Lecture Notes for Biomaths

Thomas Ouldridge

October 28, 2016

Contents

1 Equilibrium transitions

In this section we will consider biomolecular transitions. To do so, we will need to group states into macrostates of similar type; for example, we will consider a "bound state" of two biomolecules which will contain all configurations in which the molecules are in close proximity and interacting strongly. The "unbound state" will include states in which the molecules are far apart and not interacting. Clearly, actually defining these macrostates for a given problem might be subtle, but we shall simply assume that it can be done. In fact, for the purposes of thermodynamic calculations, almost any splitting of configurations into macrostates will be valid in a formal sense, but the analysis will only provide helpful insight if the splitting is well chosen. When the kinetics of transitions between states are considered, the problem is more subtle and even more care must be taken. This is because it is typical to assume that transitions between macrostates are poisson processes, and this is an approximation that only holds when macrostates are particularly well-defined metastable conformations.

Let us say that we have labelled our macrostates by the variable X; $X = 1$ defines a certain set of configurations, $X = 2$ another set and so on. We can define the partial partition function

$$Z(X) = \sum_{\text{configs in } X} \exp(-E_{\text{config}}/kT); \quad Z = \sum_{X} Z(X). \tag{1}$$

The key point here is that $Z(X)/Z$ is the equilibrium probability of finding the system in macrostate X. This follows directly from the Boltzmann distribution. Although we do not discuss the correspondence in this course, the Helmholtz free energy of an equilibrium system can be found via $F = -kT \ln Z$. It is therefore common to talk about the free energy of a system in macrostate X, $F(X) = -kT \ln Z(X)$. This alternative representation can be convenient, you can also do everything directly with $Z(X)$. We will proceed to consider how the relative values of partition functions determine the behaviour of biomolecular systems.
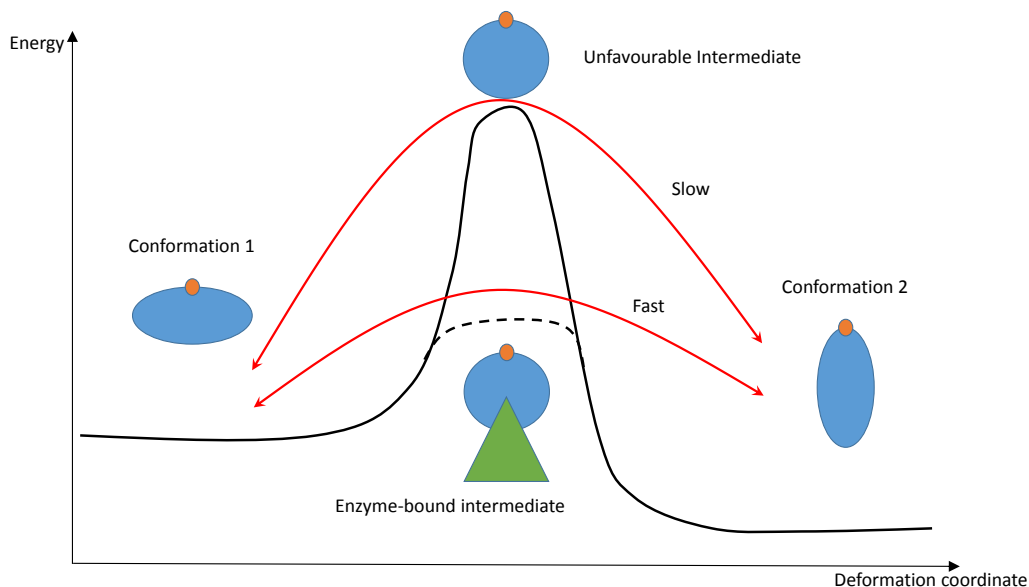
Figure 1 Simple energy landscape picture of the interconversion of a protein between two conformational macrostates. An enzyme can bind to the protein, reducing the barrier between the two conformations and enhancing the transition rates in both directions.

1.1 Two-state systems

The simplest possible biomolecular system is a two-state system, $X = 1, 2$. These macrostates could be two protein conformations, for example. Fairly trivially, the equilibrium probabilities of the system being in each macrostate are given by

$$\frac{P(1)}{P(2)} = \frac{Z(1)}{Z(2)} = \frac{\sum_{X=1} \exp(-E_{\text{config}}/kT)}{\sum_{X=2} \exp(-E_{\text{config}}/kT)} = \exp(-(F(1) - F(2))/kT); \quad P(i) = \exp(-F(i)/kT). \tag{2}$$

In other words, macrostates that have lower typical energies and a higher number of accesible configurations (higher entropy) are favoured. What about transition rates? If we assume transitions between the two states are simple Poission processes with rates k_{12} and k_{21}, then at equilibrium the flux (average flow) in either direction cancels and

$$P(1)k_{12} = P(2)k_{21}. \tag{3}$$

Therefore, from Eq. 2

$$\frac{k_{21}}{k_{12}} = \frac{Z(1)}{Z(2)} = \exp(-(F(1) - F(2))/kT). \tag{4}$$

The ratio of transition rates between two states is thus determined by the relative thermodynamic stability of two states. Importantly however, the thermodynamics says nothing about the absolute magnitude of rates.

