BIOREPS Problem Set #2
The perils of prions: mad cows and cannibals

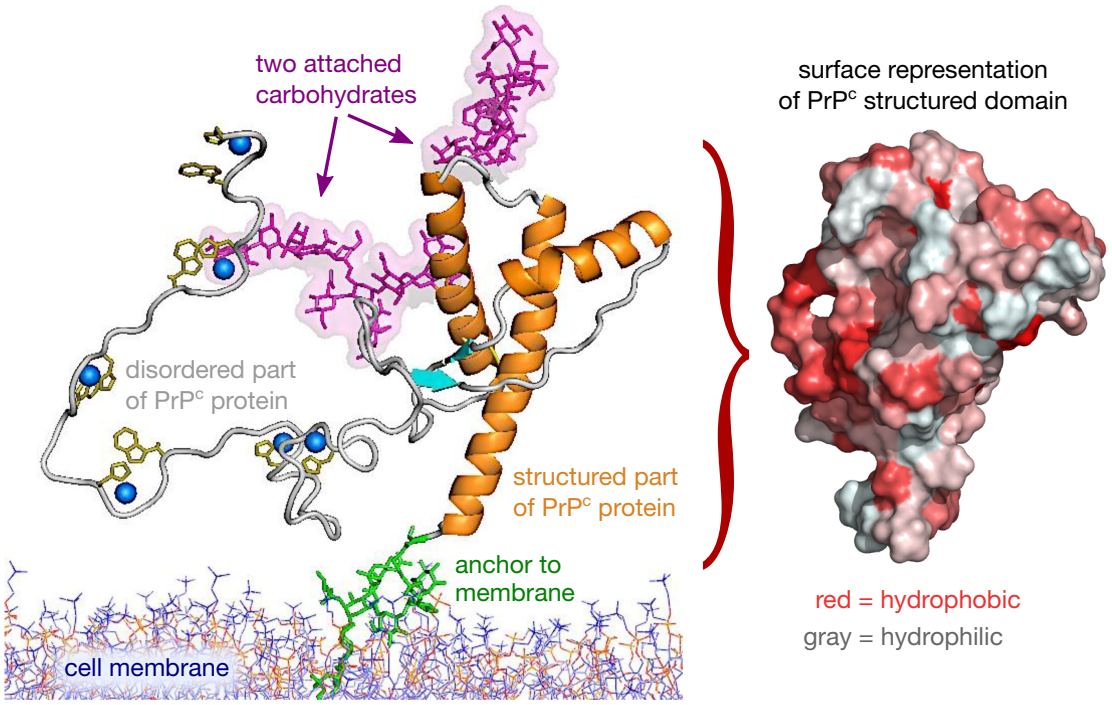

Figure 1: Left: cartoon of the human PrP^C protein in its normal configuration, anchored to the surface of a neuron cell membrane. The disordered region of PrP^C (gray chain) has binding sites for copper ions (blue), while the structured region (orange) consists of mainly amino acids arranged in alpha helices. Adapted from Ref. [1]. Right: a surface representation of the structured part, with the colors denoting the relative hydrophobicity of the amino acids near the surface.

1 Background

The crowded environment of most cellular proteins means that collisions with other molecules are frequent. Only a small subset of these collisions are a meeting of specific partners in a biologically useful reaction. The surfaces of proteins are optimized to trap such partners by carefully shaped and chemically specific binding pockets. The rest of the surface is generally as "non-sticky" as possible, to minimize binding of incorrect molecules. Since hydrophobic amino acids like to form interfaces with other hydrophobes rather than be exposed to water, the hydrophobic parts of the protein chain are usually buried in the interior, leaving mainly hydrophilic residues on the outside. This both stabilizes the three-dimensional structure of the protein, and avoids promiscuous hydrophobic attraction between different protein molecules. The alternative would be catastrophic: if protein surfaces were strongly hydrophobic, they would clump together, progressively forming large masses known as *aggregates*, much like oil droplets coalescing to phase separate from water. Though biological systems struggle to minimize the risk of this scenario,

they do not always succeed. That failure, and its frequently deadly consequences, are the subject of this problem set.

