University of California, Merced

Dissertation

Mathematical Models of Prion Protein Pathogenesis

by

Jason K. Dark

A technical report submitted
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Applied Mathematics

2017

Committee Members:
Professor Suzanne S. Sindi, Chair
Professor Harish S. Bhat
Professor Juan C. Meza

The Dissertation of Jason K. Dark is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Committee Member:

_____

Professor Harish S. Bhat

Committee Member:

_____

Professor Juan C. Meza

Committee Chair / Research Advisor:

_____

Professor Suzanne S. Sindi

_____

Date

Acknowledgments

First and foremost, I would like to thank my advisor Suzanne Sindi for being a constant source of inspiration. Without her, I would not have found myself along this path. I also owe a deep gratitude to the members of my committees, who have helped me grow into the scientist that I've become. Each has had a direct influence on how I think about mathematical modeling and on the nature of computational and applied mathematics. Finally, I couldn't have asked for better colleagues than the graduate students in the applied mathematics group. It is rare to find such a group of talented and personable folks as have been gathered in our graduate program, and it has truly been honor to work and learn alongside them.

I gratefully acknowledge funding from the University of California, Merced's School of Natural Sciences and NSF Inspire grant #1344279.

Contents

Mathematical Models of Prion Protein Pathogenesis

by

Jason K. Dark

Doctorate in Applied Mathematics

Suzanne Sindi, Chair

University of California, Merced

2017

Abstract

I develop a number of mathematical models for the study of prion phenotype propagation in *S. cerevisiae*. Prion proteins underlie a host of non-Mendelian phenotypes in yeast and several fatal, neurodegenerative diseases in mammals, most notably bovine spongiform encephalopathy ("mad-cow disease") and Creutzfeldt-Jakob disease in humans. Two fundamental questions guide this work:

1. How does the infectious prion form of a protein initially occur?
2. Once present, how do the infectious prion "aggregates" persist and amplify across multiple generations of cell division?

A significant body of literature addresses Question 2, and so I begin with a review of this literature and then develop two new models in an attempt to better address *in vivo* cellular conditions (Chapters 2 and 3). I then return to Question 1, expanding upon a Markov chain formulation of the problem developed in the literature and implement a memory-efficient, numerical solver for the estimation of the time scale of "spontaneous nucleation" (Chapter 4). Taken together, this work is (to my knowledge) the most comprehensive, stochastic treatment of the yeast prion phenotype system and is the first to permit prion strain coexistence, qualitatively match biological experiments on Hsp104 knock-out mutants, and suggest nucleation size differences between prion strains.

List of Abbreviations and Symbols

BSE	bovine spongiform encephalopathy
CME	chemical master equation
IC	initial conditions
ELNPM	enzyme-limited nucleated polymerization model
Hsp104	heat-shock protein 104
NPM	nucleated polymerization model
ODE	ordinary differential equation
Sup35	suppressor protein 35
PDE	partial differential equation
TSE	transmissible spongiform encephalopathy
α	rate of protein synthesis
β	rate of aggregation
γ	rate of fragmentation
μ	rate of dilution
$[X], [Y], \ldots$	concentration of chemical species X, Y, …
$[PSI^+]$, $[psi^-]$	the phenotypes corresponding to the presence (or absence) of the prion form of Sup35
$u_i(t)$	the concentration of aggregates of length i at time t under discrete assumptions
$u(s,t)$	the concentration of aggregates of length s at time t under continuous assumptions
$\mathbb{E}[X]$	the expectation of the random variable X
$\mathrm{Var}[X]$	the variance of the random variable X
$\Pr[A]$	the probability of event A
$\Pr[A\|B]$	the probability of event A given event B

List of Figures

List of Tables

Curriculum Vitae

Education

2012–present PhD, Applied Mathematics.
Advised by Prof. Suzanne Sindi.
University of California, Merced. Merced, CA.

2009–2011 MS, Mathematics and Statistics.
Georgetown University. Washington, DC.

2004-2008 ScB, Mathematics & AB, Education Studies.
Brown University. Providence, RI.

Publications

- Vinson, David W., **Davis, Jason K.**, Sindi, Suzanne S., and Dale, Rick. "Efficient n-gram analysis in R with cmscu." Behavior research methods 48.3 (2016): 909-921.
- **Davis, Jason K.**, and Sindi, Suzanne S. "Initial condition of stochastic self-assembly." *Physical Review E* 93.2 (2016): 022109.
- **Davis, Jason K.**, and Sindi, Suzanne S. "A mathematical model of the dynamics of prion aggregates with chaperone-mediated fragmentation." *Journal of Mathematical Biology* (2015): 1-24.
- **Davis, Jason K.**, and Sindi, Suzanne S. "A study in nucleated polymerization models of protein aggregation." *Applied Mathematics Letters* 40 (2015): 97-101.

Teaching Experience

Summer 2013 Teaching Associate and Instructor of Record,
MATH32 Probability and Statistics.
University of California, Merced.

Fall 2012, 2013 Graduate Teaching Assistant,
MATH141 Linear Analysis.
University of California, Merced.

Spring 2008 Undergraduate Teaching Assistant,
CS22 Introduction to Discrete Structures and Probability.
Brown University.

Fall 2006, 2007 Undergraduate Teaching Assistant,
ED101 The Craft of Teaching.
Brown University.

Chapter 1

Introduction

1.1 Prion Biology

The central dogma of molecular biology describes the transfer of information via nucleic acids (DNA). It has been summarized as "once (sequential) information has passed into a protein, it cannot get out again," [12, 11]. This is perhaps what puzzled biologists so much when they first studied scrapie, a disease infecting sheep. It was concluded in [1] that the disease-causing agent "is likely to be of an unusual nature," citing experimental evidence showing that the agent could amplify without containing nucleic acid.

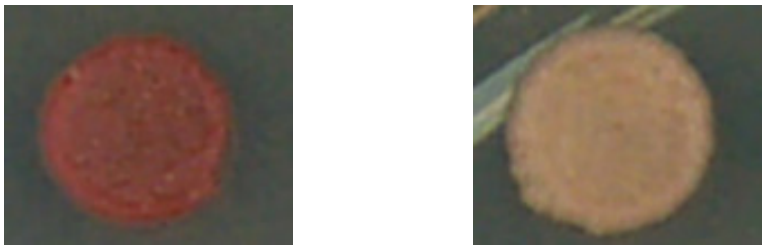
The resolution to this paradox was first suggested by a mathematician; in [24], Griffith detailed a number of plausible mechanisms for the self-replication of a protein conformation. Griffith's protein-only hypothesis was later validated by Prusiner [36] in work that would eventually earn him the Nobel Prize in Physiology or Medicine [21]. In this work, Prusiner also coined the term prion, short for "proteinacious infectious agent."

The mechanism underlying prion diseases is rooted in protein folding; while proteins are comprised of various amino acids, their function is determined by the 3-dimensional shape the protein obtains when chemical bonds are formed between the amino acids. The resulting shape is understood to correspond to an optimal configuration, in the sense that it minimizes free energy as a function of possible bonds [4]. Thus, it will require energy to change the bonds, making the configuration particularly stable. Proteins can misfold, however, and there is a host of cellular machinery dedicated to either eliminating or refolding these aberrations [31].

In a prion disease, enough misfolded protein has aggregated into a sufficiently stable structure to resist the efforts of this machinery. That is, the "cooperative" aggregate structure is believed to be somehow more energetically favorable than the normal monomer form. In mammals, these aggregates form amyloid plaques, which is also characteristic of Alzheimer's, Parkinson's, and other so-called amyloid diseases [26]. What makes prion diseases particularly devastating, and possibly distinct from the other amyloid diseases, is that these aggregates are *infectious*[1] – not only do they resist efforts

[1]There are a number of superficial similarities between amyloid diseases and prion diseases. In fact, it is

to refold or degrade, but they actively promote the conversion of correctly-folded protein into the misfolded state [37]. Thus, analogous to a virus, the prion aggregates repurpose healthy protein to spread the abnormal form. However, a single aggregate can only convert protein at a limited rate; the actual spread of the disease state comes from the fragmentation of the aggregate into smaller aggregates, which can each independently continue converting protein [32]. The fragmentation is believed to be caused by a failed attempt by the cellular repair machinery to repair the protein [8].



(a) [psi^{-}] (prion-free) yeast colony

(b) [PSI^{+}] (prion-infected) yeast colony

Figure 1.1: Yeast prions can be engineered with pigments, allowing visual inspection of the infection (images provided by the Serio lab[2]).

On the other hand, a growing body of literature points to prion proteins as being an essential tool for a number of biological functions. Certain plants, for example, exhibit a phenomenon known as vernalization where they flower only after certain weather events have occurred. The proteins involved in this process bear many structural similarities to known prion proteins [5, 7]. Other "prion-like" proteins have also been been implicated in long-term memory formation in both Aplysia (sea slugs) and Drosophila (fruit flies) [39, 20]. Thus, prion "behavior" may not simply be an accident of nature, but rather a selected-for trait providing a mechanism for proteins to dynamically change their function.

Non-mammalian prion proteins have been most studied in *S. cerevisiae*, i.e. brewer's yeast. Yeast biologists have long observed a phenotype of *S. cerevisiae* termed [PSI^{+}] that behaved in a non-Mendelian fashion [8]. It was eventually determined that the protein responsible, Sup35, was exhibiting a prion-like conformation. The Sup35/[PSI^{+}] system is now considered the yeast analog to mammalian PrP/PrP^{Sc}, the protein (and its prion conformation) underlying Creutzfeldt-Jakob disease (humans), scrapie (sheep), and bovine spongiform encephalopathy (cattle). Intriguingly, Sup35 is involved in translation termination; when [PSI^{+}] is expressed, stop-codon read-through occurs, leading to the expression of what was previously believed to be nonsense "junk" DNA. However, [PSI^{+}] cells have been shown to have higher fitness in certain, stressful environments, thus

becoming increasingly less controversial to claim that the former is simply a special case of the latter (e.g. [3, 22]), especially in light of a recent study with circumstantial evidence for the infectivity of Alzheimer's disease [29]. Nonetheless, until the "infectivity" criterion can be clinically proven this connection will remain at best conjecture.

[2] `http://mcb.arizona.edu/people/tserio`

[PSI^+] may actually be a survival strategy allowing the yeast to express "back-up" genes that aren't normally expressed [42, 25].

In yeast, the Hsp104 chaperone protein is known to serve a crucial role in the protein repair machinery – this protein's under or over-expression is capable of eliminating the presence of prions in the yeast colony [8, 30]. The understood mechanism is that Hsp104 physically fragments the aggregates. When it is completely suppressed, aggregates no longer fragment and are unable to multiply. They are then asymptotically diluted in concentration to undetectable levels as the colony grows. When Hsp104 is over-expressed, the aggregates are rapidly fragmented into very small polymers. There is evidence for a minimal, stable aggregate size [9]: when aggregates have fewer incorporated proteins than this number, the aggregates rapidly disassociate and the constituent proteins refold into the normal monomer form. Thus, the over-expressed chaperone fragments the aggregates until they are no longer stable, effecting the same outcome. Biologists have yet to identify a naturally-occurring mammalian analog to Hsp104, although it is known that Hsp104 does amplify mammalian PrP^{Sc} fragmentation [14].

Prion proteins have another surprising property: they admit more than one infectious conformation. Both mammalian PrP and yeast Sup35 have a number of *strains* [41]. These different strains still aggregate, but the respective aggregates are now competing with each other for healthy protein to convert. These strains are characterized by different kinetic rates as well as physiological measures, such as time-to-onset of symptoms [41]. In yeast colonies, it is actually possible to have different strains dominating in different spatial regions of the plate (or "weaker" strains competing against prion-free regions). In the engineered strains, this manifests visually as heterogeneous regions (sectors) of the colony with differing shades of pink [2].

This combination of biological features provides for a rich mathematical modeling opportunity. We now summarize the existing mathematical literature, then outline our own research program.

1.2 Mathematics of Nucleated Polymerization

1.2.1 Background

The nucleated polymerization model (NPM) is a mathematical model formally introduced in Nowak et al. [35] and validated in Masel, Jansen, and Nowak [34] that describes the concentration of prion proteins over time. While not the first mathematical model of prion aggregation and replication[3] it has emerged as the canonical framework from which to quantitatively study prions.

The NPM incorporates a number of biochemical assumptions. First, normal protein is synthesized at rate α. Second, concentrations are diluted at rate μ due to cell division. Third and fourth are the assumptions governing the dynamics of prion aggregates.

[3] See [17] for a review of other early attempts. As previously mentioned, the first model [24] actually predated a biological understanding and predicted the mechanism of prion diseases. As such, the modeling of prion diseases within a mathematical framework is an early success of mathematical biology.

In this model, prion aggregates are treated as linear fibers; as such, these fibers may lengthen or shorten. In the lengthening reactions, which are called aggregation events, protein monomers are incorporated into existing fibers at rate 2β – the factor of two is to explicitly indicate that either end of the fiber may add a monomer. The shortening reactions, which are called fragmentation events, are a binary fission of the aggregate. It is assumed that a cellular chaperone binds (at rate γ) to a junction between bound proteins and fragments the aggregate into two. Letting Ψ_i denote an aggregate of size i and X a protein monomer, these reactions are summarized by the chemical equations

$$X + \Psi_{i-1} \xrightarrow{2\beta} \Psi_i \xrightarrow{\gamma} \underbrace{\Psi_j + \Psi_{i-j}}_{j=1,2,\ldots,i-1}. \tag{1.1}$$

The first two biochemical assumptions are represented by

$$X \underset{\mu}{\overset{\alpha}{\rightleftharpoons}} \emptyset \qquad \text{and} \qquad \Psi_i \xrightarrow{\mu} \emptyset. \tag{1.2}$$

It is additionally assumed that fibers must be of a certain minimum size, which is denoted n, for nucleation size. If a fragmentation event yields an aggregate Ψ_i with $i < n$, then this model assumes that the aggregate immediately disassociates into its monomer constituents, e.g. $\Psi_i \to iX$. This is the assumption that gives the nucleated polymerization model its name.

1.2.2 Discrete Size Model

Following [34], let $x(t) = [X](t)$ (the time-varying concentration of protein monomer) and $u_i(t) = [\Psi_i](t)$ (the time-varying concentration of aggregates of size i). Then, the law of mass action yields an infinite-dimensional system of ordinary differential equations (ODE)

$$x' = \alpha - \mu x - 2\beta x \sum_{i \geq n} u_i + \gamma n(n-1) \sum_{i \geq n} u_i, \tag{1.3}$$

$$u_i' = -2\beta x(u_i - u_{i-1}) - \mu u_i - \gamma(i-1)u_i + 2\gamma \sum_{j>i} u_j, \tag{1.4}$$

where $u_i(t) = 0$ for $i < n$. Masel, Jansen, and Nowak [34] then observed that this infinite-dimensional ODE system admits an exact moment closure. Defining $y(t) = \sum_{i \geq n} u_i(t)$ and $z(t) = \sum_{i \geq n} i u_i(t)$, one obtains

$$x' = \alpha - \mu x - 2\beta xy + \gamma n(n-1)y, \tag{1.5}$$

$$y' = -(\mu + \gamma(2n-1))y + \gamma z, \tag{1.6}$$

$$z' = 2\beta xy - \mu z - \gamma n(n-1)y. \tag{1.7}$$

The authors completed their mathematical analysis with a local stability analysis, in which the "disease-free" fixed point ($y = z = 0$) is locally stable if $R = \frac{2\alpha\beta\gamma}{\mu(\mu+\gamma n)(\mu+\gamma(n-1))} < 1$ and the "endemic" fixed point ($y, z \neq 0$) is locally unstable. If

$R > 1$ the stability is exchanged. Due to the conservation law $(x + z)' = \alpha - \mu(x + z)$ embedded in the dynamics, all orbits are globally attracted to a 2-dimensional manifold (where $x + z = \alpha/\mu$). Thus, although [34] provided only a local analysis, their results can be made global by application of the Poincaré-Bendixson theorem to the "trapping region" $0 < ny < z < \alpha/\mu$.

1.2.3 Continuous Size Model

The NPM was later simplified in a series of papers [38, 18, 23]. The discrete size of the aggregates was relaxed, allowing for sizes of any real, positive number. In a scaling argument, [16] demonstrated that the resulting dynamics were equivalent in the limit of large average aggregate sizes. Instead of considering $u_i(t)$, one instead considers $\tilde{u}(s,t)$ which satisfies

$$\tilde{x}' = \alpha - \mu\tilde{x} - 2\beta\tilde{x}\int_a^\infty \tilde{u}(s,t)\,\mathrm{d}s + \gamma a^2 \int_a^\infty \tilde{u}(s,t)\,\mathrm{d}s, \tag{1.8}$$

$$\frac{\partial \tilde{u}}{\partial t} = -2\beta\tilde{x}\frac{\partial \tilde{u}}{\partial s} - \mu\tilde{u} - \gamma s\tilde{u} + 2\gamma\int_s^\infty \tilde{u}(s',t)\,\mathrm{d}s', \tag{1.9}$$

where $\tilde{u}(a,t) = 0$ for some $a > 0$. This version admits an exact moment closure as well; defining $\tilde{y}(t) = \int_a^\infty \tilde{u}(s,t)\,\mathrm{d}s$ and $\tilde{z}(t) = \int_a^\infty s\tilde{u}(s,t)\,\mathrm{d}s$, one obtains

$$\tilde{x}' = \alpha - \mu\tilde{x} - 2\beta\tilde{x}\tilde{y} + \gamma a^2\tilde{y}, \tag{1.10}$$

$$\tilde{y}' = -(\mu + 2\gamma a)\tilde{y} + \gamma\tilde{z}, \tag{1.11}$$

$$\tilde{z}' = 2\beta\tilde{x}\tilde{y} - \mu\tilde{z} - \gamma a^2\tilde{y}. \tag{1.12}$$

The same papers [38, 18, 23] also provided the first global stability results on the NPM (either continuous or discrete), finding a Lyapunov function when $R = \frac{2\alpha\beta\gamma}{\mu(\mu+\gamma a)^2} < 1$ and using a semi-group argument for $R > 1$. Prüss and Pujo-Menjouet [38] additionally connected the NPM to the SEIS (susceptible-exposed-infected) compartment model from mathematical epidemiology via a change of variables, which equivalently provides the same global results. The SEIS model's dynamics are fully characterized by the "basic reproductive number" [19], and this connection provides the useful interpretation of the NPM constant R as this number.

This connection to the SEIS model is not limited to the continuous-size relaxation; though not otherwise noted in the literature, the discrete equivalent is provided below (left column), alongside the continuous transformation (right column).

$$\frac{\alpha}{\mu}S(\gamma t) = x(t) \qquad \frac{\alpha}{\mu}\tilde{S}(\gamma t) = \tilde{x}(t) \tag{1.13}$$

$$\frac{\alpha}{\mu}E(\gamma t) = z(t) - (n-1)y(t) \qquad \frac{\alpha}{\mu}\tilde{E}(\gamma t) = \tilde{z}(t) - a\tilde{y}(t) \tag{1.14}$$

$$\frac{\alpha}{\mu}I(\gamma t) = (n-1)y(t) \qquad \frac{\alpha}{\mu}\tilde{I}(\gamma t) = a\tilde{y}(t) \tag{1.15}$$

1.2.4 Asymptotic Size Density

Finally, [18, 23] provided analysis of the size density itself. In particular, at steady-state with $R > 1$, the size density satisfies

$$0 = -\left(\frac{\mu}{\gamma} + a\right)^2 \frac{\partial \tilde{u}}{\partial s} - \frac{\mu}{\gamma}\tilde{u} - s\tilde{u} + 2\int_s^\infty \tilde{u}(s')\,\mathrm{d}s'. \tag{1.16}$$

Differentiating Equation (1.16) with respect to s yields a linear, 2nd order ordinary differential equation which is solved by Laplace transform. The solution was first provided in [18], where

$$\lim_{t\to\infty} \tilde{u}(s+a,t) = \left(\lim_{t\to\infty} \tilde{y}(t)\right) \frac{s(2\tilde{\zeta}+s)}{\tilde{\zeta}^3} \exp\left(-\frac{s(2\tilde{\zeta}+s)}{2\tilde{\zeta}^2}\right), \tag{1.17}$$

and $\tilde{\zeta} = a + \mu/\gamma$.

My first contribution to the literature, Davis and Sindi [13], was a closed-form expression for the discrete size density, which satisfies at steady-state

$$0 = -\left(\frac{\mu}{\gamma} + n\right)\left(\frac{\mu}{\gamma} + n - 1\right)(u_i - u_{i-1}) - \frac{\mu}{\gamma}u_i - (i-1)u_i + 2\sum_{j>i} u_j. \tag{1.18}$$

One may derive an ordinary differential equation by considering the probability generating function $f(x) = \sum_{i\geq n} u_i x^i$. The solution is readily found up to quadrature, then expanded in a Taylor series about $x = 0$ and integrated term-by-term. Doing so, one obtains

$$\lim_{t\to\infty} u_{i+n-1}(t) = \left(\lim_{t\to\infty} y(t)\right) \frac{i(2\zeta+i-1)}{\zeta^2(\zeta-1)} \frac{\Gamma(\zeta^2+1)}{\Gamma(\zeta^2+i+1)} \left(\zeta(\zeta-1)\right)^i, \tag{1.19}$$

where $\zeta = n + \mu/\gamma$ and $\lim_{t\to\infty} y(t)$ is found by the moment-closure. This derivation is now reproduced in its entirety[4] from [13].

We begin by defining $v_i = \lim_{t\to\infty} u_{i+n-1}(t)/y(t)$. These $\{v_i\}$ will satisfy the 2nd order, linear recurrence relation

$$0 = \left(\zeta^2 + i + 2\right) v_{i+2} - \left(2\zeta^2 - \zeta + i - 1\right) v_{i+1} + \left(\zeta^2 - \zeta\right) v_i, \tag{1.20}$$

with $\zeta > 2$, $v_0 = 0$ and $v_1 = \frac{2}{1+\zeta^2}$.

We construct $f(x) = \sum_{i=0}^\infty v_i x^i$; from (1.20), $f(x)$ will satisfy

$$f'(x) + \left(\frac{\zeta^2}{x} + \frac{2}{1-x} - \zeta(\zeta-1)\right) f(x) = \frac{2}{1-x}, \tag{1.21}$$

[4]With permission from Davis, Jason K., and Suzanne S. Sindi, Applied Mathematics Letters 40, 2015. Copyright (2015) by Elsevier. (RightsLink Order #4066170090924). The co-author listed in this publication directed and supervised research which forms the basis for the dissertation.

which has solution

$$f(x) = 2\frac{(1-x)^2}{x^{\zeta^2}}\mathrm{e}^{\zeta(\zeta-1)x}\int_0^x \frac{s^{\zeta^2}}{(1-s)^3}\mathrm{e}^{-\zeta(\zeta-1)s}\,ds. \tag{1.22}$$

To find a closed form solution to our recurrence, we need only find the power series for $f(x)$. We begin by expanding $1/(1-s)^3$ as an infinite sum and exchanging the order of summation and integration:

$$f(x) = \frac{(1-x)^2}{x^{\zeta^2}}\mathrm{e}^{\zeta(\zeta-1)x}\sum_{n=0}^{\infty}(n+2)(n+1)\int_0^x s^{\zeta^2+n}\mathrm{e}^{-\zeta(\zeta-1)s}\,ds. \tag{1.23}$$

The integral in (1.23) may be integrated by parts repeatedly, forming a convergent, infinite series:

$$\begin{aligned} f(x) &= (1-x)^2\frac{\mathrm{e}^{\zeta(\zeta-1)x}}{x^{\zeta^2}}\sum_{n=0}^{\infty}(n+2)(n+1)\sum_{k=1}^{\infty}(\zeta(\zeta-1))^{k-1}\frac{\Gamma(\zeta^2+n+1)}{\Gamma(\zeta^2+n+k+1)}\frac{x^{\zeta^2+n+k}}{\mathrm{e}^{\zeta(\zeta-1)x}} \\ &= (1-x)^2\sum_{n=0}^{\infty}(n+2)(n+1)\sum_{k=1}^{\infty}(\zeta(\zeta-1))^{k-1}\frac{\Gamma(\zeta^2+n+1)}{\Gamma(\zeta^2+n+k+1)}x^{n+k}. \end{aligned} \tag{1.24}$$

We then substitute $m = n + k$ into the inner summand and reverse the order of summation:

$$\begin{aligned} f(x) &= (1-x)^2\sum_{n=0}^{\infty}(n+2)(n+1)\sum_{m=n+1}^{\infty}(\zeta(\zeta-1))^{m-n-1}\frac{\Gamma(\zeta^2+n+1)}{\Gamma(\zeta^2+m+1)}x^m \\ &= (1-x)^2\sum_{m=1}^{\infty}\left[\sum_{n=0}^{m-1}(n+2)(n+1)\frac{\Gamma(\zeta^2+n+1)}{\Gamma(\zeta^2+m+1)}(\zeta(\zeta-1))^{m-n-1}\right]x^m. \end{aligned} \tag{1.25}$$

The inner sum has a closed form in terms of gamma functions,

$$\begin{aligned} &\sum_{n=0}^{m-1}(n+2)(n+1)\frac{\Gamma(\zeta^2+n+1)}{\Gamma(\zeta^2+m+1)}(\zeta(\zeta-1))^{m-n-1} \\ &\qquad = \zeta^m(\zeta-1)^{m+1}\frac{\Gamma(\zeta^2+1)}{\Gamma(\zeta^2+m+1)} + (1+m-\zeta), \end{aligned} \tag{1.26}$$

which is readily verified by induction.

To arrive at the coefficients for x^m in the power series of $f(x)$, we expand $(1-x)^2$ and simplify:

$$v_m = \begin{cases} 0 & m = 0 \\ 2/(1+\zeta^2) & m = 1 \\ [2(\zeta-1)(2\zeta+1)]/[(\zeta^2+1)(\zeta^2+2)] & m = 2 \\ m(2\zeta+m-1)\frac{\Gamma(\zeta^2)}{\Gamma(\zeta^2+m+1)}\zeta^m(\zeta-1)^{m-1} & m \geq 3. \end{cases} \tag{1.27}$$

1.2.5 Numerical Evaluation of the Transient Dynamics

This review of the NPM is now concluded with a few remarks on its numerical solution. First, mammalian prion diseases are characterized by a long "lag-time" between the introduction of an aggregate and noticeable symptoms [10, 28], which then progress rapidly – thus any good mathematical model must necessarily exhibit multi-scale behavior and stiff dynamics, making its numerical solution challenging and potentially requiring the use of implicit methods. These features are depicted in Figures 1.2 and 1.3. Second, although the continuous-size model may admit a more convenient asymptotic analysis, the PDE has the form of an advection equation with global dependencies, which presents a number of challenges related to the choice of discretization. On the other hand, the discrete model offers a straight-forward approach and efficient implementation via the inversion of a lower-triangular matrix at every time-step. The inverse operator may even be written in a matrix-free form, further reducing its memory footprint. As such, when calculation of transient behavior is required, it is recommended to consider only the discrete model. Due to the moment-closure, one can solve to numerical precision any truncation of the full $[n, \infty)$ size-density.

A numerical implementation in C is provided for the discrete system[5] that leverages the SUNDIALS [27] software suite. A non-dimensionalized and transformed representation of the system is solved for, which is described below. Though an implementation for the continuous NPM is not provided, I offer a few simplifications that may be useful in the development of a solver or additional analysis.

Discrete NPM

Rather than solving for $(x(t), y(t), z(t))$, a more fundamental quantity is the average displacement: a non-negative, non-dimensional measurement defined by

$$W(t) = \frac{z(t)}{y(t)} - n. \tag{1.28}$$

Further scaling $x(t) = \frac{\alpha}{\mu} X(\gamma t)$, $y(t) = \frac{\alpha}{\mu} Y(\gamma t)$, and $z(t) = \frac{\alpha}{\mu} Z(\gamma t)$, one may instead solve the ODE system

$$X' = r\,(1 - X) + n(n-1)Y - bXY, \tag{1.29}$$
$$Y' = (W - (n - 1 + r))\,Y, \tag{1.30}$$
$$W' = bX - W(W + 1), \tag{1.31}$$

with non-dimensional constants $r = \frac{\mu}{\gamma}$ and $b = \frac{2\alpha\beta}{\gamma\mu}$. The size density $\{u_i(t)\}$ is additionally mapped to a probability mass over the natural numbers by defining

$$f_i(\gamma t) = u_{i+n-1}(t)/y(t), \tag{1.32}$$

[5] The code is located in the Git repository `http://www.github.com/jasondark/dissertation` in the folder `dnpm`.

such that

$$f_i' = -bX(f_i - f_{i-1}) - (i - 1 + W)f_i + 2\left(1 - \sum_{j=1}^{i} f_i\right). \tag{1.33}$$

The numerical implementation solves $(X, Y, W, f_1, f_2, \dots)$ for the requested time intervals. Using parameter values from [41], typical solutions are plotted in Figures 1.2 and 1.3.

