PHYS 320 Problem Set 3
Chaperones to the rescue

1 Background

In the previous problem set, we encountered one of the great hazards of being a protein inside the cell: the small, but nontrivial, possibility of ending up in a defective state that is prone to aggregating. These misfolded states typically have more exposed hydrophobic residues on their surface than the correctly folded, functional protein. Prions represent the most pathological examples of this phenomenon, nucleating an irreversible and deadly growth of fibrillar aggregates. But misfolding of proteins is a far more widespread phenomenon, an inevitable byproduct of the fact that long chains of amino acids can find multiple ways of collapsing into three-dimensional structures. The correct state (known as the native state) will typically have the lowest enthalpy, and thus represent the most stable end-point for the folding process. Misfolded states will lack some of the interior hydrogen bonds and hydrophobic contacts of the correct state, and thus their enthalpies will be higher than the native one, but still low enough that the system may get stuck in these states for appreciable lengths of time. Since there are typically many more ways to misfold than to fold correctly, the total population of misfolded proteins may vastly outnumber the correct ones (in some cases by an order of magnitude if left unchecked). The end result would be a disaster: large concentrations of proteins with exposed hydrophobic surfaces, forming insoluble aggregates that disrupt cell function.

Happily, nature has evolved remarkable defense mechanisms to solve this problem (up to limit: prions are somehow able to evade these defenses, by means not entirely clear.) The answer is a class of proteins known as chaperones, which as their name implies, essentially keep certain other proteins, those particularly prone to misfolding, from causing trouble. The most widely studied family of chaperones are the heat shock proteins (Hsp) [1]. These were initially discovered in the context of figuring out how cells respond to the stress of elevated temperatures. As we will see in the first part of the problem set, increasing temperature increases the probability of a protein unfolding from its native state, exposing hydrophobic residues. Since the melting temperatures of many proteins may be precariously near ($10 - 20^\circ$ C above) their normal operating temperature, even a few degrees of environmental heating can have a substantial effect. Cells express more Hsp proteins in response to higher temperatures, to help bring back proteins to their native states. But in fact Hsp proteins are general-purpose fixers, responding to a variety of stresses, such as radiation damage and toxic chemicals. Since some degree of misfolding is always occurring, even in the absence of stress, Hsp proteins are always active as essential factors in keeping cells alive, though their populations are lower under non-stressful conditions. Given the extreme environments in which early life may have evolved (i.e. deep-sea hydrothermal vents), it is unsurprising the Hsp chaperones are among the most ancient proteins, highly conserved throughout evolutionary history.

As we will see below, the general theory of Hsp chaperone action can actually be understood using the ideas of nonequilibrium thermodynamics [2]: the presence of the chaperone creates a biochemical cycle, with a nonzero current driven by ATP hydrolysis. The function of the cycle is simple: take misfolded proteins, allow them to bind to chaperones, catalyze their unfolding, and