In our daily experience protein aggregation is generally benign: most proteins in our food will unfold from their compact structure ("denature") between 120–160°F, exposing the hydrophobic residues normally buried in the core. When you heat up the white of an egg, composed of albumin proteins dissolved in water, neighboring proteins begin to bind to each through these hydrophobic patches, eventually separating from the water in a white, solid mass. Heating a steak above 160°F accomplishes something similar, resulting in a dry aggregation of myosin and actin muscle proteins that tastes like cardboard. (The art of barbecue consists of *not* entirely denaturing the actin in your meat.)

At non-cooking temperatures we normally do not expect our proteins to spontaneously unfold and begin to aggregate. And yet it is precisely such exceedingly rare fluctuations that underlie the more sinister aspects of aggregation. The most notorious example is the protein PrP^{C}, found in humans and other mammals. Though present throughout the body, it is most highly expressed in the brain and the spinal cord, where it is attached to the exterior cell membrane of neurons (Fig. 1). In its normal state it performs a variety of functions that are not entirely understood, but it is likely to participate in signaling between cells, and the trafficking of copper ions, to which it has a particularly strong affinity. Like many membrane-bound proteins, it has a long disordered region, which is essentially a (mostly) unstructured chain of amino acid residues (gray in Fig. 1). These residues are mainly hydrophilic, and thus lack the hydrophobes that drive protein folding into compact structures. This chain can act like a flexible, thermally fluctuating, fishing net for the small molecules that the protein wants to bind (for example the copper ions). The rest of the PrP^{C} protein has more hydrophobic residues, and thus forms the structured region shown in orange. This region acts like scaffolding for the carbohydrate/lipid anchor (green) that keeps the protein attached to the cell membrane. The structures are mainly alpha helices, bound together in a hydrophobic core (see the cartoon on the left of Fig. 1). The more realistic surface representation (right of Fig. 1), shows the amino acids exposed to water, many of them hydrophilic (gray shading).

PrP^{C} is the Mr. Hyde persona of an elusive Dr. Jekyll: by a mechanism that is currently unknown, PrP^{C} converts to an alternative structure known as PrP^{Sc}. This almost certainly involves the partial or total unfolding of the structured region, exposing hydrophobic residues. What follows is a disaster in slow motion: PrP^{Sc}, with its sticky surface, binds to the small hydrophobic patches exposed through thermal fluctuations on any normal PrP^{C} it encounters during random diffusion on the cell membrane. The bound PrP^{C} can lower its energy by forming additional hydrophobic contacts with PrP^{Sc}. Fueled simply by thermal agitation, the PrP^{C} structure gradually rearranges, exposing hydrophobic residues to maximally bind to PrP^{Sc}. The end result of this process, known as template-assisted refolding, is that the bound PrP^{C} has turned into another copy of PrP^{Sc}. Protein by protein, the aggregate of PrP^{Sc} grows, consuming the population of normal PrP^{C}. The result is a linear structure known as an amyloid fibril, which grows by binding PrP^{C} to its exposed ends.

The story described so far is necessarily incomplete and tentative: despite advances in protein structure-detection techniques, experimentalists have still to figure out the structure of PrP^{Sc} by itself or in its fibril form. Various alternative models exist, including a leading candidate (Fig. 2) proposed by the research group of Witold Surewicz at CWRU and collaborators [2]. Here the the PrP^{Sc} takes on a hairpin shape, which can be stacked into a fibril structure like adding layers on a