1.1.1 Enzymes

The fact that the relative thermodynamic stability of two states does not determine the magnitude of transition rates between them is central to the principle underlying enzyme operation. Let us imagine a biomolecule with two macrostates, separated by unfavourable intermediates that make transitions slow. This situation is illustrated schematically in Fig. 1. Now let us imagine that an enzyme can bind transiently to the biomolecule. The enzyme does not act to make either the initial or final conformation more favourable in a thermodynamic sense – when the enzyme is not bound to the biomolecule, $Z(1)$ and $Z(2)$ are unchanged. However, if being bound to the enzyme makes the intermediate configurations less unfavourable, the existence of the enzyme allows for easier passage between the two macrostates (see Fig. 1). So the enzyme can effectively increase k_{12} and k_{12} whilst maintaining the value of k_{21}/k_{12} - this is *catalysis*. If you look closely enough, all enzymes are essentially carrying out this process. Often, they are catalysing the breakdown of chemical fuel, and using the process to do useful work; if the fuel is not explicitly discussed, the catalytic behaviour can be hard to spot.
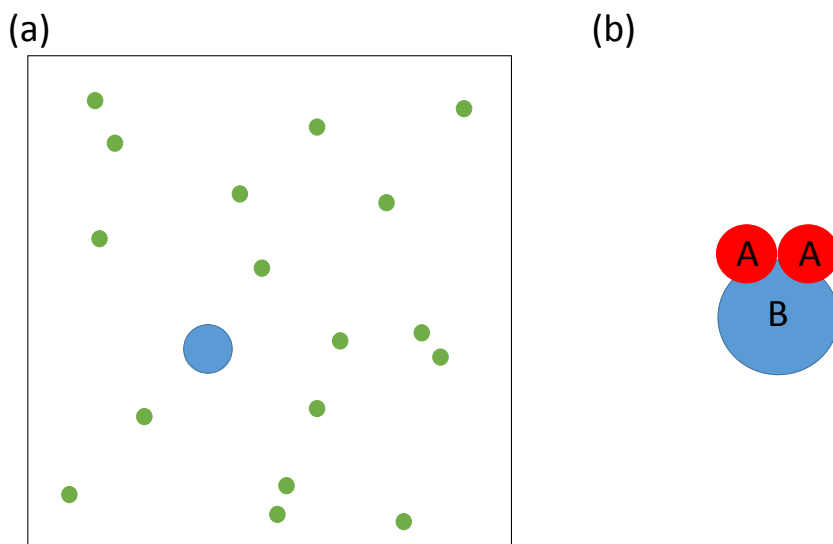
(a)   (b)

Figure 2 (a) A single receptor of type B surrounded by multiple ligands (molecules of type A) in a volume V. (b) A tri-molecular assembly A_2B with the potential for cooperativity.

1.2   Binding transitions

In this section we will assume that it is possible to divide the configurations of two molecules into two sets: one in which the molecules are strongly interacting and in close proximity (bound states), and one in which they are not in close proximity and interact weakly, if at all (unbound states).

We will proceed by thinking about partition functions. Let the partition function of an isolated biomolecule A in a large volume V be q_A. The single-particle partition function q_A will be proportional to V, as the number of states a molecule can visit will scale with the size of the box in which it is contained. Now consider N_A particles of type A; assume that they are dilute and don't interact strongly, basically forming an ideal gas. Then the partition function of N_A particles of type A is

$$Z_A(N_A) = \frac{q_A^{N_A}}{N_A!}. \tag{5}$$

To understand Eq. 5, note that for non-interacting molecules, the energy of the system is simply a sum of individual molecular energies. This suggests that the partition function can be factorized, and the total partition function $Z_A(N_A)$ is just the product of individual partition functions q_A. This is essentially true, but Eq. 5 also contains a factor of $N_A!$. This term accounts for the fact that identical molecules are *physically indistinguishable*. We can label molecules of type A differently for the purposes of calculation, but in reality there is *absolutely no difference* between molecules A_1 and A_2 being at positions r_1 and r_2, respectively, and A_1 and A_2 being at positions r_2 and r_1, respectively. In writing down $Z_A(N_A) = q_A^{N_A}$, we are mistakenly counting separately states that are actually identical; dividing by $N_A!$ corrects for this.

