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.

2 Review of current research

? 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'. Bell introduced his model in the context of cell adhesion, but it has been widely used to model mechanical unfolding in proteins due to its simplicity and ease of use. Kramers introduced his theory in the context of thermally activated barrier crossing, which is how we use it here.

There is an excellent review of Kramers’ theory in Hanggi et al. 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. 20. Properties of the Bell model receive more coverage under the name of the older and equivalent Gompertz distribution. 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 Eqn. 5 and Juckett and Rosenberg Fig. 2.

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 Universitats Munchen
  http://cell.e22.physik.tu-muenchen.de/gruppmatthias/index.html
or laser-tweezers based

- Bustamante, University of California, Berkley
  http://alice.berkeley.edu/
- Forde, Simon Fraser University
  http://www.sfu.ca/fordelab/index.html

2.2 Evolution of unfolding modeling

Evans introduced the saddle-point Kramers’ approximation in a protein unfolding context 1997 (Evans and Ritchie\(^7\) Eqn. 3). Early work on mechanical unfolding focused on. In the early ‘00’s, the saddle-point/steepest-descent approximation to Kramer’s model (Hänggi et al.\(^10\) Eqn. 4.56c) was introduced into our field\(^4,12\). By the mid ‘00’s, the full-blown double-integral form of Kramer’s model (Hänggi et al.\(^10\) Eqn. 4.56b) was in use\(^16\).

There has been some tangential attempts towards even fancier models. Dudko et al.\(^4\) attempted to reduce the restrictions of the single-unfolding-path model. Hyeon and Thirumalai\(^12\) attempted to measure the local roughness using temperature dependent unfolding.

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\(^7\) introduces a smart Monte Carlo (SMC) Kramers’ simulation.

2.4 History of experimental AFM unfolding experiments

- ?:

2.5 History of experimental laser tweezer unfolding experiments

- Izrailev et al.\(^13\):
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

\[ \nu \equiv \frac{dx}{dt} \text{ the pulling velocity} \]  
\[ k \equiv \frac{dF}{dx} \text{ the loading spring constant} \]  
\[ P_0 \equiv P(t=0) \text{ the initial number of folded proteins} \]  
\[ D \equiv P_0 - P \text{ the number of dead (unfolded) proteins} \]  
\[ \kappa \equiv -\frac{1}{P} \frac{dP}{dt} \text{ the unfolding rate} \]

The proteins are under constant loading because

\[ \frac{dF}{dt} = \frac{dF}{dx} \frac{dx}{dt} = kv, \]

a constant, since both $k$ and $\nu$ are constant (Evans and Ritchie \(^7\) in the text on the first page, Dudko et al. \(^5\) 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. \(^5\) Eqn. 2, remembering that $\dot{F} = kv$, that their probability density is not a histogram ($W = 1$), and that their pdf is normalized to $N = 1$).

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

Solving for theoretical histograms is merely a question of taking your chosen $\kappa$, solving for $P(f)$, and plugging into Eqn. 7. We can also make a bit of progress solving for $P$ in terms of $\kappa$ as follows:

\[ \kappa \equiv -\frac{1}{P} \frac{dP}{dt} \]  
\[ -\kappa dr \frac{dF}{dt} = \frac{dP}{P} \]  
\[ -\frac{1}{kv} \int \kappa dF = \ln(P) + c \]  
\[ P = C \exp \left( -\frac{1}{kv} \int \kappa dF \right), \]

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

3.1 Constant unfolding rate

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

\[ P = C \exp \left( -\frac{1}{kv} \int \kappa dF \right) = C \exp \left( -\frac{1}{kv} \kappa F \right) = C \exp \left( -\frac{\kappa F}{kv} \right) \]  
\[ P(0) \equiv P_0 = C \exp(0) = C \]  
\[ h(F) = \frac{W}{vk} P \kappa = \frac{W \kappa P_0}{vk} \exp \left( -\frac{\kappa F}{kv} \right) \]

Suprise! 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.

3.2 Bell model

Stepping up the intensity a bit, we come to Bell’s model for unfolding (Hummer and Szabo \(^11\) Eqn. 1 and the first paragraph of Dudko et al. \(^5\) and Dudko et al. \(^6\)).

\[ \kappa = \kappa_0 \cdot \exp \left( \frac{F dx}{k_B T} \right) = \kappa_0 \cdot \exp(\kappa F), \]
where we’ve defined \( a \equiv dx/k_BT \) to bundle some constants together. The unfolding histogram is then given by

\[
P = C \exp \left( \frac{-1}{kv} \int \kappa dF \right) = C \exp \left( -\frac{1}{kv} \frac{\kappa_0}{a} \exp(aF) \right) = C \exp \left( -\frac{\kappa_0}{akv} \exp(aF) \right) \tag{16}
\]

\[
P(0) \equiv P_0 = C \exp \left( -\frac{\kappa_0}{akv} \right) \tag{17}
\]

\[
C = P_0 \exp \left( \frac{\kappa_0}{akv} \right) \tag{18}
\]

\[
P = P_0 \exp \left\{ \frac{\kappa_0}{akv} [1 - \exp(aF)] \right\} \tag{19}
\]

\[
h(F) = \frac{W}{vk} P \kappa = \frac{W}{vk} P_0 \exp \left\{ \frac{\kappa_0}{akv} [1 - \exp(aF)] \right\} \kappa_0 \exp(aF) = \frac{W \kappa_0 P_0}{vk} \exp \left\{ aF + \frac{\kappa_0}{akv} [1 - \exp(aF)] \right\} . \tag{20}
\]