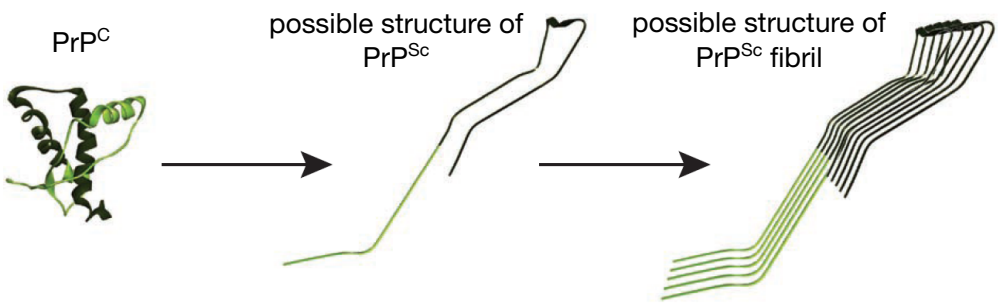

Figure 2: A possible model for the structure of the infectious PrP^{Sc} form of PrP^{C}, based on experiments by Surewicz and coworkers [2]. Figure adapted from Ref. [3].

cake. This resembles the known structures of other proteins that form amyloid fibril aggregrates. Many questions remain on how one gets from PrP^{C} to the fibril: it is entirely possible that PrP^{Sc} is not stable by itself, as a monomer. It could be that the beginning of aggregration requires the clustering of a small number (a so-called "nucleus") of PrP^{C} proteins in some intermediate state with only partial unfolding. This nucleus may then rearrange itself through fluctuations into a more stable stack in PrP^{Sc} form, and this stack then acts as the seed of fibril formation, growing quickly through templated refolding of PrP^{C}. Though the reproduction through the template is highly accurate, conserving the structure of PrP^{Sc}, it is possible that a variety of initial seed forms could exist, leading to somewhat different fibril structures. These questions are likely to persist until we know the actual structure (or structures) of PrP^{Sc}, which will be a major scientific coup.

What happens next is the most remarkable aspect of PrP^{Sc}: the fibrils can fragment, either randomly or through the action of defensive proteins that routinely attempt to clean cells of protein aggregates. Normally the aggregates can be broken into small pieces, and tagged for destruction by proteases, enzymes that break down proteins into amino acids. PrP^{Sc} evades being completely digested by proteases, and individual small fibril fragments become new seeds for fibril growth. If the fibril fragments detach from one cell membrane and diffuse to another, they will begin to consume the healthy PrP^{C} population of the new cell, in an inexorable progression from cell to cell. Eventually the fibrils cluster in giant plaques (Fig. 3), disrupting neurological function, leading to dementia and ultimately death.

Because PrP^{Sc} is likely able to form a variety of fibril structures, the end result is a variety of diseases, differing by their incubation period and symptoms: Creutzfeldt-Jakob disease (CJD), fatal familial insomnia, and kuru in humans, scrapie in sheep, bovine spongiform encephalopathy (BSE or "mad cow disease") in cattle, chronic wasting disease in deer and moose. All forms of the disease are incurable and universally fatal. Because the process of aggregration, fragmentation, and spread from cell to cell is slow, the concentration of PrP^{Sc} in tissues may be tiny and virtually undetectable for years or decades. As we will see in this problem set, this long incubation hides the distressing fact that the growth of the fibril population, while a pool of healthy PrP^{C} exists, is in reality exponential (with a large time constant). Hence once the PrP^{Sc} mass is big enough to cause symptoms and be diagnosed, death follows swiftly, usually in less than a year.

The PrP^{Sc} fibrils are robust and long-lasting, and can be transmitted from animal to animal by eating infected tissue (particularly neural tissue). The human kuru epidemic in Papua New Guinea was caused by ritual cannibalism associated with funeral practices. BSE spread by feeding

cows protein supplements derived from the remains of slaughtered cows. Cross-species transmission is more difficult, but still possible. The small variations in PrP^C proteins between different species means for example that a cow PrP^{Sc} fibril, if eaten by a human, should be relatively ineffective at recruiting human PrP^C to continue growing. Yet the probability is not strictly zero: though hundreds of thousands of BSE-infected cattle entered the human food chain up through the 1980's, the disease was passed to humans in less than 200 cases. The scariest aspect of prion diseases is that they can arise out of nowhere, thanks to the inherently malleable nature of the proteins under thermal agitation. For example, though a small fraction of human CJD is familial, caused by mutations in the PrP^C gene that make the protein more susceptible to conversion, nearly 85% of cases are due to spontaneous conversion of PrP^C to PrP^{Sc}. The rarity of this event probably explains why it occurs generally in older individuals (> 50 years) and with extremely low incidence (one case per million people per year).