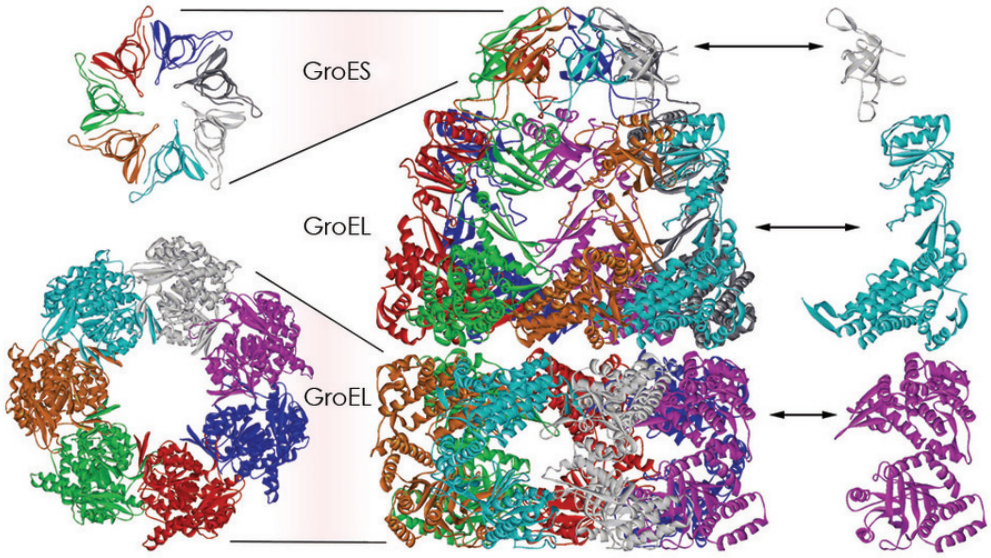

Figure 1: The structure of the GroEL/GroES chaperone system, a large complex consisting of many individual proteins. Two rings of seven GroEL proteins are stacked one on top of another, capped by a ring of seven GroES proteins. The two GroEL rings are in different conformational states (blue versus purple units shown on the right). The ability of GroEL to switch conformations during the biochemical cycle (fueled by ATP hydrolysis) is a critical part of its role in unfolding misfolded proteins. The GroES cap can bind either to the top or bottom GroEL ring, and alternates its place during the full cycle. Image credit: https://www.caymanchem.com/app/template/Article.vm/article/2130/figure/2.

give them a second chance at folding into the correct shape. Because the worst offenders will often misfold again on their second chance (and third, fourth, etc.) a single protein may undergo dozens of loops through the cycle before it succeeds at finding the native state. The end result is a stationary population of correctly folded proteins that far exceeds the equilibrium value (in the absence of chaperones/fuel). This is a brute-force and energetically expensive "if at first you don't succeed" approach, but the mechanism has to be simple and generic, because a given chaperone is responsible for helping many different kinds of misfolded proteins. The strategy is known as *iterative annealing* and was first first explained (experimentally and theoretically) by Todd, Lorimer, and Thirumalai in the context of one particular Hsp system, GroEL/GroES, employed by many bacteria [3]. You will reconstruct the basic features of their mathematical argument in working through the problem set.

Like other chaperone systems, the setup for GroEL/GroES is elaborate, involving more than one type of Hsp protein, and many copies: Fig. 1 shows how fourteen copies of the GroEL protein (in two stacked, hollow rings of seven units) and seven copies of the smaller GroES protein assemble to form what is essentially a nanoscale cocktail shaker. The rough outlines of the cycle (Fig. 2) are as follows: we will focus only on one ring of GroEL, i.e. the top one, since both rings undergo the same cycle. Initially (state T) the ring does not have an attached GroES cap, and the inner cavity of the ring has an exposed hydrophobic surface that serves as a fly-trap for misfolded proteins. The capture of the misfolded protein triggers a complex sequence of events, involving