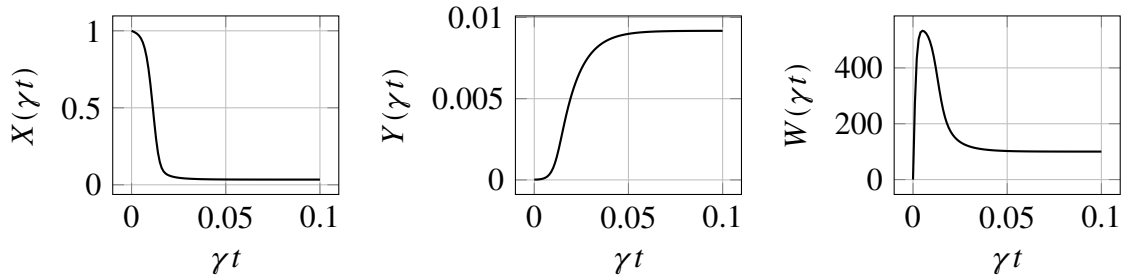


Figure 1.2: (X, Y, W) vs γt (non-dimensionalized time) with $r = 96.25$, $b = 3 \times 10^5$, and $n = 5$ (values from [41]) with $X_0 = 1 - 10^{-4}$, $Y_0 = 10^{-4}/n$, and $W_0 = 0$ corresponding to the spontaneous conversion of 0.01% of the protein into minimum-sized aggregates.

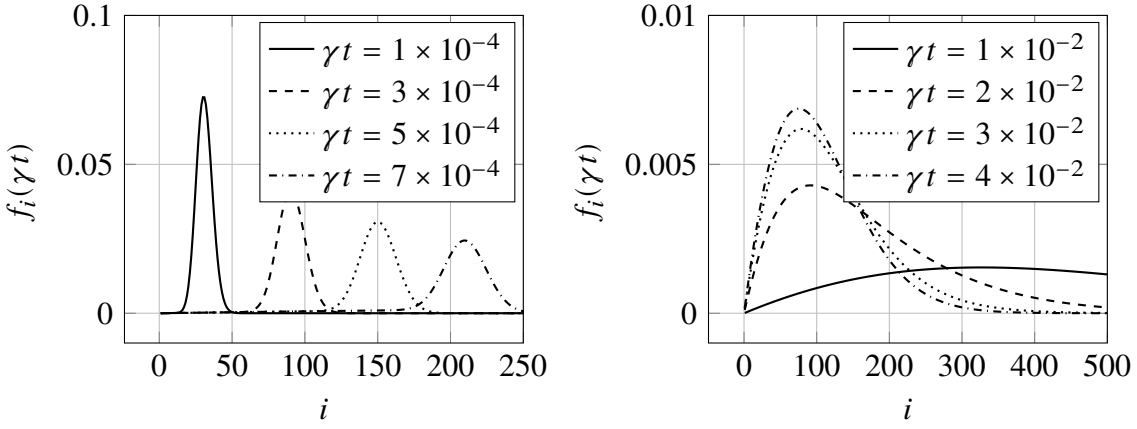


Figure 1.3: Time evolution of the probability density with parameters from [41] and initial conditions as in Figure 1.2. Note the initial fast-traveling wave (corresponding to rapid incorporation of protein monomers by the new aggregate), followed by a slower convergence to the asymptotic density.

Though this transformation is not utilized, there exists a transformation that reduces the system even further and may be useful for other analysis. Define

$$v_i = i \sum_{j=1}^{i} f_j - \sum_{j=1}^{i} j f_j. \tag{1.34}$$

Then $v_0(t) \equiv v_1(t) \equiv 0$ and for $i > 1$,

$$v_i' = -bX(v_i - v_{i-1}) - (i + W)v_i + i(i - 1). \tag{1.35}$$

The $\{f_i\}$ are recovered by the relation $f_i = v_{i+1} - 2v_i + v_{i-1}$.

Continuous NPM

Rescaling as in the discrete version, with the equivalent non-dimensional parameters, one obtains the moment-closed system

$$X' = r\,(1 - X) + a^2 Y - bXY, \tag{1.36}$$
$$Y' = (W - (a + r))\,Y, \tag{1.37}$$
$$W' = bX - W^2. \tag{1.38}$$

Defining again

$$f(t, x) = \tilde{u}(t, x + a)/\tilde{y}(t), \tag{1.39}$$

and

$$v(t, x) = x \int_0^x f(t, y)\,\mathrm{d}y - \int_0^x y f(t, y)\,\mathrm{d}y, \tag{1.40}$$

(as originally suggested in [18]), one arrives at the PDE

$$v_t = -bXv_x - (x + W)v + x^2. \tag{1.41}$$

This more canonical representation may be solved by characteristics, for example. The distribution $f(t, x)$ is recovered by the relation $f(t, x) = v_{xx}(t, x)$.

1.3 Motivation and Goals

Prion diseases are universally fatal in mammals, and in addition to their health impact, there are considerable economic costs associated with the bovine form ("mad cow disease") of these transmissible spongiform encephalopathy (TSE) diseases [40]. Understanding, preventing, and treating TSEs is an important goal in its own right, but these diseases are also symptomatically linked to other protein misfolding diseases such as Alzheimer's and Parkinson's diseases [6].

To aid our understanding, we look towards simpler, model organisms to study and understand the nature of prion disease. While experiments in transgenic mice have validated the protein-only mechanism of these diseases [43], yeast also have their own class of prion diseases. Interestingly, some yeast colonies have the ability to cure themselves of the prion infection [33, 15] – ultimately we wish to understand the underlying causes effecting this outcome, as well as ways to prevent prion diseases in the first place.

In order to do so, we appeal to mathematical modeling. However, as reviewed, the existing models have been developed and validated under *in vitro* observations. As such, they have been unable to recreate a number of important physiological characteristics, including strain coexistence and variable phenotype induction rates (where a yeast colony does not necessarily become $[PSI^+]$ after having prions introduced). Furthermore, by virtue of the *in vitro* assumptions, little work has been done on mathematically

understanding the *origin* of the prion conformation – that is, once an infectious conformation of a prion protein arises in a population it may spread via the understood mechanisms of nucleated polymerization, but what governs the *spontaneous nucleation* of the first prion aggregate?

In Chapter 2, I incorporate the Hsp104 chaperone into a model of nucleated polymerization, which yields the first mathematical model permitting prion strain coexistence. While it suggests a possible mechanism for variable phenotype induction, it occurs in a regime where stochastic, "mesoscopic" effects are also necessarily present.

Thus, in Chapter 3 I develop the first stochastic model of prion phenotype induction and provide evidence that prion strains not only have different biochemical rate parameters, but different minimum stable sizes. (An observation previously unconsidered in the literature.)

In Chapter 4, I review the coagulation literature and highlight challenges in modeling the nucleation time of a prion aggregate. I derive a number of product-form distributions for the asymptotic distributions of proteins and compute moments of nucleation time distributions under a variety of assumptions. Lacking biological data to validate our models, I conclude with a falsifiable hypothesis of the nucleation mechanism.

Finally, Chapter 5 outlines future work, with an emphasis on further generalizing the stochastic models I have developed to better model *in vivo* prion phenotype propagation. I also outline the development of general purpose software for solving the chemical master equation, inspired by my work in computing nucleation time from Chapter 4.

Bibliography

[1] T. Alper, D. Haig, and M. Clarke. “The exceptionally small size of the scrapie agent”. *Biochemical and biophysical research communications* 22.3 (1966), pp. 278–284.

[2] M. E. Bradley and S. W. Liebman. “Destabilizing interactions among [PSI+] and [PIN+] yeast prion variants”. *Genetics* 165.4 (2003), pp. 1675–1685.

[3] P. Brundin, R. Melki, and R. Kopito. “Prion-like transmission of protein aggregates in neurodegenerative diseases”. *Nature Reviews Molecular Cell Biology* 11.4 (2010), pp. 301–307.

[4] J. D. Bryngelson et al. “Funnels, pathways, and the energy landscape of protein folding: a synthesis”. *Proteins: Structure, Function, and Bioinformatics* 21.3 (1995), pp. 167–195.

[5] S. Chakrabortee et al. “Luminidependens (LD) is an Arabidopsis protein with prion behavior”. *Proceedings of the National Academy of Sciences* (2016), p. 201604478.

[6] F. Checler and B. Vincent. “Alzheimer’s and prion diseases: distinct pathologies, common proteolytic denominators”. *Trends in neurosciences* 25.12 (2002), pp. 616–620.

[7] Y. O. Chernoff. “Are there prions in plants?” *Proceedings of the National Academy of Sciences* 113.22 (2016), pp. 6097–6099.

[8] Y. O. Chernoff et al. “Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]”. *Science* 268.5212 (1995), pp. 880–884.

[9] S. R. Collins et al. “Mechanism of prion propagation: amyloid growth occurs by monomer addition”. *PLoS biology* 2.10 (2004), e321.

[10] J. H. Come, P. E. Fraser, and P. T. Lansbury. “A kinetic model for amyloid formation in the prion diseases: importance of seeding”. *Proceedings of the National Academy of Sciences* 90.13 (1993), pp. 5959–5963.

[11] F. Crick et al. “Central dogma of molecular biology”. *Nature* 227.5258 (1970), pp. 561–563.

[12] F. H. Crick. “On protein synthesis.” *Symposia of the Society for Experimental Biology*. Vol. 12. 1958, p. 138.

[13] J. K. Davis and S. S. Sindi. “A study in nucleated polymerization models of protein aggregation”. *Applied Mathematics Letters* 40 (2015), pp. 97–101. issn: 0893-9659. doi: 10.1016/j.aml.2014.09.007.

[14] S. K. DebBurman et al. “Chaperone-supervised conversion of prion protein to its protease-resistant form”. *Proceedings of the National Academy of Sciences* 94.25 (1997), pp. 13938–13943.

[15] A. Derdowski et al. “A size threshold limits prion transmission and establishes phenotypic diversity”. *Science* 330.6004 (2010), pp. 680–683.

[16] M. Doumic, T. Goudon, T. Lepoutre, et al. “Scaling limit of a discrete prion dynamics model”. *Communications in Mathematical Sciences* 7.4 (2009), pp. 839–865.

[17] M. Eigen. “Prionics or the kinetic basis of prion diseases”. *Biophysical Chemistry* 63.1 (1996). issn: 03014622. doi: 10.1016/S0301-4622(96)02250-8.

[18] H. Engler, J. Prüss, and G. F. Webb. “Analysis of a model for the dynamics of prions II”. *Journal of Mathematical Analysis and Applications* 324.1 (2006), pp. 98–117. issn: 0022247X. doi: 10.1016/j.jmaa.2005.11.021.

[19] M. Fan, M. Y. Li, and K. Wang. “Global stability of an SEIS epidemic model with recruitment and a varying total population size”. *Mathematical Biosciences* 170.2 (2001), pp. 199–208. issn: 00255564. doi: 10.1016/S0025-5564(00)00067-5.

[20] L. Fioriti et al. “The persistence of hippocampal-based memory requires protein synthesis mediated by the prion-like protein CPEB3”. *Neuron* 86.6 (2015), pp. 1433–1448.

[21] T. Frängsmyr. *The Nobel Prizes 1997*. Almqvist & Wiksell Intl, 1998.

[22] M. Goedert. “Alzheimer’s and Parkinson’s diseases: The prion concept in relation to assembled Aβ, tau, and α-synuclein”. *Science* 349.6248 (2015), p. 1255555.

[23] M. L. Greer, L. Pujo-Menjouet, and G. F. Webb. “A mathematical analysis of the dynamics of prion proliferation”. *Journal of Theoretical Biology* 242.3 (2006), pp. 598–606. issn: 00225193. doi: 10.1016/j.jtbi.2006.04.010.

[24] J. S. Griffith. “Self-replication and scrapie.” *Nature* 215.5105 (1967), p. 1043.

[25] R. Halfmann, S. Alberti, and S. Lindquist. “Prions, protein homeostasis, and phenotypic diversity”. *Trends in cell biology* 20.3 (2010), pp. 125–133.

[26] J. Hardy and D. J. Selkoe. “The amyloid hypothesis of Alzheimer’s disease: progress and problems on the road to therapeutics”. *Science* 297.5580 (2002), pp. 353–356.

[27] A. C. Hindmarsh et al. “SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers”. *ACM Transactions on Mathematical Software (TOMS)* 31.3 (2005), pp. 363–396.

[28] J. T. Jarrett and P. T. Lansbury Jr. “Seeding “one-dimensional crystallization” of amyloid: a pathogenic mechanism in Alzheimer’s disease and scrapie?” *Cell* 73.6 (1993), pp. 1055–1058.

[29] Z. Jaunmuktane et al. “Evidence for human transmission of amyloid-[bgr] pathology and cerebral amyloid angiopathy”. *Nature* 525.7568 (2015), pp. 247–250.

[30] G. Jung and D. C. Masison. “Guanidine hydrochloride inhibits Hsp104 activity in vivo: a possible explanation for its effect in curing yeast prions”. *Current microbiology* 43.1 (2001), pp. 7–10.

[31] Y. E. Kim et al. “Molecular chaperone functions in protein folding and proteostasis”. *Annual review of biochemistry* 82 (2013), pp. 323–355.

[32] D. S. Kryndushkin et al. “Yeast [PSI+] prion aggregates are formed by small Sup35 polymers fragmented by Hsp104”. *Journal of Biological Chemistry* 278.49 (2003), pp. 49636–49643.

[33] V. V. Kushnirov et al. “Chaperones that cure yeast artificial [PSI+] and their prion-specific effects”. *Current Biology* 10.22 (2000), pp. 1443–1446.

[34] J. Masel, V. A. A. Jansen, and M. A. Nowak. “Quantifying the kinetic parameters of prion replication”. *Biophysical Chemistry* 77.2-3 (1999), pp. 139–152. issn: 03014622. doi: 10.1016/S0301-4622(99)00016-2.

[35] M. A. Nowak et al. “Prion infection dynamics”. *Integrative Biology: Issues, News, and Reviews* 1.1 (1998), pp. 3–15. issn: 10934391. doi: 10.1002/(SICI)1520-6602(1998)1:1<3::AID-INBI2>3.3.CO;2-0.

[36] S. B. Prusiner. “Novel proteinaceous infectious particles cause scrapie”. *Science* 216.4542 (1982), pp. 136–144.

[37] S. B. Prusiner. “Prions”. *Proceedings of the National Academy of Sciences* 95.23 (1998), pp. 13363–13383.

[38] J. Prüss and L. Pujo-Menjouet. “Analysis of a model for the dynamics of prions”. *Discrete and Continuous Dynamical Systems - Series B* 6.1 (2006), pp. 225–235.

[39] B. L. Raveendra et al. “Characterization of prion-like conformational changes of the neuronal isoform of Aplysia CPEB”. *Nature structural & molecular biology* 20.4 (2013), pp. 495–501.

[40] W. Schlenker and S. B. Villas-Boas. “Consumer and market responses to mad cow disease”. *American Journal of Agricultural Economics* 91.4 (2009), pp. 1140–1152.

[41] M. Tanaka et al. “The physical basis of how prion conformations determine strain phenotypes”. *Nature* 442.7102 (2006), pp. 585–589.

[42] J. Tyedmers, M. L. Madariaga, and S. Lindquist. “Prion switching in response to environmental stress”. *PLoS Biol* 6.11 (2008), e294.

[43] C. Weissmann and E. Flechsig. “PrP knock-out and PrP transgenic mice in prion research”. *British medical bulletin* 66.1 (2003), pp. 43–60.

Chapter 2

Enzyme Limited Nucleated Polymerization Model (Journal Article)

This chapter originally published as "Davis, J. K., & Sindi, S. S. (2016). A mathematical model of the dynamics of prion aggregates with chaperone-mediated fragmentation. *Journal of Mathematical Biology*, 72(6), 1555-1578." Reprinted in accordance with the Creative Commons Attribution 4.0 International License (`http://creativecommons.org/licenses/by/4.0/`) with only minor formatting changes.

The co-author listed in this publication directed and supervised research which forms the basis for the dissertation.

2.1 Abstract

Prions are proteins most commonly associated with fatal neurodegenerative diseases in mammals but are also responsible for a number of harmless heritable phenotypes in yeast. These states arise when a misfolded form of a protein appears and, rather than be removed by cellular quality control mechanisms, persists. The misfolded prion protein forms aggregates and is capable of converting normally folded protein to the misfolded state through direct interaction between the two forms.

The dominant mathematical model for prion aggregate dynamics has been the nucleated polymerization model (NPM) which considers the dynamics of only the normal protein and the aggregates. However, for yeast prions the molecular chaperone Hsp104 is essential for prion propagation. Further, although mammals do not express Hsp104, experimental assays have shown Hsp104 also interacts with mammalian prion aggregates.

In this study, we generalize the NPM to account for molecular chaperones and develop what we call the enzyme-limited nucleated polymerization model (ELNPM). We discuss existence, uniqueness and stability of solutions to our model and demonstrate that the NPM represents a quasi-steady-state reduction of our model. We validate the ELNPM by demonstrating agreement with experimental results on the yeast prion $[PSI^+]$ that could not be supported by the NPM. Finally, we demonstrate that, in contrast to the NPM, the ELNPM permits the coexistence of multiple prion strains.

2.2 Introduction

The central dogma of molecular biology stipulates that phenotypes, an organism's expressed states, are determined by genotypes, the vertically transmitted DNA [11]. However, the link between genotype and phenotype is not always this direct. Today we understand that a number of phenotypes are determined epigenetically, without a change to the nucleotide sequence of DNA [20]. In 1965, a number of yeast phenotypes were found to violate the laws of Mendelian inheritance and were thus inconsistent with DNA-based transmission [10]. Further experimental studies demonstrated that the phenotypic states were not the function of the underlying DNA but were the function of a misfolded (prion) protein [44]. As such, the phenotypes were transmitted by the proteins themselves. This phenomenon of "protein only inheritance," also called the prion hypothesis, has over time gone from highly controversial to commonly accepted [43]. Today, nearly a dozen proteins in yeast have been shown to be able to behave as prions [29]. Of course, prions extend far beyond yeast. In mammals, prions are associated with a number of irreversible fatal neurological diseases such as Creutzfeldt-Jakob disease, fatal familial insomnia, chronic wasting disease and bovine spongiform encephalopathy. Mammalian prion diseases have varying modes of transmission and have been shown to be able to pass from one species to another. At present, all mammalian prion diseases are the result of a single protein, PrP [1]. In addition, prion diseases are closely related to other protein misfolding diseases such as Parkinson's, Huntington's, Alzheimer's diseases [4, 27, 3].

Although humans are vastly different from yeast, the dynamics of prion proteins in both hosts is quite similar. Both mammals and yeast have cellular machinery dedicated to identifying and removing misfolded proteins [31] – prion proteins are capable of evading such protective mechanisms and transmitting their misfolded (prion) state to other normally folded proteins. Prion proteins aggregate into complexes which act as templates for initiating further misfolding of normally folded protein. These aggregated complexes may also fragment into smaller units, each of which can template further misfolding [39, 43]. Finally, in order to spread the prion state to a colony or throughout a tissue, prion aggregates must be transmitted to other cells. In yeast colonies, prion aggregates are transmitted from mother to daughter cells during cell division [42, 13]. In mammalian prion diseases, PrP aggregates are thought to be transmitted extracellularly [9]. For other mammalian neurodegenerative diseases, there is a growing body of evidence suggesting neuron to neuron propagation of the misfolded proteins [3].

Many mathematical models have been developed to study the dynamics of prion aggregates primarily in the context of the mammalian host [30, 35, 21, 6]. Tanaka et al. [41] applied these mathematical models to the $[PSI^+]$ prion in yeast. However, experimental studies of $[PSI^+]$ have shown that the molecular chaperone Hsp104 is essential for fragmentation, and recent studies have demonstrated that Hsp104 acts in a rate limiting fashion with respect to fragmentation [36, 13]. We also note that fragmentation is important not only for efficient conversion of normal protein by providing more templates, but also to ensure there are sufficiently many templates

to allow efficient transmission; thus an accurate model of fragmentation is essential to understanding the *in vivo* dynamics of prion aggregates. As such, to accurately model prion aggregates in yeast, the dynamics of Hsp104 and its interaction with prion aggregates must be considered.

Further, modeling Hsp104 will lend insight to more general prion and protein misfolding disorders. Although no chaperones are known to be involved in mammalian prion dynamics, prion amplification in mammals necessarily requires aggregate fragmentation [30]. In addition, while mammals do not express Hsp104, recent *in vitro* work demonstrates that engineered mutants of Hsp104 suppress the toxicity of misfolded protein aggregates associated with mammalian neurodegenerative disorders [25]. At present, no mathematical model exists which considers the dynamics of protein aggregates in the presence of a chaperone mediating fragmentation.

In this study we develop a mathematical model of prion dynamics where fragmentation requires the interaction of Hsp104 with aggregated proteins. In Sections 2 and 3 we provide the mathematical background and analysis of our model, which we call the Enzyme-Limited Nucleated Polymerization Model (ELNPM). In Section 4 we illustrate the necessity of including enzyme-limited fragmentation by demonstrating important experimental properties of $[PSI^+]$ that are not described by previously published mathematical models. We also demonstrate that in contrast to models which consider only the prion aggregates, interactions with the enzyme Hsp104 permit stable co-existence of multiple prion strains. In Section 5 we provide a summary and concluding remarks.

2.3 Mathematical Models of Prion Aggregate Fragmentation

We develop our model of enzyme-mediated fragmentation by considering the key biochemical processes involved in the dynamics of prions. We first discuss the dynamics included in previous mathematical formulations and then detail the additional features necessary to depict interactions between enzymes and aggregates. Finally, we demonstrate that through a series of assumptions, consistent with the yeast prion $[PSI^+]$, our system of infinite ordinary differential equations can be analyzed with a 5-dimensional system of differential equations which approximates the full system dynamics.

2.3.1 Prion Aggregate Dynamics

While the biochemical processes depicted differs between prion model formulations [30, 35, 21, 6], in all cases aggregates change size through conversion and fragmentation. That is, a prion aggregate increases in length by actively converting and incorporating normal (healthy) protein monomers. Typically, aggregates are assumed to be linear fibrils and, as such, conversion of normal protein can only take place on one of the fibril ends. Aggregates may also fragment into 2 smaller aggregates, each of which now act as a

template to convert additional protein. It is often assumed that any aggregates smaller than the minimum stable size, n_0, immediately disassociate into healthy prion monomers (see Figure 2.1).

Such models are referred to as nucleated polymerization models (NPM); mathematical formulations of the NPM were first introduced and subsequently validated by Nowak et al. [32] and Masel, Jansen, and Nowak [30]. This model is so-named due to the assumption that there is a minimum "stable" size of a prion aggregate (a nucleus). The spontaneous formation of such an initial nucleus (or seed, as it is also called) is the time-limiting step in prion disease initialization, but once seeded, the disease progresses primarily by the processes of conversion, fragmentation and transmission.

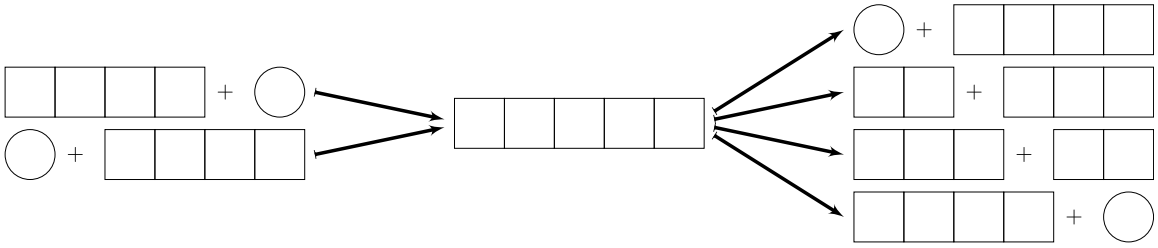


Figure 2.1: Nucleated Polymerization Model: Conversion and Fragmentation ($n_0 = 2$). Conversion of healthy protein (circles) lengthens the aggregate (squares), which may in turn fragment. If a daughter fragment is smaller than the stable nucleus size (n_0), it is immediately disassociated into healthy protein monomers.

The NPM equations are derived from the Law of Mass Action applied to a minimal set of kinetic rate equations. Masel, Jansen, and Nowak [30] give them as

$$s' = \alpha_s - \mu_s s(t) - 2\beta s(t) \sum_{i=n_0}^{\infty} u_i(t) + \gamma (n_0 - 1) n_0 \sum_{i=n_0}^{\infty} u_i(t), \tag{2.1}$$

$$u'_m = -2\beta s(t)[u_m(t) - u_{m-1}(t)] - [\mu_0 + \gamma(m-1)]u_m(t) + 2\gamma \sum_{i=m+1}^{\infty} u_i(t), \tag{2.2}$$

where $s(t)$ denotes the concentration of healthy protein and $u_m(t)$ the density of aggregates of size m. Many authors [21, 35, 17] have studied the more analytically-tractable equations that come from a continuous relaxation of aggregate sizes:

$$s' = \alpha_s - \mu_s s(t) - 2\beta s(t) \int_{x_0}^{\infty} u(t,x) + \gamma {x_0}^2 \int_{x_0}^{\infty} u(t,x), \tag{2.3}$$

$$\frac{\partial u}{\partial t} = -2\beta s(t) \frac{\partial u}{\partial x} - [\mu_0 + \gamma x] u(t,x) + 2\gamma \int_x^{\infty} u(t,y)\, dy. \tag{2.4}$$

Though it is known that this latter system converges weakly to the former in the limit of large average aggregate size under very general assumptions [15, 14], we choose to generalize the discrete model for simplicity.

In adapting this prion model to yeast, we identify the kinetic parameters as representing the following physical quantities:

- α_s the basal rate of transcription of Sup35,
- μ_s the decay (or dilution) rate of Sup35,
- n_0 the minimum stable aggregate size (aggregates of size smaller than n_0 immediately disassociate into soluble Sup35),
- μ_0 the decay (or dilution) rate of aggregated protein,
- β the rate of conversion of healthy protein (from the end of a prion filament), and
- $\gamma(m-1)$ the fragmentation rate of a prion aggregate of size m.

2.3.2 Enzyme-Mediated Fragmentation

We now draw attention to an underlying assumption in these equations that we will modify: for the mammalian prion PrP, it was assumed that the fragmentation rate is an intrinsic function of the aggregate size itself. However, with yeast prion systems, it has been demonstrated that fragmentation requires the additional presence of heat-shock protein 104 (Hsp104) [36]. Its under- or over-expression can eliminate prion aggregates entirely [7]. Additionally, over-expressing Sup35 results in a translational shift in the aggregate-size density [13] – the model equations of Masel, Jansen, and Nowak [30] do not admit such behavior (see Section 2.5.3), suggesting the possibility of a rate-limited fragmentation mechanism rooted in the Hsp104 interactions.

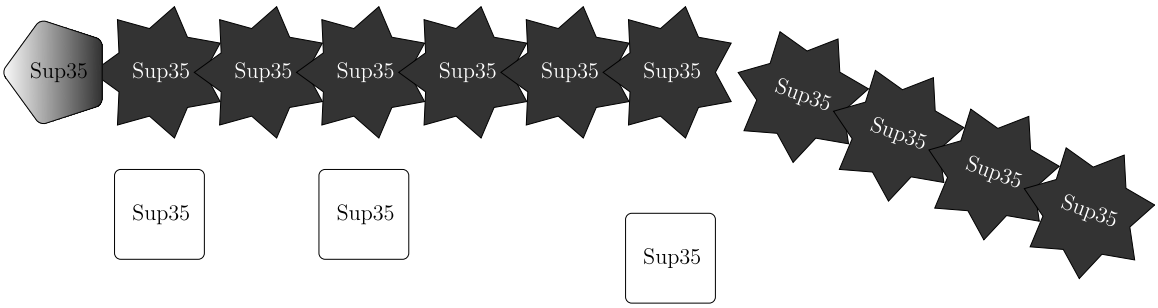
Though yeast prion systems have been studied with the NPM [41], we believe the impact of Hsp104 to be nonnegligible and explicitly consider the Hsp104 concentration and dynamics. We assume a prion aggregate of size i has $i-1$ sites to which a hexamer (the active unit) of Hsp104 can bind and subsequently fragment – we denote such an aggregate with j bound hexamers as $X_{i,j}$. We note that there are actually $\binom{i-1}{j}$ unique configurations of bound Hsp104 that $X_{i,j}$ could refer to, but the proposed kinetic equations will best be described by the amount of Hsp104 bound, not their configuration. We use standard terminology for the enzyme kinetics (parameters k_{on} and k_{off}), and for simplicity, we do not model the formation of Hsp104 hexamers from monomers explicitly. We additionally define α_h and μ_h for Hsp104 similarly as α_s and μ_s for Sup35. Lastly, while we write the dilution rates μ_s, μ_h, and μ_0 separately, we will take them as all equal to the rate of growth of the yeast cell in our numerical experimentation. We now propose our generalization in the form of the following kinetic relations (and illustrate in Figure 2.2).