The spread of PrP^{Sc} is the simplest biochemical process that achieves a kind of simulacrum of life: growth and templated reproduction, even the ability to "mutate" based on small changes in fibril structure over time that lead to different disease phenotypes. This is accomplished entirely without nucleic acids like DNA or RNA, and without the complex, ATP-driven machinery of transcription and translation that is necessary for virus- or bacteria-based disease. The "fuel" is the supply of normal PrP^C. Stanley Prusiner's 1982 discovery of PrP^{Sc} as the cause of scrapie in sheep led to his coining the term *prion*, a protein infectious agent [4]. This work was so controversial at the time that it nearly led to Prusiner being denied tenure at UCSF, since it contradicted a forty year consensus that nucleic acids are the only carriers of reproducible information in biology. Prusiner would go on to win the Nobel Prize in Medicine in 1997, and our understanding of prions has expanded far beyond disease. For example, certain transcription factor proteins in yeast and other fungi can be converted to a prion form (one well studied example is the yeast protein Sup35). This leads to the sequestration of the protein into fibrils and prevents it from affecting gene expression by binding to DNA. The prion fibrils fragment and spread to daughter cells, and thus pass on the altered gene expression behavior. This is only one of the many ways in which gene expression can be modified without any changes to the genetic code of the organism—the rapidly growing modern field of epigenetics.

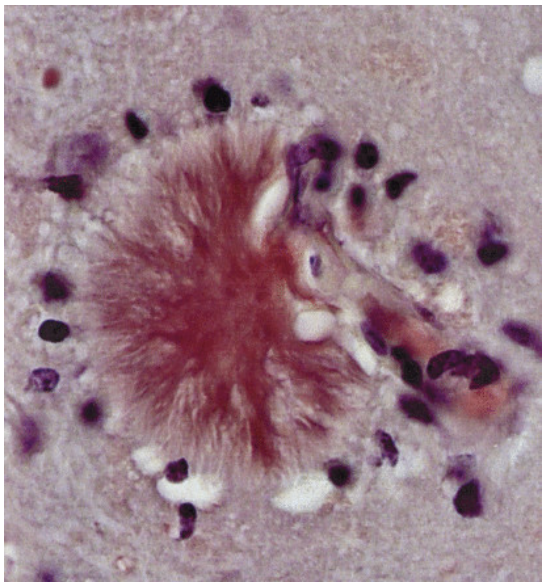

Figure 3: An aggregrate of prion fibrils (dark red cluster) in the brain of a mouse infected with BSE. From: http://jid.oxfordjournals.org/content/195/7.cover-expansion.

Prions are also part of a much broader category of protein aggregation processes linked to disease, most of which are not infectious (in the sense of being able to spread from individual to individual). Many neurodegenerative disorders in humans—Alzheimer's, Parkinson's, Huntington's diseases, as well as chronic traumatic encephalopathy linked to repetitive brain trauma in sports—are marked by deadly accumulations of aggregrated proteins. Prions are distinct in their

stability, and the efficiency of their recruitment of healthy proteins into aggregrates, allowing them to spread not only between cells, but between organisms and even to some extent between species. But the physical processes that govern the growth of fibrils in all these cases share many similarities. The analytical model of fibril dynamics we will explore in this problem set, based on the recent work of Knowles *et al.* [5], thus has a broad range of potential applications.

Note: All references are available in the Resources section of the course website.

References

[1] Acevedo-Morantes, C. Y. and Holger, W. The Structure of Human Prions: From Biology to Structural Models—Considerations and Pitfalls. *Viruses* **6**, 3875–3892 (2014).