If we have multiple non-interacting molecules, A,B,C..., then the total partition function is given by

$$Z(N_A, N_B, N_C...) = Z_A(N_A)Z_B(N_B)Z_C(N_C).... \tag{6}$$

This whole construction becomes useful when we allow for interconversions – for example, $A + B \rightleftharpoons C$, where $C = AB$ is then the bound configuration of AB. Eq. 6 still holds for a given N_A, N_B, N_{AB}, but now the total partition function is

$$Z = \sum_{\{N_A, N_B, N_{AB}\}} Z(N_A, N_B, N_{AB}). \tag{7}$$

in which the sum runs over the set of possible N_A, N_B, N_{AB} given the total number of molecules in the system. Most importantly, $Z(N_A, N_B, N_{AB})/Z$ is the probability of observing a macrostate with molecule numbers N_A, N_B and N_{AB}. We will be using this fact

in what follows. Note that we are still assuming ideal behaviour of the separate molecules – the molecule's don't interact *except to form bound states*.

1.2.1 Single receptor

Let us imagine we have a single molecule of type B (a receptor) in a sea of molecules of type A (Fig. 2 (a)). What is the probability that it is bound in thermodynamic equilibrium? We have two macrostates of the system, and so the probability that the receptor is bound ($N_{AB} = 1$) is

$$P_{\text{bound}}^{B} = \frac{Z_A(N_A^{\text{tot}} - 1)Z_{AB}(1)}{Z_A(N_A^{\text{tot}})Z_B(1) + Z_A(N_A^{\text{tot}} - 1)Z_{AB}(1)} = \frac{N_A^{\text{tot}} q_{AB}}{q_A q_B + N_A^{\text{tot}} q_{AB}} = \frac{[A]K_{AB}}{1 + [A]K_{AB}}. \tag{8}$$

In the final equality, $[A]$ is the concentration of particles of type A, and K_{AB} is a concentration-independent equilibrium constant that characterises the reaction. Here, we have made use of the fact that q_A, q_B and q_{AB} all scale with V so that

$$K_{AB} = \frac{V q_{AB}}{q_A q_B} \tag{9}$$

is a constant that depends on the molecular details, but not the arbitrary box of size V in which we consider the system. K_{AB} is typically determined by the balance between attractive interaction energies (which increase q_{AB} relative to $q_A q_B$) and entropic costs of binding (which increase $q_A q_B$ relative to q_{AB}).

1.2.2 General bimolecular association

Now consider an identical system, but allow for multiple particles of type B. In this case,

$$P(N_{AB}) = \frac{q_{AB}^{N_{AB}} q_A^{(N_A^{\text{tot}} - N_{AB})} q_B^{(N_B^{\text{tot}} - N_{AB})} / N_{AB}!(N_A^{\text{tot}} - N_{AB})!(N_B^{\text{tot}} - N_{AB})!}{Z} \tag{10}$$

is the probability of finding N_{AB} bound pairs. Let us assume that N_A, N_B and N_{AB} are all large; $P(N_{AB})$ is then very sharply peaked about its maximum, and we can find the expected value of N_{AB} by maximising $P(N_{AB})$ (this is known as the saddle point approximation). For convenience, we maximise $\ln P(N_{AB})$

$$\frac{d \ln P(N_{AB})}{dN_{AB}} = \ln \frac{q_{AB}}{q_A q_B} - \frac{d}{dN_{AB}} \left(\ln N_{AB}! + \ln(N_A^{\text{tot}} - N_{AB})! + \ln(N_B^{\text{tot}} - N_{AB})! \right) = 0. \tag{11}$$

To progress further, we use Stirling's approximation $\ln x! \approx x \ln x - x$, which is accurate for large x. The result is

$$\ln \left(\frac{q_{AB}(N_A^{\text{tot}} - N_{AB})(N_B^{\text{tot}} - N_{AB})}{N_{AB} q_A q_B} \right) = 0. \tag{12}$$

Rearranging,

$$\frac{q_{AB}}{q_A q_B} = \frac{N_{AB}}{N_A N_B}, \tag{13}$$

which, using Eq. 9, implies

$$\frac{[AB]}{[A][B]} = K_{AB}. \tag{14}$$

Results of this form (including those for more complicated reactions) are know as "the law of mass action". One can find the same relationship between concentrations if a kinetic model of the following form is assumed:

- $A + B \rightarrow AB$ at a rate $k_+[A][B]$.

- $AB \rightarrow A + B$ at a rate $k_-[AB]$.

You should satisfy yourselves that this rate model gives an equilibroum relationship consistent with Eq. 14, with $K_{AB} = k_+/k_-$. Our derivation is more general, as no rate model is assumed, and easily transferrable to more complex systems. Note that Eq. 14 can be rearranged to give the fraction of B that are bound to A:

$$P_{\text{bound}}^B = \frac{[A]K_{AB}}{1 + [A]K_{AB}}. \tag{15}$$

This result appears to be identical to Eq. 8. In fact, it is more general as we have not assumed that $N_A^{\text{tot}} \gg N_B^{\text{tot}}$ during the derivation.

1.3 Cooperativity

Two interactions between biomolecules are cooperative when the presence of one interaction makes the presence of the other more likely. Cooperativity is extremely widespread within biological systems; we shall consider two classic examples.