Translation and Dilution

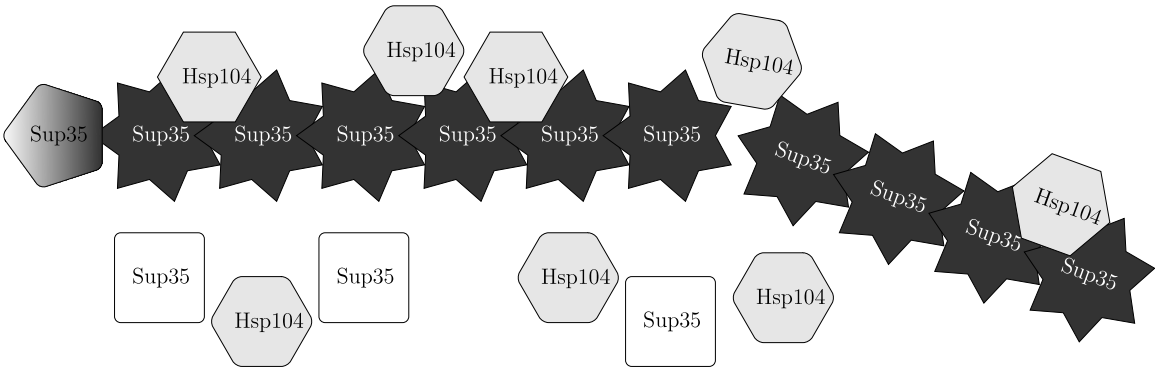
$$\emptyset \underset{\alpha_s}{\overset{\mu_s}{\rightleftharpoons}} \text{Sup35}, \qquad \emptyset \underset{\alpha_h}{\overset{\mu_h}{\rightleftharpoons}} \text{Hsp104}, \qquad \emptyset \underset{\mu_0}{\longleftarrow} X_{i,j},$$

Aggregation (Polymerization/Coagulation)

$$\text{Sup35} + X_{i,j} \xrightarrow[2\beta]{} X_{i+1,j},$$



(a) Nucleated Polymerization Model (stochastic fragmentation)



(b) Enzyme-Limited Polymerization Model (mechanistic fragmentation via Hsp104 chaperone)

Figure 2.2: NPM is based on a random breaking of the prion aggregate (stars), while our model introduces the Hsp104 enzyme (hexagons) that mechanistically fragments the aggregate. Healthy Sup35 (squares) are converted by the ends of the prion aggregate in both models; proteins undergoing conversion are represented with pentagons to demonstrate the active change in conformation.

Enzyme Kinetics

$$\text{Hsp104} + X_{i,j} \underset{k_{\text{on}}(i-j-1)}{\overset{k_{\text{off}}(j+1)}{\rightleftharpoons}} X_{i,j+1},$$

Fragmentation (with unspecified probability $\kappa(m,n;i,j)$)

$$X_{i,j} \xrightarrow[\gamma j]{} \begin{cases} X_{m,n} + \text{Hsp104} + X_{i-m,j-n-1} & m, i-m \geq n_0 \\ m\text{Sup35} + (n+1)\text{Hsp104} + X_{i-m,j-n-1} & m < n_0, i-m \geq n_0 \\ X_{m,n} + (j-n)\text{Hsp104} + (i-m)\text{Sup35} & m \geq n_0, i-m < n_0 \\ i\text{Sup35} + j\text{Hsp104} & m, i-m < n_0. \end{cases}$$

The key unknown in our model is the density of configurations of bound Hsp104, which is incorporated into the fragmentation kernel $\kappa(m,n;i,j)$. Typically, all fragmentation sites in an aggregate are taken to be equally likely [30, 35, 21]:

$$\sum_{n=0}^{m-(i-j)} \kappa(m,n;i,j) = \frac{1}{i-1}. \tag{2.5}$$

Furthermore, we require total Sup35 and Hsp104 to be conserved across fragmentation events, which corresponds to

$$\sum_{m=1}^{i-1}\sum_{n=0}^{m-1} m\kappa(m,n;i,j) = \frac{i}{2} \quad \text{and} \quad \sum_{m=1}^{i-1}\sum_{n=0}^{m-1} n\kappa(m,n;i,j) = (j-1)/2. \tag{2.6}$$

We claim that

$$\kappa(m,n;i,j) = \frac{1}{i-1}\frac{\binom{m-1}{n}\binom{i-m-1}{j-n-1}}{\binom{i-2}{j-1}} = \frac{1}{j}\frac{\binom{m-1}{n}\binom{i-m-1}{j-n-1}}{\binom{i-1}{j}} \tag{2.7}$$

is the natural choice, which follows from taking each $\binom{i-1}{j}$ configuration of X_{ij} to be equally likely, which in turn corresponds to enzyme binding acting on a faster time-scale than conversion and fragmentation (refer to the Supplemental Materials for the argument).

2.3.3 Enzyme-Limited Nucleated Polymerization Model

With the biochemical processes defined, we are now able to formally derive our enzyme-limited nucleated polymerization model (ELNPM). We define $s(t)$ as the concentration of soluble Sup35, $\eta(t)$ as the concentration of aggregates, $z(t)$ as the concentration of bound Sup35 ($z(t) \geq n_0\eta(t)$), $h(t)$ as the concentration of unbound Hsp104, and $z_b(t)$ as the concentration of bound Hsp104. Using $[S]$ to denote the concentration of chemical species S, these definitions correspond to $s(t) = [\text{Sup35}](t)$, $h(t) = [\text{Hsp104}](t)$, and

$$\eta(t) = \sum_{i=n_0}^{\infty}\sum_{j=0}^{i-1} u_{ij}(t), \qquad z(t) = \sum_{i=n_0}^{\infty}\sum_{j=0}^{i-1} iu_{ij}(t), \qquad z_b(t) = \sum_{i=n_0}^{\infty}\sum_{j=0}^{i-1} ju_{ij}(t), \tag{2.8}$$

where we have let $u_{ij}(t) = [X_{ij}](t)$. Let us define a new quantity $p(t)$ by the relation $z_b(t) = p(t)[z(t) - \eta(t)]$ and apply the Law of Mass Action to our proposed kinetic

equations. We obtain

$$s'(t) = \alpha_s - \mu_s s(t) - 2\beta s(t)\eta(t) + \gamma(n_0-1)n_0\eta(t)\sum_{i=n_0}^{\infty}\sum_{j=0}^{i-1}\frac{j}{i-1}\frac{u_{ij}(t)}{\eta(t)} \tag{2.9a}$$

$$h'(t) = \alpha_h - \mu_h h(t) + \left[(k_{\text{off}}+\gamma)p(t) - k_{\text{on}}h(t)(1-p(t))\right](z(t)-\eta(t)) + \gamma(n_0-1)(n_0-2)\eta(t)\sum_{i=n_0}^{\infty}\sum_{j=0}^{i-1}\frac{j(j-1)}{(i-1)(i-2)}\frac{u_{ij}(t)}{\eta(t)} \tag{2.9b}$$

$$\eta'(t) = -(\mu_0+\gamma p(t))\eta(t) + \gamma p(t)z(t) - 2\gamma(n_0-1)\eta(t)\sum_{i=n_0}^{\infty}\sum_{j=0}^{i-1}\frac{j}{i-1}\frac{u_{ij}(t)}{\eta(t)} \tag{2.9c}$$

$$z'(t) = 2\beta s(t)\eta(t) - \mu_0 z(t) - \gamma n_0(n_0-1)\eta(t)\sum_{i=n_0}^{\infty}\sum_{j=0}^{i-1}\frac{j}{i-1}\frac{u_{ij}(t)}{\eta(t)} \tag{2.9d}$$

$$p'(t) = k_{\text{on}}h(t)(1-p(t)) - (k_{\text{off}}+\gamma)p(t) + p(t)^2 - \frac{2\beta s(t)\eta(t)p(t)}{z(t)-\eta(t)} + \frac{\gamma(n_0-1)(n_0-2)\eta(t)}{z(t)-\eta(t)}\left(\sum_{i=n_0}^{\infty}\sum_{j=0}^{i-1}\frac{j}{i-1}\frac{u_{ij}(t)}{\eta(t)} - \frac{1}{p(t)}\sum_{i=n_0}^{\infty}\sum_{j=0}^{i-1}\frac{j(j-1)}{(i-1)(i-2)}\frac{u_{ij}(t)}{\eta(t)}\right). \tag{2.9e}$$

We note that u_{ij}/η defines a probability mass function for the joint random variable (I,J) over $\{(i,j) : 0 \le j < i, i \ge n_0\}$, thus the sums may be interpreted as an expected value. In the interest of obtaining a simple set of equations, we approximate

$$\mathbb{E}\left[\frac{J}{I-1}\right] \approx \mathbb{E}[J]/\mathbb{E}[I-1] = z_b/(z-\eta) = p \tag{2.10}$$

and

$$\mathbb{E}\left[\frac{J(J-1)}{(I-1)(I-2)}\right] \approx \mathbb{E}[J]^2/\mathbb{E}[I-1]^2 = p^2. \tag{2.11}$$

The error of this approximation is on the order of the inverse square of the average aggregate size, a size which is typically large by assumption [35, 21, 15]. (This analysis is provided in the Supplemental Materials.)

Making these substitutions, we obtain the enzyme-limited, nucleated polymerization model (ELNPM):

$$s'(t) = \alpha_s - \mu_s s(t) - 2\beta s(t)\eta(t) + \gamma p(t)(n_0-1)n_0\eta(t) \tag{2.12a}$$

$$h'(t) = \alpha_h - \mu_h h(t) + \left[(k_{\text{off}}+\gamma)p(t) - k_{\text{on}}h(t)(1-p(t))\right](z(t)-\eta(t)) \tag{2.12b}$$

$$\eta'(t) = -\left[\mu_0 + \gamma p(t)(2n_0-1)\right]\eta(t) + \gamma p(t)z(t) \tag{2.12c}$$

$$z'(t) = 2\beta s(t)\eta(t) - \mu_0 z(t) - \gamma p(t)n_0(n_0-1)\eta(t) \tag{2.12d}$$

$$p'(t) = k_{\text{on}}h(t)(1-p(t)) - (k_{\text{off}}+\gamma)p(t) - p(t)\left(\frac{2\beta s(t)\eta(t)}{z(t)-\eta(t)} - \gamma p(t)\right), \tag{2.12e}$$

and marginal density equations

$$\frac{du_m}{dt} = -2\beta s(t)[u_m(t) - u_{m-1}(t)] - [\mu + \gamma p(t)(m-1)]u_m + 2\gamma p(t)\sum_{i=m+1}^{\infty}u_i(t). \tag{2.13}$$

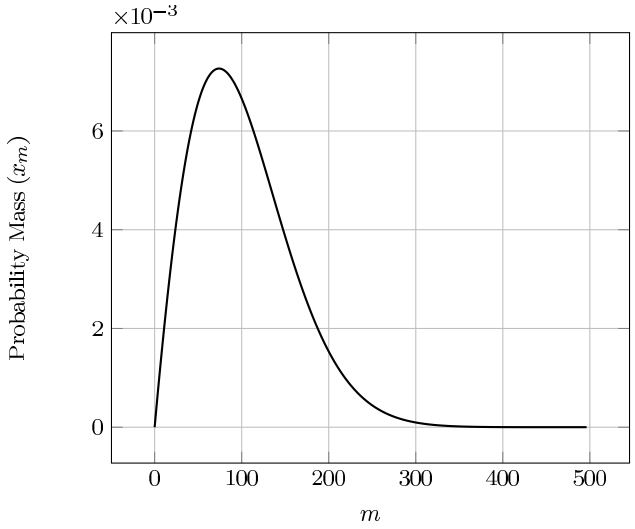


Figure 2.3: Plot of steady-state $x_m = u_{m+n_0-1}/\eta$ with parameter values chosen from [41], appropriately modified to match the steady-state, effective fragmentation rate with the paper's constant rate.

We note that these equations may alternatively be derived by prescribing the binomial form $u_{ij}(t) = \binom{i-1}{j} p(t)^j [1 - p(t)]^{i-j-1} u_i(t)$, then finding the unique $p(t)$ that preserves Hsp104 conservation across fragmentation events.

We further observe that the systems $(s, \{u_m\})$ and (s, η, z) are identical to that of the original NPM [30], except the constant fragmentation rate γ has been replaced by the time-varying quantity $\gamma p(t)$. This provides the interpretation of $p(t)$ as the measure of how effectively the system is fragmenting at a given time. Furthermore, this formulation suggests a quasi-steady-state interpretation of our approximation, since now every aggregate is always bound with Hsp104 proportionally as $p(t)$; that is, the enzyme binding reaches equilibrium before any conversion or fragmentation events occur. We finally note that the first 2 terms in Equation (2.12e) reflect a Michaelis-Menten simplification of the enzyme kinetics [37]; however, since Hsp104 off-binding results in a change in the amount of binding substrate with probability $\gamma/(\gamma + k_{\text{off}})$, we may view the last term as the correction to preserve Hsp104 conservation.

Before detailed analysis of the ELNPM, we briefly examine the qualitative form of the aggregate size distribution. In Figure 2.3 we plot a typical equilibrium solution to Equation (2.13), where we have defined $x_m = u_{n_0-1+m}/\eta$ to be the corresponding probability mass function over the natural numbers. As expected, given the asymptotic similarity of our system to [30] (in that $\gamma p(t)$ presumably converges to a fixed $\tilde{\gamma}$), the equilibrium size density is of the same distribution.

2.4 Analysis of the ELNPM

We first prove a few results on existence and uniqueness for this system and then provide a non-dimensionalized, transformed system we will use to study stability. We analytically demonstrate the stability of the disease-free state and derive conditions which will ensure aggregate persistence.

2.4.1 Existence and Uniqueness of Solutions

Theorem 1. *Trajectories of Equation* (2.12) *remain invariant under a bounded, "feasible" subset of the non-negative cone* $\mathbb{R}^5_+$.

Proof. Let us define the feasible subset to be the set of all (s, h, η, z, p) where $s, h, \eta, z \geq 0$ and $0 \leq p \leq 1$, with the further restriction that $z \geq n_0\eta$, $0 \leq s + z \leq \alpha_s/\mu_0$, and $0 \leq h + p(z - \eta) \leq \alpha_h/\mu_0$. For analytical convenience, and in line with typical parameter regimes, we assume $\mu_0 \leq \mu_s, \mu_h$ and $\frac{\gamma}{2\beta} < \frac{\alpha_s}{\mu_s}$.

Now consider the violation of any single constraint. It is straight-forward to show $s'|_{s<0}, h'|_{h<0}, \eta'|_{\eta<0}, p'|_{p<0} \geq 0$. Similarly, $p'|_{p>1} \leq 0$. Next, $(z - n_0\eta)'|_{z<n_0\eta} = 2\beta s\eta - \mu_0(z - n_0\eta) - n_0\gamma p(z - n_0\eta) \geq 0$, which also demonstrates the non-negativity of z since $z \geq n_0\eta \geq 0$. Writing $\mu_s/\mu_0 = 1+\epsilon/\mu_0$, we have $(s+z)'|_{s+z>\alpha_s/\mu_0} = \alpha_s - \mu_0(s+z) - \epsilon s \leq 0$. With a similar approach, we also find $(h + p(z - \eta))'|_{h+p(z-\eta)>\alpha_h/\mu_0} \leq 0$. □

Theorem 2. *Solutions satisfy the existence and uniqueness criteria within the invariant, feasible region.*

Proof. With the exception of the $\frac{\eta}{z-\eta}$ term in p', the derivatives are polynomial in the dependent variables, which yields continuous partial derivatives. Considering now this last term, $z \geq n_0\eta$ in our feasible region so we have $0 \leq \frac{\eta}{z-\eta} \leq \frac{1}{n_0-1} \leq 1$. This term's partial derivatives are also continuous in this region; let $q(\eta, z) = \eta/(z - \eta)$. Then, $z/\eta = 1 + 1/q$ and

$$\begin{aligned} q_\eta &= \frac{\eta'}{z-\eta} + \frac{\eta(z-\eta)'}{(z-\eta)^2} = \frac{\eta'}{\eta}q + \frac{z'-\eta'}{\eta}q^2 \\ &= q(q-1)\left(\mu_0 + \gamma p(2n_0 - 1) - \gamma p(1/q + 1)\right) \\ &\quad + q^2\left(2\beta s - \mu_0(1 + 1/q) - \gamma p n_0(n_0 - 1)\right) \\ &= (q-1)\left(q\mu_0 + q\gamma p(2n_0 - 1) - \gamma p(q + 1)\right) \\ &\quad + q\left(2q\beta s - \mu_0(1 + q) - q\gamma p n_0(n_0 - 1)\right), \end{aligned} \tag{2.14}$$

and

$$\begin{aligned} q_z &= -\frac{\eta}{(z-\eta)^2}z' = -q^2\frac{z'}{\eta} = -q^2\left(2\beta s - \mu_0(1 + 1/q) - \gamma p n_0(n_0 - 1)\right) \\ &= -q\left(2q\beta s - \mu_0(1 + q) - q\gamma p n_0(n_0 - 1)\right). \end{aligned} \tag{2.15}$$

Thus, q, q_η, q_z are continuous in our region, even as $\eta \to 0^+$. This establishes existence and uniqueness. □

2.4.2 Non-Dimensionalized Equations

We now reduce the ELNPM equations to non-dimensional form,

$$s' = A_s(1-s) - Bs\eta + (n_0-1)n_0 p\eta \tag{2.16a}$$
$$h' = A_h(1-h) + r\left[(\omega+n_0-1)\left(k_{-1}p - k_1 h[1-p]\right) + (n_0-1)(n_0-2)p^2\right]\eta \tag{2.16b}$$
$$\eta' = (\omega - A_0/p - n_0 + 1)\, p\eta \tag{2.16c}$$
$$\omega' = Bs - p(\omega+1)\omega \tag{2.16d}$$
$$p' = k_1 h(1-p) - k_{-1}p - p\left(\frac{Bs}{\omega+n_0-1} - p\right) \tag{2.16e}$$

where we have replaced $z(t)$ by the displacement of the average aggregate size from the minimum size $\omega(t) = z(t)/\eta(t) - n_0 \geq 0$. We have scaled time by γ, $s(t)$ and $\eta(t)$ by α_s/μ_s and $h(t)$ by α_h/μ_h, and used the following non-dimensional constants

$$\begin{gathered} A_s = \mu_s/\gamma, \quad A_h = \mu_h/\gamma, \quad A_0 = \mu_0/\gamma, \quad B = \frac{2\alpha_s\beta}{\gamma\mu_s}, \\ k_{-1} = (k_{\text{off}}+\gamma)/\gamma, \quad k_1 = \frac{k_{\text{on}}\alpha_h}{\gamma\mu_h}, \quad r = \frac{\alpha_s/\mu_s}{\alpha_h/\mu_h}. \end{gathered} \tag{2.17}$$

All subsequent analysis will done with respect to these non-dimensional equations. We note that by construction $k_{-1} > 1$ and by assumption, $B > 1$.

2.4.3 Asymptotic Behavior of ELNPM

With the nondimensional equations established, we next consider the asymptotic behavior of the ELNPM. We call any trajectory satisfying $\lim_{t\to\infty}\eta(t) = 0$ *disease-free*; otherwise, we call the prion state *persistent*. [35] observed that an appropriate transformation could reduce the NPM equations to the standard SEIS model of mathematical epidemiology – a model which is governed entirely by a single parameter R_0 (the basic reproductive number). If $R_0 < 1$, the only equilibrium is disease-free and is globally stable. If $R_0 > 1$, a unique endemic equilibrium appears and exchanges stability with the disease-free state; that is, the endemic equilibrium is asymptotically globally stable [28]. Applying the transformation from Prüss et al. [35], our R_0 will vary in time through its dependence on $p(t)$:

$$R_0(p) = \frac{B/p}{(A_0/p + n_0 - 1)(A_0/p + n_0)}. \tag{2.18}$$

It is convenient to think of the quantity $R_0(p(t)) = R_t$ as the *effective reproductive number* of the disease system, where typically $R_t < R_0$. Though $p(t)$ appears to always converge to a fixed steady-state value, it does so in a non-trivial way making it difficult to provide a Lyapunov analysis. Instead, we provide a local analysis of the disease-free state and offer numerical evidence in support of $R_0(p_{\text{disease-free}})$ determining global stability.

Disease-Free Steady State

At a disease-free equilibrium, we have $\eta = 0$ and subsequently, $s = h = 1$. Thus, we need only study solutions to

$$\begin{aligned} 0 &= B - p\omega(\omega + 1) \\ 0 &= k_1(1-p) - k_{-1}p + p^2 - \frac{Bp}{\omega + n_0 - 1}. \end{aligned} \tag{2.19}$$

This system has 5 solutions in general, though we shall show there is only a single solution inside our feasible region.

Theorem 3. *There is a unique solution to Equation* (2.19) *in our feasible region of trajectories.*

Proof. We may write $p = p(\omega) = B/[\omega(\omega+1)]$; then ω satisfies

$$\begin{aligned} 0 = f(\omega) = B^2(n_0 - 1) - B(n_0 - 1)(k_1 + k_{-1})\omega - [B^2 + Bk_{-1}n_0 + \{1 + (B-1)n_0\}k_1]\omega^2 \\ [-B(k_1 + k_{-1}) + k_1(2n_0 - 1)]\omega^3 + k_1(n_0 + 1)\omega^4 + k_1\omega^5. \end{aligned} \tag{2.20}$$

Since $\frac{k_1}{k_1 + k_{-1}} < 1 < B$, we observe 2 sign changes in the coefficients of $f(\omega)$, implying 0 or 2 real roots by Budan's theorem [2]. We may also write $\omega = \omega(p) = -1/2 + \sqrt{1/4 + B/p}$. Since $p < 1$, we have $\omega > \omega_{\min} = -1/2 + \sqrt{1/4 + B}$. However, $f(0) = B^2(n_0 - 1) > 0$, $f(\omega_{\min}) = -\frac{B^2}{2}(2B + (k_{-1} - 1)(2n_0 - 3 + \sqrt{1 + 4B})) < 0$, and $\lim_{\omega \to \infty} f(\omega) > 0$. Thus, by sign analysis we have an infeasible root $\omega \in (0, \omega_{\min})$ and a feasible root $\omega > \omega_{\min}$. □

We now establish local stability criteria of this root; let p_0 and ω_0 be the unique solution to Equation (2.19).

Theorem 4. *The unique disease-free equilibrium is locally stable when $R_0(p_0) < 1$ and unstable when $R_0(p_0) > 1$.*

Proof. The eigenvalues of the localized Jacobian will satisfy

$$\begin{aligned} 0 = (A_s + \lambda)(A_h + \lambda)(p_0(\omega_0 + n_0 - 1) - A_0 - \lambda) \\ \times \left[\begin{array}{l} \lambda^2 + \left(2\omega_0 p_0 + k_1 + k_{-1} - p_0 + \dfrac{B}{\omega_0 + n_0 - 1}\right)\lambda \\ \quad + p_0(1 + 2\omega_0)\left(k_{-1} + k_1 - 2p_0 + \dfrac{B}{\omega_0 + n_0 - 1}\right) + \dfrac{Bp_0\omega_0(\omega_0 + 1)}{(\omega_0 + n_0 - 1)^2}. \end{array} \right] \end{aligned} \tag{2.21}$$

Of the five roots, the first 2 are clearly negative. The quadratic factor will also admit 2 stable roots: we see that its quadratic and linear coefficients are strictly positive and now show that the constant term is as well. If $k_1 > 0$, then $k_1 + k_{-1} - 2p_0 > 2(1 - p_0) > 0$. If not, then consider p' at $p = k_1$:

$$p' < k_1(1 - k_1) - k_{-1}k_1 + k_1^2 = k_1(1 - k_{-1}) < 0. \tag{2.22}$$

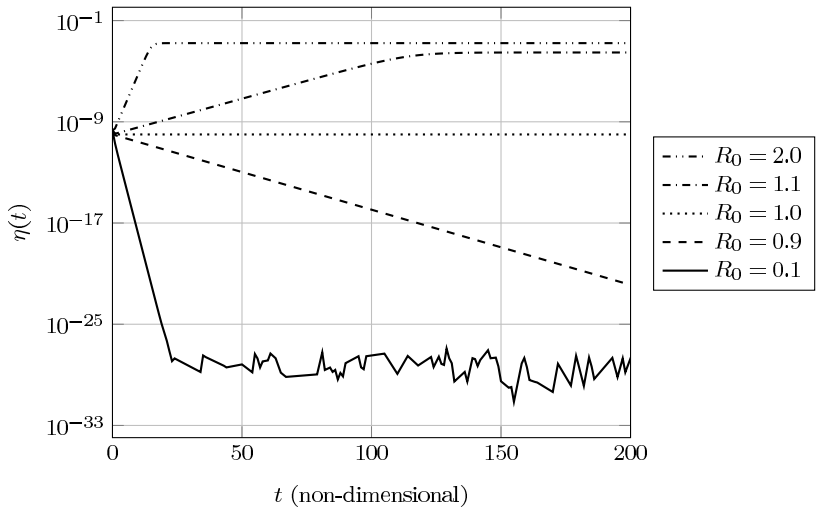


Figure 2.4: Non-dimensionalized plots of aggregate density over time with varying $R_0 = R_0(p_0)$. The system is initialized with a 10^{-9} perturbation of aggregated protein from an otherwise healthy initial state.

So, $p(t) < k_1$ for all time, implying $p_0 < k_1$ and $k_1 + k_{-1} - 2p_0 > k_{-1} - p_0 > 1 - p_0 > 0$.

Finally, the remaining root will be negative when $p_0(\omega_0 - n_0 + 1) - A < 0$. Substituting $\omega_0 = -1/2 + \sqrt{1/4 + B/p_0}$ and simplifying, our expression reduces to

$$R_0(p_0) = \frac{B/p_0}{(A_0/p_0 + n_0)(A_0/p_0 + n_0 - 1)} < 1. \tag{2.23}$$

□

We support this claim numerically in Figure 2.4, where we vary $R_0(p_0)$ across a range of parameters (described in Table 2.1). We observe that when $R_0(p_0) > 1$, there appears to be an attractive *endemic* equilibrium.

Endemic Steady State

The local instability of the disease-free equilibrium yields a persistent disease state; however, the numerical experimentation in Figure 2.4 suggests further that there is an attractive, endemic equilibrium. Generally speaking, steady-state solutions to our system will be solutions of a quintic polynomial in 5 variables, thus preventing closed-form descriptions of such states. However, one can parameterize these values in terms of a

fixed (but unknown) value $\tilde{p}$ corresponding to $p = \tilde{p}$:

$$\tilde{s} = \frac{(A_0/\tilde{p} + n_0 - 1)(A_0/\tilde{p} + n_0)}{B/\tilde{p}} \tag{2.24a}$$

$$\tilde{h} = 1 - r(A_s/A_h)(1 - \tilde{s})\tilde{p}\left(1 - \frac{1}{A_0/\tilde{p} + 2n_0 - 1}\right) \tag{2.24b}$$

$$\tilde{\eta} = (A_s/A_0)\frac{1 - \tilde{s}}{A_0/\tilde{p} + 2n_0 - 1} \tag{2.24c}$$

$$\tilde{\omega} = A_0/\tilde{p} + n_0 - 1, \tag{2.24d}$$

where $\tilde{p}$ satisfies

$$0 = k_1\tilde{h}(1 - \tilde{p}) - k_{-1}\tilde{p} + \tilde{p}^2 - \frac{B\tilde{s}\tilde{p}}{\tilde{\omega} + n_0 - 1}. \tag{2.25}$$

We note that this quadratic is uniquely invertible within our feasible region, which yields the recursive relation

$$\begin{aligned} 2\tilde{p} = &\left(k_{-1} + k_1\tilde{h} + \frac{B\tilde{s}}{\tilde{\omega} + n_0 - 1}\right) \\ &- \sqrt{\left(k_{-1} + k_1\tilde{h} + \frac{B\tilde{s}}{\tilde{\omega} + n_0 - 1}\right)^2 - 4k_1\tilde{h}}. \end{aligned} \tag{2.26}$$

Incidentally, if we suppose $k_1 h + k_{-1} \gg \frac{B\tilde{s}}{\tilde{\omega}+n_0-1}, k_1 h$, then we obtain

$$\tilde{p} = \frac{k_1 h}{k_1 h + k_{-1}} + O\left(\frac{1}{(k_1 h + k_{-1})^2}\right). \tag{2.27}$$

We draw attention to the similarity between Equation (2.27) and what would be the Michaelis-Menten value for $\tilde{p}$.

Our feasible region requires $s, h, \eta, \omega > 0$; $\tilde{s}$ and $\tilde{\omega}$ are always positive, while $\tilde{\eta} > 0 \implies \tilde{s} < 1 \implies R_0(\tilde{p}) > 1$. This also gives us $\tilde{h} > 0$, since we typically assume $\alpha_h \leq \alpha_s \implies r(A_s/A_h) \leq 1$, which implies

$$\tilde{h} > 1 - (1 - \tilde{s})\tilde{p}\left(1 - \frac{1}{A_0/\tilde{p} + 2n_0 - 1}\right) > 0. \tag{2.28}$$

Thus, we will have endemic equilibria when Equation (2.26) has solutions satisfying $R_0(\tilde{p}) > 1$. Based on considerable numerical studies, we conjecture that this solution uniquely exists and is globally asymptotically stable when $R_0(p_0) > 1$.

2.5 Discussion

Formulating the ELNPM allows us to consider aspects of prion aggregate dynamics that cannot be explained by prior mathematical approaches that neglected the role of the Hsp104 chaperone in fragmentation. We first demonstrate that the NPM is a limiting case

of the ELNPM and comment on implications of the ELNPM to the larger question of the appearance of prion strains in a population. We then demonstrate that the ELNPM is the first model capable of supporting two experimentally observed phenomena. First, the ELNPM is the first model capable of reproducing shifts in the aggregate densities associated with increases in synthesis of Sup35. Finally, we demonstrate that the binding kinetics of Hsp104 in the ELNPM allows the possibility of multiple co-existing prion strains. Intriguingly, the co-existence of multiple strains is thought to be crucial towards understanding the transmission of prion diseases between species [8].