[2] Cobb, N. J., Sonnichsen, F. D., McHaourab, H. and Surewicz, W. K. Molecular architecture of human prion protein amyloid: a parallel, in-register β-structure. *Proc. Natl. Acad. Sci. USA* **104**, 18946–18951 (2007).

[3] Diaz-Espinoza, R. and Soto, C. High-resolution structure of infectious prion protein: the final frontier. *Nat. Struct. Mol. Biol.* **19** 370–377 (2012).

[4] Bolton, D. C., McKinley, M. P. and Prusiner, S. B. Identification of a protein that purifies with the scrapie prion. *Science* **218**, 1309–1311 (1982).

[5] Knowles, T. P. *et al.* An analytical solution to the kinetics of breakable filament assembly. *Science*, **326**, 1533–1537 (2009).

2 Analytical model for fibril assembly

Assume we have a cellular volume, with a concentration $m(t)$ of normal protein monomers at time t. Pairs of these can randomly collide, and in most cases they will not bind together. However under rare circumstances (with a tiny nucleation rate k_n) they will collide and spontaneously change form to make an infectious dimer. The concentration of dimers is $c_2(t)$. These dimers can recruit and convert normal proteins with rate k_+, growing into longer polymers. In general, $c_j(t)$ represents the concentration of polymers of length $j \geq 2$. The polymers can break randomly at any point along their length with a fragmentation rate k_-. Graphically we will represent these different chemical species as shown in Fig. 4.

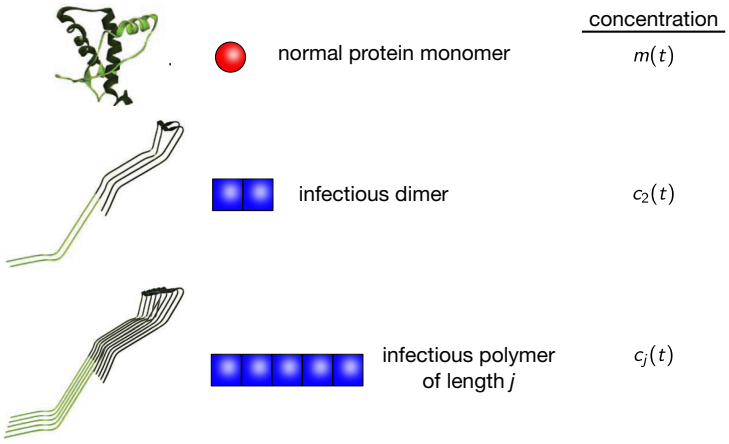

Figure 4: Graphical representation of the different chemical species in our problem.

To capture the dynamics of fibril growth and fragmentation, we will write down approximate chemical rate equations for the concentrations $m(t)$ and $c_j(t)$, $j \geq 2$, as outlined in Ref. [5]. Note that this corresponds to focusing on mean populations, ignoring fluctuations about the mean. In principle we could start with a more accurate master equation approach, but the rate equations are more analytically tractable, and sufficient to describe the important physical processes. Before writing the rate equations, let us define the mass $M(t)$ of all the infectious polymers in our system, per unit volume. Using units where a single monomer has mass 1, and hence a polymer of length j has mass j, the total infectious mass concentration is:

$$M(t) = \sum_{j=2}^{\infty} j c_j(t). \tag{1}$$

Assume the total mass overall (infectious and normal) per unit volume is M_{tot}, and this is a fixed constant in the cell. Then by conservation of mass we have

$$M_{\text{tot}} = m(t) + M(t). \tag{2}$$

Thus in principle we only need to solve for the dynamics of the polymer concentrations $c_j(t)$ for $j \geq 2$. Once we know these, we can calculate $M(t)$ from Eq. (1) and the monomer concentration is given by $m(t) = M_{\text{tot}} - M(t)$.