1.3.1 Multimolecular association

Consider the assembly of the structure illustrated in Fig. 2 (b), which is formed from one molecule of type B and two molecules of type A. It is possible that the system is stabilised not only by the $A-B$ interactions but also by an additional $A-A$ attraction. If so, the system is likely to behave cooperatively; the second A molecule to enter the complex benefits from two attractive interactions rather than just one.

We shall model the system assuming perfect cooperativity; in other words, we shall assume that the state AB is so unfavourable it doesn't contribute to the partition function. We must therefore consider the equilibrium between A, B, and A_2B. We can follow the same reasoning as in Section 1.2.2 to obtain

$$P(N_{A_2B}) = \frac{q_{A_2B}^{N_{A_2B}} q_A^{(N_A^{\text{tot}} - 2N_{A_2B})} q_B^{(N_B^{\text{tot}} - N_{A_2B})} / N_{A_2B}! (N_A^{\text{tot}} - 2N_{A_2B})! (N_B^{\text{tot}} - N_{A_2B})!}{Z} \tag{16}$$

Maximising $P(N_{A_2B})$, and pressing the Stirling approximation into service once more,

$$\ln\left(\frac{q_{A_2B}(N_A^{\text{tot}} - 2N_{A_2B})^2(N_B^{\text{tot}} - N_{A_2B})}{N_{A_2B}q_A^2 q_B}\right) = 0. \tag{17}$$

Rearranging then gives

$$\frac{q_{A_2B}}{q_A^2 q_B} = \frac{N_{A_2B}}{N_A^2 N_B}. \tag{18}$$

As in the bimolecular case, we can define an equilibrium constant:

$$K_{A_2B} = \frac{V^2 q_{A_2B}}{q_A^2 q_B} = \frac{[A_2B]}{[A]^2[B]}. \tag{19}$$

This is another law of mass action result. It can be rearranged to give

$$P_{\text{bound}}^B = \frac{[A]^2 K_{A_2B}}{1 + [A]^2 K_{A_2B}}. \tag{20}$$

Later in the course you will study a mechanism through which cooperativity between ligands can arise without direct interactions between the two.

1.3.2 Hill functions

Eqs. 8 and 20, the probabilities that a receptor B is ligand-bound for single-ligand and cooperative-double-ligand binding, can both be written as "Hill functions". A Hill function is a general expression for dose-response:
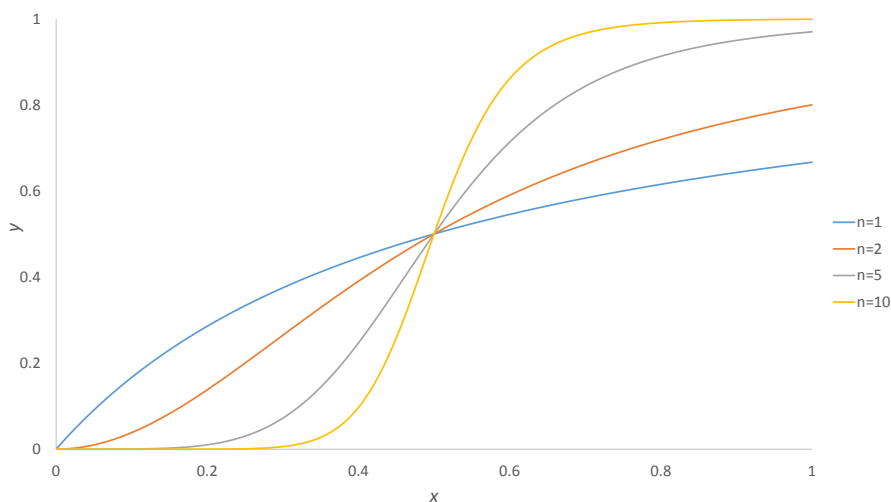
$$y = \frac{x^n}{x_{1/2}^n + x^n} \tag{21}$$

Figure 3 Hill functions $y = x^n/(x_{1/2}^n + x^n)$ plotted for a range of n at fixed $x_{1/2} = 0.5$.

Here, y is the response (P_{bound} in our systems), x is the dose ($[A]$ in our systems) and $x_{1/2}$ is a constant that quantifies the dose required to give half the maximum response ($1/K_{AB}$ and $1/K_{A_2B}$ in our systems). The Hill coefficient $n = 1$ for the single-ligand system, and $n = 2$ for the cooperative 2-ligand system.

The response of a Hill function is sharper as n is increased; different hill functions at fixed $x_{1/2}$ are compared in Fig. 3. Cooperative dual-ligand binding therefore shows a sharper response, and cooperative multi-ligand binding is sharper still (cooperative n-ligand binding gives a response curve with a hill coefficent of n). Such sharp, almost digital responses can be highly desirable for certain regulatory circuits or receptors that need to trigger a dramatic change in cellular behaviour in response to ligand concentrations. Note that the transition is only sharp for a large number of ligands when it is cooperative; if binding is uncooperative, the transition gets broader with increasing numbers of ligands.