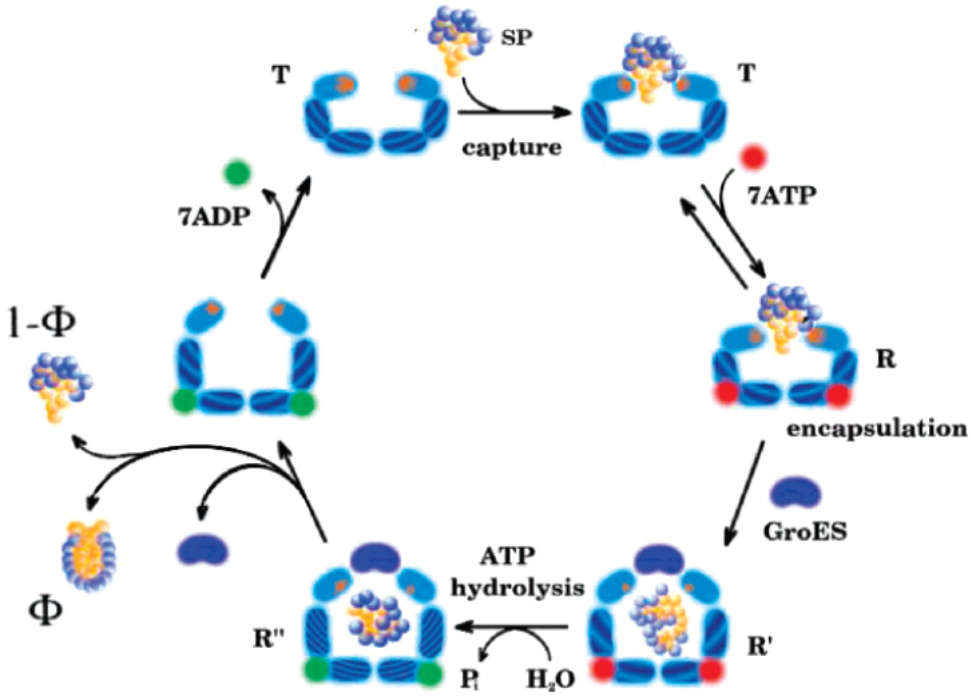

Figure 2: The biochemical cycle of the GroEL/GroES system, focusing only on one GroEL ring. SP refers to the "substrate protein" captured by the system, which is generally a protein in a misfolded conformation with exposed hydrophobic residues. The end result of the cycle, which requires hydrolysis of seven ATP molecules, is that the misfolded protein is unfolded, after which it can transition to the correct folded state with probability Φ, or return to a misfolded conformation with probability $1 - \Phi$. For some proteins prone to misfolding (small Φ), it may require many repetitions of this cycle before the correct folded state is reached, a process known as *iterative annealing*.

binding of seven ATP molecules, inducing conformational changes of the GroEL that expand the volume of the inner cavity, pulling the protein into the cavity, and sealing the entrance by the GroES cap. During this process the protein structure is disrupted, and the unfolded protein is released from the cavity walls because they undergo a transition from hydrophobic to hydrophilic. After ATP hydrolysis the GroES cap comes off and the unfolded protein can drift away to try its chances at folding again. Since the protein spends a long time trapped inside (≈ 10 s) it might even fold again (or misfold) while still encapsulated, depending on how fast or slow it folds (folding times can vary from less than a millisecond to minutes, depending on the protein). Success or failure is irrelevant to the chaperone: regardless of the protein's state when the GroES cap comes off, the crucial function of the GroEL/GroES system is to unfold the initial misfolded state. If the protein returns to being misfolded, another round of chaperone binding and remodeling awaits.

How exactly does GroEL unfold the misfolded protein after binding? Unfortunately, the mechanical details of the unfolding are still open questions for most chaperones, including GroEL. What happens inside the cocktail shaker is difficult to probe. But a recent study by Kellner *et al.* using fluorescent markers [5] has shed some light on this question in a completely different Hsp system, DnaK / DnaJ / GrpE. The key player in this cycle, DnaK, attacks the misfolded protein en