Let us start with the rate equation for $j = 2$:

$$\frac{dc_2(t)}{dt} = k_n m^2(t) + 2k_- \sum_{i=3}^{\infty} c_i(t) - 2k_+ m(t) c_2(t) - k_- c_2(t) \tag{3}$$

The right-hand side describes processes that either produce or destroy dimers (see Figs. 5 and 6). First term: dimers can be produced with rate k_n by collisions and conversion of monomer pairs, which can happen in approximately $m^2(t)$ different combinations (for large numbers of monomers). Second term: dimers can also be produced when any polymer of length $i \geq 3$ breaks with rate k_- into a piece of length 2 and a piece of length $i-2$, and such a break can happen in one of two places on the polymer. Third term: dimers can be destroyed (turning into polymers of length 3) when a monomer attaches and converts at either of the dimer ends. Fourth term: dimers can be destroyed (turning back into monomers) when their bond breaks with rate k_-.

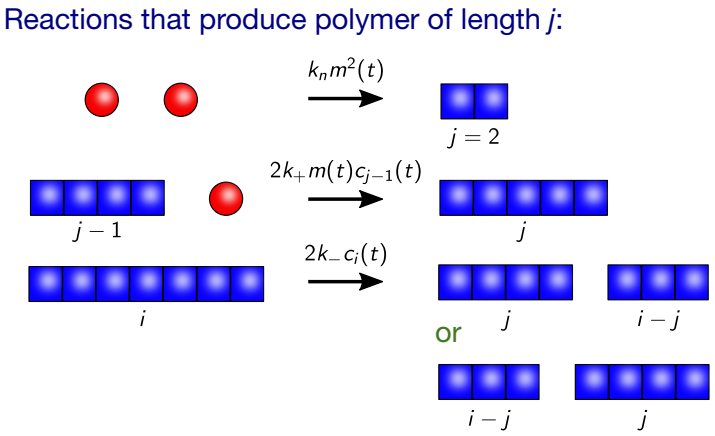

Figure 5: Production reactions.

The corresponding rate equation for $j > 2$ is:

$$j > 2: \qquad \frac{dc_j}{dt} = 2k_+ m(t) c_{j-1}(t) + 2k_- \sum_{i=j+1}^{\infty} c_i(t) - 2k_+ m(t) c_j(t) - (j-1) k_- c_j(t) \tag{4}$$

The terms are similar, except that the first term describes production of a polymer of length j by a binding/conversion of a monomer at either end of a polymer of length $j-1$. And the last term now is multiplied by $j-1$, since there are $j-1$ bonds that can break in a polymer of length j. Notice that there are no terms corresponding to two polymers of length $j-i$ and $i \geq 2$ merging into a polymer of length j. These are left out for mathematical simplicity (and because they do not contribute as much as monomer recruitment), but in principle could exist. Eqs. (3)-(4) represent a large, complicated system of coupled differential equations, because each equation for $c_j(t)$ depends on the concentrations of polymers for lengths smaller and larger than j. To make sense of this model, let us introduce $P(t)$, the total number of polymers per unit volume:

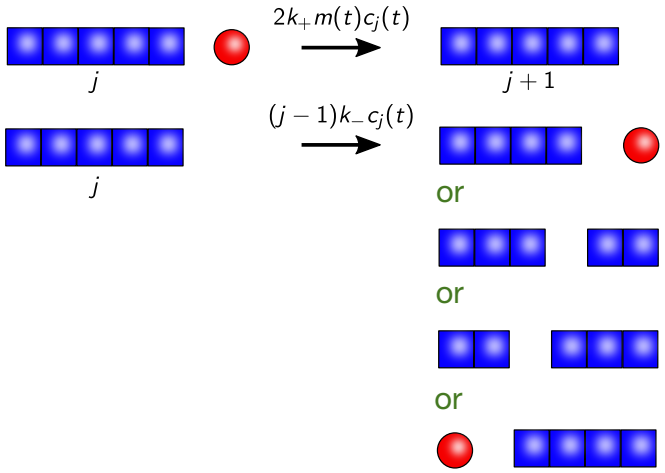

Figure 6: Destruction reactions.

$$P(t) = \sum_{j=2}^{\infty} c_j(t) \tag{5}$$

As it turns out, we will be able to convert Eqs. (3)-(4) into a much simpler system of differential equations just involving $P(t)$ and $M(t)$. These in turn will be the starting point for analyzing the dynamics of fibril assembly.

