

1 Simple FBA analysis of Crabtree/Warburg effect

Crabtree effect describes the phenomena observed for some yeast species that when yeast cells are put in an aerobic culture with high concentration of glucose, the yeast cells still mostly do fermentation, rather than respiration. This is weird because fermentation generates less ATP per glucose molecule, while respiration generates much more. A similar effect is observed for cancer cells, and named Warburg effect. The exact reason for this effect is still not conclusive. But we provide a bit of exploration of this phenomenon through very simple analysis using FBA. If interested, see [2] as a reference for a review and possible explanation on Crabtree effect.

Problem A: Stoichiometry and the fluxes.

For simplicity, let us consider just two metabolites, intermediates x_1 and ATP x_2 . Glucose import and conversion into intermediate cannot be changed very fast, and is not actively regulated in some yeast species (this is hypothesized as one possible reason for Crabtree effect in those species). So we consider the glucose conversion into intermediate as an external flux.

Hence, there are three fluxes, two internal and one external. Denote v_1 the flux consuming one intermediate and producing one ATP, representing fermentation. Denote v_2 the flux consuming one intermediate and producing two ATP, representing respiration. Denote v_3 the flux consuming one ATP, representing maintenance. Denote w the flux generating intermediate, representing glucose import and conversion into intermediate.

Write down the stoichiometry and fluxes of the system, in the form $\frac{d}{dt}x = Sv + S^w w$, where x is a column vector of the metabolite concentrations, S is the stoichiometry matrix for the internal fluxes, v is the flux variables, S^w is stoichiometry for the external fluxes, and w is external flux variable.

Solution A:

$$\frac{d}{dt} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 \\ 1 & 2 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} w.$$

Problem B: Steady state fluxes.

i. Write down the steady state relation between internal and external fluxes. It should be of the form $Sv = -S^w w$.

Solution B.i:

$$\begin{bmatrix} -1 & -1 & 0 \\ 1 & 2 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = - \begin{bmatrix} 1 \\ 0 \end{bmatrix} w$$

ii. Is the stoichiometry matrix invertible? If not, is the system underdetermined or overdetermined?

Solution B.ii: Not invertible, since it is not square. Underdetermined, since more rows than columns.

iii. Use pseudo-inverse to write down the set of all possible solutions to the above steady state equations. Write it in a form of $v^* = Bw^* + B^c c$, for some matrices or vectors B and B^c , and c is a vector or scalar with same dimension as the degrees of freedom left. c serves as coefficients that can take any real value.

Recall that for $\mathbf{Ax} = \mathbf{b}$ with \mathbf{A} wide rectangular. The set of all possible solutions can be found as $\mathbf{x} = \mathbf{A}^\dagger \mathbf{b} + \mathbf{x}^\perp$, where \mathbf{A}^\dagger is pseudo-inverse of \mathbf{A} , and \mathbf{x}^\perp is any vector satisfying $\mathbf{Ax}^\perp = 0$. One way to find \mathbf{x}^\perp vectors is to use projectors based on pseudoinverses: compute the projector $\mathbf{I} - \mathbf{A}^\dagger \mathbf{A}$, where \mathbf{I} is identity matrix with same dimension as the number of columns of \mathbf{A} . Then this projector should have the number of linearly independent columns exactly equal to the degrees of freedom left. If there is only one degree of freedom left, then all columns should be scalar multiples of each other. Then you can take any column as a basis for what \mathbf{x}^\perp can take, multiplied by a real number as coefficient. See the wikipedia page on pseudoinverse for more detail (there is a section on obtaining all solutions of linear equations).

Pseudo-inverses can be easily computed via computer code. You can use any language of your preference. If you are not familiar with this, here is how to calculate pseudo-inverse using WolframAlpha: <https://www.wolframalpha.com/input/?i=pseudoinverse+%5B%5B1%2C-1%2C0%5D%2C%5B-1%2C2%2C-1%5D%5D>

For you to check answers, one way to parameterize the set of solutions is the following.

$$\mathbf{v}^* = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} w^* + \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} c.$$

Solution B.iii: Solution is given as the last sentence above. The calculation step is that, the pseudo-inverse of stoichiometry matrix \mathbf{S} is

$$\mathbf{S}^\dagger = \begin{bmatrix} -1 & -1/3 \\ 0 & 1/3 \\ -1 & -2/3 \end{bmatrix}.$$

So the part from pseudo-inverse is

$$\mathbf{S}^\dagger(-\mathbf{S}^w)\mathbf{w}^* = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} w^*.$$

Now to find the vectors \mathbf{x}^\perp such that $\mathbf{Ax}^\perp = 0$, we use the projector method as suggested. In this case we get the vectors $(-1, 1, 1)c$, where c is some real scalar.

Problem C: Crabtree effect.

Now, assume all the fluxes are non-negative. We consider how Crabtree effect might be rationalized based on the set of steady state fluxes we see. Check that the balance in consumption of ATP results in the following relation: $v_1 + 2v_2 = v_3$. In other words, the ATP consumed comes from both fermentation and respiration. We can compare their fraction in contribution to ATP consumption by looking at $\frac{v_1}{v_3}$ and $\frac{2v_2}{v_3}$. Express these two ratios in terms of w^* and c .

Next, argue that when w^* is very large compared to c , which flux dominates in contribution to ATP consumption. Then argue that if w^* is small, so close to c , then which flux dominates.

Link the above observation to possibly explain the Crabtree effect.

