} \)

\[ \dot{s} \] First derivative of the time-series \( s(t) \) with respect to time. \( \dot{s} = \frac{ds}{dt} \)

\[ \equiv \] Defined as (i.e. equivalent to)

\[ \gamma \] Damped harmonic oscillator drag coefficient \( F_{\text{drag}} = \gamma \dot{x} \)

\[ \infty \] Infinity

\[ \omega \] Angular frequency (radians per second)

\[ \omega_0 \] Resonant angular frequency (radians per second)

\[ F \] Force (newtons)

\[ f \] Frequency (hertz)

\[ k \] Spring constant (newtons per meter)

\[ k_B \] Boltzmann's constant, \( k_B = 1.38065 \cdot 10^{-23} \text{ J/K} \)

\[ t \] Time (seconds)

\[ x \] Displacement (meters)

\[ \mathcal{F}\{s(t)\} \] Fourier transform of the time-series \( s(t) \). \( s(f) = \mathcal{F}\{s(t)\} \)

AFM Atomic Force Microscope (or Microscopy)

DNA Deoxyribonucleic Acid

FJC Freely-Jointed Chain

I27 Immunoglobulin-like domain 27 from human Titin

PSD Power spectral density in angular frequency space

PSD\(_f\) Power spectral density in frequency space

\( \mathbb{R} \) Real numbers

WLC Wormlike Chain
### Index

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>47, 48</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>47, 48</td>
</tr>
<tr>
<td>$\omega_0$</td>
<td>50</td>
</tr>
<tr>
<td>AFM</td>
<td>2</td>
</tr>
<tr>
<td>beta</td>
<td>50</td>
</tr>
<tr>
<td>biotin</td>
<td>1</td>
</tr>
<tr>
<td>cosine half-angle</td>
<td>49</td>
</tr>
<tr>
<td>Damped harmonic oscillator</td>
<td>47, 50</td>
</tr>
<tr>
<td>extremely overdamped</td>
<td>49</td>
</tr>
<tr>
<td>damped harmonic oscillator</td>
<td>48</td>
</tr>
<tr>
<td>FJC</td>
<td>6</td>
</tr>
<tr>
<td>Fourier transform</td>
<td>48</td>
</tr>
<tr>
<td>gamma</td>
<td>50</td>
</tr>
<tr>
<td>I27</td>
<td>30</td>
</tr>
<tr>
<td>Parseval’s theorem</td>
<td>49</td>
</tr>
<tr>
<td>PSD</td>
<td>47, 49-51</td>
</tr>
<tr>
<td>in frequency space</td>
<td>48</td>
</tr>
<tr>
<td>residue theorem</td>
<td>52</td>
</tr>
<tr>
<td>streptavidin</td>
<td>1</td>
</tr>
<tr>
<td>Streptomyces avidini</td>
<td>1</td>
</tr>
<tr>
<td>Wiener-Khinchin theorem</td>
<td>49</td>
</tr>
<tr>
<td>WLC</td>
<td>4, 6, 17, 18, 30, 32</td>
</tr>
<tr>
<td>relaxation time</td>
<td>15</td>
</tr>
</tbody>
</table>
Vita