3 Questions

a) Show that Eqs. (3)-(4) imply the following two equations for $P(t)$ and $M(t)$:

$$\begin{aligned} \frac{dP}{dt} &= k_- M(t) - 3k_- P(t) + k_n m^2(t) \\ \frac{dM}{dt} &= 2k_+ m(t) P(t) - 2k_- P(t) + 2k_n m^2(t) \end{aligned} \tag{6}$$

Note that with the relation $m(t) = M_{\text{tot}} - M(t)$ we now have a closed set of coupled differential equations for $P(t)$ and $M(t)$. *Hint:* Write out the definitions of $P(t)$ and $M(t)$ term by term for the first few terms:

$$P(t) = c_2(t) + c_3(t) + c_4(t) + \cdots, \qquad M(t) = 2c_2(t) + 3c_3(t) + 4c_4(t) + \cdots$$

and then take the time derivative, plugging in Eqs. (3)-(4). You should notice a pattern in the first few terms of each sum which corresponds to Eq. (6).

b) To get an idea of how the system behaves, let us numerically solve the differential equations for $P(t)$ and $M(t)$ from part a). Use a simple iterative scheme: start with $P(0) = 0$, $M(0) = 0$ at time $t = 0$. For every t, get the set of values $P(t + \delta t)$ and $M(t + \delta t)$ at the next time step $t + \delta t$ using the equations

$$\begin{aligned} P(t+\delta t) &= P(t) + \delta t \left[k_- M(t) - 3k_- P(t) + k_n (M_{\text{tot}} - M(t))^2 \right] \\ M(t+\delta t) &= M(t) + \delta t \left[2k_+ (M_{\text{tot}} - M(t)) P(t) - 2k_- P(t) + 2k_n (M_{\text{tot}} - M(t))^2 \right]. \end{aligned} \tag{7}$$

Iterate this, saving the values of $P(t)$ and $M(t)$, until a certain time t_{max}. The parameter values you should use are: $\delta t = 0.1$ yr, $t_{\text{max}} = 100$ yr, $k_+ = 2 \times 10^7$ M^{-1} yr^{-1}, $k_- = 7 \times 10^{-5}$ yr^{-1}, $k_n = 10^{-2}$ M^{-1} yr^{-1}, $M_{\text{tot}} = 5 \times 10^{-6}$ M. With these parameters, time is measured in years (yr) and concentrations $M(t)$ and $P(t)$ in molars (M). Note that $P(0) = 0$ and $M(0) = 0$ corresponds to the case of no infectious polymers at $t = 0$. The only way for them to appear is through the spontaneous conversion/nucleation process that makes dimers out of monomers (akin to the spontaneous appearance of Creutzfeldt-Jakob disease).

Plot $M(t)$ and $P(t)$ versus t from $t = 0$ to t_{max}. Also plot the ratio $M(t)/P(t)$ versus t, which approximately represents the average mass per polymer (i.e. the typical length of one of the infectious polymers). Observe that $M(t)$ and $P(t)$ initially increase slowly, then start to grow much faster after a certain time. This initial incubation period is called the *lag phase*, and we will estimate its duration below. After the lag phase, $M(t)$ increases until it reaches M_{tot}, completely consuming all the available normal protein. On the other hand, $P(t)$ continues to increase even when $M(t)$ has saturated, which means that the infectious polymers are continually fragmenting into smaller pieces. The typical polymer length $M(t)/P(t)$ initially increases during the lag phase, but then gradually decreases at longer times.