The Hill function is so widely known in biological circles that it is often fitted to generic sigmoidal curves, even if there is no underlying theory that predicts Hill-like behaviour (there are plenty of functional forms that give a sigmoidal curve). This can be pretty misleading, as a single n may fail to describe the data over the whole range. So beware of this, but don't be surprised!

1.3.3 DNA hybridisation

Another situation in which cooperativity arises is when multiple interactions occur between two macromolecules. For example, a DNA molecule of N base pairs contains N separate base-pairing interactions that hold the molecule together. Let us consider a single pair of complementary bases in a box of volume V. Unless the box is absolutely tiny, these bases will not be bound for a significant fraction of the time in equilibrium. In coming together, they pay a large entropic penalty; instead of both bases being able to freely diffuse around the whole simulation volume independently, they must diffuse around as a unit. The energetic gain from a single base pair is not enough to overcome this cost. Now consider the same box, but containing two 2-base DNA strands. To form the first base pair, we once again must pay a large entropic penalty. But the entropic cost of forming the second base pair is far smaller; the two bases are already held in close proximity by the first contact – in forming the next base pair, we aren't losing as much freedom, and so it is much easier to go from one to two base pairs than from zero to one.

This physical picture can be encoded into a mathematical model. Let $Z_N(n)$ by the contribution to the partition function of a state with n base pairs, for two complementary N-base strands in a box of volume V. Let us normalize so that $Z_N(0) = 1$. For $Z_N(n > 0)$,

$$Z_N(n) = \sigma t^n. \tag{22}$$

In Eq. 22, t represents the stability of base pairs. if $t > 1$, states with more base pairs are more likely than states with fewer base pairs. $\sigma < 1$ encodes the additional cost of forming the first base pair.
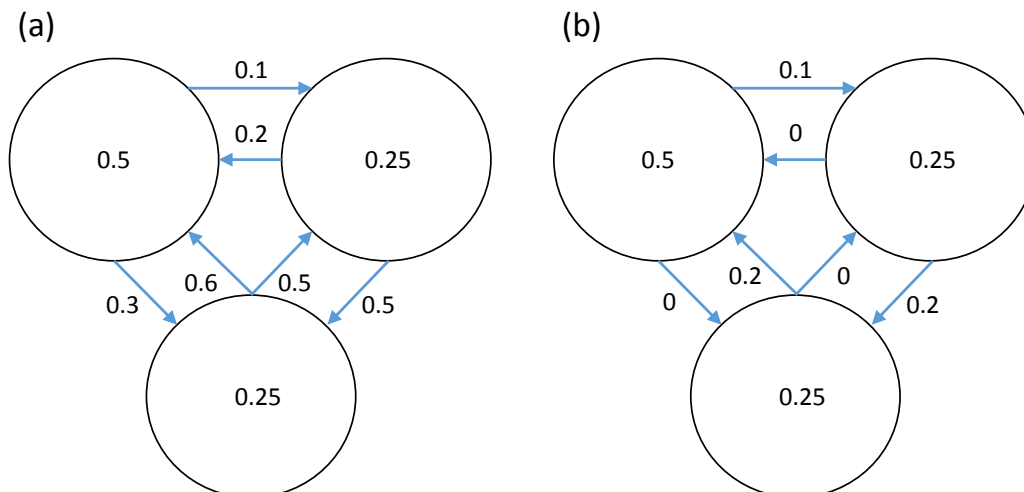
Figure 4 Two 3-state systems in steady state. Transition rates (in arbitrary units) label the arrows, $P(X)$ are shown inside the circles. $P(X)$ is identical in both cases, but in (a) the system obeys detailed balance whereas in (b) there is a persistent clockwise flow.

What is the stability of the duplex within this model? We can calculate

$$P_{\text{bound}} = \frac{\sum_{n=1}^{N} \sigma t^n}{1 + \sum_{n=1}^{N} \sigma t^n} = \frac{\sigma t(t^N - 1)}{t - 1 + \sigma t(t^N - 1)} \approx \frac{\sigma t^N}{1 + \sigma t^N}, \tag{23}$$

where the approximation is valid for large t. We see that if $t^N < \sigma^{-1}$, the two strands are mostly unbound. The stability of the duplex, however, grows exponentially with N. Thus at large N, the bound state is incredibly stable, even if σ is small and t is only marginally greater than unity. This cooperativity explains how very long duplex DNA is stable enough to function as genetic storage material, whilst individual base pairs can be disrupted relatively easily by proteins to allow access to the genetic material.

2 Out-of-equilibrium systems

In the calculations hitherto we have always focussed on calculating the equilibrium steady state. Not being in thermodynamic equilibrium is, however, a key feature of biological systems. Some systems in particular show features that are qualitatively distinct from that which can be achieved in equilibrium. In this section we discuss the meaning of being out-of-equilibrium in general, before focussing on a specific system (kinetic proofreading).