2.5.1 The NPM is a Limiting Case of ELNPM

Our moment-closed ELNPM model, Equations (2.9), is nearly identical to the original NPM model but with a time-varying, effective fragmentation rate $\gamma p(t)$ instead of the constant γ. When $p(t) \to \tilde{p}$, the dynamics of ELNPM will mirror that of the NPM with γ replaced by $\gamma\tilde{p}$. As such, it is convenient to think of the NPM as a quasi-steady-state approximation of the full enzyme kinetics we have considered in our model. We informally used this observation in Section 2.4.3 to motivate (but not prove) global stability based on known results of the NPM system.

We plot in Figure 2.5 $p(t)$ over the same parameters in Figure 2.4 and note that – since yeast has a doubling time of roughly 90 minutes [22] – $p(t)$ will not reach its asymptotic value for a few cell-divisions. As such, even though the NPM represents a quasi-steady-state approximation it may not represent the aggregate dynamics during the early cell divisions following the introduction of prion aggregates.

2.5.2 Transient Fragmentation Efficiency May Impact Prion Stability

Although the NPM may be viewed as an asymptotic simplification of the ELNPM, we remark that the differences in transient behavior may provide insight into the underlying stochastic dynamics that arise when a single prion seed is introduced into a healthy yeast colony. The ELNPM predicts an initial fragmentation rate that can be larger or smaller than the asymptotic rate – this is because the availability of enzyme (Hsp104) is much larger than the availability of substrate (binding sites) in this initial configuration. (See Figure 2.5.)

Since aggregate amplification is essential to spreading a prion disease these transient fragmentation rates may impact prion stability. For example, a higher transient fragmentation rate for a prion strain with low R_0 would represent a barrier to successful "seeding" of the prion state than would otherwise be predicted by a constant fragmentation rate. This provides a plausible mechanism for the removal of an initial prion aggregate appearing in a population and therefore the low frequency of spontaneous appearance of the prion state.

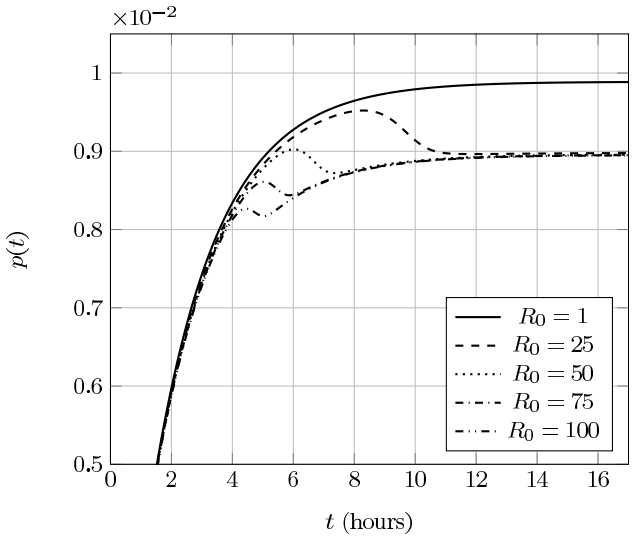


Figure 2.5: $p(t)$ over time with varying $R_0 = R_0(p_0)$; parameters are the same as in Figure 2.4. The transient fragmentation efficiency may be higher (for small R_0) or lower (for large R_0) than the asymptotic efficiency.

2.5.3 Hsp104 Acts as a Rate Limiter for Fragmentation

We noted in Section 2.3 that the NPM does not admit translational shifts in the aggregate density as a function of increasing synthesis of the prion protein (α_s), we now formally demonstrate this is the case. Let us revisit the quantity $u_m(t)$, the density of aggregates of size m. We write $x_m = u_{m+n_0-1}/\eta$ – this quantity defines a probability mass function that is independent of the amount of aggregated protein. In our rescaled variables, we have

$$x'_m = -Bs(x_m - x_{m-1}) - p(m+\omega+1)x_m + 2p - 2p\sum_{i=0}^{m-1} x_i. \tag{2.29}$$

Davis and Sindi [12] gave a closed-form for x_m at steady-state:

$$x_m = m(2\zeta + m - 1)\frac{\Gamma(\zeta^2)}{\Gamma(\zeta^2 + m + 1)}\zeta^m(\zeta - 1)^{m-1}, \tag{2.30}$$

where $\zeta = \frac{\mu}{\gamma\tilde{p}} + n_0$.

The size distribution's dependence on α_s can only occur through its relationship with the steady-state fragmentation efficiency $\tilde{p}$. This is fixed in the NPM, thus the size distribution will not change in response to changes in α_s. This is in contradiction to the experimental results described by Derdowski et al. [13]. Since our model does allow $\tilde{p}$ to vary as a function of the kinetic parameters, we are able to numerically investigate qualitative shifts in the distribution. We demonstrate these shifts in Figure 2.6, which are in qualitative agreement with the experiments of Derdowski et al. [13]. As such, experimental evidence supports that fragmentation can not be purely a function of the number of available fragmentation sites and must depend on the amount of Hsp104 in the system.

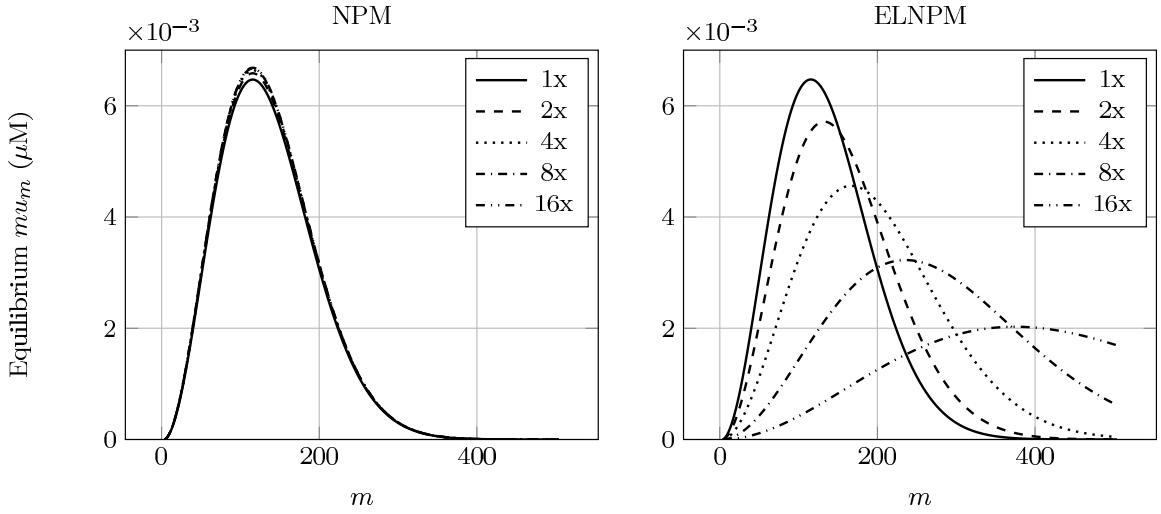


Figure 2.6: Theoretical shifts in the steady-state concentration of aggregate size distributions $\{mu_m\}$ by increasing the synthesis of normal protein (α_s). Left, $mu_m = \eta m x_{m-n_0+1}$ from the NPM. Though x_m is invariant to α_s, η does have a dependency, resulting in the small changes in scaling. Right, mu_m from our ELNPM. Both the scaling and translation are affected by α_s. Initial kinetic parameters are as in Figure 2.3.

2.5.4 Prion Extinction and Hsp104 Expression Levels

Beyond translational shifts in aggregate size distribution, the $[PSI^+]$ prion phenotype in yeast has been shown to be very sensitive to the amount of Hsp104 in the system [23, 16, 38]. Our mathematical formulation correctly captures the dependency of all prions to the under-expression of Hsp104. In contrast, and in agreement with recent experimental studies [26], over-expression of Hsp104 does not necessarily drive prions to extinction.

First, sufficiently high concentrations of guanidine hydrochloride GdnHCl have been shown to severely disrupt the fragmentation process by inactivating Hsp104 [18, 5]. Since fragmentation is halted the total number of aggregates will not change and aggregates will eventually be lost through dilution in the population due to cell division. Over time the population will be cured of the prion disease as the fraction of cells with aggregates approaches zero.

Quantitatively, we treat the inactivation of Hsp104 as letting $k_{\text{on}} \to 0$. In the limit, we'll obtain $p' = -k_1 p + p^2 - \frac{Bsp}{\omega+n_0-1} < -\frac{Bsp}{\omega+n_0-1} \leq 0 \implies p \to 0^+$. With $p = 0$, we have $\eta' = -A_0\eta \leq 0 \implies \eta \to 0^+$, which corresponds to the elimination of prion aggregates. We demonstrate this in Figure 2.7.

While earlier studies seemed to indicate over-expression of Hsp104 could cause loss of the prion state [7], more recent experimental evidence indicates that Hsp104 over-expression is not sufficient to drive prions to extinction [26]. In our formulation over-expressing Hsp104 drives $p \to 1$. This is readily demonstrated by assuming $k_1 h = 1/\epsilon \gg 1$ and rescaling time by this quantity; then $p' = 1 - p + O(\epsilon)$. However, as is experimentally, this alone is not mathematically sufficient to cure the prion state.

Consider an endemic state with $R_0(\tilde{p}) = \frac{B/\tilde{p}}{(A_0/\tilde{p}+n_0)(A_0/\tilde{p}+n_0-1)} > 1$. Depending on

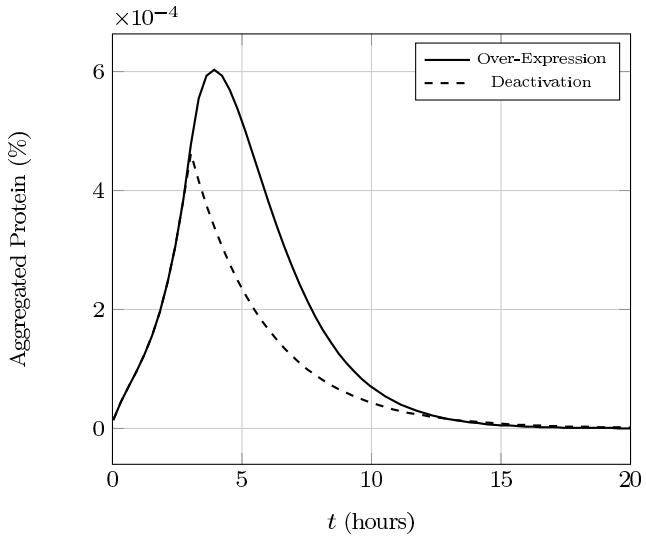


Figure 2.7: Hsp104 production is up-regulated or Hsp104 is deactivated after 3 hours, both by a factor of 10^4. The new system in either case is unable to stably support the presence of prion aggregates with our engineered parameters.

the other kinetic parameters, R_0 may be either increasing or decreasing with respect to $\tilde{p}$ – prion extinction would only occur if $R_0(1) < 1$. Specifically, $\text{sgn}(R_0'(p)) = \text{sgn}(A_0^2 - (n_0 - 1)n_0p^2)$. If $A_0 > n_0 - 1$, R_0 is always increasing and extinction is impossible under our model. This is consistent with the belief that there are other (unmodeled) factors more likely to contribute to prion phenotype loss [33, 26]; nonetheless, we do provide an engineered parameter set (described in Table 2.1) that demonstrates prion extinction by maximizing fragmentation efficiency in Figure 2.7. We additionally plot the dependence of R_0 on α_h in this parameter set as well as the original set we've used in Figure 2.8.

2.5.5 Stability of the Coendemic Prion Strains

Up until this point we have considered the aggregates of a single prion species; however, a prion protein is capable of adopting a host of misfolded confirmations each of which is capable of the biochemical processes of conversion of normal protein and fragmentation [39, 42, 43]. That is, $[PSI^+]$ or PrP^{Sc} does not refer to a single prion phenotype, but many related ones, each characterized by different pathology (implying different kinetic parameter values). These distinct prion states are referred to as prion "strains."

Biologists have observed multiple coexisting strains [40, 34], but there has been limited mathematical modeling of multiple prion strains. Previously, Tanaka et al. [41] considered the NPM, under the continuous relaxation of aggregate sizes, with $n_0 = 1$ and demonstrated that if two strains were present then, asymptotically, one strain would dominate and drive the other to extinction. The outcome was determined by the strain which had maximized $\beta\gamma$ (which is proportional to the reproductive number in the case of continuous-size, $n_0 = 1$ NPM). Since level curves of $\beta\gamma$ represent a set of measure

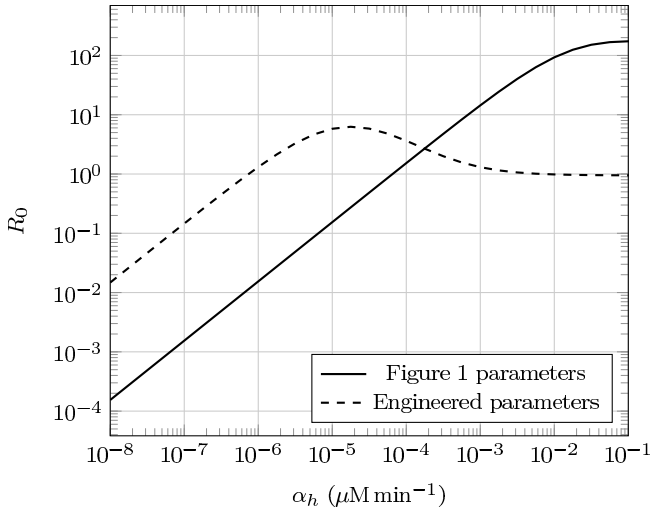


Figure 2.8: The reproductive number R_0 as a function of α_h. Prion strains will only be driven to extinction by Hsp104 over-expression if $\lim_{\alpha_h \to \infty} R_0(\alpha_h) < 1$.

0 in parameter space, realistically this prevents asymptotic prion strain coexistence. By coexistence, we mean that there exists $i \neq j$ such that $\lim_{t\to\infty} \eta_i(t), \eta_j(t) > 0$ when $\eta_i(0), \eta_j(0) > 0$ (where η_i is the concentration of aggregates of strain i).

We now generalize the ELNPM to include multiple prion strains, each capable of converting the same normal protein. Because aggregation is based on conversion to a particular prion strain conformation, we consider an aggregate as consisting of misfolded protein of a single strain. We write the equations for k strains with similar constants as before, but scale time by $\Gamma = \sum_{i=1}^{k} \gamma_i$ and write $\Gamma_i = \gamma_i/\Gamma$:

$$s' = A_s(1-s) - s\sum_{i=1}^{k} B_i\eta_i + \sum_{i=1}^{k} \Gamma_i(n_i-1)n_i p_i \eta_i \tag{2.31a}$$

$$h' = A_h(1-h) + r\sum_{i=1}^{k} \left[(\omega_i + n_i - 1)\left(k_{i,-1}p_i - k_{i,1}h[1-p_i]\right) + (n_i-1)(n_i-2)\Gamma_i p_i^2\right]\eta_i \tag{2.31b}$$

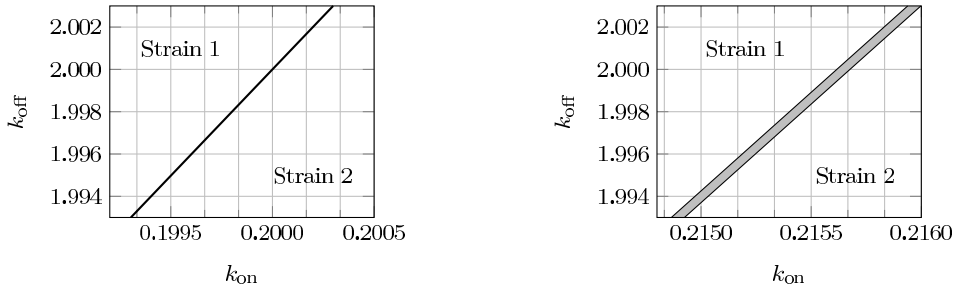
$$\eta_i' = \left[\Gamma_i p_i \omega_i - A_i - (n_i-1)\Gamma_i p_i\right]\eta_i \tag{2.31c}$$

$$\omega_i' = B_i s - \Gamma_i p_i(\omega_i+1)\omega_i \tag{2.31d}$$

$$p_i' = k_{i,1}h[1-p_i] - k_{i,-1}p_i - p_i\left(\frac{B_i s}{\omega_i + n_i - 1} - \Gamma_i p_i\right). \tag{2.31e}$$

Because our mathematical formulation also requires the molecular chaperone Hsp104, this opens up the possibility for an alternative pathway to prion strain coexistence – rather than out-compete solely on conversion β and fragmentation γ, a second strain may be more efficient at sequestering Hsp104 ($k_{\text{on}}/k_{\text{off}}$). Increasing this ratio improves the strain's own fragmentation efficiency, as well as decreases the other strain's efficiency by decreasing the amount of available Hsp104.

It is helpful to think of prion strain competition and coexistence in terms of the



(a) No stable coexistence (b) Region of stable coexistence

Figure 2.9: Different parameter regimes exhibit fundamentally different behavior with respect to coendemic stability. The labeled regions denote the surviving strain, and gray regions denote areas of mutual coexistence. Refer to Table 2.1 for the parameter values

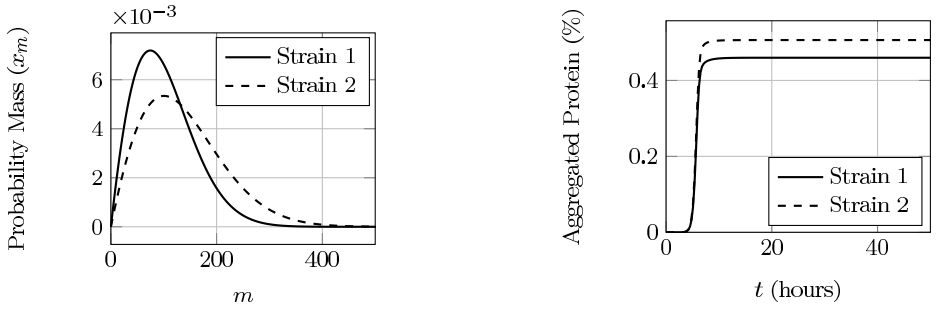
reproductive numbers described in Section 2.4.3. Intuitively, the strain with the highest reproductive number will dominate and drive others to extinction. As already stated, this is exactly what [41] observed – however, recall that with the NPM, the reproductive number is fixed, so there will not be any dependency on the kind of or number of strains present. Our model, however, has an effective reproductive number dependent on the current fragmentation efficiency. In terms of strain-specific constants, this number is given by

$$R_i(p_i) = \frac{B_i/p_i}{(A_i/p_i + n_i - 1)(A_i/p_i + n_i)}. \tag{2.32}$$

The fragmentation efficiency of strain i, p_i, is dependent on the current concentration of soluble Sup35 and free Hsp104 (Equation (2.31e)), which in turn depend on all of the strains' concentrations. As such, the reproductive numbers of the strains are coupled to one another. These nonlinear, secondary interactions make analytic determinations of coexistence difficult. However, we are able to numerically demonstrate coexistence of prion strains (Figure 2.9).

Remarkably, Figure 2.9b demonstrates that strain coexistence is possible for parameters lying in a non-zero area of parameter space. We note that, in contrast to prior models, this type of coexistence is biologically feasible because is it robust to small perturbations in parameter space. Thus, our numerical studies demonstrate that strains with different reproductive numbers (in isolation) can coexist. Further, at the coendemic state each strain's "cooperative" reproductive number is different from their isolated value, but equal to each of the other strains' cooperative numbers.

We choose 2 specific parameter sets from Figure 2.9b and plot of their steady-state, cooperative size densities in Figure 2.10a and concentration of aggregated Sup35 in each strain over time in Figure 2.10b.



(a) Steady-state aggregate densities (b) Concentration of Sup35 in aggregates

Figure 2.10: Plots of a specific parameter set admitting asymptotically stable, coendemic behavior. We note that the size densities and aggregated protein take on distinct values, despite very similar reproductive numbers.

2.6 Conclusion

In this work we successfully developed a mathematical formulation of aggregate dynamics where fragmentation occurs through the molecular chaperone Hsp104. We demonstrate that, under certain restrictions, our model reduces to a numerically tractable form which we call the Enzyme-Limited Nucleated Polymerization Model (ELNPM). By including chaperone-mediated fragmentation, this work represents an important step towards a more complete understanding of prion and protein misfolding *in vivo*.

We derived a unique disease-free steady-state of the ELNPM and analyzed its stability. We demonstrated that the ELNPM supports experimentally observed results such as shifts in the aggregate-size distribution with increasing Sup35 synthesis and response to over- and under- expression of Hsp104. Additionally, it represents a first step towards quantifying prion strain coexistence.

While the ELNPM successfully describes the effects of varying amounts of Sup35 and Hsp104 in the system, we note that there are factors common to enzyme-substrate kinetics that were not included in our model. First, in many biochemical systems there is evidence of cooperation between binding sites [31]. Since there is no evidence of interaction between binding sites for Hsp104, we have modeled binding events as purely a function of the free enzyme and available binding sites.

Second, by assuming that k_{off} was large we were able to assume that for an aggregate consisting of i Sup35 monomers with j sites bound by Hsp104, all possible configurations of bound Hsp104 are equally likely (see the Supplemental Material). Since under normal expression Hsp104 is observed to be only minimally bound to $[PSI^+]$ aggregates [26], we interpret this to indicate that k_{off} must indeed be large relative to k_{on}. Third, we considered a generalization of the uniform fragmentation kernel which corresponds to equality in fragmentation at all binding sites. Together, these three assumptions allowed the use of analytical approaches previously employed in the analysis of the NPM to demonstrate existence, uniqueness and asymptotic stability of the disease-free steady-state.

Beyond Hsp104, other enzymes have been identified as important players in the

Table 2.1: A table of values used for the plots. In all cases, $\alpha_s = 0.0154\ \mu\text{M}\,\text{min}^{-1}$, $n_0 = n_1 = n_2 = \ldots = 5$, and $\mu_s = \mu_h = \mu_0 = \mu_1 = \mu_2 = \ldots = 0.0077\,\text{min}^{-1}$

Figures	α_h ($\mu\text{M}\,\text{min}^{-1}$)	β ($\mu\text{M}^{-1}\text{min}^{-1}$)	γ (min^{-1})	k_{on} ($\mu\text{M}^{-1}\text{min}^{-1}$)	k_{off} (min^{-1})
2.3, 2.6, 2.8	0.002567	6.0	0.00294	0.20000	2.00000
2.4, 2.5	0.00142	-	0.00294	0.15964	2.93706
2.7	4.16503×10^{-6}	0.216456	0.0421667	81.8524	4.38533
2.8	-	0.216456	0.0421667	81.8524	4.38533
2.9a	0.002567	(6.0,6.0)	(0.00294,0.00294)	(0.2,-)	(2.0,-)
2.9b	0.002567	(6.0,8.0)	(0.00294,0.002)	(0.2,-)	(2.0,-)
2.10	0.002567	(6.0,8.0)	(0.00294,0.002)	(0.2,0.21567)	(2.0,2.0)

dynamics of prion aggregate fragmentation [24, 38]. As such, our mathematical formulation may be taken as the representing collective impact of enzymes on fragmentation. However, compared to the other enzymes, Hsp104 occurs in the lowest molecular number [19] and is thus likely to represent a rate limiting step. In addition, we have evaluated our model by comparison to experimental results on the $[PSI^+]$ prion which has shown to have greatest sensitivity to Hsp104 [23]. Lastly, again note that the form of Hsp104 is a hexamer [16] – we have assumed the kinetics of hexamer formation are not relevant to the aggregate dynamics.

In addition to including additional biological complexities, in future studies we plan to investigate global asymptotic stability and explore the conditions underlying prion strain coexistence.

2.7 Derivation of ELNPM

2.7.1 Full Model

We examine the full kinetics implied by the rate equations in Section 2.3.2, where we consider each possible configuration of bound Hsp104 separately. We define $u_{i(b)}$ to be the density of aggregates of size i with Hsp104 configuration b, where $0 \leq b < 2^{i-1}$ and b's binary expansion is taken to reflect the states of possible fragmentation sites: a 0 appearing in the location of an unbound site, and a 1 in the location of a bound site. Let c_b denote the Hamming weight of b, or the number of 1's appearing in the binary expansion of b. Finally, we assume that fragmentation occurs at a rate proportional to the amount of bound Hsp104 and that the resulting daughters are uniformly distributed amongst the possible configurations. Thus, for $u_{i(b)}$, fragmentation occurs with rate γc_b into daughter aggregates with probability $1/c_b$. Finally, let $b_1 \otimes b_2$ denote the bit-wise and

of integers b_1 and b_2 and $\delta(x) = 1$ if $x = 0$ and 1 otherwise. Then,

$$\begin{aligned}\frac{du_{m(b)}}{dt} &= -2\beta s(t) u_{m(b)} + \beta s(t)\left[u_{m-1,(b)}\delta(b \otimes 2^{m-2}) + u_{m-1,(b/2)}\delta(b \otimes 1)\right] - (\gamma c_b + \mu_0)u_{m(b)} \\ &\quad + \gamma \sum_{i=m+1}^{\infty} \sum_{k=0}^{2^{i-m-1}-1} \left[u_{i(b+2^{m-1}+2^m k)} + u_{i(k+2^{i-m-1}+2^{i-m}b)}\right] \\ &\quad - k_{\text{on}} h(t) \left[(m-1-c_b)u_{m(b)} - \sum_{\substack{b' \text{ s.t.} \\ c_{b'}=c_b-1 \\ b' \otimes b = b'}} u_{m(b')}\right] + k_{\text{off}} \left[\sum_{\substack{b' \text{ s.t.} \\ c_{b'}=c_b+1 \\ b' \otimes b = b}} u_{m(b')} - c_b u_{m(b)}\right].\end{aligned} \tag{2.33}$$

We now sum over every $b < 2^{i-1}$ such that $c_b = j$, only assuming symmetry in the aggregate configuration densities ($u_{i(b)} = u_{i(b')}$ where b' is the reversed bitstring of b). We write $u_{mn} = \sum_{b \text{ s.t. } c_b = n} u_{m(b)}$ and carefully count bitstrings and simplify to obtain

$$\begin{aligned}\frac{du_{mn}}{dt} &= -2\beta s(t)\left[u_{mn} - u_{m-1,n}\right] - (\gamma n + \mu_0)u_{mn} \\ &\quad + 2\gamma \sum_{i=m+1}^{\infty} \sum_{j=n+1}^{i-(m-n)} \sum_{\substack{b \text{ s.t.} \\ b<2^{m-1} \\ c_b = n}} \sum_{\substack{b' \text{ s.t.} \\ b'<2^{i-m-1} \\ c_{b'} = j-n-1}} u_{i(b+2^{m-1}+2^m b')} \\ &\quad - k_{\text{on}} h(t)\left[(m-n-1)u_{mn} - (m-n)u_{m,n-1}\right] + k_{\text{off}}\left[(n+1)u_{m,n+1} - n u_{mn}\right].\end{aligned} \tag{2.34}$$

We simplify the remaining recovery term with a claim: since conversion effectively biases Hsp104 configurations towards the center of the aggregate, a relatively fast enzyme off-binding will restore the configuration distribution to approximate uniformity. Let us proceed by formally assuming $u_{i(j)} = u_{i(j')}$ if $c_j = c_{j'}$. Then (2.34) reduces to

$$\begin{aligned}\frac{du_{mn}}{dt} &= -2\beta s(t)\left[u_{mn} - u_{m-1,n}\right] - (\gamma n + \mu_0)u_{mn} + 2\gamma \sum_{i=m+1}^{\infty} \sum_{j=n+1}^{i-(m-n)} \frac{\binom{m-1}{n}\binom{i-m-1}{j-n-1}}{\binom{i-1}{j}} u_{ij} \\ &\quad - k_{\text{on}} h(t)\left[(m-n-1)u_{mn} - (m-n)u_{m,n-1}\right] + k_{\text{off}}\left[(n+1)u_{m,n+1} - n u_{mn}\right].\end{aligned} \tag{2.35}$$

2.7.2 Reduced Equations

Equipped with these assumptions, let us define the moments of our aggregate density:

$$\eta = \sum_{i=n_0}^{\infty} \sum_{j=0}^{i-1} u_{ij}, \quad z = \sum_{i=n_0}^{\infty} \sum_{j=0}^{i-1} i u_{ij}, \quad z_b = \sum_{i=n_0}^{\infty} \sum_{j=0}^{i-1} j u_{ij}. \tag{2.36}$$

Then,

$$\frac{ds}{dt} = \alpha_s - \mu_s s(t) - 2\beta s(t)\eta(t) + \gamma(n_0 - 1)n_0 \sum_{i=n_0}^{\infty}\sum_{j=1}^{i-1} \frac{j}{i-1} u_{ij}(t) \tag{2.37}$$

$$\begin{aligned}\frac{dh}{dt} &= \alpha_h - \mu_h h(t) - k_{\text{on}} h(t)[z(t) - \eta(t) - z_b(t)] + (k_{\text{off}} + \gamma) z_b(t) \\ &\quad + \gamma(n_0 - 1)(n_0 - 2) \sum_{i=n_0}^{\infty}\sum_{j=1}^{i-1} \frac{j(j-1)}{(i-1)(i-2)} u_{ij}(t)\end{aligned} \tag{2.38}$$

$$\frac{d\eta}{dt} = -\mu_0 \eta(t) + \gamma z_b(t) - 2\gamma(n_0 - 1) \sum_{i=n_0}^{\infty}\sum_{j=0}^{i-1} \frac{j}{i-1} u_{ij}(t) \tag{2.39}$$

$$\frac{dz}{dt} = 2\beta s(t)\eta(t) - \mu_0 z(t) - \gamma(n_0 - 1)n_0 \sum_{i=n_0}^{\infty}\sum_{j=0}^{i-1} \frac{j}{i-1} u_{ij}(t) \tag{2.40}$$

$$\begin{aligned}\frac{dz_b}{dt} &= -(\mu_0 + \gamma + k_{\text{off}}) z_b(t) + k_{\text{on}} h(t)[z(t) - \eta(t) - z_b(t)] \\ &\quad - \gamma(n_0 - 1)(n_0 - 2) \sum_{i=n_0}^{\infty}\sum_{j=1}^{i-1} \frac{j(j-1)}{(i-1)(i-2)} u_{ij}(t)\end{aligned} \tag{2.41}$$

While simplified, we still lack moment-closure. However, we note that the unclosed terms are of a very particular form, motivating a discrete transformation of u_{ij}. Define

$$v_{mn} = \sum_{j=n}^{m-1} \frac{\binom{j}{n}}{\binom{m-1}{n}} u_{mj}. \tag{2.42}$$

Then,

$$v_{i,0} = \sum_{j=0}^{i-1} u_{ij}, \qquad v_{i,1} = \sum_{j=1}^{i-1} \frac{j}{i-1} u_{ij}, \qquad v_{i,2} = \sum_{j=2}^{i-1} \frac{j(j-1)}{(i-1)(i-2)} u_{ij}, \tag{2.43}$$

and more generally,

$$\begin{aligned}\frac{dv_{mn}}{dt} &= -2\beta s(t)\left[v_{mn}(t) - \left(1 - \frac{n}{m-1}\right) v_{m-1,n}\right] - \mu_0 v_{mn}(t) \\ &\quad - \gamma(m - n - 1) v_{m,n+1}(t) + 2\gamma \sum_{i=m+1}^{\infty} v_{i,n+1}(t) \\ &\quad - n\left[(k_{\text{on}} h(t) + k_{\text{off}} + \gamma) v_{mn}(t) - k_{\text{on}} h(t) v_{m,n-1}(t)\right].\end{aligned} \tag{2.44}$$

This recurrence is 2nd order in n, so we see that it is ill-posed. We can argue 1 boundary condition without additional assumptions: since $v_{m,n+1} \leq v_{mn}$ by construction, and prion aggregates biologically must have an upper size limit, we have $\lim_{n\to\infty} v_{mn} = 0$. We are now left to find 1 more condition to make the problem well-posed.