c) To estimate the length of the lag phase, we first need to find an approximate analytical solution to Eq. (6) at small times t. Here we can make several approximations: we assume that $m(t) \approx M_{\text{tot}}$, since there is very little infectious polymer mass $M(t)$ early on. We also assume that fragmentation is slow: $k_- \ll k_+ M_{\text{tot}}$. Because of this polymers tend to become long in the lag phase, so we can assume $M(t) \gg P(t)$. All of these approximations are justified for our parameter values by the numerical results of part b). Plugging these assumptions into Eq. (6), we get:

$$\begin{aligned} \frac{dP}{dt} &\approx k_- M(t) + k_n M_{\text{tot}}^2 \\ \frac{dM}{dt} &\approx 2k_+ M_{\text{tot}} P(t) + 2k_n M_{\text{tot}}^2 \end{aligned} \tag{8}$$

Solve these equations for $P(t)$ and $M(t)$, with the initial conditions $P(0) = 0$, $M(0) = 0$. *Hint:* Take the time derivative of the left and right-hand sides of the $P(t)$ equation, and then plug in the $M(t)$ equation. This will give you a differential equation just in terms of $P(t)$. You can solve it by guessing a form,

$$P(t) = A_1 e^{\kappa t} + A_2 e^{-\kappa t} + A_3$$

and figuring out the constants κ and A_3 by comparing the two sides of the equation (looking at the $e^{\kappa t}$, $e^{-\kappa t}$, and time-independent terms). Once you know $P(t)$, you can plug into the first line of Eq. (8) to get $M(t)$. The initial conditions allow you to solve for the remaining constants A_1 and A_2.

d) Plot the short-time solution for $M(t)$ from part c) versus the numerical results (for the same parameter values). Note that the approximate solution always increases exponentially at larger times, while the actual solution initially increases exponentially, but then slows down and saturates at M_{tot}. Because the approximation and the actual solution diverge over a relatively short period, we can make a rough estimate of the lag phase duration by defining t_{lag} as the point of intersection of the approximate $M(t)$ curve and M_{tot}. Use the expression for $M(t)$ from part c), ignoring the $\exp(-\kappa t)$ term since this vanishes for large times, and solve $M(t_{\text{lag}}) = M_{\text{tot}}$ to find t_{lag}. Check that the analytical result for t_{lag} gives a reasonable estimate of the lag phase when you plug in the parameters, compared to the numerical result. We thus have arrived at a compact formula for how it long it takes (roughly) for any spontaneous fibril growth process to consume its monomer supply (and kill you along the way). You should find that the formula has the form:

$$t_{\text{lag}} = \frac{1}{\kappa} \log F(k_+, k_-, k_n, M_{\text{tot}})$$

where F is some function.

Make a log-log plot of t_{lag} versus the fragmentation rate, k_-, for $k_- = 10^{-5}$ yr^{-1} to 10^{-1} yr^{-1} (keeping all other parameters as above). Note how t_{lag} *decreases* as k_- increases. Faster fragmentation substantially hastens the consumption of healthy proteins, because with more polymers there are more "sticky" ends that can recruit monomers. This is a result that has been verified in lab measurements of protein aggregation: if you agitate or stir the mixture, breaking up the fibrils, you get faster aggregation. There is a potential (though yet unconfirmed) link to the role of frequent brain trauma in the onset of chronic traumatic encephalopathy (CTE), the disease that

has received major attention recently in former National Football League players. CTE is a neurodegenerative disease associated with the aggregation of certain proteins like tau and amyloid beta. Do the mechanical forces of repetitive collisions in sports increase the fragmentation rate of protein fibrils? We do not know for sure, but this is certainly a possible mechanism.

e) Finally, let us consider the case where the cell gets invaded by some infectious polymers at $t = 0$ (for example, you wittingly or unwittingly ate brain tissue of an infected organism). Repeat the numerical calculation of part b), using the same parameters except that $M_{\text{tot}} = 5.1 \times 10^{-6}$ M, and the initial conditions are $M(0) = 10^{-7}$ M, $P(0) = 10^{-8}$ M. This corresponds to a small initial dose of infectious polymers of average length 10. Plot $M(t)$ versus t in this case, and note how quickly $M(t)$ reaches M_{tot}, compared to the slow growth of part b). Given a large enough initial seed for aggregation, the lag phase disappears. (Do not eat brains!)