2.1 Equilibirum and the principle of detailed balance

A system that is in thermodynamic equilibrium has reached a steady distribution of states that persists indefinitely; the distribution is such that it maximises the combined entropy of the system in question and any reservoir to which it is connected. However, there is another characteristic of equilibrium that is not implied by these statements. Namely, it is postulated/observed that in thermodynamic equilibrium there are no steady currents; the flow of systems from state X_1 to state X_2 balances the flow of systems from state X_2 to state X_1. Systems with the same $P(X)$ that either do or don't obey this rule are illustrated in Fig. 4. This is the principle of *detailed balance*.

If we have two states X and Y that are connected by a transition, detailed balance implies that the rates of transition $r(X_1 \rightarrow X_2)$ and $r(X_2 \rightarrow X_1)$ between the two states in equilibrium are related:

$$r(X_1 \rightarrow X_2)P(X_1) = r(X_2 \rightarrow X_1)P(X_2), \tag{24}$$

in which $P(X)$ is the probability of being in state X in equilibrium. This implies

$$\frac{r(X_1 \to X_2)}{r(X_2 \to X_1)} = \frac{P(X_2)}{P(X_1)} = \frac{Z(X_2)}{Z(X_1)} = \exp(-(F(X_2) - F(X_1))/kT). \tag{25}$$

Importantly, we would not expect the transition rate $r(X_1 \to X_2)$ to depend on the actual distribution $\rho(X)$ (which may be different from $P(X)$, the equilibrium distribution). We would expect it to depend on temperature, biochemical details and the nature of states X_1 and X_2, but not $P(X)$. This implies that we would expect Eq. 24 to hold *even if we were out of equilibrium and $\rho(X) \neq P(X)$.*

2.2   Chemical fuel and persistent currents

As mentioned previously, biological systems are out of equilibrium, in that the states occupied are often not representative of the equilibrium distribution. But they are also out of equilibrium in a second sense, and one which is perhaps more relvant to life. Biological systems often have persistent currents, in that if you watch the system for a long time, you will see $X_1 \to X_2$ more than $X_2 \to X_1$ and this will carry on essentially indefinitely. Clearly this system is not in equilibrium, but it is also not simply an out-of-equilibrium system relaxing towards equilibrium. As a concrete example of a system relaxing to equilibrium, consider a gas that is confined to half of a box that is suddenly released and allowed to fill the whole box. Initially, there will be a net flow but this flow will decrease over time as the gas becomes more evenly distributed, and eventually the system will settle into equilibrium. This is not what is going on in biochemical systems that exhibit peristent currents.

Biological systems avoid the fate of the gas in the box by constantly taking in chemical fuel and releasing chemical waste to the environment. This allows them to maintain a persistent out-of-equilibrium state with net currents. One of the most common chemical fuels that are used by biological systems is adenosine triphosphate (ATP), a molecule that can be broken down into adenosine diphosphate and a single phosphate. Cells use the (free) energy from food to keep the $\text{ATP} \rightleftharpoons \text{ADP} + \text{P}$ reaction out of equilibrium. Sub-systems within the cell are then coupled to the $\text{ATP} \rightleftharpoons \text{ADP} + \text{P}$ reaction, allowing them to be systematically pushed in one direction. From the perspective of the sub-system, and ignoring what is happening to the ATP, it can then seem like Eq. 24 is violated (although it is not when the ATP is considered explicitly as well). To be specific, coupling a reaction $X_1 \to X_2$ to the breakdown of ATP allows

$$\frac{r(X_1 \to X_2)}{r(X_2 \to X_1)} \neq \exp(-(F(X_2) - F(X_1))/kT), \tag{26}$$

if the chemical fuel is ignored in the description of the state of the system X. The existence of implicit chemical fuel will allow us to write down reactions for biochemical systems that apparently violate Eq. 24. In doing so, it is assumed that the system functions in an environemnt (the rest of the cell) that supplies fuel and removes waste at the rate required to keep the system ticking over.

2.3   Kinetic proofreading during RNA translation

During the translation of messenger RNA (mRNA) into proteins, amino acids are recognised and added to the growing polypeptide chain due to the complementarity of a three-base sequence in the mRNA (a 'codon') and a three-base stretch of RNA attached to the amino acid (an 'anti-codon' loop that is part of a larger RNA structure called 'tRNA'). This process ocurrs in the ribosome, an enzyme that will be discussed later; see Fig. 5 for a very schematic sketch.