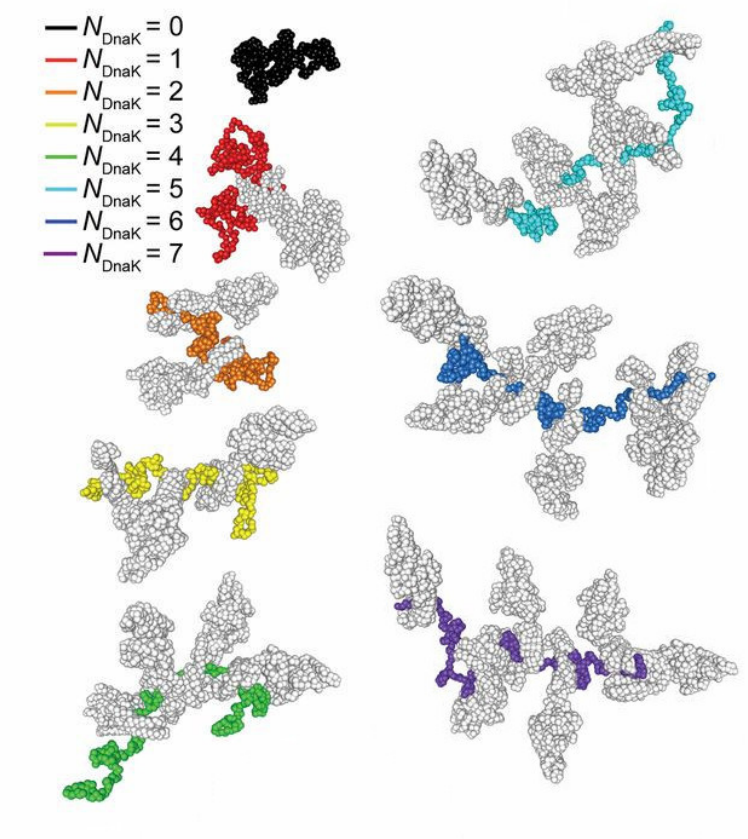

Figure 3: The chaperone protein DnaK can bind in multiple copies to a misfolded protein: a sequence of simulation snapshots (taken from Ref. [5]) showing the misfolded protein (in color) and an increasing number of bound DnaK proteins (gray). The DnaK molecules act like a series of vises, clamping down on the protein and forcing it into a progressively longer, more unfolded structure. The full biochemical cycle involves eventual dissociation of DnaK from the protein, giving it a second chance to fold. Like other chaperone systems, this cycle involves multiple proteins (DnaJ, DnaK, GrpE) and the hydrolysis of ATP to create a nonequilibrium current.

masse, with DnaK copies binding at multiple locations (Fig. 3). Each DnaK is like a vise, clamped on the protein, and the only way to accommodate all the DnaK molecules is to unwind the free parts of the protein in-between. This progressive expansion unfolds the protein bit by bit with each additional binding. This is of course only one part of an overall cycle involving DnaJ and GrpE proteins, which ultimately leads to release of the protein from the DnaK. However Kellner *et al.* marks the first time we could see the key unfolding step of a chaperone system in action.

Though this problem set is focused on chaperones, the biological error-correction principles it illustrates apply more widely: for example the polymerases that copy DNA can make random mistakes, adding the wrong nucleotide. Proofreading and correction takes the form of an ATP-driven nonequilibrium biochemical cycle, giving the polymerase additional chances to fix the copying error before moving on to the next base. The underlying mathematics are quite similar to the chaperone case.

References

[1] Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and ecological physiology. *Annu. Rev. Physiol.* **61**, 243–282 (1999).

[2] Barducci, A. and De Los Rios, P. Non-equilibrium conformational dynamics in the function of molecular chaperones. *Curr. Opin. Struct. Biol.* **30** 161–169 (2015).

[3] Todd, M. J., Lorimer, G. H. and Thirumalai, D. Chaperonin-facilitated protein folding: Optimization of rate and yield by an iterative annealing mechanism. *Proc. Natl. Acad. Sci. USA* **93**, 4030–4035 (1996).

[4] Thirumalai, D. and Hyeon, C. RNA and Protein Folding: Common Themes and Variations. *Biochemistry* **44**, 4957–4970 (2005).

[5] Kellner *et al.* Single-molecule spectroscopy reveals chaperone-mediated expansion of substrate protein. *Proc. Natl. Acad. Sci. USA* **111**, 13355–13360 (2014).