2.7.3 Approximation

We perform the same non-dimensionalization as in Equations (2.16a)-(2.16e), writing again $\omega(t) = z(t)/\eta(t) - n_0$ and $p(t) = z_b(t)/[z(t) - \eta(t)]$:

$$s' = A_s(1-s) - Bs\eta + (n_0-1)n_0 \sum_{i=n_0}^{\infty} v_{i,1} \tag{2.45}$$

$$h' = A_h(1-h) + r\left[(\omega + n_0 - 1)(k_{-1}p - k_1 h[1-p])\eta + (n_0-2)(n_0-1)\sum_{i=n_0}^{\infty} v_{i,2}\right] \tag{2.46}$$

$$\eta' = (-A_0 + p(\omega + n_0 - 1))\eta - 2(n_0-1)\sum_{i=n_0}^{\infty} v_{i,1} \tag{2.47}$$

$$\omega' = Bs - p(\omega+1)\omega + (n_0-1)(2\omega + n_0)\left(\frac{1}{\eta}\sum_{i=n_0}^{\infty} v_{i,1} - p\right) \tag{2.48}$$

$$p' = k_1 h(1-p) - k_{-1}p + p^2 - \frac{Bsp}{\omega + n_0 - 1} + \frac{(n_0-1)(n_0-2)}{p\eta(\omega+n_0-1)}\left(p\sum_{i=n_0}^{\infty} v_{i,1} - \sum_{i=n_0}^{\infty} v_{i,2}\right). \tag{2.49}$$

Written in this way, it becomes clear that enforcing $\sum_{i=n_0}^{\infty} v_{i,n+1} = p\sum_{i=n_0}^{\infty} v_{i,n}$ will yield moment closure. Interpreting u_{ij}/η as a probability mass and (I, J) as a joint random variable modeling aggregate size and bound Hsp104, we write $p = \mathbb{E}[J]/\mathbb{E}[I-1]$ and ultimately understand the nature of our approximation to be the validity of the approximations

$$\mathbb{E}\left[\frac{J}{I-1}\right] \approx \frac{\mathbb{E}[J]}{\mathbb{E}[I-1]}, \tag{2.50}$$

and

$$\mathbb{E}\left[\frac{J(J-1)}{(I-1)(I-2)}\right] \approx \left(\frac{\mathbb{E}[J]}{\mathbb{E}[I-1]}\right)^2. \tag{2.51}$$

Consider the 2nd order Taylor expansion of an arbitrary function $f(x,y)$ about $(\bar{x}, \bar{y})$:

$$\begin{aligned} f(x,y) \approx\ & f(\bar{x},\bar{y}) + (x-\bar{x})f_x(\bar{x},\bar{y}) + (y-\bar{y})f_y(\bar{x},\bar{y}) \\ & + \frac{1}{2}\left[(x-\bar{x})^2 f_{xx}(\bar{x},\bar{y}) + 2(x-\bar{x})(y-\bar{y})f_{xy}(\bar{x},\bar{y}) + (y-\bar{y})^2 f_{yy}(\bar{x},\bar{y})\right] \end{aligned} \tag{2.52}$$

Treating X and Y as random variables and $\bar{x}$ and $\bar{y}$ as their means, then

$$\mathbb{E}[f(X,Y)] \approx f(\bar{x},\bar{y}) + \frac{1}{2}\left(f_{xx}(\bar{x},\bar{y})\mathrm{Var}[X] + 2f_{xy}(\bar{x},\bar{y})\mathrm{Cov}[X,Y] + f_{yy}(\bar{x},\bar{y})\mathrm{Var}[Y]\right). \tag{2.53}$$

In our case $X = J, \bar{x} = z_b/\eta$, and $Y = I - 1, \bar{y} = z/\eta - 1$. Let $\bar{u} = z/\eta$; for Approximation (2.50), we have $f(x,y) = x/y$ and

$$\mathbb{E}\left[\frac{J}{I-1}\right] = p + \frac{1}{\bar{u}^2}\left(p\mathrm{Var}[I-1] - \mathrm{Cov}[J, I-1]\right) + O(1/\bar{u}^3). \tag{2.54}$$

For Approximation (2.51), we have $f(x, y) = \frac{x(x-1)}{y(y-1)}$ and

$$\begin{aligned}\mathbb{E}\left[\frac{J(J-1)}{(I-1)(I-2)}\right] = p^2 &- \frac{p(1-p)}{\bar{u}} \\ &+ \frac{\mathrm{Var}[J] - 4p\mathrm{Cov}[J, I-1] + 3p^2\mathrm{Var}[I-1] - 2p(1-p)}{\bar{u}^2} \\ &+ O(1/\bar{u}^3).\end{aligned} \tag{2.55}$$

Assuming the variances are dominated by the average aggregate size, we have error terms in the first approximation of $O(1/\bar{u}^2)$ and $O(1/\bar{u})$ in the second; when multiplied against the terms' coefficients, we obtain $O(1/\bar{u}^2)$ in either case.

Bibliography

[1] A. Aguzzi and M. Polymenidou. “Mammalian prion biology: one century of evolving concepts”. *Cell* 116.2 (2004), pp. 313–327.

[2] A. G. Akritas. “Reflections on a pair of theorems by Budan and Fourier”. *Mathematics Magazine* (1982), pp. 292–298.

[3] J. Brettschneider et al. “Spreading of pathology in neurodegenerative diseases: a focus on human studies”. *Nature Reviews Neuroscience* 16.2 (2015), pp. 109–120.

[4] P. Brundin, R. Melki, and R. Kopito. “Prion-like transmission of protein aggregates in neurodegenerative diseases”. *Nature Reviews Molecular Cell Biology* 11.4 (2010), pp. 301–307.

[5] L. J. Byrne et al. “Cell division is essential for elimination of the yeast [PSI^+] prion by guanidine hydrochloride”. *Proceedings of the National Academy of Sciences* 104.28 (2007), pp. 11688–11693.

[6] V. Calvez et al. “Prion dynamics with size dependency–strain phenomena”. *Journal of biological dynamics* 4.1 (2010), pp. 28–42.

[7] Y. O. Chernoff et al. “Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]”. *Science* 268.5212 (1995), pp. 880–884.

[8] P. Chien, J. S. Weissman, and A. H. DePace. “Emerging principles of conformation-based prion inheritance”. *Annual review of biochemistry* 73.1 (2004), pp. 617–656.

[9] J. Collinge. “Prion diseases of humans and animals: their causes and molecular basis”. *Annual review of neuroscience* 24.1 (2001), pp. 519–550.

[10] B. Cox, M. Tuite, and C. McLaughlin. “The *psi* factor of yeast: a problem in inheritance”. *Yeast* 4.3 (1988), pp. 159–178.

[11] F. H. Crick. “On protein synthesis.” *Symposia of the Society for Experimental Biology*. Vol. 12. 1958, p. 138.

[12] J. K. Davis and S. S. Sindi. “A study in nucleated polymerization models of protein aggregation”. *Applied Mathematics Letters* 40 (2015), pp. 97–101.

[13] A. Derdowski et al. “A size threshold limits prion transmission and establishes phenotypic diversity”. *Science* 330.6004 (2010), pp. 680–683.

[14] M. Doumic and P. Gabriel. “Eigenelements of a general aggregation-fragmentation model”. *Mathematical Models and Methods in Applied Sciences* 20.05 (2010), pp. 757–783.

[15] M. Doumic, T. Goudon, T. Lepoutre, et al. “Scaling limit of a discrete prion dynamics model”. *Communications in Mathematical Sciences* 7.4 (2009), pp. 839–865.

[16] S. M. Doyle and S. Wickner. “Hsp104 and ClpB: protein disaggregating machines”. *Trends in biochemical sciences* 34.1 (2009), pp. 40–48.

[17] H. Engler, J. Prüss, and G. F. Webb. “Analysis of a model for the dynamics of prions II”. *Journal of mathematical analysis and applications* 324.1 (2006), pp. 98–117.

[18] P. C. Ferreira et al. “The elimination of the yeast [PSI^+] prion by guanidine hydrochloride is the result of Hsp104 inactivation”. *Molecular microbiology* 40.6 (2001), pp. 1357–1369.

[19] S. Ghaemmaghami et al. “Global analysis of protein expression in yeast”. *Nature* 425.6959 (2003), pp. 737–741.

[20] A. D. Goldberg, C. D. Allis, and E. Bernstein. “Epigenetics: a landscape takes shape”. *Cell* 128.4 (2007), pp. 635–638.

[21] M. L. Greer, L. Pujo-Menjouet, and G. F. Webb. “A mathematical analysis of the dynamics of prion proliferation”. *Journal of theoretical biology* 242.3 (2006), pp. 598–606.

[22] L. H. Hartwell and M. W. Unger. “Unequal division in Saccharomyces cerevisiae and its implications for the control of cell division.” *The Journal of cell biology* 75.2 (1977), pp. 422–435.

[23] T. Higurashi et al. “Specificity of the J-protein Sis1 in the propagation of 3 yeast prions”. *Proceedings of the National Academy of Sciences* 105.43 (2008), pp. 16596–16601.

[24] Y. Inoue et al. “Hsp104 binds to yeast Sup35 prion fiber but needs other factor (s) to sever it”. *Journal of Biological Chemistry* 279.50 (2004), pp. 52319–52323.

[25] M. E. Jackrel and J. Shorter. “Potentiated Hsp104 variants suppress toxicity of diverse neurodegenerative disease-linked proteins”. *Disease models & mechanisms* 7.10 (2014), pp. 1175–1184.

[26] C. L. Klaips et al. “Spatial quality control bypasses cell-based limitations on proteostasis to promote prion curing”. *eLife* 3 (2014), e04288.

[27] T. P. Knowles, M. Vendruscolo, and C. M. Dobson. “The amyloid state and its association with protein misfolding diseases”. *Nature Reviews Molecular Cell Biology* 15.6 (2014), pp. 384–396.

[28] M. Y. Li et al. “Global dynamics of a SEIR model with varying total population size”. *Mathematical biosciences* 160.2 (1999), pp. 191–213.

[29] S. W. Liebman and Y. O. Chernoff. “Prions in yeast”. *Genetics* 191.4 (2012), pp. 1041–1072.

[30] J. Masel, V. A. Jansen, and M. A. Nowak. “Quantifying the kinetic parameters of prion replication”. *Biophysical chemistry* 77.2 (1999), pp. 139–152.

[31] D. L. Nelson, A. L. Lehninger, and M. M. Cox. *Lehninger principles of biochemistry*. Macmillan, 2008.

[32] M. A. Nowak et al. “Prion infection dynamics”. *Integrative Biology Issues News and Reviews* 1.1 (1998), pp. 3–15.

[33] K. Palmer, M. Ridout, and B. Morgan. “Kinetic models of guanidine hydrochloride-induced curing of the yeast [PSI$^+$] prion”. *Journal of theoretical biology* 274.1 (2011), pp. 1–11.

[34] M. Polymenidou et al. “Coexistence of multiple PrPSc types in individuals with Creutzfeldt-Jakob disease”. *The Lancet Neurology* 4.12 (2005), pp. 805–814.

[35] J. Prüss et al. “Analysis of a model for the dynamics of prions”. *Discrete Contin. Dyn. Syst. Ser. B* 6.1 (2006), pp. 225–235.

[36] P. Satpute-Krishnan, S. X. Langseth, and T. R. Serio. “Hsp104-dependent remodeling of prion complexes mediates protein-only inheritance”. *PLoS biology* 5.2 (2007), e24.

[37] L. A. Segel and M. Slemrod. “The quasi-steady-state assumption: a case study in perturbation”. *SIAM review* 31.3 (1989), pp. 446–477.

[38] J. Shorter and S. Lindquist. “Hsp104, Hsp70 and Hsp40 interplay regulates formation, growth and elimination of Sup35 prions”. *The EMBO journal* 27.20 (2008), pp. 2712–2724.

[39] S. S. Sindi and T. R. Serio. “Prion dynamics and the quest for the genetic determinant in protein-only inheritance”. *Current opinion in microbiology* 12.6 (2009), pp. 623–630.

[40] M. Strbuncelj. “The impact of genetic and epigenetic interactions on propagation of the [PSI+] prion”. PhD thesis. Brown University, 2009.

[41] M. Tanaka et al. “The physical basis of how prion conformations determine strain phenotypes”. *Nature* 442.7102 (2006), pp. 585–589.

[42] M. F. Tuite and B. S. Cox. “Propagation of yeast prions”. *Nature Reviews Molecular Cell Biology* 4.11 (2003), pp. 878–890.

[43] M. F. Tuite and T. R. Serio. “The prion hypothesis: from biological anomaly to basic regulatory mechanism”. *Nature Reviews Molecular Cell Biology* 11.12 (2010), pp. 823–833.

[44] R. B. Wickner. “[*URE*3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*”. *Science* 264.5158 (1994), pp. 566–569.

Chapter 3

Stochastic Models of Nucleated Polymerization

3.1 Introduction

The appearance of the $[PSI^+]$ prion phenotype, referred to as induction, is rare and depends on the particular prion strain. The mathematical model developed in Chapter 2, which explicitly considered the role of a molecular chaperone in the prion fragmentation process, provides a potential explanation for this variability in $[PSI^+]$ induction [10, 17]: the transient, variable fragmentation rate may under- or over-perform relative to the asymptotic fragmentation rate, permitting phenotype extinction when it would otherwise be stable [4].

However, the derivation of our enzyme-limited generalization of nucleated polymerization requires large numbers of enzyme binding sites, which in turn requires a large amount of aggregated protein. Because the regime of interest for $[PSI^+]$ induction is for (initially) small concentrations of aggregated proteins, the dynamics need to be validated with a stochastic model. That is, while models derived from the law of mass action and quasi-steady assumptions (such as those described in Chapters 1 and 2) tend to be fairly robust with respect to their assumptions [16], one can imagine a single aggregate *in vivo* per chance fragmenting prematurely, rather than asymptotically amplifying into an endemic infection.

As has been established in earlier chapters, the probability a single aggregate amplifies into a full prion "infection" is mediated by the presence and concentration of the Hsp104 chaperone. In order to study this probability, I reformulate the nucleated polymerization dynamics as a Markov chain and solve the corresponding chemical master equation. I then use the solution to to generalize strain stability from the deterministic framework and endow this stability with a probabilistic interpretation.

3.2 Intuition

Let us assume that biological observations roughly correspond to asymptotic behavior. Thus, we are interested specifically in the event that a prion strain (with basic reproductive number $R_0 > 1$) capable of endemic "infection" fails to do so due to the unmodeled, mesoscopic effects. To appreciate how these dynamics influence this outcome, we revisit the continuous-size, moment-closed nucleated polymerization model [15, 9, 7] (reviewed in Chapter 1), where

$$\begin{aligned} x'(t) &= \alpha - \mu x(t) - 2\beta x(t)y(t) + \gamma a^2 y(t), \\ y'(t) &= -(\mu + 2\gamma a)y(t) + \gamma z(t), \\ z'(t) &= 2\beta x(t)y(t) - \mu z(t) - \gamma a^2 y(t). \end{aligned} \tag{3.1}$$

The parameters and their meaning are summarized in Table 3.1. The system will be initialized with a small amount of prion aggregate, mirroring a spontaneous "nucleation" event or the seeding of a cell with pre-existing aggregated protein, such that $x(0) = \frac{\alpha}{\mu}(1 - \epsilon)$, $y(0) = \frac{\alpha}{\mu}\frac{\epsilon}{a}$, and $z(0) = \frac{\alpha}{\mu}\epsilon$.

Table 3.1: Parameters of the Nucleated Polymerization Model

α	rate of transcription of protein monomer
μ	rate of volume dilution due to cell division
β	rate of monomer aggregation by an aggregate end
γ	rate of polymer junction fragmentation
a	minimum stable size of prion aggregate
ϵ	initial percentage of aggregated protein

By initializing the total Sup35 concentration $x(t) + z(t)$ at its steady-state value α/μ, one variable may be eliminated. We additionally replace $z(t)$ with the displacement of the average size from the minimum size $\omega(t) = z(t)/y(t) - a$ to obtain the reduced two-dimensional system

$$y'(t) = \left(\gamma\omega(t) - (\mu + \gamma a)\right) y(t) \tag{3.2}$$

$$\omega'(t) = 2\beta\left(\frac{\alpha}{\mu} - (\omega(t) + a)y(t)\right) - \gamma\omega(t)^2. \tag{3.3}$$

The system is fully non-dimensionalized with the choice of variables

$$t = \frac{\tau}{\mu + \gamma a}, \tag{3.4}$$

$$y(t) = \frac{\alpha}{\mu a} Y(\tau), \tag{3.5}$$

$$\omega(t) = \left(a + \frac{\mu}{\gamma}\right) W(\tau), \tag{3.6}$$

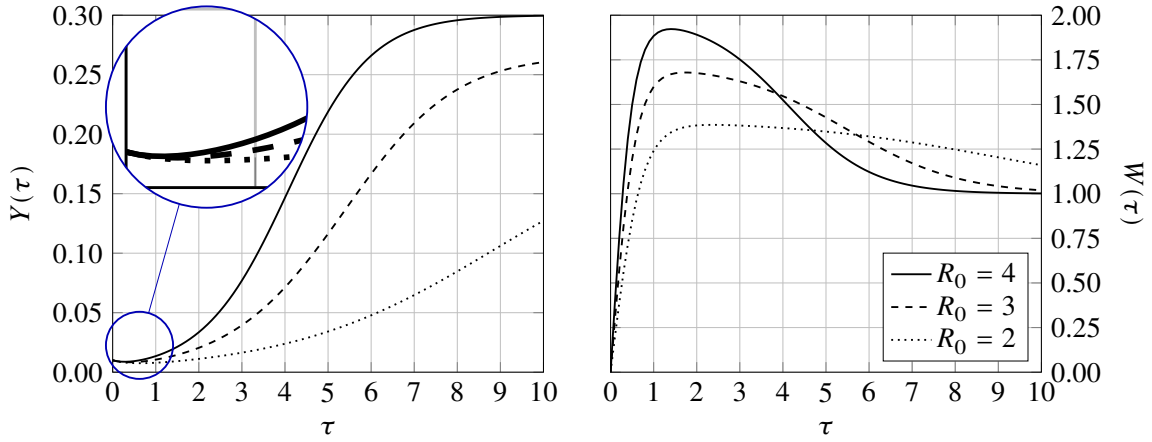


Figure 3.1: The scaled concentration $Y(\tau)$ and displacement $W(\tau)$ versus (dimensionless) time τ with $Y(0) = 0.01$, $W(0) = 0$, and $\kappa = 1.5$. Note that $Y'(\tau)$ remains negative (and more generally, $Y(\tau)$ remains small) for increasingly large τ as $R_0 \to 1^+$.

which yields

$$Y' = (W - 1)Y, \tag{3.7}$$

$$W' = R_0 \left(1 - (1 + \kappa W)\, Y\right) - W^2, \tag{3.8}$$

where, in addition to $R_0 = \frac{2\alpha\beta\gamma}{\mu(\mu+\gamma a)^2}$, one last dimensionless constant $\kappa = 1 + \frac{\mu}{\gamma a}$ has been defined.

Several solutions to this scaled system are plotted in Figure 3.1. Clearly, independent of the concentration $Y(\tau)$, we see that $Y'(\tau) < 0$ when $W(\tau) < 1$ (see Equation (3.7) and Figure 3.1, left inset) . Thus, if the system is initialized with a small perturbation of sufficiently small aggregates (e.g. $Y(0) = \epsilon$ and $W(0) < 1$), then there is instantaneously a further decrease of concentration. If ϵ is to be interpreted as to correspond to a single aggregate, $Y(\tau)$ for the regime $0 < \tau < W^{-1}(1)$ must be interpreted in a population-level or *frequentist* sense – that is, out of many phenotype-inducing experiments, at least some of them failed (despite a prion strain with $R_0 > 1$). Given this population-level understanding, it is of interest to quantify the probability that an individual experiment fails, i.e. that a prion fails to amplify into an infection.

Using a pertubation expansion, the length-scale of this *mesocopic* regime can be computed and is solely determined by R_0. Suppose $Y(0) = \epsilon \ll 1$ and $W(0) < 1$. Then, for $0 \le \tau \le W^{-1}(1)$, $W(\tau)$ is well-approximated by $\tilde{W}(\tau)$, where

$$\tilde{W}' = R_0 - \tilde{W}^2 \implies \tilde{W}(\tau) = \sqrt{R_0}\tanh\left(\sqrt{R_0}\tau + \operatorname{arctanh}\left(\frac{W(0)}{\sqrt{R_0}}\right)\right). \tag{3.9}$$

$\tilde{W}(\tau)$ is then inverted to find $\tau^* = \tilde{W}^{-1}(1) \approx W^{-1}(1)$:

$$\tau^* = \frac{1}{2\sqrt{R_0}} \log\left|\frac{\sqrt{R_0}+1}{\sqrt{R_0}-1} \times \frac{\sqrt{R_0}-W(0)}{\sqrt{R_0}+W(0)}\right| \approx \frac{1-W(0)}{(\sqrt{R_0}-1)^2} + O(\sqrt{R_0}-1)^{-3}. \tag{3.10}$$

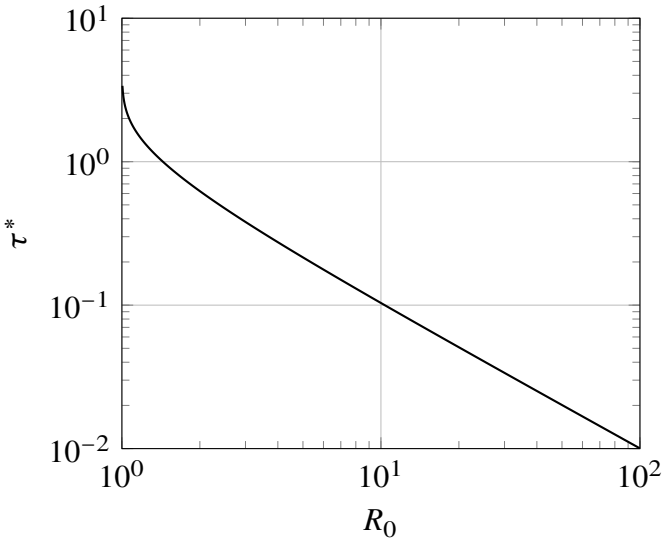


Figure 3.2: A plot of τ^* (the duration of the mesoscopic regime) vs R_0 for $W(0) = 0$ and $Y(0) = O(\epsilon)$. For $0 < \tau < \tau^*$, random, mesoscopic effects contribute to the overall dynamics of the aggregate concentration.

This relationship is depicted in Figure 3.2 for $W(0) = 0$.

It is clear in both Figure 3.2 and the series approximation in Equation (3.10) that for $R_0 \gg 1$ τ^* is negligable; however, as $R_0 \to 1^+$, τ^* diverges. Thus, the unmodeled mesoscopic effects persist for increasingly long periods of time as $R_0 \to 1^+$. To more precisely explore the dynamics in this parameter regime, I develop a stochastic model to capture these mesoscopic effects and recast the formerly qualitative observations as quantitative probabilities.

3.3 Stochastic Formulation

I next consider the stochastic dynamics of prion aggreagetes within a cell. For simplicity, protein synthesis and degradation is neglected, fixing the number of proteins per cell. For *S. cerevisiae*, estimates on the abundance of prion protein Sup35 range from 10^3-10^6 [3, 8]. More generally, the integer parameter m is used to denote this number. Supposing that an aggregate must be of at least size k (such that smaller aggregates will immediately disassociate into their monomer constituents, as described in Chapters 1 and 2 and originally proposed in [14]), the aggregate configuration of a cell $\vec{n} = (n_k, n_{k+1}, \dots, n_m)$ describing the number of aggregates of each size i (n_i) in a host cell will belong to the set

$$\mathcal{N} = \left\{ \vec{n} \geq 0 : \sum_{i \geq k} i n_i = m \right\}. \tag{3.11}$$

A full stochastic description of nucleated polymerization will define continuous-time transitions of a cell between the states in $\mathcal{N}$. Assuming that these transitions occur at

exponential rates will lead to a Markov chain formulation. However, even for moderate k and m, the size of $\mathcal{N}$ quickly exceeds computational feasibility. This size may be precisely computed with the use of generating functions [19] – by counting principles, the size of $\mathcal{N}$ corresponds to the m-th coefficient in a Taylor series expansion about $x = 0$ of $\prod_{i=k}^{m}(1-x^i)^{-1}$. A few of these values corresponding to lower bounds of estimates of m for Rnq1 and Sup35 prion proteins [3] and k [17] are tabulated in Table 3.2.

Table 3.2: $|\mathcal{N}|$ for various protein abundance m and minimum stable aggregate size k. For sufficiently large m, varying k does not affect the leading order magnitude.

Protein	m	$k = 5$	$k = 10$	$k = 15$	$k = 20$
Rnq1	265	7.23×10^{14}	7.23×10^{14}	7.22×10^{14}	7.12×10^{14}
Sup35	1440	1.86×10^{38}	1.86×10^{38}	1.86×10^{38}	1.86×10^{38}

In order to study a stochastic model of these dynamics, we must develop a computationally feasible and biologically meaningful reduction, as the full system is numerically intractable. It is here that the previous observation from the continuous model is leveraged: the primary source of uncertainty in the model is the length of time that otherwise unmodeled mesoscopic effects perturb the deterministic behavior of the system. This is most drastic when $R_0 \approx 1$, which in turn corresponds to slow amplification of the disease state. Thus, instead of considering the full aggregate configuration of a cell, only a single aggregate and its size is considered. This aggregate may successfully divide or disintegrate before the first cell division event. The former is a necessary (but not sufficient) condition for the overall stability of a prion strain.

The single absorbing state of the "full model" corresponding to $\vec{n} = (0, 0, 0, \dots)$ is replaced with two labeled absorbing states: extinction, where the single aggregate has been destroyed, and persistence, where the aggregate has doubled. For now, the possibility that a "doubled" aggregate's daughters (or granddaughters, or great-granddaughters, etc) are destroyed before cell division is neglected, but this simplification is validated with a full Monte Carlo simulation later (Figure 3.4). If formulated as a Markov chain with appropriate transition rates, with probability 1 one of these two states is reached. The arrival, or absorption time of the second, doubled state is of interest – if it occurs before a separate, uncoupled process corresponding to cell division is activated, then the prion phenotype has at least persisted across 1 division. This Markov chain is now formulated.