We can model the process through the state of the mRNA codon in question. In state R, there is nothing bound to the mRNA. In the state R_c, the correct amino acid is bound via its tRNA anchor. In the state R_i, an incorrect amino acid is bound (we shall assume a single competing amino acid is our most troubling competitor - extending the derivation to all of them is not very illuminating). We shall assume that the rates of binding to a given codon are the same in both cases. Unbinding, however, is much slower for the correct amino acid with the correct tRNA anchor: $k_-^i \gg k_-^c$. As a result, the correct amino acid is much more likely to be bound; in equilibrium

$$\frac{P_{\text{eq}}(R_c)}{P_{\text{eq}}(R_i)} = \frac{k_-^c}{k_-^i} = \frac{K_c}{K_i}, \tag{27}$$

where K is the equilibrium constant of the reaction in question. So far we have ignored the ribosome. It catalyses the formation of covalent bonds between successive amino acids and eventually the removal of the tRNA tag. Thus at some rate k_{cat}, R_c or R_i can
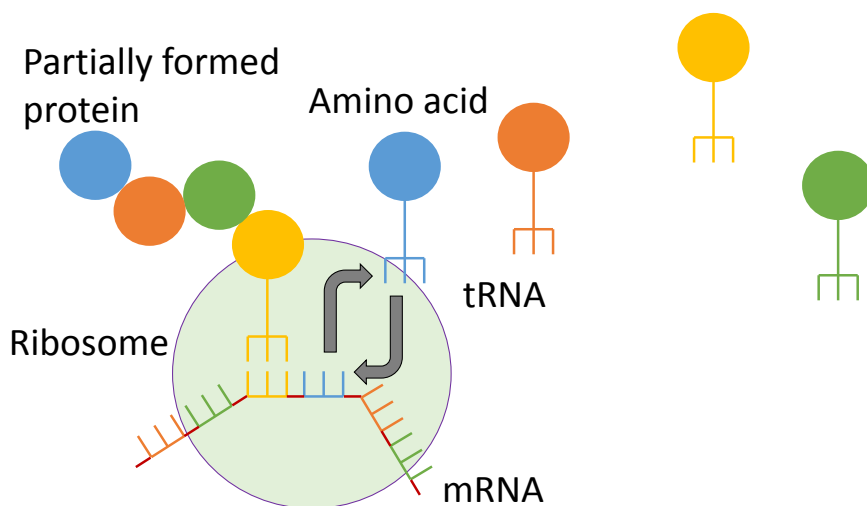
Figure 5 Schematic illustration of the translation of mRNA into proteins. The *tRNA* on the amino acid recognises the mRNA codon and binds to it, bringing the appropriate amino acid into close proximity witht he growing protein. The Ribosome then catalyses the addition of the amino acid to the covalently-bonded chain.

be converted into the final product. If $k_{\text{cat}} \ll k_-^c, k_-^i$, then the codon/amino acid binding reaction will approximately equilibrate before each enzymatic action and the ratio of correct to incorrect additions will be

$$\frac{P_c}{P_i} = \frac{P_{\text{eq}}(R_c)}{P_{\text{eq}}(R_i)} = \frac{K_c}{K_i}. \tag{28}$$

If k_{cat} is larger, then this ratio is reduced - in the limit $k_{\text{cat}} \gg k_-^c, k_-^i$, for example, any amino acid that binds will be converted and selectivity is negligible. If we plug in the numbers, it turns out that $\frac{K_c}{K_i} \sim 100$. This sounds good, but it means that 1% of amino acids incorporated into proteins would be a typo - on the order of one per protein. Such a high failure rate is not tolerable; indeed, the typo rate is closer to 0.01% in cells. How is this achieved?

2.3.1   Picasso lovers

Understanding how this can be achieved is helped by an analogy. Let us imagine that an art gallery director wants to give away a painting to someone who really loves Picasso. She sets up a room containing only a Picasso painting, connected to a corridor by a single door (Fig. 6 (a)). She finds that all people are equally likely to enter the room, but Picasso lovers stay 10 times as long. If she randomly picks someone in the room to receive the painting, she is enriching for Picasso lovers by a factor of 10.

Can she do any better? The gallery director opens a second room connected to the first and the corridor, and puts another Picasso painting in it (Fig. 6 (b)). She finds that, unfortunately, she simply ends up with a second room that is enriched by a factor of 10. Finally, she tries shutting off the second room from general access, locking the door between rooms and making the door from the second room to the corridor open from the inside only. Every hour, however, an attendent wil unlock the door between rooms and usher the occupants of the first room into the second room, locking the door behind them (Fig. 6 (c)).