Solution C: The steady state fluxes satisfy

$$\begin{bmatrix} v_1^* \\ v_2^* \\ v_3^* \end{bmatrix} = \begin{bmatrix} w^* - c \\ c \\ w^* + c \end{bmatrix}.$$

So we have $v_1^* + 2v_2^* = w^* - c + 2c = w^* + c = v_3^*$. The ratios are

$$\frac{v_1^*}{v_3^*} = \frac{w^* - c}{w^* + c}, \quad \frac{2v_2^*}{v_3^*} = \frac{2c}{w^* + c}.$$

When w^* is much larger than c , we have $w^* - c$ is close to $w^* + c$, therefore $\frac{v_1^*}{v_3^*}$ is close to 1, while $\frac{2v_2^*}{v_3^*}$ is close to zero. This corresponds to most of the ATP flux comes from fermentation, not respiration.

When w^* close to c , this means $w^* - c$ is small compared to w^* or c , therefore $\frac{v_1^*}{v_3^*}$ is close to zero, while $\frac{2v_2^*}{v_3^*} = \frac{2c}{w^* + c} \approx 1$. So most of the ATP flux comes from respiration, not fermentation.

Recall from lecture that w^* is the externally determined flux, while c is the internally adjustable degree of freedom. From the above calculation, we see that how w^* and c balances, i.e. how external and internal degrees of freedom balances govern whether fermentation or respiration is dominant for ATP flux.

Relating this to Crabtree effect, if we consider the internal degree of freedom c is limited by an upper bound (since $c = v_2^*$, this means respiration is limited by an upper bound), then if w^* is very large, much higher than c 's upper limit, then as we calculated above, most of the ATP flux will come from fermentation flux. This could be one explanation that the Crabtree effect is a result of simply balancing the steady state fluxes when respiration has limited maximum flux while the glucose uptake flux is very large.

2 Dynamics of life is stick balancing

One unavoidable part of life is growth, which in terms of chemical reactions is autocatalysis. Below, we formulate a few simple models of autocatalysis to see the generic structure involved. Namely, the system has an intrinsically unstable positive feedback on itself. Then, we show that a perhaps familiar mechanical control problem, stick balancing, also has the same structure. Therefore our intuition from stick balancing can be used to intuitively understand what it is like to maintain homeostasis when controlling metabolism.

Problem A: A simple model of glycolysis as an example of autocatalysis.

We begin by looking at what simple models of autocatalysis might have in common. We look at simple lumped models of three autocatalytic processes: glycolysis, amino acid-ribosome, and RNAP-ribosome.

One common example of autocatalysis is energy regeneration in glycolysis, where ATP is used to activate glucose, with more ATP generated eventually. To model this in a minimal way, we can lump detailed reaction steps together and consider just two reactions, one consuming ATP to activate glucose, resulting in an intermediate, and the other consuming an intermediate, producing more ATP. We can denote the two molecular species involved generically as X_1 and X_2 , with X_2 denoting the target species of autocatalysis such as ATP, and X_1 denoting intermediate. If we normalize the unit of concentration for ATP or energy charge such that the net production through an autocatalytic cycle generates one unit of X_2 , then we have the following dynamics of the metabolite concentrations.

$$\frac{d}{dt} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 1+q & -q \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} w. \quad (1)$$

Here q denotes the amount of ATP needed to activate glucose to produce the intermediate. We also added an external flux w that consumes ATP. This is generic for autocatalytic processes, since the target autocatalytic species is always produced for a consuming goal, such as an energy source or accumulated as part of biomass. Accumulation into biomass is effectively consuming the species under exponential growth.

Answer the following: how would v_1 and v_2 depend on x_1 and x_2 by default, without active regulation? Answer this in terms of a 2×2 sign matrix $\text{sgn } V$. For example, $\text{sgn } V_{11}$ is $+$ means v_1 increases with x_1 , and $\text{sgn } V_{11}$ is $-$ means v_1 decreases with x_1 , and $\text{sgn } V_{11}$ is 0 means v_1 does not change with x_1 . For the more mathematically minded, V can be considered as the derivative matrix of the fluxes v to the metabolites $\frac{\partial v}{\partial x}$, and then sign is taken to produce the sign matrix $\text{sgn } V$.

Solution A: By default, since v_1 is a reaction that consumes x_1 , then higher x_1 should increase the flux v_1 . Therefore V_{11} should be positive, $+$. Similarly, v_2 consumes x_2 , so V_{22} should be positive. So the sign matrix is

$$\text{sgn } V = \begin{bmatrix} + & 0 \\ 0 & + \end{bmatrix}.$$

Problem B: A simple model of amino acids-ribosome as an example of autocatalysis.

Other types of autocatalytic processes can be obtained based on whether the target species X_2 and the intermediate X_1 are catalytic or consumed in the reactions. For example, if we write a simple lumped equation for ribosomes catalyzing the production of itself, we may write

$$\frac{d}{dt} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -q & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} w. \quad (2)$$

Here the second reaction is ribosomes X_2 catalyzing the production of its ribosomal parts or intermediates X_1 , and the first reaction is parts assembled into ribosomes. The parameter q here correspond to the units of intermediates needed to produce one unit of ribosome. We choose the normalizations so that one autocatalytic cycle still produces one unit of the target species X_2 . Here the external consumption w now corresponds to ribosomes degrading or becoming dysfunctional over time.

Answer the following: compared to glycolysis, what is different about this amino acid-ribosome autocatalysis process? (Hint: why is the (2, 2) entry of the stoichiometry matrix zero?) What is sign $\text{sgn } V$ in this case? Compare it with the glycolysis case.

Solution B: This amino acid-ribosome example of autocatalysis is different from glycolysis in that x_2 , the ribosome here, is not consumed to make x_1 , but instead catalyzes the production of x_1 . In other words, although it is till the case that x_1 positively influence x_2 , and x_2 positively influence x_1 , the x_2 -to- x_1 influence is through catalysis rather than conversion here. The (2, 2) entry of the stoichiometry is zero here captures this fact, that x_2 is not consumed in v_2 reaction.