Let $p_i(t)$ be the probability an undivided host cell has a single aggregate of size i at time t, and $p_0(t)$ and $p_*(t)$ be the probability of being in the extinct or persisted states at time t, respectively. These probabilities satisfy

$$\begin{aligned} p_0'(t) &= \gamma \sum_{i=k}^{2(k-1)} (2k-1-i)p_i(t), \\ p_i'(t) &= 2\beta(m-i+1)p_{i-1} - (\gamma(i-1) + 2\beta(m-i))p_i(t) + 2\gamma \sum_{j=i+1}^{i+k-1} p_j(t), \\ p_*'(t) &= \gamma \sum_{i=2k}^{\infty} (i-2k+1)p_i(t). \end{aligned} \tag{3.12}$$

Writing $\vec{p} = (p_k, p_{k+1}, \dots, p_m)^{\mathrm{T}}$, this chemical master equation may be more concisely written $\frac{d\vec{p}}{dt} = A\vec{p}$, where A has appropriate entries and has an upper Hessenberg structure. The other variables $p_0(t)$ and $p_*(t)$ are then easily computed by taking a dot-product with the solution vector $\vec{p}(t)$. Being a linear differential equation, $\vec{p}(t) = \mathrm{e}^{At}\vec{p}(0)$ for some initial condition $\vec{p}(0)$, which is taken to be $(1, 0, 0, \dots)$. This corresponds to a "worst-case" where the single aggregate is of minimal size.

The random variable $p_*(T)$, where T is a model of cell division, is now considered. T is typically assumed to be distributed as a $\Gamma(s, \theta)$ random variable [1], in which case a simple calculation yields

$$\mathbb{E}[p_*(T)] = \gamma \left(\sum_{i=2k}^{\infty} (i - 2k + 1)\vec{e}_i \right)^{\mathrm{T}} [(I - \theta A)^{-s} - I] A^{-1}\vec{p}(0), \tag{3.13}$$

where $\{\vec{e}_i\}$ are the standard, canonical basic vectors. The quantity $\mathbb{E}[p_*(T)]$ is the probability that an initial minimally sized aggregate successfully persists and amplifies into two aggregates at a rate at least that of cell division, which is a necessary condition for endemic infection.

By construction, A is diagonalizable, as is $(I - \theta A)$. While exploiting that diagonalization is one means of evaluating $(I - \theta A)^{-s}$, a simpler and cheaper method is just to use an integer value for s and invert $(I - \theta A)$ s-times. Using the sample moments of cell division times published in [1], s is chosen to be the nearest integer to the point estimate within the margin of error and θ is adjusted to preserve the sample mean. The rest of this work assumes $s = 43$ and $\theta = 1.91$, where the units of $\Gamma(s, \theta)$ are minutes. The mass constant $m = 78900$ is chosen based on the larger estimates of Sup35 abundance provided by [8] and the minimum size $k = 5$ unless otherwise specifies.

A parameter sweep of $\mathbb{E}[p_*(T)]$ over a range of (β, γ) is plotted in Figure 3.3.[1] Point estimates of (β, γ) for various prion strains, computed from data reported by the Serio lab [6] are also plotted and they exist exactly on the manifold where there is a sharp transition from probability 0 to probability 1. For β significantly smaller or larger, the overall probability of persisting would be insensitive to vertical shifts in γ (e.g. fluctuations in the availability of Hsp104 chaperone). However, the biological systems exist not only where there is non-binary probability of amplification, but also exactly where there is sensitivity to the active fragmentation rate. Were the strains to not exist in this region, then the enzymatic effects modeled in Chapter 2 would not have been biologically observable. These results are also consistent with the previously reported stochastic simulation from [6].

Lastly, I verify that neglecting the daughters and granddaughters of the initial aggregate does not introduce meaningful error with a Monte Carlo simulation. This simulation tracks any number of aggregates and their sizes starting from a single minimally-sized aggregate and terminates after a randomly-sampled cell division time. If there are at least two aggregates at this time the simulation reports a success. This verification is plotted in Figure 3.4 over the same parameter regime as Figure 3.3.

[1] The code to perform these calculations is available at `https://github.com/jasondark/dissertation` in the `persist` folder.

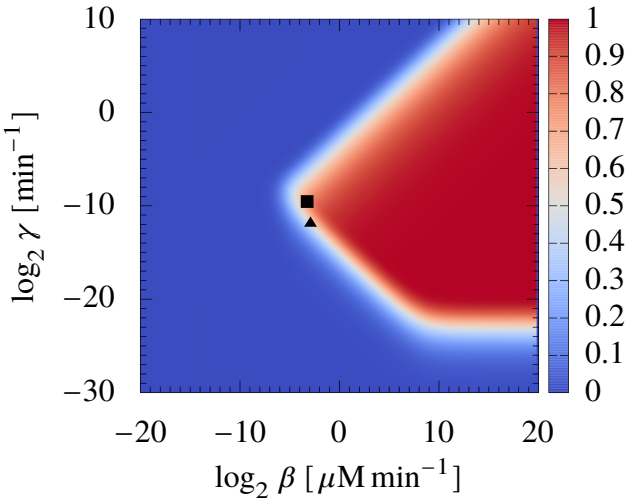


Figure 3.3: $\mathbb{E}[p_*(T)]$ over a range of (β, γ), plotted with parameter estimates for $[PSI^+_{\text{strong}}]$ (square) and $[PSI^+_{\text{weak}}]$ (triangle). Both strains were assumed to have a minimum stable size of 5 [17].

3.4 Discussion

This work represents the first formal stochastic treatment of nucleated polymerization. A direct "translation" of the NPM reactions to a stochastic framework would define a multi-type branching process [11] with a virtually impossible number of types (every mass-preserving configuration $\vec{n}$, described in Table 3.2). By carefully studying the deterministic model, a reduced stochastic model was able to be formulated, valid only in the initial transient dynamics of a single aggregate whose strain has basic reproductive number $R_0 \approx 1$. As demonstrated in Figures 3.3 and 3.4, this region is precisely where a number of prion strains are believed to exist, making this work immediately relevant to the study of "seeding" experiments where the prion phenotype is induced from a small number of initial aggregates.

Additionally, a novel observation is drawn from Figure 3.3. Over-expression of Hsp104 is known to suppress the rate of $[PSI^+]$ induction and stability [2, 12]. However, as predicted by this stochastic model, the weak strain should actually see an increase in its induction rate under this scenario, which in turn suggests it would be more stable to perturbations in the fragmentation rate. This contradiction is resolved by allowing the weak strain to have a larger minimum size than the strong strain: instead of size 5, Figure 3.3 is recreated with a minimum size of 15 in Figure 3.5, with the new "corrected" weak strain estimate plotted as before. This modification places the weak strain back into agreement with the biology, where elevated amounts of Hsp104 (larger values of γ) eliminate the prion phenotype. While this constitutes a minor mathematical point (varying an integer parameter), until now the idea that prion strains can have different minimum sizes was completely unconsidered in the biology community (e.g. [6, 17]),

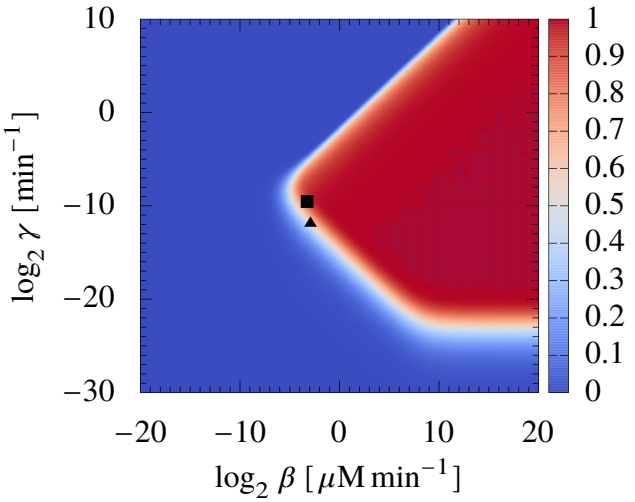


Figure 3.4: A replication of Figure 3.3 but with a full Monte Carlo simulation (10^4 iterations per point) instead of the approximate Markov chain.

even though previously published gels support this hypothesis. For example, lanes 1 and 2 of Figure 3.6 (where the vertical axis corresponds to aggregate size), originally published in [13] and reproduced with permission, strongly suggest this difference.

3.5 Appendix: Parameter Estimation

The cell division constants $s = 43$ and $\theta = 1.91163$ are obtained (and slightly perturbed) from the values in [1]. For the purposes of converting to and from kinetic rate constants and stochastic rate constants, the cellular volume of an *S. cerevisiae* cell is taken to be 37 μm^3 [18]. Finally, the abundance of Sup35 was fixed to $m = 78900$ [8], yielding kinetic rate estimates

$$\alpha = 0.0286665\ \mu\text{M}\,\text{min}^{-1}, \qquad \mu = 0.00809562\,\text{min}^{-1}. \tag{3.14}$$

With respect to the point estimates for the strong and weak $[PSI^+]$ phenotypes, the equilibrium behavior of the deterministic NPM model from Chapter 1 is used to determine the kinetic rates. It can be shown [15, 5] that the endemic steady-state concentrations of Sup35, denoted X, and size i aggregates, denoted Ψ_i, satisfy

$$[X] = \frac{\alpha/\mu}{R_0}, \qquad \sum_{i \geq n_0} i[\Psi_i] + [X] = \alpha/\mu, \qquad \frac{\sum_{i \geq n_0} i[\Psi_i]}{\sum_{i \geq n_0} [\Psi_i]} = 2n_0 - 1 + \frac{\mu}{\gamma}. \tag{3.15}$$

For strong, values from [6] were used, where the percentage of Sup35 in monomer

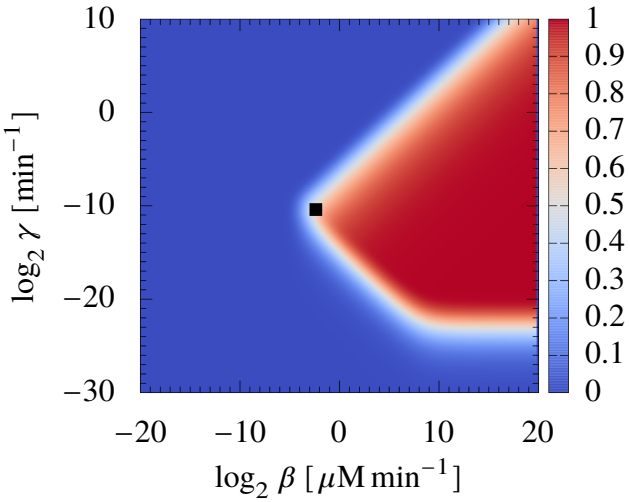


Figure 3.5: $\mathbb{E}[p_*(T)]$ over a range of (β, γ), plotted with a parameter estimate for $[PSI^+_{\text{weak}}]$ (square) with a minimum aggregate size of 15 instead of 5.

form and the average aggregate sizes were described approximately as

$$\frac{[X]}{[X] + \sum_{i \geq n_0} i[\Psi_i]} = 0.2, \qquad \frac{\sum_{i \geq n_0} i[\Psi_i]}{\sum_{i \geq n_0} [\Psi_i]} = 15. \tag{3.16}$$

For weak, these values were

$$\frac{[X]}{[X] + \sum_{i \geq n_0} i[\Psi_i]} = 0.35, \qquad \frac{\sum_{i \geq n_0} i[\Psi_i]}{\sum_{i \geq n_0} [\Psi_i]} = 40. \tag{3.17}$$

These values were then used to solve for the unknown kinetic rates (β, γ).

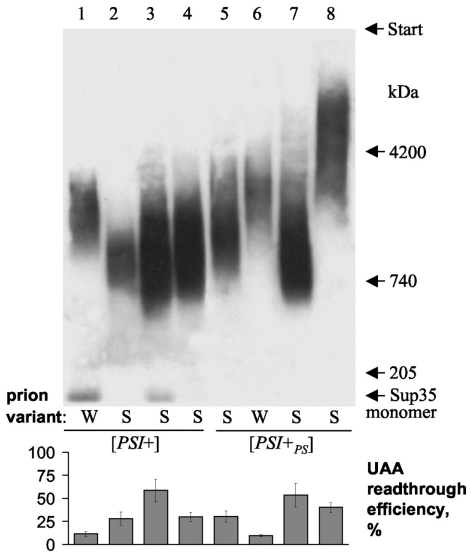


Figure 3.6: This figure was originally published in Kryndushkin et al. [13] and reproduced in accordance with the American Society for Biochemistry and Molecular Biology's Copyright and Permission Policy. Figure and caption © 2003 the American Society for Biochemistry and Molecular Biology. Original caption text: Comparison of the size of Sup35 prion polymers by SDD-AGE and of the efficiency of UAA nonsense codon read-through in different $[PSI^+]$ and $[PSI^+_{PS}]$ variants of the 5V-H19 and PS-5V-H19 strains. Prion variants: W, weak; S, strong. Lanes: 1, $[PSI^+_6]$; 2, $[PSI^+_1]$; 3, $[PSI^+_{11}]$; 4, $[PSI^+_{37}]$; 5; $[PSI^+_{PS-20}]$; 6, $[PSI^+_{PS-7}]$; 7, $[PSI^+_{PS-2}]$; 8, $[PSI^+_{PS-1}]$.

Bibliography

[1] L. J. Byrne et al. “The number and transmission of [PSI+] prion seeds (propagons) in the yeast Saccharomyces cerevisiae”. *PLoS One* 4.3 (2009), e4670.

[2] Y. O. Chernoff et al. “Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]”. *Science* 268.5212 (1995), pp. 880–884.

[3] Y. T. Chong et al. “Yeast proteome dynamics from single cell imaging and automated analysis”. *Cell* 161.6 (2015), pp. 1413–1424.

[4] J. K. Davis and S. S. Sindi. “A mathematical model of the dynamics of prion aggregates with chaperone-mediated fragmentation”. *Journal of mathematical biology* 72.6 (2016), pp. 1555–1578.

[5] J. K. Davis and S. S. Sindi. “A study in nucleated polymerization models of protein aggregation”. *Applied Mathematics Letters* 40 (2015), pp. 97–101. issn: 0893-9659. doi: 10.1016/j.aml.2014.09.007.

[6] A. Derdowski et al. “A size threshold limits prion transmission and establishes phenotypic diversity”. *Science* 330.6004 (2010), pp. 680–683.

[7] H. Engler, J. Prüss, and G. F. Webb. “Analysis of a model for the dynamics of prions II”. *Journal of Mathematical Analysis and Applications* 324.1 (2006), pp. 98–117. issn: 0022247X. doi: 10.1016/j.jmaa.2005.11.021.

[8] S. Ghaemmaghami et al. “Global analysis of protein expression in yeast”. *Nature* 425.6959 (2003), pp. 737–741.

[9] M. L. Greer, L. Pujo-Menjouet, and G. F. Webb. “A mathematical analysis of the dynamics of prion proliferation”. *Journal of Theoretical Biology* 242.3 (2006), pp. 598–606. issn: 00225193. doi: 10.1016/j.jtbi.2006.04.010.

[10] R. Kao et al. “The potential size and duration of an epidemic of bovine spongiform encephalopathy in British sheep”. *Science* 295.5553 (2002), pp. 332–335.

[11] M. Kimmel and D. E. Axelrod. “Multitype Processes”. *Branching Processes in Biology*. Springer, 2015, pp. 19–36.

[12] C. L. Klaips et al. “Spatial quality control bypasses cell-based limitations on proteostasis to promote prion curing”. *eLife* 3 (2014), e04288.

[13] D. S. Kryndushkin et al. “Yeast [PSI+] prion aggregates are formed by small Sup35 polymers fragmented by Hsp104”. *Journal of Biological Chemistry* 278.49 (2003), pp. 49636–49643.

[14] J. Masel, V. A. A. Jansen, and M. A. Nowak. “Quantifying the kinetic parameters of prion replication”. *Biophysical Chemistry* 77.2-3 (1999), pp. 139–152. issn: 03014622. doi: 10.1016/S0301-4622(99)00016-2.

[15] J. Prüss and L. Pujo-Menjouet. “Analysis of a model for the dynamics of prions”. *Discrete and Continuous Dynamical Systems - Series B* 6.1 (2006), pp. 225–235.

[16] L. A. Segel and M. Slemrod. “The quasi-steady-state assumption: a case study in perturbation”. *SIAM review* 31.3 (1989), pp. 446–477.

[17] M. Tanaka et al. “The physical basis of how prion conformations determine strain phenotypes”. *Nature* 442.7102 (2006), pp. 585–589.

[18] C. B. Tyson, P. G. Lord, and A. E. Wheals. “Dependency of size of Saccharomyces cerevisiae cells on growth rate”. *Journal of Bacteriology* 138.1 (1979), pp. 92–98.

[19] H. S. Wilf. *generatingfunctionology*. Elsevier, 2013.

Chapter 4

Spontaneous Nucleation

4.1 Introduction

A rich literature for the mathematical modeling of the prion phenotype exists. This literature was reviewed in Chapter 1 and expanded upon in Chapters 2 and 3; however, key to most of this work is the assumption of a minimum stable size to a prion aggregate. The nucleated polymerization model (NPM) describes the dynamics of aggregates larger than this size and assumes that aggregates smaller than this size, generated by a fragmentation event, immediately disassociate into the normal monomer form. Therefore, these models are necessarily initialized with some preexisting concentration or number of aggregates. This is akin to growing a yeast colony in a medium already containing prion fibers [16] or a human consuming BSE-tainted beef [14]. This process is known as *induction*, where the primary question of interest is "What are the conditions and initial concentrations required to induce a prion phenotype?" The NPM and its generalizations provide insight into this question.

The prion phenotype is also known to occur naturally and spontaneously in a population. For example, the $[PSI^+]$ phenotype occurs spontaneously in *S. cerevisiae* at a rate of every million cell divisions [19] with roughly 90 minutes between each cell division [7]. Creutzfeldt-Jakob disease is a late-onset, neurodegenerative disease (as are Alzheimer's and Parkinson's diseases, if they too are considered prion diseases). Taken together, it shall be assumed for modeling purposes that *spontaneous nucleation*, under normal circumstances, must be an exceedingly rare event. However, the rate of spontaneous $[PSI^+]$ occurrence drastically increases when the yeast cells are exposed to heat shock [24], and the mammalian diseases are believed to occur due to a somatic mutation in the protein translation machinery [25]. Thus, a perturbation to the normal cellular environment significantly increases this rate of nucleation. This process is specifically referred to as nucleation in order to connect it directly to the NPM: a slow process gradually assembles increasingly large aggregates out of available protein monomers until a particle has reached the nucleation size. Then, NPM dynamics operate on a much faster time-scale. The instantaneous dissolution of small aggregates in NPM is then understood to be a return to the slower time-scale, which is effectively instantaneous

relative to the NPM dynamics.

Mathematically, there is a vast literature corresponding to coagulation processes. More generally, an assembly model is a process that permits both coagulation (lengthening of aggregates) and fragmentation (shortening of aggregates). Given an assembly model with some initial condition, the length of time until a sufficiently large particle is formed is called the *nucleation time*, and the characterization and efficient calculation of this nucleation time will be the primary aim of this chapter. Much of the existing literature focuses on the deterministic regime (e.g. [5, 6, 4, 2, 3]), where there are enough reactants to treat all quantities as continuous concentrations. However, a number of recent papers [35, 12] demonstrated that these continuous models fail in their ability to estimate the nucleation time and instead they propose the use of Markov chains with appropriate *absorbing states* as a better, more accurate technique.

I begin with a review of the Markov chain formulation and then detail an implementation of software I wrote to solve the associated problem. I then comment on appropriate initial conditions for *in vivo* modeling of the nucleation problem. This work, combined with induction model from Chapter 3, permits an investigation into a "two-hit" model of prion pathogenesis, outlined as future work in Chapter 5.

4.2 Markov Chain Models of Nucleation

The most immediate way to formulate a nucleation model is to specify assembly rules as biochemical reactions. For example, one can imagine a "proto-aggregate" as a linear fiber which grows by monomer addition (polymerization) and shrinks by monomer removal (depolymerization). This is represented by the biochemical reaction scheme

$$X_i + X_1 \underset{k_{i+1}^-}{\overset{k_i^+}{\rightleftharpoons}} X_{i+1}, \tag{4.1}$$

with prescribed rate constants $\{k_i^\pm\}$, which is the widely studied Becker-Döring model [5, 4, 2, 35, 12]). A more general scheme, which includes Becker-Döring as a special case, is the coagulation-fragmentation model [28, 6, 3, 11], which has generalized binary fusion and fission rules represented by the biochemical scheme

$$X_j + X_{i-j} \xrightarrow{a_{j,i-j}} X_i \xrightarrow{b_{k,i-k}} X_k + X_{i-k} \tag{4.2}$$

for each triplet (i, j, k) where $1 \leq j, k < i$ with rate constants $\{a_{i,j}\}$ and $\{b_{i,j}\}$. It is assumed that $a_{ij} = a_{ji}$ and $b_{ij} = b_{ji}$. Much like the nucleated polymerization models from the previous chapters, these equations have been analyzed under assumptions leading to differential equations in the time-varying concentrations $x_i(t)$ for each X_i, e.g.

$$\frac{dx_i}{dt} = \frac{1}{2}\sum_{j=1}^{i-1} a_{j,i-j}x_j x_{i-j} - x_i \sum_{j\geq 1} a_{i,j}x_j + 2\sum_{j>i} b_{i,j-i}x_j - x_i \sum_{j=1}^{i-1} b_{j,i-j} \tag{4.3}$$

for each $i > 0$.

The issue with this approach, as noted by [35, 12, 11], is that there does not exist accurate techniques to infer nucleation times from the time-varying concentrations of $\{x_i\}$. To illustrate this point, consider an assembly model with $a_{i,j} = b_{i,j} = 1$ for all i, j and initial conditions $x_1(0) = 1$ and $x_i(0) = 0$ for $i > 1$. Let us consider only $\{x_1, \dots, x_k\}$ and determine the time at which a particle x_i where $i > k$ first appears. We construct $y = \sum_{i \geq 1} x_i$ such that $y' = 1 - y(1 + y/2)$ and $y(0) = 1$. Using this solution, the individual $\{x_i\}$ will satisfy

$$x_i' = \frac{1}{2} \sum_{j=1}^{i-1} (x_{i-j} - 4)x_j + 2y - (i + 1 + y)x_i, \tag{4.4}$$

Solutions are depicted in Figure 4.1 for $\{x_2, x_3, x_4\}$. Suppose that $k = 3$ and we are interested in the time of first appearance of X_4 – in terms of concentrations, what does this actually mean? If it corresponds to a detectable level of X_4, e.g. $t = x_4^{-1}(\epsilon)$, what is ϵ? What is the variance of t? These are not questions that this ODE formulation can answer. Instead, one must employ a full stochastic description.

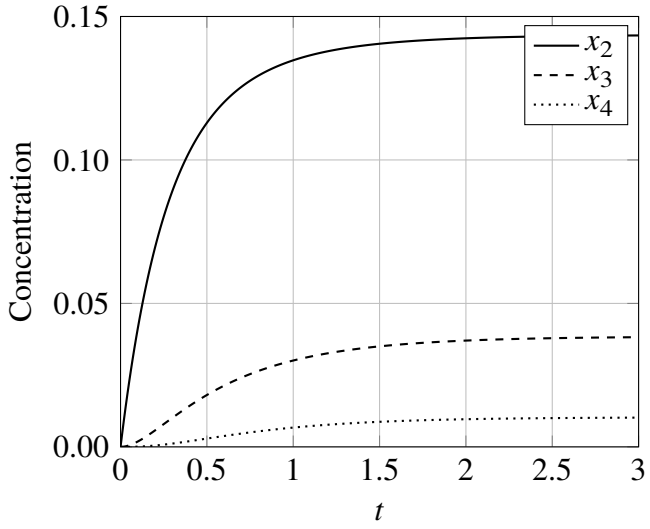


Figure 4.1: Concentrations of X_2, X_3, X_4 over time under a coagulation-fragmentation assembly model with $a_{ij} = b_{ij} = 1$ and $x_1(0) = 1$, $x_i(0) = 0$ for $i > 1$.

The most obvious approach is to employ a Monte Carlo method (e.g. the Gillespie algorithm or tau-leaping) to simulate sequences of the biochemical reactions corresponding to the assembly model of interest, terminating when a sufficiently large particle is formed. However, since the nucleation event is assumed to be extremely rare, these simulations will in turn take an extremely long time to complete. This prevents the efficient construction of an accurate, empirical distribution of nucleation times. The framework outlined in [35, 12, 11] instead numerically characterizes the *absorption time* of a corresponding Markov chain.

4.2.1 Nucleation Time as a Linear System

Suppose generally that we have a directed graph with a discrete enumeration of nodes $\{0, 1, 2, \dots\}$ and edges with weights λ_{ij} from node i to j, where the weight $\lambda_{ij} \geq 0$ is interpreted as the rate at which a random walker at node i transitions to node j. We define the total outward rate $\lambda_i = \sum_{j \neq i} \lambda_{ij}$. Any node k satisfying $\lambda_k = 0$ is said to be an absorbing state. For any directed graph that is strongly connected, aperiodic, and finite, all stochastic trajectories will eventually arrive at an absorbing state (if at least one is present) with probability 1 [10]. Such a chain is depicted in Figure 4.2, with node 0 as an absorbing state.

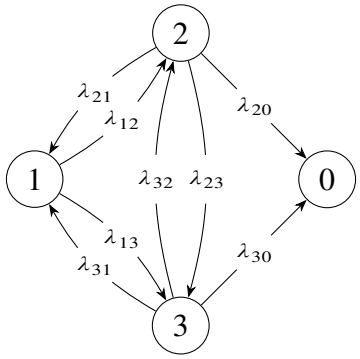


Figure 4.2: A graphical depiction of a Markov chain with an absorbing state labeled "0".

There are a number of classical derivations (e.g. [21]) for computing the moments of the more general *first-passage time* of an arbitrary state. The first-passage time of an absorbing state is also its absorption time, since by definition the trajectory will then never leave the state. Making use of the assumption that the state is absorbing, a simpler argument can be made to characterize this time. Consider a stochastic trajectory starting at node i. The functions $f_i(t)$ and $F_i(t)$ are the probability density and cumulative distribution functions of the time T_i at which the trajectory arrives at node 0, having started at node i. Noting that $f_0(t) = \delta(t)$, an application of the law of total probability yields

$$\begin{aligned} f_i(t) &= \sum_{j \geq 0} \frac{\lambda_{ij}}{\lambda_i} \int_0^t f_j(t-s)\, \lambda_i \mathrm{e}^{-\lambda_i s}\, ds \\ &= \sum_{j \geq 1} \frac{\lambda_{ij}}{\lambda_i} \int_0^t f_j(t-s)\, \lambda_i \mathrm{e}^{-\lambda_i s}\, ds + \frac{\lambda_{i,0}}{\lambda_i} \lambda_i \mathrm{e}^{-\lambda_i t}. \end{aligned} \tag{4.5}$$

The moment generating function $M_i(x) = \int_0^\infty \mathrm{e}^{tx} f_i(t)\, dt$ will satisfy

$$(\lambda_i - x) M_i(x) - \sum_{j \geq 1} \lambda_{ij} M_j(x) = \lambda_{i,0}. \tag{4.6}$$

Collecting $\{M_i(x)\}$ for $i > 0$ into the vector $\vec{M}(x)$, we obtain the linear equation

$$(\Lambda - xI)\vec{M}(x) = \Lambda \vec{1} \tag{4.7}$$

where $\Lambda_{ii} = \lambda_i$ for $i \in \mathbb{N}$ and $\Lambda_{ij} = -\lambda_{ij}$ otherwise.

This relation can be used to write a closed form for the uncentered moments, since

$$\vec{M}(x) = (\Lambda - xI)^{-1}\Lambda\vec{1} = \sum_{n=0}^{\infty} \Lambda^{-n}\vec{1}x^n = \sum_{n=0}^{\infty} \mathbb{E}[\vec{T}^n]\frac{x^n}{n!}$$
$$\implies \mathbb{E}[\vec{T}^n] = n!\,\Lambda^{-n}\vec{1} = n\Lambda^{-1}\mathbb{E}[\vec{T}^{n-1}], \quad (4.8)$$

where $\vec{T}^n \equiv (T_1^n, T_2^n, \dots)$. Rearranging, the following recurrence relation is obtained for the computation of the uncentered moments:

$$\Lambda\,\mathbb{E}[\vec{T}^n] = n\mathbb{E}[\vec{T}^{n-1}]. \quad (4.9)$$

By construction, Λ is a weakly diagonally-dominant matrix with at least one row where the dominance is strict, guaranteeing the existence of an inverse.

4.2.2 Curse of Dimensionality

Returning to the coagulation-fragmentation models of assembly, let us more formally describe the state space to be endowed with stochastic transitions via the Markov chain. Consider the set of non-negative, integer-valued vectors

$$\Omega_m = \left\{ \vec{n} \in (\{0\} \cup \mathbb{N})^m : \sum_{i=1}^{m} i n_i = m \right\}. \quad (4.10)$$

Ω_m is the set of all mass-preserving configurations of particle sizes; e.g. $\Omega_m = \{(m, 0, 0, 0, \dots), (m-2, 1, 0, 0, \dots), \dots\}$. Now consider the subset $\mathscr{A}_{k,m} \subset \Omega_m$ such that

$$\mathscr{A}_{k,m} = \{\vec{n} \in \Omega_m : \exists\, i > k \text{ where } n_i > 0\}. \quad (4.11)$$

For the purposes of nucleation, all $\vec{n} \in \mathscr{A}_{k,m}$ are equivalent, thus a full, stochastic description of the nucleation problem will describe transitions between each $\vec{n} \in \Omega_m \backslash \mathscr{A}_{k,m}$ and transitions from those vectors to $\mathscr{A}_{k,m}$. In our formulation of the nucleation time computation, we were able to omit the "0" state from the linear system. Therefore, the dimensionality of the nucleation time problem for coagulation-fragmentation will be $\left|\Omega_m \backslash \mathscr{A}_{k,m}\right|$.