2 Questions

2.1 Simple kinetic model for protein folding

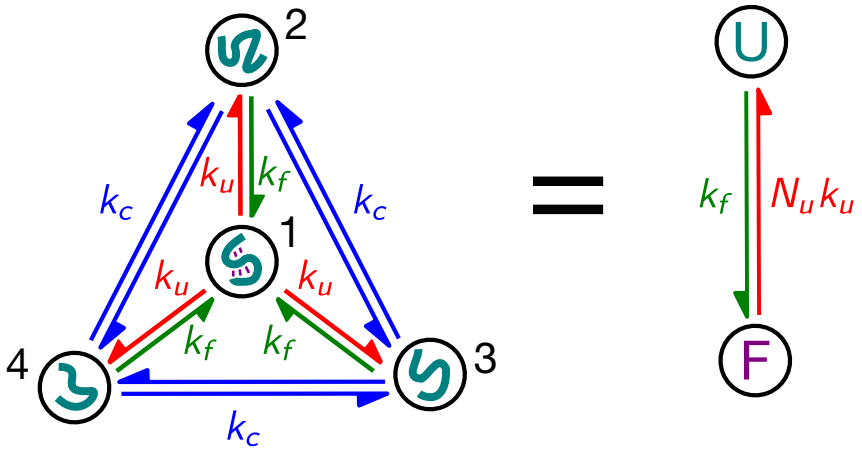

Figure 4: Left: kinetic network for $N = 4$, $N_u = 3$. Right: equivalent two state mapping, gathering all the unfolded states $n \geq 2$ into a single state U, and the folded state 1 into F.

We start our exploration of protein folding by introducing one of the simplest, exactly solvable models for the folding process. Assume the protein has a unique folded structure, which we will label as state $n = 1$, and a large number $N_u = N - 1$ of unfolded states, $n = 2, 3, \ldots, N$. There are many possible configurations of the unfolded chain, but only one correctly folded three-dimensional structure. For now we will ignore misfolded states, which we will consider later in the problem set. The unfolded states all have the same enthalpy $H_n = 0$, $n = 2, \ldots, N$, while the folded state has a lower enthalpy $H_1 = -\epsilon$, where $\epsilon > 0$, reflecting the intramolecular bonds

stabilizing the folded configuration. The transition matrix Ω for the system has the following form:

$$\Omega_{nm} = \begin{cases} k_c & n \geq 2, m \geq 2, n \neq m \\ k_f & n = 1, m \geq 2 \\ k_u & n \geq 2, m = 1 \\ -\sum_{n' \neq n} \Omega_{n'n} & m = n \end{cases} \tag{1}$$

Here k_c is the rate at which unfolded states interconvert between each other (note that the backward and forward rates between any two unfolded states are the same because all unfolded states have the same enthalpy). k_f is the rate at which a particular unfolded state can transition to the folded state 1, and k_u is rate of the reverse (unfolding) transition. Since any unfolded state has an enthalpy larger than the folded state by an amount ϵ, the ratio of k_u to k_f is given by $k_u/k_f = \exp(-\beta\epsilon)$ by detailed balance, where $\beta = 1/k_B T$. Fig. 4 (left) shows the kinetic network for $N = 4$.

a) To solve the dynamics of this system, note that the unfolded states are all equivalent to each other, and equally likely to visit (or be visited from) state 1. Let us start with an initial condition at $t = 0$ corresponding to an unfolded ensemble where each unfolded state has an equal probability, $p_n(0) = N_u^{-1}$, $n = 2, \ldots, N$, and the folded state has zero probability, $p_1(0) = 0$. Then by the symmetry of the network we know the probabilities $p_n(t)$ for $n \geq 2$ must always stay equal to each other, $p_n(t) = N_u^{-1} p_U(t)$, where

$$p_U(t) \equiv \sum_{n=2}^{N} p_n(t) = 1 - p_F(t)$$

is the total probability of being in any unfolded state, and $p_F(t) \equiv p_1(t)$ is the probability of being folded. Show that $p_U(t)$ satisfies the differential equation

$$\frac{dp_U(t)}{dt} = N_u k_u p_F(t) - k_f p_U(t) \tag{2}$$

Hint: Look at the Ω matrix given by Eq. (1) for a small value of N_u, for example $N_u = 3$, and write down the explicit master equations for dp_2/dt, dp_3/dt, ..., dp_N/dt. Sum the equations for dp_n/dt for $n = 2$ through N, and notice the pattern of term cancellations. You should be able to rewrite the sum as Eq. (2).