Here, the sign matrix $\text{sgn } V$ is the same as in the previous problem. The reason is although x_2 is not consumed for v_2 , it is the catalyst, so by default the flux should increase with higher number of catalysts.

Problem C: A simple model of RNAP-ribosome as an example of autocatalysis.

As another example, we can consider ribosomes as X_2 and RNAP (RNA polymerases) as X_1 , although in this case both may serve as the autocatalytic target. In this case, ribosomes produce RNAPs , while RNAP also help produce ribosomes, both through catalysis.

Write the dynamics equation for RNAP-ribosome like we did for glycolysis and amino acid-ribosome.

Solution C:

$$\frac{d}{dt} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & q \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} w. \quad (3)$$

Here q is used to capture the difference in catalytic activities of RNA polymerase and ribosomes. Both S_{11} and S_{22} are zero because both RNAP and ribosomes are catalysts. In other words, the nucleic acids and amino acids that are consumed in these catalysis reactions are hidden, therefore assumed abundant.

Problem D: Dynamics of stick balancing.

Without active regulations, autocatalytic systems are intrinsically unstable, in the sense that increase in x_1 above the steady state value will cause increase in x_2 , which in turn cause increase in x_1 , thus an upward spiral blowing up to infinity. Alternatively, decrease in x_1 below its steady state value will cause decrease in x_2 which in turn cause decrease in x_1 , thus forming a downward spiral crashing to zero. As regulatory motifs, we can consider x_1 and x_2 positively influence each other, creating a positive feedback loop. However, unlike gene regulatory circuits where there exists saturation or safety valve by default since material can run out, in metabolism running out means cells crash and die.

We can intuitively understand the situation by comparing with a familiar intrinsically unstable system: balancing a stick. Below, through standard calculations, we show that the dynamic equations of balancing a stick has similar form as the autocatalytic equations. As a result, our intuitions about the hardness in balancing a stick, and the oscillatory movements of hand and stick when disturbed, all pass down to dynamic properties of autocatalytic metabolism.

Balancing a stick by hand is the same problem as balancing an inverted pendulum on a moving cart. Consider a stick of length ℓ , with mass m at the head, and we control the bottom of the stick on a horizontally moving cart (or hand), which has mass M . Let θ be the angle of the stick, so that $\theta = 0$ is straight down, and $\theta = \pi$ is straight up.

The equations of motion governing an inverted pendulum on a cart, the same as a hand balancing a stick, is the following (see Wikipedia page on inverted pendulum if you want detailed derivations).

$$(M+m)\ddot{x} + m\ell(\ddot{\theta} \cos \theta - \dot{\theta}^2 \sin \theta) = u,$$

$$m(\ddot{x} \cos \theta + \ell\ddot{\theta} - g \sin \theta) = 0.$$

Here x is the horizontal position of the cart or hand, with positive x in direction to the right. u is the force applied to the cart or hand horizontally, also to the right.

- Linearize this around the upright position with zero angular velocity, i.e. around $\theta^* = 0$, so that $\theta = \theta^* + \delta\theta$ for small deviation $\delta\theta$, and similarly for higher order time derivatives, such as $\dot{\theta} = \dot{\theta}^* + \delta\dot{\theta}$ for small deviation in angular velocity $\delta\dot{\theta}$. Check that you obtain the following.

$$(M+m)\delta\ddot{x} + m\ell\delta\ddot{\theta} = u,$$

$$\delta\ddot{x} + \ell\delta\ddot{\theta} - g\delta\theta = 0.$$

Solution D.i: For linearization, we throw away any term that is higher than first order. Recall from Taylor expansion to first order, linearize a function $f(x)$ at a point x^* is $f(x) \approx f(x^*) + \delta x \frac{df}{dx}(x^*)$, where $\delta x = x - x^*$. So just looking at the first order term for small deviation around the reference point, we denote $\delta(f(x)) \approx \delta x \frac{df}{dx}(x^*)$. Here we have three variables, θ , $\dot{\theta}$ and $\ddot{\theta}$, and we linearize around the reference point $\theta = \dot{\theta} = \ddot{\theta} = 0$, so the Taylor expansion formula for a generic function $f(\theta, \dot{\theta}, \ddot{\theta})$ is $\delta f(\theta, \dot{\theta}, \ddot{\theta}) \approx \delta\theta \frac{\partial f}{\partial \theta}(0, 0, 0) + \delta\dot{\theta} \frac{\partial f}{\partial \dot{\theta}}(0, 0, 0) + \delta\ddot{\theta} \frac{\partial f}{\partial \ddot{\theta}}(0, 0, 0)$. Recall that trigonometric functions linearize as the following $\cos \theta \approx 1 - \delta\theta$, $\sin \theta \approx \delta\theta$.

So for small deviations, $\delta(\ddot{\theta} \cos \theta) \approx \delta\ddot{\theta}$, because the only partial derivative nonzero at the reference point is with respect to $\ddot{\theta}$, which is $\cos \theta = 1$ at the reference point. Similarly, $\delta(\dot{\theta}^2 \sin \theta) \approx 0$ since all partial derivatives are zero at the reference point. We also have, $\delta(\ddot{x} \cos \theta) \approx \delta\ddot{x}$, since \ddot{x} is zero at the reference point as well (the reference point is static). All this together gives the desired linearization given in the problem.

- ii. Next, focus on just the angle $\delta\theta$, combine the equations to get rid of $\delta\ddot{x}$. Show that you can obtain the following system of equations relating control input u from the hand, and the angle.

$$(1 + \frac{m}{M}) \frac{g}{\ell} \delta\theta - \frac{u}{M\ell} = \delta\ddot{\theta}.$$

Solution D.ii: Plug in the expression for $\delta\ddot{x}$ from the second equation into the first equation, we obtain $(M + m)g\delta\theta - M\ell\delta\ddot{\theta} = u$. Divide by $M\ell$ we obtain the desired equation.