What happens? The group of people in the first room are enriched for Picasso lovers by a factor of 10. This means that, at the moment the attendant ushers them into the second room, Picasso lovers are 10 times more likely to be present than normal. Then, however, people start to leave the room through the one-way door – and Picasso lovers will leave more slowly. The last person to leave will be 100 times more likely to be a Picasso lover than average. Why does this work when the earlier strategy didn't? The first case is like an equilibrium system, and the flow of people into each room from the corridor must balance the flow out; the occupancy of both rooms is then simply determined by the "equilibrium constants" of the visitors – and Picasso lovers "bind" 10 times more strongly to either room. In the second case, we are not in equilibrium. Detailed blance is violated
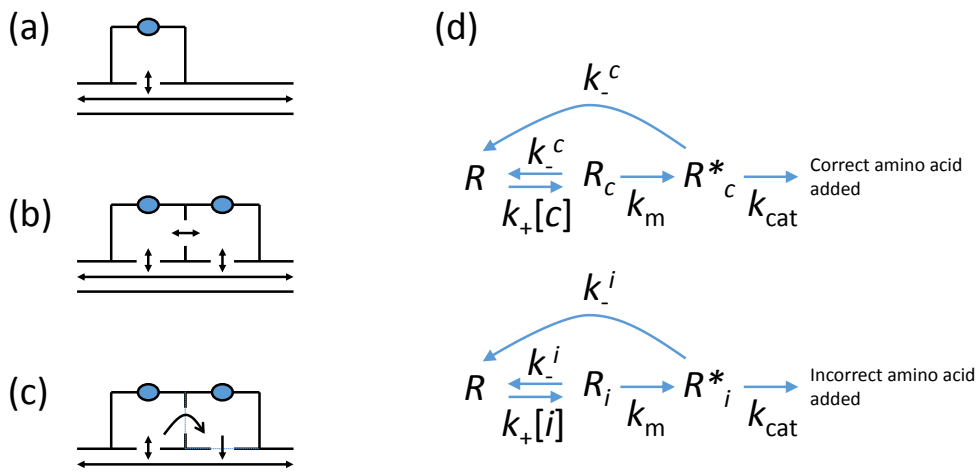
Figure 6 Kinetic proofreading. (a-c) The art gallery director's attempt to identify Picasso lovers. (a) A room containing a Picasso is opened up off a corridor. (b) A second adjacent room is opened up. (c) The second room is made inaccessible except when an attendent ushers people into it, locking the door benind them. The patrons can return to the corricor by a one-way door. (d) Set of reactions for modelling kinetic proofreading. Incorrect and correct amino acids are only differentiated by their unbinding rates. The implicit use of chemical fuel allows the apparently irreversible reactions, and the flow of systems around the $R \to R_x \to R_x^* \to R$ loop.

– you can see this easily because there is a one-way flow of people through the second room (remember: fluxes must balance in equilibrium). This means that the occupancy of the room is not fixed by equilibrium constants, and the director can achieve a greater enrichment for Picasso lovers.

2.3.2 Kinetic proofreading in detail

Kinetic proofreading during RNA translation works in essentially the same way as selecting for Picasso lovers. The initial binding of amino acids via their tRNA anchors to codons of the mRNA is like the discrimination provided by the first room with a Picasso painting. The R_i and R_c states do not proceed directly to incorporation of the amino acid into the growing protein, however. Instead the tRNA first undergoes chemical mdoification, converting to R_i^* or R_c^*. This step is coupled to breakdown of GTP, a chemical fuel like ATP, and therefore can be essentially irreversible; we effectively never see $R_i^* \to R_i$ or $R_c^* \to R_c$. This step is anologous to the action of the attendant in the art gallery. Finally, the amino acid can be added to the chain from the new state R_i^* or R_c^*, which is like selecting the prizewinner from the second room (having allowed time for some of the patrons to leave the room). It was initially thought that the first fuel-consuming step served no purpose, until J.J. Hopfield pointed out its capacity for improving specificity.

A detailed mathematical model is outlined in Fig. 6 (d). The key features of the model are the apparent irreversibility of certain reactions (made possible by the drive of chemical fuel) and the fact that the detachment of incorrect amino acids is faster at both the pre-modification and post-modification stages (the incorrect amino acid is still bound by a mismatched sequence after the initial modification). If $k_{\text{cat}} \ll k_-^c, k_-^i$, then the system reaches a non-equilibrium steady state between the R, $R_{c/i}$ and $R_{c/i}^*$ states; the molecules in the $R_{c/i}^*$ state are then slowly converted into the product. If $k_m \ll k_-^c, k_-^i$, the rate at which correct amino acids are added is given by

$$r_{\text{correct}} = k_{\text{cat}} R_c^* = k_{\text{cat}} \frac{k_m R_c}{k_-^c} = k_{\text{cat}} \frac{k_m}{k_-^c} \frac{k_+[c]}{k_-^c} R. \tag{29}$$

Similarly,

$$r_{\text{incorrect}} = k_{\text{cat}} R_i^* = k_{\text{cat}} \frac{k_m R_i}{k_-^i} = k_{\text{cat}} \frac{k_m}{k_-^i} \frac{k_+[i]}{k_-^i} R. \tag{30}$$

Thus, if $k_+[c] = k_+[i]$,

$$\frac{r_{\text{correct}}}{r_{\text{incorrect}}} = \left(\frac{k_-^i}{k_-^c}\right)^2 = \left(\frac{K_c}{K_i}\right)^2.$$

(31)

Kinetic proofreading is an excellent example of how biological systems can use chemical fuel to drive another reaction out of equilibrium. We will explore another instance in this week's reading.