b) The procedure in part a) effectively maps the complicated, many-state system into a simpler two-state form (Fig. 4 right), with all the unfolded states gathered into a single state U with probability $p_U(t)$, and state 1 renamed F with probability $p_F(t) = 1 - p_U(t)$. Eq. (2) shows that the effective total rate from F to U is $N_u k_u$ (since there are N_u alternative ways to unfold), while the total rate from U to F is k_f (since there is only one folded state). Solve Eq. (2) for $p_U(t)$ with the initial condition $p_U(0) = 1$, $p_F(0) = 0$. Make sure your solution has mathematically sensible limiting cases: if $k_f = 0$ (the protein cannot fold) then you should find $p_U(t) = 1$ for all t. On the other extreme, if $k_u = 0$ (the protein cannot unfold), then $p_U(t)$ should decay to zero as $t \to \infty$, since all the unfolded states will eventually fold and never unfold again. Plot $p_U(t)$ versus t for the parameters $\epsilon = 315$ kJ/mol, $N_U = 3^{100}$, $k_f = 10$ s^{-1}, at $T = 310$ K, typical for a small protein

at body temperature. The value of Boltzmann's constant is 8.314×10^{-3} kJ/(mol·K). The large size of N_U reflects the astronomical number of possible unfolded states: for a chain of 100 amino acid residues, with each residue capable of adopting three distinct configurations, the total number of configurations would be $N_U = 3^{100}$.

c) Another way of checking that the solution from part b) makes sense is to look at the long time limit $t \to \infty$. Write down expressions for $p_1(\infty) = p_F(\infty)$ and $p_n(\infty) = N_u^{-1} p_U(\infty)$, $n \geq 2$, using the results of part b). Verify that they are equivalent to what you would expect in equilibrium from the Boltzmann distribution, $p_n^{\text{eq}} = \exp(-\beta H_n)/Z$. *Hint:* Remember the detailed balance relations!

d) Plot the equilibrium total unfolded probability p_U^{eq} as a function of temperature T, with the other parameters as listed in part b). You should notice that $p_U^{\text{eq}} \approx 0$ for small T, and becomes ≈ 1 at large T. Increasing temperature will eventually unfold (denature) any protein. Find an analytical expression for T_m, the protein melting temperature, defined as the transition temperature when $p_U^{\text{eq}} = 1/2$. Make sure the numerical value of T_m when you plug in the parameters is physically reasonable: it should be something achievable with a backyard grill. At body temperature, $T = 310$ K, what fraction of this protein is unfolded in equilibrium?

2.2 Misfolded states

The picture above is in many cases too rosy: in addition to a folded state, proteins may be able to adopt many alternative partially folded or misfolded configurations. These are not quite as stable as the true folded state (and so individually are less likely to occur in equilibrium), but if there are enough misfolded alternatives, their total equilibrium population can significantly outnumber the correctly folded ones, in some cases by an order of magnitude. Let us modify the above model to incorporate N_m misfolded states, each with enthalpy $-\nu$, where $0 < \nu < \epsilon$. For simplicity, we will gather all the misfolded states into a total misfolded probability $p_M(t)$. The corresponding network diagram is shown in Fig. 5. The transition rate from any unfolded state to a misfolded state is k_m, which satisfies the detailed balance condition $k_u/k_m = \exp(-\beta\nu)$. Since there are N_m misfolded states, the total rate of transitions from U to M in the simplified network picture (Fig. 5) is $N_m k_m$.

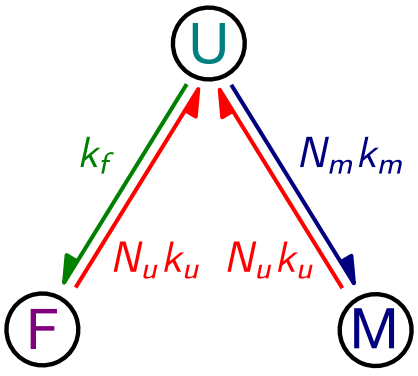

Figure 5: Protein kinetic network, including the possibility of misfolded conformations.

e) What is the ratio $p_F^{\text{eq}}/p_M^{\text{eq}}$ of folded to misfolded proteins in equilibrium? Evaluate this ratio for $\nu = 305$ kJ/mol and $N_m = 1000$, with the other parameters as in part b). As you can see, in this case the misfolded states greatly outnumber the correctly folded ones.