- iii. Then, take the stick angle θ and its velocity $\dot{\theta}$ as the state variables, just like x_1 and x_2 in autocatalysis equations. Write the dynamics into a linear system form. In other words, fill in the ? matrix below.

$$\frac{d}{dt} \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} = ? \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{1}{M\ell} \end{bmatrix} u. \quad (4)$$

Compare this with the autocatalytic systems.

Solution D.iii:

$$\frac{d}{dt} \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ (1 + \frac{m}{M}) \frac{g}{\ell} & 0 \end{bmatrix} \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{1}{M\ell} \end{bmatrix} u$$

This has exactly the same form as the RNAP-ribosome example. We can make this more explicit by linearizing the RNAP-ribosome example. Since v_1 positively varies with x_1 , $\delta v_1 \approx k_1 \delta x_1$, and similarly $\delta v_2 \approx k_2 \delta x_2$, for some positive coefficients k_1 and k_2 . So the linearization of RNAP-ribosome example is

$$\frac{d}{dt} \begin{bmatrix} \delta x_1 \\ \delta x_2 \end{bmatrix} = \begin{bmatrix} 0 & k_2 q \\ k_1 & 0 \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} \delta w.$$

Exactly the same form as the stick balancing dynamics.

The other two examples of autocatalysis, namely glycolysis and amino acid-ribosome, can be shown to have the same form via a stranformation of states. For example, the amino acid-ribosome system in variable $\delta y_1 = \delta x_1 + q\delta x_2$, $\delta y_2 = \delta x_2$, has the same form as the stick balancing system.

3 Reaction order in a simple binding network

Binding reactions regulate catalysis fluxes. The simplest binding network is just one binding reaction. Luckily, in this case everything can be solved explicitly. We analyze this here.

Problem A: We denote one binding reaction as $E + S \xrightleftharpoons[k^-]{k^+} C$. The dynamics from this binding reaction for the concentration of the complex C is

$$\frac{d}{dt}C = k^+ES - k^-C.$$

Here the symbols denote their concentrations. Write the steady state equation that C has to satisfy. Use $K = \frac{k^-}{k^+}$.

Solution A: $CK = ES$.

Problem B: Regulation through the binding network is achieved through adjusting the total concentrations of enzyme and substrates. Here the definition for the totals are total enzyme $t_E = E + C$ and total substrate $t_S = S + C$.

Take the steady state equation, which is in terms of (E, S, C, K) , and transform it to an equation relating C to the totals, in terms of (t_E, t_S, C, K) .

Solution B: $CK = (t_E - C)(t_S - C)$.

Problem C: The resulting equation is quadratic. Solve it for C in terms of the totals, and check that the solution is

$$C(t_E, t_S) = \frac{1}{2} \left(t_E + t_S + K - \sqrt{(t_E + t_S + K)^2 - 4t_E t_S} \right).$$

Solution C: Apply the formula to solve quadratic equations, or use Mathematica or Wolfram alpha...

Problem D: This formula is exact, but a bit complicated. One simplification is by Michaelis-Menten assumption that there is much more substrate than enzyme, $t_S \gg t_E$. This asymptotic condition can be used to simplify the expression, such as $t_E + t_S \approx t_S$, since t_E is negligible.

Check that apply $t_S \gg t_E$ to the exact solution above yield the Michaelis-Menten formula $C \approx t_E \frac{t_S}{t_S + K}$.

Solution D: Under $t_S \gg t_E$, we have

$$\begin{aligned} \sqrt{(t_E + t_S + K)^2 - 4t_E t_S} &= \sqrt{(t_S + K)^2 - 4t_E t_S} \approx (t_S + K) \sqrt{1 - 4 \frac{t_E}{t_S} \frac{t_S^2}{(t_S + K)^2}} \\ &\approx (t_S + K) \left(1 - 2 \frac{t_E}{t_S} \frac{t_S^2}{(t_S + K)^2} \right) = t_S + K - 2 \frac{t_E t_S}{t_S + K}, \end{aligned}$$

where we used the approximation $(1 + \varepsilon)^a \approx 1 + a\varepsilon$, for small ε and $a > 0$. Since $t_E + t_S + K \approx t_S + K$, together we

have

$$C(t_E, t_S) \approx \frac{1}{2} \left(t_S + K - \left(t_S + K - 2 \frac{t_E t_S}{t_S + K} \right) \right) = \frac{t_E t_S}{t_S + K},$$

as desired.

Problem E: Since we can apply the scenario that substrates are much more than the enzyme, we could apply other scenarios as well. Consider the case that enzyme and substrate binds very tightly, so that $t_E, t_S \gg K$.

Show that apply the above condition to the exact solution yields the following result.

$$C(t_E, t_S) \approx \frac{(t_E + t_S) - |t_E - t_S|}{2} = \min\{t_E, t_S\}.$$

Does this tight binding formula make sense?

Just to give some context. Although this tight binding formula is not as popular as Michaelis-Menten, it should be because of its high biological relevance. In [1], a circuit motif achieving robust perfect adaptation is proposed that is based on the strong binding of two molecules, such as sense and anti-sense RNA strands or sigma and anti-sigma factors. This tight binding formula is perfectly appropriate in this situation, and is used in analyzing the circuit motif in [3]. In [4], this tight binding formula taking minimum between enzyme and substrate is used to unify several bacterial growth laws in distinct growth conditions. It is highly likely that the tight binding formula is used and applied in many scenarios that we do not know, simply because of the lack of a common name like Michaelis-Menten or Hill to enthrone the tight binding formula into the canon of quantitative biology.