With a counting argument similar to that from Chapter 3, the number of such states will be the coefficient of x^m in the Taylor series expansion of $\prod_{i=1}^{k}(1-x^i)^{-1} = \sum_{i\geq 0} a_{k,i}x^i$. These coefficients are tabulated in Table 4.1 for nucleation sizes $k+1 = 5, 10, 15, 20$ and protein numbers $m = 265$ and $m = 1140$ (estimates for the numbers of Rnq1 and Sup35 proteins in a typical *S. cerevisiae* cell [8]). For the smaller nucleation sizes, the associated linear problem is tractable even on commodity computer hardware, provided that care is taken with the choice of schemes to solve the linear system and the representation of the linear operator itself.

Table 4.1: $\left|\Omega_m \backslash \mathscr{A}_{k,m}\right|$ for various m and k.

$a_{k,m}$	$k = 4$	$k = 9$	$k = 14$	$k = 19$
$m = 265$	1.37×10^5	3.16×10^9	5.75×10^{11}	1.03×10^{13}
$m = 1440$	1.04×10^7	2.28×10^{14}	1.80×10^{19}	5.62×10^{22}

4.2.3 A Matrix-Free Method

By restricting the assembly mechanism to binary fusion and fission reactions, it is ensured that Λ is a sparse matrix. The degree of that sparsity, of course, depends on the specific allowable transitions: for example, the Becker-Döring matrix will be sparser than a general coagulation-fragmentation matrix. Nonetheless, sparse storage of the matrix will increase memory usage of a solver by at least a constant factor, which may push it beyond the memory capacity of a commodity computer for the parameter regimes of interest. I instead outline and implement a matrix-free method to computing the action of Λ on a vector, permitting the use of iterative methods to invert the linear system as well as enabling a limited number of preconditioners.

First, a natural ordering is imposed over the state space. States with larger particles will have a higher enumerative index than states with smaller particles. An example of this ordering is depicted in Figure 4.3. This ordering permits a straight-forward way of enumerating over entries in the solution vector.

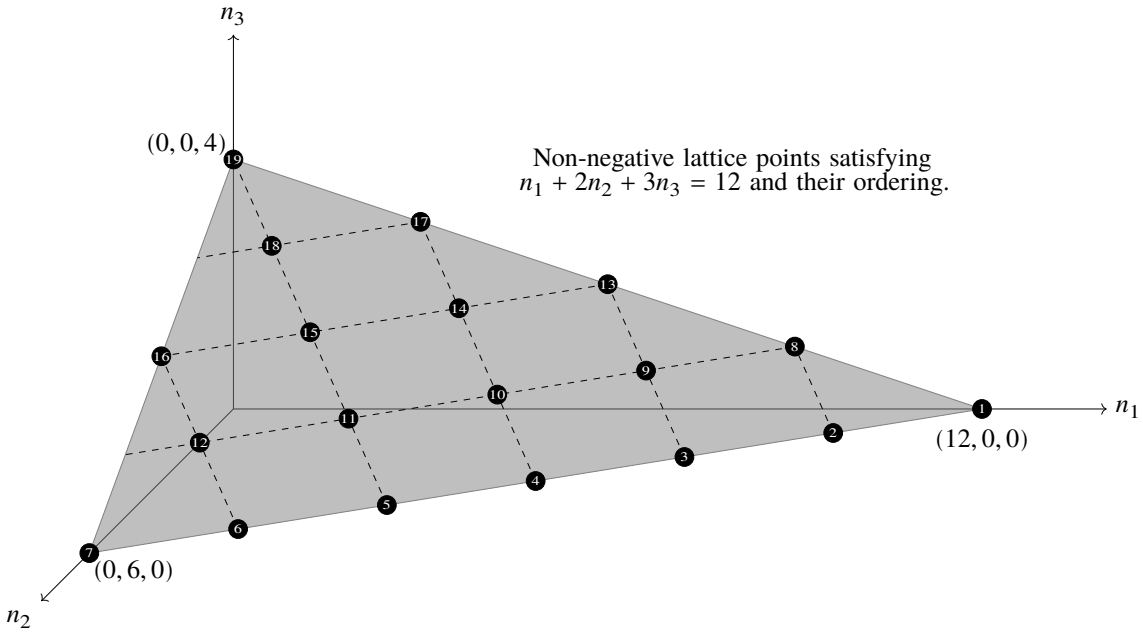


Figure 4.3: An example of the "natural ordering" over states $\vec{n} \in \Omega_{12} \backslash \mathscr{A}_{3,12}$.

Second, given a state $\vec{n}$, the assembly model specifies each of its neighbors $\{\vec{n}'\}$ and the rates at which trajectories transition from $\vec{n}$ to $\vec{n}'$. Thus, as $(\vec{n}, i)$ is enumerated along the solution vector, the corresponding enumerative index i' for every neighbor $\vec{n}'$ needs

to be found. Fortunately, by defining the enumerative scheme in this manner, the index of a state may be written as the sum of several disjoint sets, each whose sizes can in turn be computed recursively with the same partitioning. This indexing scheme can be written as

$$\begin{aligned}\mathrm{ind}(n_1, n_2, \ldots, n_k) = &\sum_{i=0}^{n_k-1} a_{k,m-ki} + \sum_{i=0}^{n_{k-1}-1} a_{k-1,m-kn_k-(k-1)i} \\ &+ \ldots + \sum_{i=0}^{n_2-1} a_{2,m-kn_k-(k-1)n_{k-1}-\ldots-3n_3-2i} + 1.\end{aligned} \tag{4.12}$$

Each index calculation requires a bounded number of integer operations and is subsequently cheap relative to the overall cost of matrix multiplication. Taken together, these three principles allow us to compute the action of Λ on any vector, permitting the application of iterative methods to solving the linear system. These same principles also enable the computation of the action of Λ^{T} on a vector so that the solution scheme is not necessarily limited to transpose-free methods such as TFQMR [26].

4.2.4 Preconditioners for Nucleation

Given the matrix-free representation of the operator and assumption of high-dimensional solution vectors, we are limited to iterative schemes with small recurrences. This includes schemes such as BiCGSTAB, TFQMR, and GMRES(k) for small k [26]. Further improvements then come from the selection of preconditioners. The construction of the enumeration scheme additionally permits some insight into this choice: since states with larger particles are sorted after states with smaller particles, entries in Λ corresponding to fission are in the lower-triangular portion of the matrix and those corresponding to fusion are in the upper-triangular region. Thus, the matrix Λ can be split such that

$$\Lambda = D - L - U, \tag{4.13}$$

where D is a positive, diagonal matrix, L is a non-negative, lower triangular matrix that is is solely a function of the fission reactions, and U is a non-negative, upper triangular matrix and solely a function of the fusion reactions.

Insight into the assembly model then translates into insight into effective preconditioners: if fission is *a priori* known to be the dominant reaction type, then $|L| \gg |U|$ and one should choose a forward Gauss-Seidel or more general forward SOR preconditioner. Conversely, if fusion is the dominant reaction type, then $|U| \gg |L|$ the backward Gauss-Seidel or backward SOR should perform better. The application of either preconditioner requires only 1 extra matrix-multiply per iteration, but may quickly recoup that cost in actual use. This is demonstrated this in Figure 4.4, where an 8037-dimensional linear system corresponding to $m = 100$ proteins and $k = 4$ maximum size is solved for with both Becker-Döring assembly (with $a_{1,j} = a_{i,1} = \sigma$ and $a_{ij} = 0$ otherwise and $b_{ij} = 1$ or 0 similarly) and coagulation-fragmentation assembly (with $a_{ij} = \sigma$ and $b_{ij} = 1$), varied over a range of parameter values to highlight how the dominant reaction type can suggest a preconditioner.

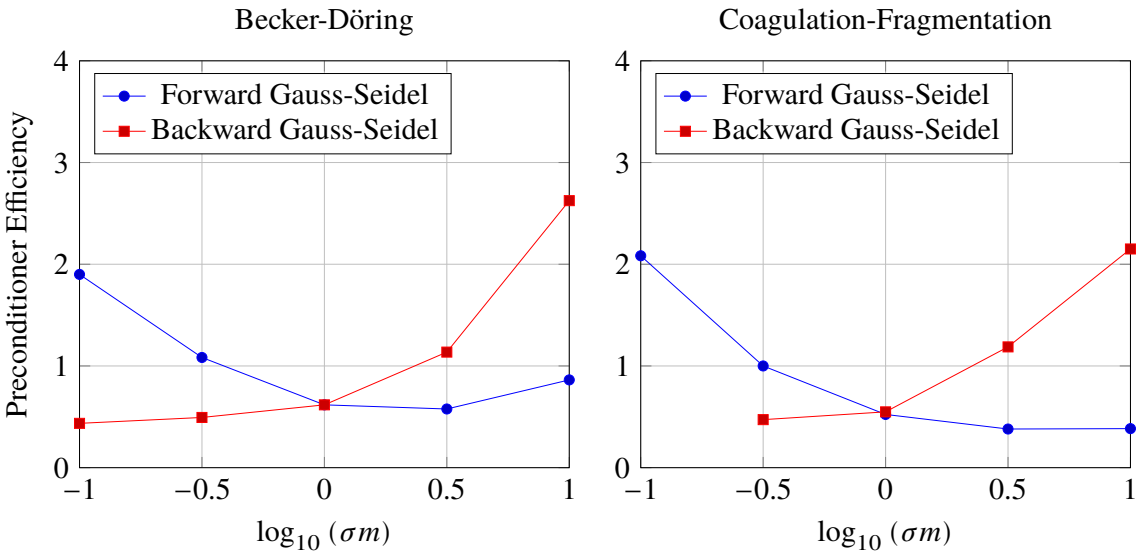


Figure 4.4: Each iteration of a preconditioned solve requires 2 matrix multiplications rather than 1. Using the default convergence criteria in `Eigen`'s BiCGSTAB implementation, the ratio of the number of matrix multiplies in an unpreconditioned solve to the number of matrix multiplies in a preconditioned solve is plotted against a varying parameter that roughly corresponds to the preference of fusion reactions over fission reactions. If this ratio is larger than 1 the preconditioned solve was more efficient than the unpreconditioned solve. We see that $\sigma m \ll 1$ or $\sigma m \gg 1$ is a good heuristic for the choice of preconditioner in these models of assembly. (A point in the coagulation-fragmentation plot was omitted where the backward preconditioner failed to converge within 4×8037 iterations at $\sigma m = 10^{-1}$.)

I provide a `C++` implementation of the above algorithms that computes an arbitrary number of moments for the nucleation time in this dissertation's software repository[1]. The software does not yet perform any sort of parallel optimization as it has mainly been used in parameter sweep explorations, which are themselves embarrassingly parallel. Nonetheless, the software has allowed the preliminary exploration of the otherwise computationally difficult nucleation problem at problem sizes larger than previously published in the literature – such explorations are depicted in Figures 4.6, 4.7, and 4.8.

There are a number of remaining considerations, however, before one can start drawing biological conclusions from the output of the solver. One such consideration is detailed in the next section, where the choice of a prior probability density over the state space is contemplated, since an *in vivo* cell is likely to have some uncertainty over which protein-configuration it possesses.

[1]The code is available at `http://www.github.com/jasondark/dissertation` in the folder `moment-solver`.

4.3 Initial Conditions of Homogeneous Assembly (Journal Article)

The co-author listed in this publication directed and supervised research which forms the basis for the dissertation.

4.3.1 Abstract

The formation of a stable protein aggregate nucleus is regarded as the rate limiting step in the establishment of prion diseases. In these systems, once aggregates reach a critical size the growth process accelerates and thus the waiting time until the appearance of the first critically-sized aggregate is a key determinant of disease onset. In addition to prion diseases, aggregation and nucleation is a central step of many physical, chemical and biological process. Previous studies have examined the first-arrival time at a critical nucleus size during homogeneous self-assembly under the assumption that at time $t = 0$ the system was in the all-monomer state. However, in order to compare to *in vivo* biological experiments where protein constituents inherited by a newly born cell likely contain intermediate aggregates, other possibilities must be considered. We consider one such possibility by conditioning the unique ergodic size distribution on sub-critical aggregate sizes – this "least-informed" distribution is then sampled from to provide initial conditions. We demonstrate that these initial conditions can lead to significantly different, averaged waiting times relative to the all-monomer case under various models of assembly.

4.3.2 Introduction

The self-assembly of particles into aggregates is fundamental to many physical, chemical, and biological processes [29, 32, 33]. In particular, protein aggregation is critical for prion diseases which encompass a number of fatal neurodegenerative diseases in mammals such as Creutzfeldt-Jacob and Kuru in humans, bovine spongiform encephalopathies (BSE) in cows, scrapie in sheep, and chronic wasting in elk [13, 17]. In these diseases a misfolded form of the protein appears and forms aggregates. These small aggregates of misfolded protein are thought to reproduce slowly until reaching a critical size (nucleation), at which point the aggregates are able more efficiently amplify [18, 20, 15, 22]. As such, a key limiting step in the onset of prion diseases is the waiting time until the formation of this nucleus, which we will also call the assembly time. Beyond mammalian diseases, prions have been associated with a number of harmless heritable phenotypes in yeast [31, 27]. Both the harmless nature of the prions and the experimental tractability of yeast have made it an ideal model system to study the appearance of prion disease *in vivo*, and whose considerations will in turn affect our choice of *in silico* models.

Most prior models for computing the waiting times until nucleus formation have assumed that the initial condition of the stochastic process is the all monomer state. That is, no aggregates of any kind exist at time $t = 0$. However, in many systems an individual realization of protein aggregation represents a sample drawn from a preexisting pool of proteins. Since the experimental system has existed for some time, it is possible that some aggregates already exist at the start and, as such, the all monomer initial condition may not be the accurate choice. In particular, since newly born yeast daughter cells inherit protein constituents from their mothers, it is likely that both monomers and small aggregates are transmitted. In this manuscript, we take a step towards more accurate modeling of nucleation in prion diseases by considering protein aggregation under more realistic initial conditions that we call "least-informed." We first give an overview of common models used to study molecular assembly and then establish the least-informed distribution as a consequence of ergodicity. Lastly, we demonstrate that the initial condition chosen for stochastic self-assembly problems can have a significant impact on the expected assembly time.

4.3.3 Modeling Molecular Assembly

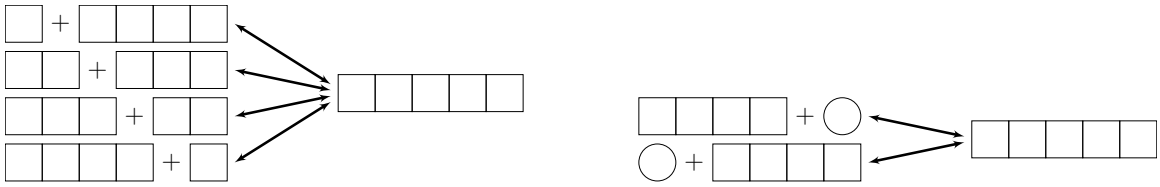
Mathematical models of assembly have been studied for nearly a century, and in particular two systems have received considerable attention: the Smoluchowski coagulation system [28] and the Becker-Döring equations [5]. The former system (with its generalization in [6]) models the coagulation (or coalescence) and fragmentation of polymers consisting of monomeric units, represented by the chemical equations

$$\underbrace{X_j + X_{i-j}}_{j=1,2,\ldots,i-1} \xrightarrow{k^+_{j,i-j}} X_i \xrightarrow{k^-_{k,i-k}} \underbrace{X_k + X_{i-k}}_{k=1,2,\ldots,i-1}, \tag{4.14}$$

where X_i denotes a polymer (oligomer) or aggregate of size i. We refer to this model more generally as the discrete coagulation-fragmentation model. The Becker-Döring model restricts the reactions to just monomer polymerization and depolymerization, represented by the chemical equations

$$X_1 + X_i \underset{k^-_{i+1}}{\overset{k^+_i}{\rightleftharpoons}} X_{i+1}. \tag{4.15}$$

Both models are depicted graphically in Figure 4.5. From these chemical equations, mathematical equations may be derived with additional assumptions. For example, use of the Law of Mass Action and the corresponding rate equations have been thoroughly studied for these models [5, 6, 4, 2, 3]. However, this approximation requires a large number of particles to be valid and fails to capture mesoscopic effects [12]. When these effects are non-negligible, one must instead assume a Markov property and obtain a continuous-time Markov chain. We may then study the *first-passage time* for the subset of states containing a sufficiently large polymer (nucleus). For small numbers of particles or certain asymptotic regimes, these equations are numerically tractable [35, 11], but in general their high-dimensionality often requires the use of Monte Carlo methods to estimate distributions.



(a) Coagulation-Fragmentation Model (b) Becker-Döring Model

Figure 4.5: A depiction of 2 models of polymer assembly.

With any differential or stochastic system, however, an initial condition (IC) must be specified. We observe that in all of the cited works, the IC has exclusively been the all-monomer state, where all the system's mass is initially bound in X_1 particles. Though convenient, it is not realistic for many *in vivo* processes. Indeed, with our yeast prion model, cells transfer material from mother to daughter during budding, including partially formed aggregates [30, 9]. Even within a single cell, these processes of protein assembly and disassembly are always occurring, even before the "experimental clock" begins. Thus, we believe it more accurate to think of the IC as being sampled from some distribution of states which reflects our uncertainty regarding the precise configuration of the polymers. We will propose one such distribution and demonstrate that it can lead to significant differences in the first-passage time versus explicitly assuming an all-monomer IC.

4.3.4 Least-Informed Distribution for Stochastic Assembly

Let us define $\Omega_m = \{n : \sum_{i>0} in_i = m\}$, where each $n \in \Omega_m$ defines a state with total mass m. A particular model of assembly will induce a directed graph on this set of states with edges weighted by the rates of transition between states. We require that the directed graph be *strongly connected* and *aperiodic* – this will imply that the corresponding Markov chain will asymptotically reach a unique ergodic distribution over Ω_m [10].

It is this ergodic distribution that we claim is a more natural source of IC, with one addition: we condition it on every polymer being of size k or less (for some fixed positive integer k). We will refer to this conditional distribution as the "least-informed" distribution (though this label is not intended in a Bayesian sense), since it corresponds to the claim that the physical system has existed long enough to deviate substantially from the all-monomer state, but with the knowledge that it has yet to fully assemble a size $k + 1$ (or larger) polymer. We now provide the novel, least-informed distributions for the constant-rate, coagulation-fragmentation and Becker-Döring models, then demonstrate differences in the expected assembly time between the least-informed distribution and the all-monomer IC.

Notation

We use standard notation for the ith canonical basis vector e_i and the Kronecker delta function δ_{ij}. We let $\mathscr{P} : [0, \infty) \times \Omega_m$ denote the solution to an appropriately-defined master equation, where $\mathscr{P}(t, n) = \Pr[N_1(t) = n_1, N_2(t) = n_2, \dots]$ (and the random

variable $N_i(t)$ is understood to represent the number of X_i particles at time t). The use of the master equation is convenient since the initial condition is itself a distribution – the all-monomer IC state corresponds to a point-mass, or degenerate distribution centered about that state, while our least-informed distribution will be more generally some non-negative vector with entries summing to 1.

Coagulation-Fragmentation Model

A coagulation-fragmentation model allows polymers of any size to freely coagulate (or coalesce) as well as fragment into any size. When these rates are assumed to be constant (size-independent) the reactions are succinctly represented:

$$\underbrace{X_j + X_{i-j}}_{j=1,2,\ldots,i-1} \xrightarrow{\beta} X_i \xrightarrow{\gamma} \underbrace{X_k + X_{i-k}}_{k=1,2,\ldots,i-1}. \tag{4.16}$$

We assume polymers can coagulate on either end; however, we double-count pairs (X_j, X_{i-j}) and (X_{i-j}, X_j). Thus, the unordered pair (X_j, X_{i-j}) coagulates at rate 2β to form X_i. We define the operators $W^{\pm}_{i,j}$ such that $W^{\pm}_{i,j}\mathscr{P}(t,n) = \mathscr{P}(t, n \pm (e_i + e_j - e_{i+j}))$ and write the master equation:

$$\begin{aligned}\frac{d\mathscr{P}}{dt} = &-\beta \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} n_i(n_j - \delta_{ij})\mathscr{P} + \beta \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(n_i + 1 + \delta_{ij})(n_j + 1)W^{+}_{i,j}\mathscr{P} \\ &- \gamma \sum_{i=1}^{\infty}(i-1)n_i\mathscr{P} + \gamma \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(n_{i+j} + 1)W^{-}_{i,j}\mathscr{P}.\end{aligned} \tag{4.17}$$

It can be verified by inspection that

$$\mathscr{P}(n) \propto \frac{1}{\prod (\beta/\gamma)^{n_i} n_i!} \tag{4.18}$$

is a steady-state solution to (4.17). (For completeness, we derive the normalizing constant in the Appendix.) Since the ergodic distribution is unique, we conclude that Equation (4.17) asymptotically converges to this solution. This gives us the least-informed distribution:

$$\Pr\left[N = n | N_i = 0 \; \forall \; i > k\right] = \frac{1}{\prod_{i\le k}(\beta/\gamma)^{n_i} n_i!} \Big/ \sum_{n'} \frac{1}{\prod_{i\le k}(\beta/\gamma)^{n'_i} n'_i!}. \tag{4.19}$$

We use this distribution to sample initial conditions for the corresponding stochastic model of assembly and compute the expected first passage-time (μ) to any state containing a nucleus of size greater than k. This statistic is then compared to that generated by the all-monomer IC in Figures 4.6 and 4.7.

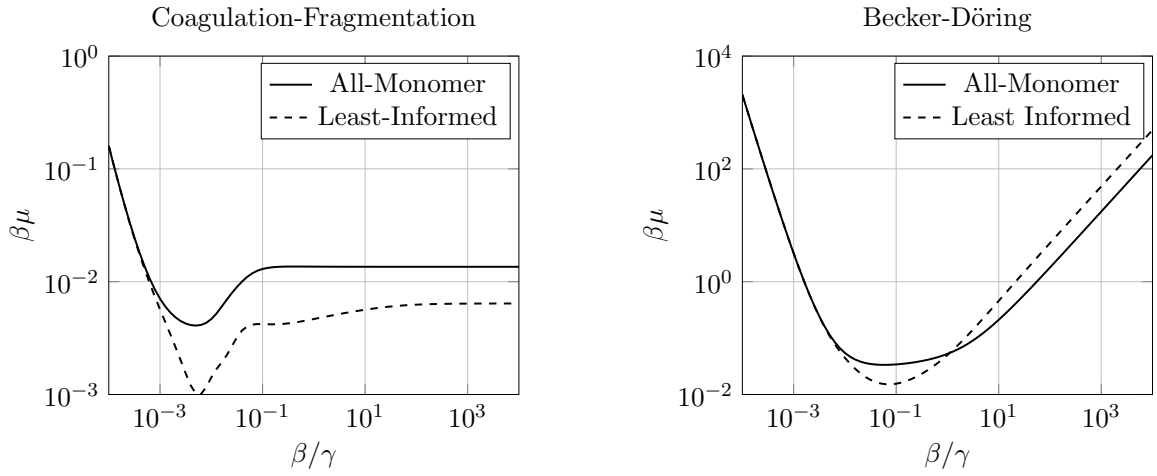


Figure 4.6: Non-dimensionalized, expected assembly times ($\beta\,\mu$) of a particle of size at least 5 ($k = 4$) over a range of β/γ with fixed mass $m = 50$.

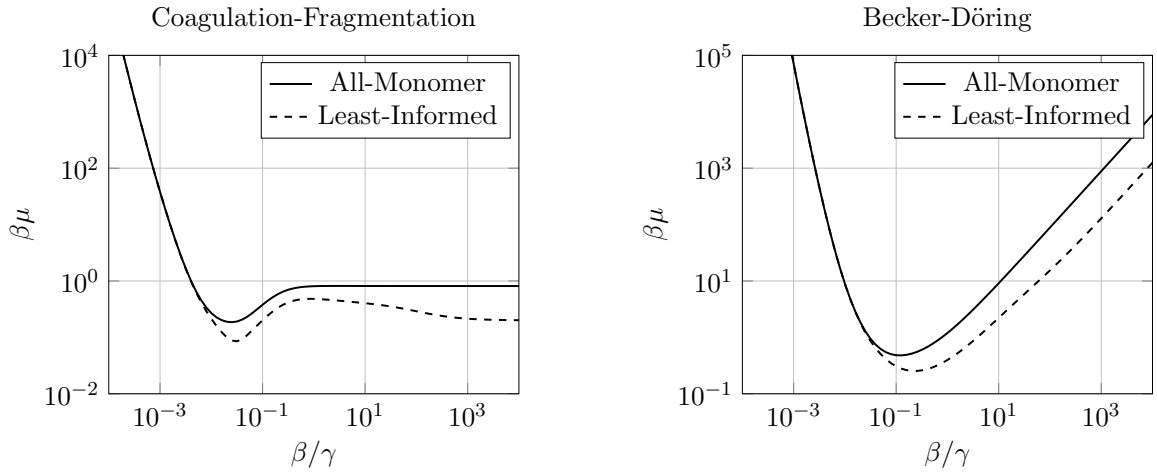


Figure 4.7: Non-dimensionalized, expected assembly times ($\beta\,\mu$) of a particle of size at least 7 ($k = 6$) over a range of β/γ with fixed mass $m = 30$.

Becker-Döring Model

The Becker-Döring model with constant rate coefficients is described by the chemical equations

$$X_1 + X_i \underset{2\gamma}{\overset{2\beta}{\rightleftharpoons}} X_{i+1}. \tag{4.20}$$

It describes an assembly mechanism where monomers attach (polymerize) to a polymer-end with rate β and detach (depolymerize) with rate γ. Defining the operators $W_i^{\pm}\mathscr{P}(t,n) = \mathscr{P}(t, n \pm (e_1 + e_i - e_{i+1}))$, the corresponding master equation is

$$\begin{aligned}\frac{d\mathscr{P}}{dt} = &-\beta n_1(n_1-1)\mathscr{P} - 2\beta n_1 \sum_{i=2}^{\infty} n_i \mathscr{P} - 2\gamma \sum_{i=2}^{\infty} n_i \mathscr{P} + 2\gamma \sum_{i=1}^{\infty}(n_{i+1}+1) W_i^{-}\mathscr{P} \\ &+ \beta(n_1+2)(n_1+1)W_i^{+}\mathscr{P} + 2\beta(n_1+1)\sum_{i=2}^{\infty}(n_i+1)W_i^{+}\mathscr{P}.\end{aligned} \tag{4.21}$$

One may verify that

$$\mathscr{P}(n) \propto \frac{2^{n_1}}{\prod (2\beta/\gamma)^{n_i} n_i!} \tag{4.22}$$

is a steady-state solution to (4.21). Then, as before, we obtain the least-informed distribution

$$\Pr\left[N = n | N_i = 0 \; \forall \; i > m\right] = \frac{2^{n_1}}{\prod_{i\le m}(2\beta/\gamma)^{n_i} n_i!} \Bigg/ \sum_{n'} \frac{2^{n_1'}}{\prod_{i\le m}(2\beta/\gamma)^{n_i'} n_i'!}. \tag{4.23}$$

The comparison of this distribution to the all-monomer IC is plotted alongside the previous Coagulation-Fragmentation comparison in Figures 4.6 and 4.7.

4.3.5 Discussion

We see in Figures 4.6 and 4.7 that when γ is large relative to β, there is virtually no difference between the standard all-monomer IC and our least-informed distribution. However, that quickly changes, and we further note that the axes are non-dimensionalized – any small difference between the 2 distributions may be scaled over orders of magnitude as β varies. This is demonstrated by considering the relative difference between the two, depicted in Figure 4.8.

We also observe that while the all-monomer IC is independent of β/γ, the least-informed distribution is not. As β/γ increases, the least-informed IC will become increasingly biased towards configurations that favor large polymer sizes. While this allows the coagulation-fragmentation model to always assemble at least as fast as with the all-monomer IC, we note that it can have adverse effects under Becker-Döring assembly: if there are no monomers available, then transitions to larger polymers cannot occur. While this explains the interesting, non-monotonic behavior of the assembly time (also observed in [35]), it additionally explains why sampling the least-informed distribution may result in faster or slower assembly times than the all-monomer IC under Becker-Döring assembly.