2.3 Chaperones

To rescue the protein, we introduce a chaperone C that can bind the misfolded states, making a complex MC, which can then catalyze unfolding of the misfolded conformation. We thus have a new state MC in our kinetic network, shown in Fig. 6, with corresponding probability $p_{MC}(t)$. The rates g_1, r_1, g_2, r_2 will depend on the details of the chaperone biochemistry, and generally will include some coupling to an external fuel source (ATP hydrolysis) that induces a nonequilibrium steady state, driving a counterclockwise current $J > 0$ around the loop formed by U, M, and MC. Remarkably, without knowing any specific details of the chaperone interactions, we can prove that the chaperone will always have a beneficial effect (increasing the ratio of folded to misfolded states relative to the equilibrium case) so long as $J > 0$.

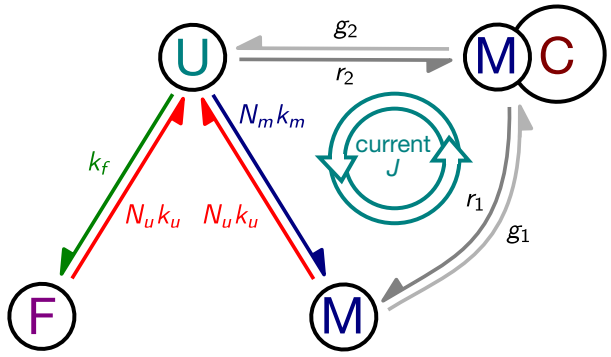

Figure 6: Protein kinetic network, with misfolded states and a chaperone.

f) Assume that in the nonequilibrium steady state the probabilities of the four states are p_F^s, p_U^s, p_M^s, and p_{MC}^s. (The precise values are unimportant, but we know that all four probabilities are greater than 0 and less than 1.) Given a certain current J in the U, M, MC loop, prove that the folded-misfolded ratio p_F^s/p_M^s can be written as:

$$\frac{p_F^s}{p_M^s} = \frac{e^{\beta(\epsilon-\nu)}}{N_m} + \alpha\frac{J}{p_M^s}$$

where α is an expression that is always positive. Find an analytical form for α. Hence $J > 0$ always increases the folded-to-misfolded ratio. For $J = 0$, how does this result compare to the equilibrium ratio you found in part e)? To see the degree to which the chaperone can help, let us assume $J > 0$ and $g_1 \gg r_1$ for simplicity, so the current from M to MC is $J \approx g_1 p_M^s$. For $g_1 = 1.0$ s^{-1}, find the numerical value of p_F^s/p_M^s (all other parameters the same as above). As you will see, this is definitely an improvement! *Hint:* Write down expressions for the current between each pair of connected states in the system. Is there a nonzero current between U and F?

g) The improvement in the ratio of folded-to-misfolded states comes at a cost: for an unfolded protein to reach the folded state, it may have to undergo many cycles of misfolding, binding to the chaperone, and unfolding again, before it finally gets lucky and makes the transition from U to F. This can make the average time to correctly fold quite long (each complete loop for the GroEL/GroES system for example, takes roughly 15 seconds). Calculate the mean first passage time τ_U to reach the F state for a protein starting in state U. You can assume the chaperone cycle is optimized so that $g_1 \gg r_1$ and $g_2 \gg r_2$, and hence we can set $r_1 = r_2 = 0$ for simplicity. For $g_1 = 1.0$ s^{-1} and $g_2 = 0.07$ s^{-1} (which gives approximately a total time of 15 s around the loop), find the numerical value of τ_U. How many loops on average must the protein complete before it can successfully fold? Given that a complete loop of GroEL/GroES consumes seven ATP molecules, that is a quite a lot of chemical energy expenditure to produce a single folded protein! *Hint:* Let τ_n be the mean first passage time from a state n to F. By definition $\tau_F = 0$, and we are interested in solving for τ_U. Recall that mean first passage times τ_n satisfy a set of equations, $\sum_{n'} \tau_{n'} \Omega_{n'n} = -1$ for any $n \neq F$. All you need to do is to figure out the transition matrix Ω corresponding to Fig. 6. This gives a total of three separate equations, which is sufficient to figure out τ_U, τ_M, and τ_{MC}, though only τ_U is of interest.