Solution E: Under the asymptotic condition $t_E, t_S \gg K$, we have $t_E + t_S + K \approx t_E + t_S$, and

$$\sqrt{(t_E + t_S + K)^2 - 4t_E t_S} \approx \sqrt{(t_E + t_S) - 4t_E t_S} = \sqrt{(t_E - t_S)^2} = |t_E - t_S|.$$

So together

$$C(t_E, t_S) \approx \frac{1}{2}(t_E + t_S - |t_E - t_S|),$$

as desired.

This formula makes sense as the tight binding limit. Whenever one of the total concentrations (t_E, t_S) is higher than the other, every molecule of the less abundant species is sequestered by the more abundant species, therefore the minimum.

Problem F: (optional). You can also directly apply log derivative to the exact solution to obtain the formula for the reaction order or log derivative of C with respect to t_E and t_S , and see that it is constrained in the triangle with vertices $(1, 0)$, $(0, 1)$ and $(1, 1)$. Relate the Michalis-Menten and tight binding formula with the triangle. Show that they are edges of the triangle.

Solution F: We only need to do $\frac{\partial \log C}{\partial \log t_E}$ to get everything, since t_E and t_S are symmetric. We calculate this here via linear derivatives. So we first relate log derivative with linear derivative. (You could also calculate using log derivative formula.)

$$\frac{\partial \log C}{\partial \log t_E} = \frac{t_E}{C} \frac{\partial C}{\partial t_E}.$$

The linear derivative satisfy

$$2 \frac{\partial C}{\partial t_E} = 1 - \frac{t_E + K - t_S}{\sqrt{(t_E + t_S + K)^2 - 4t_E t_S}}.$$

We simplify the result by expressing this in terms of (E, S, K) instead of (t_E, t_S, K) . Use the steady state equation $KC = ES$ to simplify the formula. As a result, $\sqrt{(t_E + t_S + K)^2 - 4t_E t_S} = E + S + K$, and $t_E + K - t_S = E + K - S$. So we have

$$2 \frac{\partial C}{\partial t_E} = \frac{2S}{E + S + K}.$$

Together, since $\frac{t_E}{C} = \frac{E + ES/K}{ES/K} = \frac{S+K}{S}$, we have

$$\frac{\partial \log C}{\partial \log t_E} = \frac{S+K}{S} \frac{S}{E+S+K} = \frac{S+K}{E+S+K}.$$

To get $\frac{\partial \log C}{\partial \log t_S}$ we just exchange the labels of E and S in $\frac{\partial \log C}{\partial \log t_E}$. So we have the end result

$$\frac{\partial \log C}{\partial \log(t_S, t_E)} = \left[\frac{E+K}{E+S+K} \quad \frac{S+K}{E+S+K} \right].$$

This formula can be written as convex combination of three vertices,

$$\frac{\partial \log C}{\partial \log(t_S, t_E)} = \lambda_E [1 \ 0] + \lambda_S [0 \ 1] + \lambda_K [1 \ 1],$$

where $\lambda_E = \frac{E}{E+S+K}$, $\lambda_S = \frac{S}{E+S+K}$, and $\lambda_K = \frac{K}{E+S+K}$. These three λ 's are convex coefficients, in the sense that they are non-negative and they sum to 1. Therefore, varying all these coefficients gives the triangle spanning the three vertices. Each vertex corresponds to its coefficient is 1, such as $(1, 0)$ vertex is when $\lambda_E = 1$, so the other coefficients are zero.

The Michaelis-Menten edge between $(0, 1)$ and $(1, 1)$ therefore is when $\lambda_E = 0$. Indeed, the assumption $t_S \gg t_E$ implies λ_E is small. To see this, $t_S \gg t_E$ written in (E, S, K) is $S + ES/K \gg E + ES/K$, which is equivalent to $S \gg E$, so $E \ll E + S + K$, therefore λ_E is small.

The tight binding limit is the edge between $(1, 0)$ and $(0, 1)$, i.e. when λ_K is small. Indeed, $t_S, t_E \gg K$ implies this.

4 Stochastic, discrete, statistical mechanics of simple binding

Our analysis and arguments are in the bulk scenario where concentrations of molecules are used, rather than discrete counts. For biophysics students this might be unacceptable, since distributions on states accounting for every molecule seem important. To show that the deterministic results from concentrations are intimately related to the stochastic, discrete case, in this problem we walk through the full calculation for steady state distribution of molecule counts in the simple binding network, and show that the highest probability state becomes the deterministic case when molecule numbers are high.

Let us consider the binding reaction $E + S \xrightleftharpoons[k^-]{k^+} C$ when the number of molecules are discrete and the binding and unbinding reactions are stochastic. The discreteness and stochasticity become important when the molecule amount is small. We want to obtain the steady state distribution of variables (E, S, C) , which are the number of free enzymes, free substrates, and complexes.

Problem A: In the discrete, stochastic case, the state of the system is the number of free enzyme, free substrate, and bound complexes, denoted (E, S, C) . There are two reactions, corresponding to two types of transitions between states. The binding reaction describes the transition rate from state (E, S, C) to $(E - 1, S - 1, C + 1)$, with rate k^+ES . The unbinding reaction describes the transition rate from state (E, S, C) to $(E + 1, S + 1, C - 1)$ with rate k^-C .