The \( F \) dependent behavior reduces to

\[
h(F) \propto \exp [aF - b \exp(aF)] , \tag{21}
\]

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

This looks an awful lot like the the Gompertz/Gumbel/Fisher-Tippett distribution, where

\[
p(x) \propto z \exp(-z) \tag{22}
\]

\[
z \equiv \exp \left( -\frac{x - \mu}{\beta} \right) , \tag{23}
\]

but we have

\[
p(x) \propto z \exp(-bz) . \tag{24}
\]

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 echoes my result (Wu et al. \textsuperscript{18} Eqn. 1).

Oh wait, we can do this:

\[
p(x) \propto z \exp(-bz) = \frac{1}{b} z' \exp(-z') \propto z' \exp(-z') , \tag{25}
\]

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

\[
P(x) = \frac{1}{\beta} \exp \left[ \frac{x - \alpha}{\beta} - \exp \left( \frac{x - \alpha}{\beta} \right) \right] \tag{26}
\]

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 \( \beta = 1/a = k_BT/dx \), \( \alpha = -\beta \ln(kB/kv) \), and \( F = x \) we have

\[
P(F) = \frac{1}{\beta} \exp \left[ \frac{F + \beta \ln(kB/kv)}{\beta} - \exp \left( \frac{F + \beta \ln(kB/kv)}{\beta} \right) \right] \tag{27}
\]

\[
= \frac{1}{\beta} \exp(F/\beta) \exp[\ln(kB/kv)] \exp[- \exp(F/\beta) \exp[\ln(kB/kv)]] \tag{28}
\]

\[
= \frac{1}{\beta} \kappa B \exp(F/\beta) \exp[-\kappa B/kv \exp(F/\beta)] \tag{29}
\]

\[
= \frac{\kappa}{kv} \exp(F/\beta - \kappa B/kv \exp(F/\beta)) \tag{30}
\]

\[
= \frac{\kappa}{kv} \exp(F/\beta - \kappa B/kv \exp(F/\beta)) \tag{31}
\]

\[
= \frac{\kappa}{kv} \exp[aF - \kappa/akv \exp(aF)] \tag{32}
\]

\[
= \frac{\kappa}{kv} \exp[aF - b \exp(aF)] \propto h(F) . \tag{33}
\]

So our unfolding force histogram for a single Bell domain under constant loading does indeed follow the Gumbel distribution.
For the saddle-point approximation for Kramers’ model for unfolding (Evans and Ritchie\textsuperscript{7} Eqn. 3, \textsuperscript{7} Eqn. 4.56c, van Kampen\textsuperscript{17} Eqn. XIII.2.2).

\[ \kappa = \frac{D}{l_b l_{ts}} \cdot \exp \left( \frac{-E_b(F)}{k_BT} \right) , \tag{34} \]

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 \( l_{ts} \) is the characteristic length of the transition state.

\[ l_{ts} = \text{TODO} \tag{35} \]

Evans and Ritchie\textsuperscript{7} solved this unfolding rate for both inverse power law potentials and cusp potentials.

### 3.3.1 Inverse power law potentials

\[ E(x) = \frac{-A}{x^n} \tag{36} \]

(e.g. \( n = 6 \) for a van der Waals interaction, see Evans and Ritchie\textsuperscript{7} in the text on page 1544, in the first paragraph of the section \textit{Dissociation under force from an inverse power law attraction}). Evans then gets funky with diffusion constants that depend on the protein’s end to end distance, and I haven’t worked out the math yet…

### 3.3.2 Cusp potentials

\[ E(x) = \frac{1}{2} \kappa_a \left( \frac{x}{x_a} \right)^2 \tag{37} \]

(see Evans and Ritchie\textsuperscript{7} in the text on page 1545, in the first paragraph of the section \textit{Dissociation under force from a deep harmonic well}).

## 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 sawsim simulation, 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 \( k(F) \to \text{histogram} \) portion of the simulation.

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

### 4.1 Cusp-like potentials

### 4.2 Quartic potentials

## References


[12] Changbong Hyeon and D. Thirumalai. Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? *Proc Natl Acad Sci U S A*, 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](http://www.pnas.org/cgi/content/abstract/100/18/10249). Derives the major theory behind my thesis. The Kramers rate equation is Hanggi Eq. 4.56c (page 275).


[15] 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](http://www.biophysj.org/cgi/content/abstract/90/4/L33). The inspiration behind my sawtooth simulation. Bell model fit to $f_{unfolding}(v)$, but Kramers model fit to unfolding distribution for a given $v$. Eqn. 3 in the supplement is Evans-Ritchie 1999's Eqn. 2; but it is just “[dying percent] * [surviving population] = [deaths]” (TODO, check). $v \equiv k$ is the force/time-dependent off rate... (TODO) The Kramers’ rate equation (second equation in the paper) is Hanggi Eq. 4.56b (page 275). It is important to extract $k_0$ and $\Delta x$ using every available method.