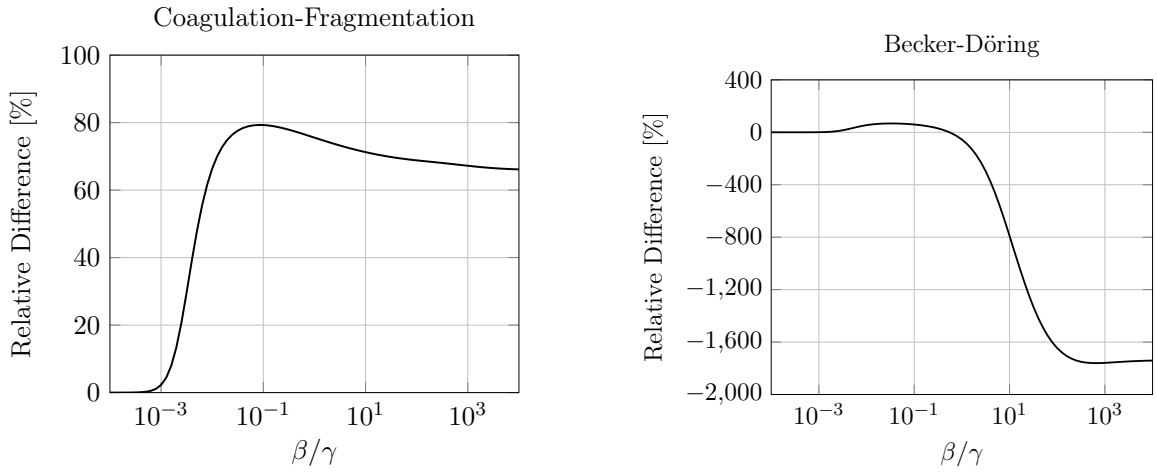


Figure 4.8: Relative difference of expected assembly times under the different IC, $(1-\mu_{\mathrm{LI}}/\mu_{\mathrm{AM}})\times 100\%$, of a particle of size at least 5 ($k=4$) over a range of β/γ with fixed mass $m=100$.

Our analysis demonstrates that the mathematical result is sensitive not only to the choice of assembly model, but also the choice of IC. We argue that for *in vivo* experiments, sampling from the least-informed distribution is more in line with our physical knowledge. Conversely, for *in vitro* experiments (where, for example, protein is heated and denatured), the all-monomer IC may be most appropriate.

We lastly note that our least-informed distribution is but one of many plausible distributions to sample the IC from. In the case of our yeast system, for example, one might consider sampling polymer configurations from some appropriately defined branching process [34, 23] to model the mother-daughter budding aspect. The point remains, however, that only in very special circumstances should the initial condition be treated as a known quantity – doing so glosses over important biological or physical details of the system under study.

Acknowledgments

The authors would like to acknowledge funding from NSF grant #1344279.

4.3.6 Appendix: Normalization Constants

When specifying the full probability distributions for ergodic assembly we omitted the normalizing constant which divides out when finding the conditional probability, making its determination unnecessary. However, we note that there is a convenient form for these constants in terms of the well-studied family of complete Bell polynomials $\{Y_m(x)\}$, where $Y_m=\sum_{k=1}^{m} B_{m,k}$ [1].

Recall, $\Omega_m=\{n : \sum in_i = m\}$. We wish to find generally

$$\kappa_m(x)=\sum_{n\in\Omega_m}\prod_i \frac{x_i^{n_i}}{n_i!}=Y_m(1!\,x_1,2!\,x_2,\ldots,k!\,x_m)/m!\,, \tag{4.24}$$

Letting $\rho = \beta/\gamma$, the Becker-Döring constant would be

$$1/\kappa_N(1/\rho, 1/(2\rho), 1/(2\rho), \dots, 1/(2\rho)) = N!/Y_N(1!/\rho, 2!/(2\rho), 3!/(2\rho), \dots, N!/(2\rho)) \tag{4.25}$$

and the coagulation-fragmentation constant would be

$$1/\kappa_N(1/\rho, 1/\rho, 1/\rho, \dots, 1/\rho) = N!/Y_N(1!/\rho, 2!/\rho, 3!/\rho, \dots, N!/\rho). \tag{4.26}$$

Though we were unable to find a closed form for the former expression, we are able to explicitly find the latter. We begin by writing the probability generating function for Equation (4.17):

$$f_t = \sum_{i=2}^{\infty} \sum_{j=1}^{i-1} (x_i - x_j x_{i-j}) \left(\rho \frac{\partial^2 f}{\partial x_j x_{i-j}} - \frac{\partial f}{\partial x_i} \right). \tag{4.27}$$

It is clear by inspection that

$$f(x) = \exp\left(\frac{1}{\rho} \sum_{i=1}^{\infty} \alpha^i (x_i - 1) \right) \tag{4.28}$$

is a steady-state solution, where α is a free constant which we will assume for convergence purposes to satisfy $|\alpha| < 1$. This constant is a result of disregarding mass conservation – to restore this property, we will need to recondition our probability density function.

Let us define $g(x) = f(x, x^2, x^3, \dots) = \sum_{n=1}^{\infty} p_n x^n$, where p_n is (by construction) the probability associated with a total system mass of n. Then,

$$\begin{aligned}
g(x) &= \exp\left(-\frac{1/\rho}{1-\alpha} \right) \exp\left(\frac{1}{\rho} \frac{1}{1-\alpha x} \right) \\
&= \exp\left(-\frac{1/\rho}{1-\alpha} \right) \left[1 + \sum_{n=1}^{\infty} \frac{1}{\rho^n n!} \left(\frac{1}{1-\alpha x} \right)^n \right] \\
&= \exp\left(-\frac{1/\rho}{1-\alpha} \right) \left[1 + \sum_{n=1}^{\infty} \frac{1}{\rho^n n!} \frac{1}{\alpha^{n-1}(n-1)!} \frac{d^{n-1}}{dx^{n-1}} \frac{1}{1-\alpha x} \right] \\
&= \exp\left(-\frac{1/\rho}{1-\alpha} \right) \left[1 + \sum_{n=1}^{\infty} \frac{1}{\rho^n n!} \sum_{m=n-1}^{\infty} \binom{m}{n-1} (\alpha x)^{m-n+1} \right] \\
&= \exp\left(-\frac{1/\rho}{1-\alpha} \right) \left[1 + \sum_{m=0}^{\infty} \left(\sum_{n=1}^{\infty} \binom{m+n-1}{n-1} \frac{1}{\rho^n n!} \right) (\alpha x)^m \right] \\
&= \exp\left(-\frac{1/\rho}{1-\alpha} \right) \left[1 + \frac{1}{\rho} \sum_{m=0}^{\infty} {}_1F_1(m+1; 2; 1/\rho)(\alpha x)^m \right] \\
&= \exp\left(-\frac{\alpha/\rho}{1-\alpha} \right) + \frac{1}{\rho} \exp\left(-\frac{1/\rho}{1-\alpha} \right) \sum_{m=1}^{\infty} {}_1F_1(m+1; 2; 1/\rho)(\alpha x)^m.
\end{aligned} \tag{4.29}$$

We may now find $\mathscr{P}$:

$$\begin{aligned}\mathscr{P} &= \left(\frac{\partial^{n_1+n_2+\dots}}{\partial x_1^{n_1} \partial x_2^{n_2} \dots} f \Big|_{x=0} \Big/ \prod n_i! \right) / p_N \\ &= \left(\frac{\alpha^N}{\rho^{n_1+n_2+\dots}} \exp\left(-\frac{\alpha/\rho}{1-\alpha} \right) \Big/ \prod n_i! \right) / p_N \\ &= \frac{\rho \exp(1/\rho)}{{}_1F_1(N+1;2;1/\rho)} \times \frac{1}{\prod \rho^{n_i} n_i!}.\end{aligned} \tag{4.30}$$

Thus the identity

$$Y_n(1!\ x, 2!\ x, \dots, n!\ x)/n! = {}_1F_1(n+1;2;x)x\mathrm{e}^{-x}. \tag{4.31}$$

4.4 Discussion

Having provided a computational framework to compute spontaneous nucleation times under Markovian assumptions and a set of reasonable initial conditions to endow the solution with a biological interpretation, we may now begin to apply computational insights to *S. cerevisiae*. This application is in its infancy, however, due to the difficulty in finding and subsequently validating against experimental data. I propose the conclusion of this line of analysis as future work to be discussed in Chapter 5.

Bibliography

[1] G. E. Andrews. *The theory of partitions*. 2. Cambridge university press, 1998.

[2] J. Ball and J. Carr. "Asymptotic behaviour of solutions to the Becker-Döring equations for arbitrary initial data". *Proc. Roy. Soc. Edinburgh Sect. A* 108 (1988), pp. 109–116.

[3] J. Ball and J. Carr. "The discrete coagulation-fragmentation equations: existence, uniqueness, and density conservation". *Journal of statistical physics* 61.1-2 (1990), pp. 203–234.

[4] J. Ball, J. Carr, and O. Penrose. "The Becker-Döring cluster equations: basic properties and asymptotic behaviour of solutions". *Communications in mathematical physics* 104.4 (1986), pp. 657–692.

[5] R. Becker and W. Döring. "Kinetische behandlung der keimbildung in übersättigten dämpfen". *Annalen der Physik* 416.8 (1935), pp. 719–752.

[6] P. Blatz and A. Tobolsky. "Note on the kinetics of systems manifesting simultaneous polymerization-depolymerization phenomena". *The journal of physical chemistry* 49.2 (1945), pp. 77–80.

[7] L. J. Byrne et al. "The number and transmission of [PSI+] prion seeds (propagons) in the yeast Saccharomyces cerevisiae". *PLoS One* 4.3 (2009), e4670.

[8] Y. T. Chong et al. "Yeast proteome dynamics from single cell imaging and automated analysis". *Cell* 161.6 (2015), pp. 1413–1424.

[9] A. Derdowski et al. "A size threshold limits prion transmission and establishes phenotypic diversity". *Science* 330.6004 (2010), pp. 680–683.

[10] J. L. Doob. *Stochastic processes*. Vol. 101. New York Wiley, 1953.

[11] M. R. D'Orsogna, Q. Lei, and T. Chou. "First assembly times and equilibration in stochastic coagulation-fragmentation". *The Journal of chemical physics* 143.1 (2015), p. 014112.

[12] M. D'Orsogna, G. Lakatos, and T. Chou. "Stochastic self-assembly of incommensurate clusters". *The Journal of chemical physics* 136.8 (2012), p. 084110.

[13] D. M. Fowler and J. W. Kelly. "Aggregating knowledge about prions and amyloid". *Cell* 137.1 (2009), pp. 20–22.

[14] A. C. Ghani et al. “Predicted vCJD mortality in Great Britain”. *Nature* 406.6796 (2000), pp. 583–584.

[15] J. T. Jarrett and P. T. Lansbury Jr. “Seeding “one-dimensional crystallization” of amyloid: a pathogenic mechanism in Alzheimer’s disease and scrapie?” *Cell* 73.6 (1993), pp. 1055–1058.

[16] C.-Y. King and R. Diaz-Avalos. “Protein-only transmission of three yeast prion strains”. *Nature* 428.6980 (2004), pp. 319–323.

[17] T. P. Knowles, M. Vendruscolo, and C. M. Dobson. “The amyloid state and its association with protein misfolding diseases”. *Nature Reviews Molecular Cell Biology* 15.6 (2014), pp. 384–396.

[18] T. P. Knowles et al. “An analytical solution to the kinetics of breakable filament assembly”. *Science* 326.5959 (2009), pp. 1533–1537.

[19] A. K. Lancaster et al. “The spontaneous appearance rate of the yeast prion [PSI+] and its implications for the evolution of the evolvability properties of the [PSI+] system”. *Genetics* 184.2 (2010), pp. 393–400.

[20] J. Masel, V. A. Jansen, and M. A. Nowak. “Quantifying the kinetic parameters of prion replication”. *Biophysical chemistry* 77.2 (1999), pp. 139–152.

[21] R. Metzler, G. Oshanin, and S. Redner. *First-Passage Phenomena and Their Applications*. World Scientific, 2014.

[22] Y. Ohhashi et al. “Differences in prion strain conformations result from non-native interactions in a nucleus”. *Nature chemical biology* 6.3 (2010), pp. 225–230.

[23] P. Olofsson, S. S. Sindi, et al. “A Crump-Mode-Jagers branching process model of prion loss in yeast”. *Journal of Applied Probability* 51.2 (2014), pp. 453–465.

[24] K.-W. Park et al. “De novo appearance and “strain” formation of yeast prion [PSI+] are regulated by the heat-shock transcription factor”. *Genetics* 173.1 (2006), pp. 35–47.

[25] S. Prusiner. “Molecular biology of prions causing infectious and genetic encephalopathies of humans as well as scrapie of sheep and BSE of cattle.” *Developments in biological standardization* 75 (1990), pp. 55–74.

[26] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.

[27] S. S. Sindi and T. R. Serio. “Prion dynamics and the quest for the genetic determinant in protein-only inheritance”. *Current opinion in microbiology* 12.6 (2009), pp. 623–630.

[28] M. v. Smoluchowski. “Drei vortrage uber diffusion, brownsche bewegung und koagulation von kolloidteilchen”. *Zeitschrift fur Physik* 17 (1916), pp. 557–585.

[29] K. Thorkelsson, P. Bai, and T. Xu. “Self-assembly and applications of anisotropic nanomaterials: A review”. *Nano Today* 10.1 (2015), pp. 48–66.

[30] M. F. Tuite and B. S. Cox. “Propagation of yeast prions”. *Nature Reviews Molecular Cell Biology* 4.11 (2003), pp. 878–890.

[31] M. F. Tuite and T. R. Serio. “The prion hypothesis: from biological anomaly to basic regulatory mechanism”. *Nature Reviews Molecular Cell Biology* 11.12 (2010), pp. 823–833.

[32] G. M. Whitesides and M. Boncheva. “Beyond molecules: Self-assembly of mesoscopic and macroscopic components”. *Proceedings of the National Academy of Sciences* 99.8 (2002), pp. 4769–4774.

[33] G. M. Whitesides and B. Grzybowski. “Self-assembly at all scales”. *Science* 295.5564 (2002), pp. 2418–2421.

[34] P. Whittle. “Statistical processes of aggregation and polymerization”. *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 61. 02. Cambridge Univ Press. 1965, pp. 475–495.

[35] R. Yvinec, M. R. D’Orsogna, and T. Chou. “First passage times in homogeneous nucleation and self-assembly”. *The Journal of chemical physics* 137.24 (2012), p. 244107.

Chapter 5

Conclusions and Future Work

5.1 Conclusion

The once controversial "prion hypothesis," the ability of proteins to act as genetic determinants, has now all but been universally accepted by the biological community [22]. More recently, prions have become viewed as an essential epigenetic switch associated with the normal functioning of various biological processes [19] instead of solely a determinant of disease. It should come as no surprise that such an unconventional biological mechanism also provides a rich mathematical modeling opportunity.

My own contributions to the field, as detailed in this dissertation, constitute a significant step towards ecologically valid, mathematical modeling of *in vivo* prion protein pathogenesis. This work was guided and organized by two fundamental questions:

1. How does the infectious prion form of a protein initially occur?

2. Once present, how do the infectious prion "aggregates" persist and amplify across multiple generations of cell division?

Chapter 2 detailed the development of the first prion model to incorporate both the nucleated polymerization assumptions and the known interactions of the Hsp104 chaperone with the prion aggregates. This new model successfully recapitulates known qualitative behavior with Sup35 over-expression and Hsp104 knock-outs [3]. By explicitly considering the chaperone concentrations, this is also the first model to permit prion strain coexistence, which is a particularly exciting result and suggests a number of biological experiments with respect to the Hsp104 expression that could falsify or confirm the model.

Next, Chapter 3 provided a stochastic treatment of the initial introduction of a prion aggregate to an otherwise normal cell. This is the first such fully stochastic model of nucleated polymerization, made tractable via a number of insights derived from the deterministic modeling that had been done prior. Coupled with data provided published by the Serio lab [4, 13], this model also provides the first quantitative evidence for differing minimal stable sizes between prion strains.

Lastly, Chapter 4 expanded upon the work of [6, 23, 5] and developed a numerical solver capable of efficiently characterizing the nucleation time distribution for general models of assembly. This solver was used to study and ultimately propose a novel set of initial conditions for the general inference of assembly mechanisms under *in vivo* conditions. Notably, while many stochastic processes admit effective approximations (e.g. [11, 18]), this particular problem of spontaneous nucleation does not lend itself to these techniques. Thus, my development of a memory-efficient, matrix-free solver and preconditioners is, at this time, the only means of quantitative exploration of these assembly models in biologically-valid parameter regimes.

Nonetheless, considerable work remains. By developing the mathematical models and tools to explore the *in vivo* dynamics of prion aggregates, I am now positioned to answer, or at least provide insight into, a number of long-standing, open biological questions. These questions are outlined in the next section.

5.2 Future Work

5.2.1 Enzyme-Limited Models of Nucleated Polymerization

Chapter 3 detailed the generalization of the nucleated polymerization model (NPM) to include known chaperone effects on the fragmentation process. It is the first such model to recapitulate a number of observed but previously unexplained behaviors, including changes in the aggregate size density and prion strain coexistence [3]. Missing from the work, however, is a global analysis of the nonlinear dynamics. The potentially non-monotonic trajectories of the fragmentation efficiency $p(t)$ make such an analysis difficult, but the same non-monotonicity also casts doubt on the generalizability of the local stability results. Thus, I will continue to search the literature for applicable global stability techniques and look for a Lyapunov function.

With such an analysis, it may be possible to further elucidate the restrictions on prion strain coexistence as well. Notably, [3] only provided a heuristic understanding of why strain coexistence is possible and did not provide any algebraic restrictions on the kinetic parameters. Since prion strains are known to both coexist and compete [17], often with clinical implications [20], the mathematical determination of such algebraic mechanisms have immediate relevance to the biology. The first steps towards better understanding prion strain coexistence include an exhaustive parameter sweep and exploration of a two-strain system, which can be represented as a 6-dimensional, nonlinear ODE system. This is sufficiently small so that for any parameter combination, the equilibria and their local stability is easily calculated. It is expected that insights from the one-strain global stability argument will generalize to the n-strain stability in some manner.

5.2.2 A Two-Hit Model of Prion Pathogenesis

Chapter 4 outlined a mathematical framework to explore the spontaneous nucleation of a *de novo* prion aggregate. By formulating transient dynamics as transitions on a

Markov chain between mass-preserving configurations and studying the absorption time of a "nucleated" state, the problem is recast into the well-studied framework of solving large sparse linear systems.

The difficulty is in reconciling the numerical experimentation with biological data. From the work in Chapter 3, it is known that forming an aggregate, while necessary, is not sufficient to induce a prion phenotype. The aggregate must persist and amplify *in vivo*; thus measurements of the spontaneous $[PSI^+]$ rate are really a combined measurement of both the nucleation rate and the persistence probability. Supposing, for convenience only, that spontaneous nucleation occurs according to a Poisson process characterized by rate parameter λ and the measured induction probability is given by p, note the following:

- λ is determined by an unknown assembly process ($\{a_{ij}\}$ and $\{b_{ij}\}$ from Chapter 4).
- p is determined by a known and well-validated process (nucleated polymerization, Chapter 3).
- Both λ and p are necessarily dependent on the availability of the Hsp104 (and related) chaperones through their respective fragmentation dynamics.

Estimates of spontaneous $[PSI^+]$ appearance will reflect combined parameter $\hat{r} = \lambda p$; however, when the fragmentation dynamics are perturbed, both λ and p will respond yielding $\hat{r}' = \lambda' p'$. Since the response of p' (relative to p) is governed by a known model, this permits an indirect study into the nucleation rate, locally quantifying the sensitivity of λ to changes in fragmentation efficiency. Equipped with both the value and gradient information of λ, one can then begin to eliminate plausible assembly mechanisms.

For example, in unpublished work from the Serio lab[1], it was found that decreasing the availability of Hsp104 increases the rate of $[PSI^+]$. The parameter inference from Chapter 3 shows that decreased Hsp104 also decreases the persistence probability, implying that the spontaneous nucleation rate λ' must significantly increase, both to accommodate the decrease in p' and the increase in $\hat{r}'$. From Chapter 4, Figures 4.6 and 4.7 suggest than that a monomer restriction on coagulation (Becker-Döring) is more capable of recapitulating this behavior than no restrictions (the more general coagulation-fragmentation model).

It is easy enough to make qualitative claims such as these; however, considerable work is required for quantitative evidence. For an assembly model with a maximum particle size of k, there are $k(k+1)$ parameters to infer, in addition to any parameters required for the computation of the persistence probability p. To approach this problem in a principled way, a Bayesian technique will need to be developed in order to propagate uncertainty from the measurements of $\hat{r}$ and the persistence probability into the assembly parameters. This inference has a number of challenges to be addressed, even ignoring what will likely be minimal and noisy experimental data: first and foremost, the maximum particle size k will itself be unknown, in addition to the $k(k+1)$ unknown parameters associated with it. Lastly, the dimensionality of the forward problem grows exponentially in k, thus computing the evidence in favor of a value of k grows exponentially difficult in k itself.

[1] `http://mcb.arizona.edu/people/tserio`

With that caveat, evidence in favor of any particular model of assembly will constitute completely new work. This work has immediate application to the clinical study of prion diseases: [6] observed that monomer-only assembly (Becker-Döring) often finds itself in "pauses" where the monomer population has depleted, preventing the intermediate particles from further growing. In this case, treating a prion infection by increasing the cell's ability to fragment aggregates may worsen the infection by freeing monomers and allowing particles to lengthen again. Conversely, if aggregates of any size may freely coagulate (coagulation-fragmentation model), then an increase in fragmentation should unilaterally improve the infection, due to a decrease in the average length of aggregates. This speaks to the need to understand the origin of spontaneous nucleation before we can effectively treat it.

5.2.3 Automated Approaches to Solving the Chemical Master Equation

Cross-disciplinary communication is one of the largest barriers to effective mathematical biology. Practitioners on either side use their own domain-specific language, and these languages in turn influence how we think about and formulate our problems. Bridging this divide is the key to successful collaboration, and in my work I have come to appreciate the universality of the *chemical equation* as a basic unit of communication.

That is to say, chemical reactions like $A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C$ can be understood very precisely by both mathematicians and biologists and thus serves effectively as a "linguistic tool" to communicate quantitative ideas back and forth. This universality is apparent: virtually all networks and diagrams in systems biology at least include chemical reactions as a fundamental component [14].

There exists a rich theory for the *deterministic* analysis of concentrations of chemical species governed by such reaction networks, known as deficiency theory [9, 10]. However, as noted throughout this dissertation (and more prominently noted in [15]), deterministic models fail when the noise in the system is as fundamental as the mean states themselves. Work has begun to emerge that finds stochastic analogs to the deterministic theorems (e.g. [8]), though in practice there remains a significant difficulty in numerically solving such stochastic formulations.

This was demonstrated in Chapter 4 where considerable effort was invested in describing the states in the Markov chain, prescribing a meaningful ordering, and providing a means of evaluating the chemical master equation (CME) operator. While generalized for all types of binary fission and binary fusion assembly, any sort of additional modification requires significant alteration. For example, the nucleated polymerization model does not have only binary fission: the reaction $X_i \to iX_1$ is possible. This would require an entirely different implementation of the CME operator. It may also be reasonable to assume that monomers do not freely assemble. For example, a monomer may first need to become susceptible, then the susceptible form can be aggregated [2]. This would augment the Becker-Döring model with the reaction $P \underset{k_-}{\overset{k_+}{\rightleftharpoons}} X_1$; in addition

to defining meaningful, rate-limiting constants $k_{\pm}$, an entirely new description of and enumeration over the state space is needed.

These sorts of changes are relatively minor when solving deterministic representations but they create an entirely new problem when treated stochastically. With that said, using very similar techniques from the nucleation time solver in Chapter 4, an automatic framework for setting up and numerically solving the chemical master equation can be developed.

Given any family of mass-preserving chemical reactions, one can construct the stoichiometric matrix: each column represents a unique chemical species involved in the reactions and each row a chemical reaction. The corresponding matrix entry is the number of that species produced or consumed by that reaction. Conservation laws are merely non-negative integer vectors in the nullspace of this matrix. For example, the Michaelis-Menten enzyme model [12] is given by the chemical equations

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SE \xrightarrow{k_2} P + E. \tag{5.1}$$

The stoichiometric matrix R is

$$R = \begin{array}{c} \begin{array}{cccc} S & SE & P & E \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & -1 \\ 1 & -1 & 0 & 1 \\ 0 & -1 & 1 & 1 \end{pmatrix} \end{array} \tag{5.2}$$

The vectors $(1, 1, 1, 0)$, corresponding to $S + SE + P = \text{constant}$, and $(0, 1, 0, 1)$, corresponding to $SE + E = \text{constant}$, are in the nullspace of R. Taking these two vectors as a basis, any state describing the numbers of S, SE, P, and E can be identified by the linear combination of these basis vectors. Endowing a natural ordering to the coefficients of the linear combination, an enumeration over all such states is then defined, just as in Chapter 4. Finally, the mass-action kinetics are easily expressed in the matrix-free application of the CME operator.

I have developed a number of preliminary programs towards this end, but three open problems remain:

1. Inspired by the work on assembly, can a heuristic to automatically select an effective preconditioner based on knowledge of the chemical reaction scheme be found?

2. Given the high dimensionality of these linear systems, can good subsets of states for *finite state projection* (FSP) approximation methods [18, 21, 1] automatically be discovered?

3. Interesting chemical systems often produce and destroy reactants as part of the reactions, making the state-space countably infinite. Can the FSP approximations also accommodate this countably infinite regime?

Without satisfactory resolutions to all of these questions, a generalized solver will be of extremely limited use. However, direct numerical solution of the chemical master equation can drastically out-perform Monte Carlo simulation in certain circumstances [16] and thus serves as a complementary tool for the simulation-based software packages (such as [7]).

Bibliography

[1] K. Burrage et al. “A Krylov-based finite state projection algorithm for solving the chemical master equation arising in the discrete modelling of biological systems”. *Proc. of The AA Markov 150th Anniversary Meeting*. 21-37. 2006.

[2] J. H. Come, P. E. Fraser, and P. T. Lansbury. “A kinetic model for amyloid formation in the prion diseases: importance of seeding”. *Proceedings of the National Academy of Sciences* 90.13 (1993), pp. 5959–5963.

[3] J. K. Davis and S. S. Sindi. “A mathematical model of the dynamics of prion aggregates with chaperone-mediated fragmentation”. *Journal of mathematical biology* 72.6 (2016), pp. 1555–1578.

[4] A. Derdowski et al. “A size threshold limits prion transmission and establishes phenotypic diversity”. *Science* 330.6004 (2010), pp. 680–683.

[5] M. R. D’Orsogna, Q. Lei, and T. Chou. “First assembly times and equilibration in stochastic coagulation-fragmentation”. *The Journal of chemical physics* 143.1 (2015), p. 014112.

[6] M. D’Orsogna, G. Lakatos, and T. Chou. “Stochastic self-assembly of incommensurate clusters”. *The Journal of chemical physics* 136.8 (2012), p. 084110.

[7] B. Drawert et al. “Stochastic Simulation Service: Bridging the Gap between the Computational Expert and the Biologist”. *PLoS computational biology* 12.12 (2016), e1005220.

[8] G. A. Enciso. “Transient absolute robustness in stochastic biochemical networks”. *Journal of The Royal Society Interface* 13.121 (2016), p. 20160475.

[9] M. Feinberg. “Complex balancing in general kinetic systems”. *Archive for rational mechanics and analysis* 49.3 (1972), pp. 187–194.

[10] M. Feinberg. “Necessary and sufficient conditions for detailed balancing in mass action systems of arbitrary complexity”. *Chemical Engineering Science* 44.9 (1989), pp. 1819–1827.

[11] D. T. Gillespie. “The chemical Langevin equation”. *The Journal of Chemical Physics* 113.1 (2000), pp. 297–306.

[12] K. A. Johnson and R. S. Goody. “The original Michaelis constant: translation of the 1913 Michaelis–Menten paper”. *Biochemistry* 50.39 (2011), pp. 8264–8269.

[13] C. L. Klaips et al. “Spatial quality control bypasses cell-based limitations on proteostasis to promote prion curing”. *eLife* 3 (2014), e04288.

[14] N. Le Novere et al. “The systems biology graphical notation”. *Nature biotechnology* 27.8 (2009), pp. 735–741.

[15] I. Lestas, G. Vinnicombe, and J. Paulsson. “Fundamental limits on the suppression of molecular fluctuations”. *Nature* 467.7312 (2010), pp. 174–178.

[16] S. MacNamara, K. Burrage, and R. B. Sidje. “Multiscale modeling of chemical kinetics via the master equation”. *Multiscale Modeling & Simulation* 6.4 (2008), pp. 1146–1168.

[17] R. Morales, K. Abid, and C. Soto. “The prion strain phenomenon: molecular basis and unprecedented features”. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1772.6 (2007), pp. 681–691.

[18] B. Munsky and M. Khammash. “The finite state projection algorithm for the solution of the chemical master equation”. *The Journal of chemical physics* 124.4 (2006), p. 044104.

[19] G. A. Newby and S. Lindquist. “Blessings in disguise: biological benefits of prion-like mechanisms”. *Trends in cell biology* 23.6 (2013), pp. 251–259.

[20] K. P. R. Nilsson et al. “Prion strain interactions are highly selective”. *Journal of Neuroscience* 30.36 (2010), pp. 12094–12102.

[21] S. Peleš, B. Munsky, and M. Khammash. “Reduction and solution of the chemical master equation using time scale separation and finite state projection”. *The Journal of chemical physics* 125.20 (2006), p. 204104.

[22] C. Soto. “Prion hypothesis: the end of the controversy?” *Trends in biochemical sciences* 36.3 (2011), pp. 151–158.

[23] R. Yvinec, M. R. D’Orsogna, and T. Chou. “First passage times in homogeneous nucleation and self-assembly”. *The Journal of chemical physics* 137.24 (2012), p. 244107.