For a given initial condition, the total number of enzymes $t_E = E + C$ and substrates $t_S = S + C$ are fixed, since they are not changed by the reactions. Define $c_{\max} = \min\{t_E, t_S\}$. Because the number of complexes is always less than the total of enzyme and the total of substrates, i.e. $C \leq c_{\max}$, the states that C can jump to are $\{0, 1, \dots, c_{\max}\}$. The states that E and S can jump to are more complicated. For example, if $t_S \geq t_E$, then S can reach states $\{t_S - t_E, t_S - t_E + 1, \dots, t_S\}$.

Because of the constraints, we can simplify the state of the system by considering just one variable, C , together with constants t_E and t_S from initial conditions. Then $S = t_S - C$, and $E = t_E - C$. So the system becomes a one-dimensional stochastic jump on state space $C \in \{0, 1, \dots, c_{\max}\}$. Transition $C \rightarrow C + 1$ has rate $k^+(t_E - C)(t_S - C)$, and transition $C \rightarrow C - 1$ has rate k^-C .

Now, we solve for the steady state distribution of C in this jump chain.

Check that the following ODE describes the dynamics of $p(C, t)$, the probability that the system is in state C at time t :

$$\frac{d}{dt}p(C, t) = k^-(C + 1)p(C + 1, t) + k^+(t_E - C + 1)(t_S - C + 1)p(C - 1, t) - (k^-C + k^+(t_E - C)(t_S - C))p(C, t),$$

when $C \in \{0, \dots, c_{\max}\}$. Here $p(C, t)$ is assumed zero always for $C < 0$ or $C > c_{\max}$.

Solution A: The ODE is correct...

At state C , there are two outflux rates. The rate to jump up from state C is k^+ES , written in (t_E, t_S, C) is $k^+(t_E - C)(t_S - C)$. The rate to jump down from state C is k^-C .

At state C , there are two influx rates: the rate that state $C - 1$ jumps up to this state is $k^+E(C - 1)S(C - 1)$, where $E(C - 1)$ denotes the value of E in state $C - 1$, i.e. $E(C - 1) = t_E - (C - 1)$; the rate that state $C + 1$ jumps down to this state is $k^-(C + 1)$.

Problem B: At steady state, because this chain is one-dimensional, the forward and backward jump rates are equal. In other words, this chain satisfies detailed balance. This corresponds to the following equation:

$$k^+(t_E - C)(t_S - C)p(C) = k^-(C + 1)p(C + 1), \quad C = 0, \dots, c_{\max} - 1.$$

where $p(C)$ is the steady state probability at state C . Show this detailed balance condition holds. (Hint: begin with the boundary case $C = 0$.)

Solution B: At $C = 0$, the ODE for probability is

$$\frac{d}{dt}p(0, t) = k^-p(1, t) - k^+t_E t_S p(0, t).$$

So the detailed balance equation is indeed satisfied at steady state. Now look at the ODE for $C = 1$,

$$\frac{d}{dt}p(1, t) = k^-(2)p(2, t) + k^+t_E t_S p(0, t) - (k^- + k^+(t_E - 1)(t_S - 1))p(1, t).$$

Apply the condition from $C = 0$ steady state, we have that the steady state condition for $p(1, t)$ becomes $k^-(2)p(2, t) - k^+(t_E - 1)(t_S - 1)p(1, t) = 0$, which is exactly the detailed balance condition. Iterate upwards through all the states gives that the detailed balance condition is satisfied for all states, as desired.

This fact that a one-dimensional jump chain's steady state satisfy detailed balance condition holds in general, even when the chain is infinite. See van Kempen's book on *Stochastic Processes in Physics and Chemistry* or James Norris's book on *Markov Chains* if you want to learn more.

Problem C: Using the above detailed balanced condition, show that this implies the steady state distribution satisfies

$$p(C) \propto \frac{K^{-C}}{E!S!C!}, \quad E = t_E - C, \quad S = t_S - C, \quad K = \frac{k^-}{k^+}.$$

Here the proportional sign \propto omits all factors that do not depend on C . In other words, the steady state distribution is a product of Poisson distributions truncated to state space $C \in \{0, \dots, c_{\max}\}$.

Solution C: We can re-write the detailed balance condition as

$$p(C+1) = \frac{K(t_E - C)(t_S - C)}{(C+1)} p(C).$$

Now iterate this formula to go from C to 0, we have

$$\begin{aligned} p(C) &= \frac{K(t_E - C + 1)(t_S - C + 1)}{C} p(C-1) = K^2 \frac{(t_E - C + 2)(t_S - C + 2)}{C-1} \frac{(t_E - C + 1)(t_S - C + 1)}{C} p(C-2) \\ &= \dots = K^C \frac{[t_E(t_E - 1) \cdots (t_E - C + 1)][t_S(t_S - 1) \cdots (t_S - C + 1)]}{C!} p(0) \\ &= K^C \frac{t_E! t_S!}{E! S! C!} p(0). \end{aligned}$$

since the detailed balance condition at the end is $p(1) = \frac{K}{t_E t_S} p(0)$, and $(t_E)(t_E - 1) \cdots (t_E - C + 1) = \frac{t_E!}{(t_E - C)!} = \frac{t_E!}{E!}$. Now since $t_E! t_S! p(0)$ are independent of C , we take them as a proportional constant to obtain the desired formula.

Problem D: We can connect this steady state distribution $p(C)$ to the deterministic steady state solution $C = \frac{ES}{K}$ by looking at the mode, i.e. C that achieves the maximum of $p(C)$, in the limit that molecule numbers E, S, C are large. Use Stirling's formula that $\log N! \approx N \log N - N$ when N is large to show that when E, S, C are large,

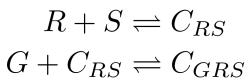
$$\log p(C) \approx \text{const} - C \log K - E \log E + E - S \log S + S - C \log C + C,$$

where const denotes an additive constant independent of C . Then take derivative of this expression with respect to C to show that

$$\frac{d}{dC} \log p(C) = \log \frac{(t_E - C)(t_S - C)}{CK}.$$

Conclude that the maximum of $p(C)$ is achieved $C = \frac{(t_E - C)(t_S - C)}{K} = \frac{ES}{K}$, which is the deterministic steady state solution.

- ① Specify a binding network



- ② Write down stoichiometry matrix

$$\mathbf{N} = \begin{bmatrix} G & R & S & C_{RS} & C_{GRS} \\ 0 & -1 & -1 & 1 & 0 \\ -1 & 0 & 0 & -1 & 1 \end{bmatrix}$$

atomic species

$$\mathbf{L} = \begin{bmatrix} G & R & S & C_{RS} & C_{GRS} \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{matrix} t_G \\ t_R \\ t_S \end{matrix}$$

- ③ Compute conservation law matrix

- ④ Compute reaction orders by formula

$$\mathbf{x} = (G, R, S, C_{RS}, C_{GRS})$$

$$\mathbf{t} = (t_G, t_R, t_S) = \mathbf{L}\mathbf{x}$$

$$\frac{\partial \log \mathbf{x}}{\partial \log(\mathbf{t}, \mathbf{k})} = \begin{bmatrix} \Lambda_t^{-1} \mathbf{L} \Lambda_x \\ \mathbf{N} \end{bmatrix}^{-1}$$

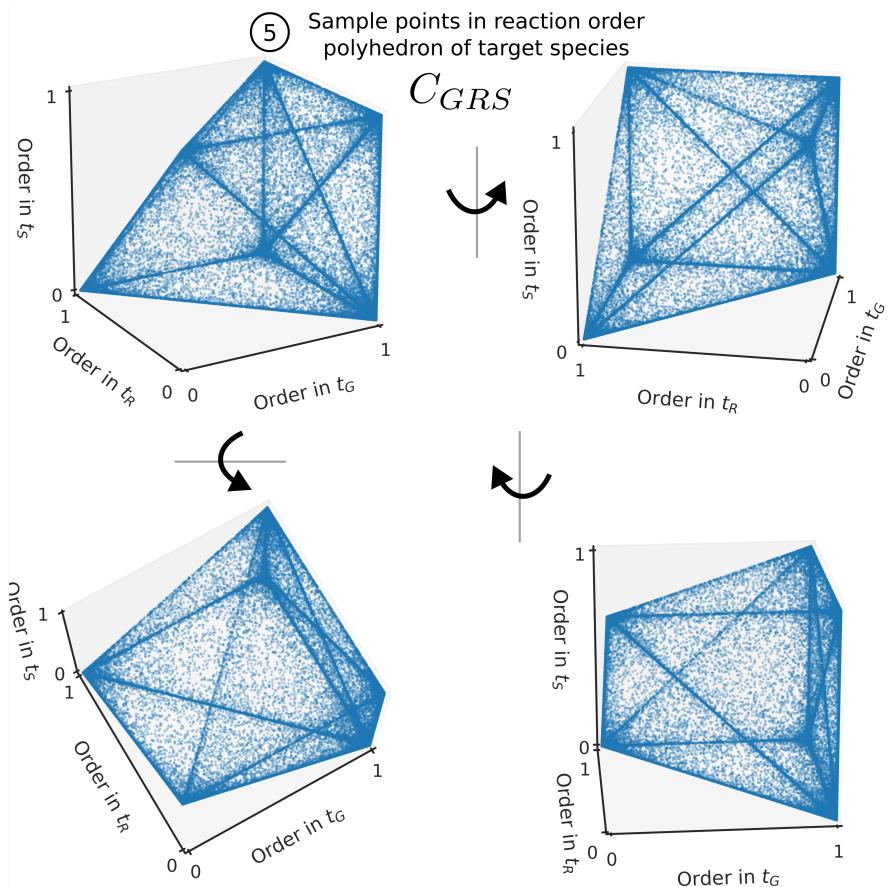


Figure 1 The procedure to computationally sample reaction order polyhedra of a binding network, illustrated with the binding network of an induced activator. Step 1, specify a binding network. Step 2, write down the stoichiometry matrix for the binding reactions. For each binding reaction, use the stoichiometry in the binding direction. For the order of the species, put the free form of the species first. These species are called atomic species, with the special property that conserved quantities represent totals of these species. Step 3, compute the conservation law matrix \mathbf{L} from the stoichiometry matrix \mathbf{N} . Step 4, compute reaction orders using the formula. Here Λ_t denote a diagonal matrix with the vector of totals \mathbf{t} on the diagonal. Same for Λ_x . Step 5, for a target species whose reaction order is to be studied, we can sample points of its reaction order polyhedron by taking random values of \mathbf{x} , calculate \mathbf{t} and pass through the formula to compute the reaction order, and then plot the points. The figure listed here shows the reaction order polytope of the induced activator, rotated in four different angles to show the 3D shape. The arrow around a line denotes in which direction is the 3D shape rotated from the upper left figure. Sampling of the C_{GRS} reaction order polytopes is done by log-uniformly sampling the values of each variable in $(G, R, S, C_{RS}, C_{GRS})$ between 10^{-6} and 10^6 with 100000 points in total.

Solution D: The calculation to get $\log p(C)$ and the derivatives are straightforward. The maximum is achieved at setting the derivative to zero. This is indeed the maximum since the second order derivative is $\frac{d^2}{dC^2} \log p(C) = -\frac{E+S+K}{ES}$ is always negative.

5 Computational sampling of reaction order polyhedra

Binding networks regulate catalysis fluxes by varying the reaction orders within a polyhedra set. One motivation for using reaction orders instead of directly solving for the flux in terms of the total concentrations is that this results in high degree polynomial problems that are intractable to solve, analytically or computationally. In contrast, although less familiar, reaction orders can be solved computationally at scale. We describe the formula that allow us to sample reaction orders by solving simple matrix algebra. We will not discuss how this is derived, but simply walk through how to use it. You are encouraged to implement this with coding language of your choice. At the end, there is a Jupyter notebook already written for you to play around.

The computational procedure is summarized in Figure 1. We walk through how to use it below, and ask you to write out how to do the computation for an enzyme binding with multiple substrates.

Problem A: Stoichiometry matrix.

The key of the formula is the equation showing how to calculate reaction order, or log derivative, in step 4. We see it requires the input of the stoichiometry matrix \mathbf{N} of dimension $r \times n$ and the conservation law matrix \mathbf{L} of dimension $d \times n$, where n is the number of species involved, r is the number of binding reactions with linearly independent stoichiometry, and d is the number of conserved quantities or the number of totals. In the example in the figure, $n = 5$, $r = 2$, and $d = 3$.

The stoichiometry matrix can be directly obtained from the binding network once it is specified, as shown in Step 2 of Figure 1. It might occur that there are binding reactions with linearly dependent stoichiometry vector. In that case, compute the rank r of the stoichiometry matrix, and select r of the linearly independent reactions to form the stoichiometry matrix \mathbf{N} . Since binding reactions are reversible, as a convention, we choose the stoichiometry vector of the binding direction to form matrix \mathbf{N} . When possible, it is helpful if a subset of d of the species, considered atomic species, can be distinguished and arranged first in the ordering of the species. Atomic species often correspond to the free form of the molecules, so that their total amount is conserved through the binding reactions.

Consider an enzyme that needs to bind with multiple substrate molecules to perform catalysis reaction. $E_0 + n_s S \rightleftharpoons E_{n_s}$, where n_s is the number of substrate molecules bound. Write down the stoichiometry matrix for it, with species order (E_0, S, E_{n_s}) . What are the atomic species?

Solution A: The stoichiometry matrix \mathbf{N} is

$$\mathbf{N} = [-1 \quad -n_s \quad 1].$$

The atomic species are E_0 and S .

Problem B: Conservation law matrix.

The conservation law matrix \mathbf{L} defines the totals (see Step 3 of Figure 1). Often we can directly write down this by inspection, based on the physical interpretation of the species. Once \mathbf{L} is obtained, the total \mathbf{t} of dimension d can be defined as $\mathbf{t} = \mathbf{Lx}$, where \mathbf{x} , a vector of dimension n , is the concentrations of all the species involved.

Write down the conservation law matrix for enzyme binding with multiple substrates.

Solution B: The conservation law matrix is the following, with totals ordered by (t_S, t_E)

$$L = \begin{bmatrix} 0 & 1 & n_s \\ 1 & 0 & 1 \end{bmatrix}.$$

Problem C: (Optional) Computing conservation law matrix from stoichiometry.

The conservation law matrix L can also be computed from the stoichiometry matrix N . If the atomic species are ordered first, then this computation is simple. We can split L into two submatrices: $L = [I_d \ L_2]$, with the first submatrix a $d \times d$ identity matrix. Similarly split $N = [N_1 \ N_2]$, then the submatrix L_2 can be computed as $L_2 = (-N_2^{-1}N_1)^T$.

Solution C: Here $N_2 = 1$, $N_1 = [-1 \ -n_s]$, therefore $L_2 = [1 \ n_s]^T$. So the full stoichiometry matrix is

$$L = \left[\begin{array}{cc|c} 0 & 1 & n_s \\ 1 & 0 & 1 \end{array} \right],$$

as obtained in the previous problem.

Problem D: Implement the algorithm to sample.

With both the stoichiometry matrix N and the conservation law matrix L written down, we can apply the reaction order formula in Step 4 of Figure 1. There Λ_t denotes a diagonal matrix with t on the diagonal, and similarly for Λ_x .

This formula can calculate the reaction orders for any given concentration vector of the species x . This allows us to sample points in the reaction order polyhedra for any species of interest. Namely, we sample some points x , compute the totals by $x = Lt$, then plug into the formula to obtain the reaction orders.

Implement the algorithm, or use the google colab notebook with link at the end. Sample the reaction order polyhedron in the enzyme binding with multiple substrate network. Focus on reaction order of E_{n_s} to total substrate and total enzyme. Suggested parameters: sample x_i between 10^6 and 10^6 , with 10^4 points or more.

What shape do you observe? Note that for $n_s = 1$, we should recover the triangle from single binding network $E + S \rightleftharpoons C$. What is the difference when n_s is increased, say to 3?

Solution D: The implementation can be found in the jupyter notebook at the end of this problem. You should find a triangle with vertices $(0, 1)$, $(1, 0)$ and $(n_s, 1)$, with coordinates as reaction order of E_{n_s} in t_S and t_E .

Problem E: Alternative binding process.

Our previous formulation of the multiple substrate binding assumed everything is formed in one step. What happens if we assume this instead happens by the substrates first form a multi-mer, then bind with the enzyme? This is the case when the substrate corresponds to multi-subunit proteins that work when in dimer, trimer or quadrimer form.

The binding network then is $n_s S \rightleftharpoons S_{n_s}$, and $E_0 + S_{n_s} \rightleftharpoons E_{n_s}$. Write down the stoichiometry matrix with species order (E_0, S, S_{n_s}, E_{n_s}). What are the atomic species? Also write down the conservation law matrix. Then sample the polyhedra. Compare with the previous case where the substrates are bound directly from monomers. What is the difference?

A jupyter notebook implementing the code for sampling already written for you to play around. You can click the "google colab" button to directly run the notebook on the cloud.

https://github.com/chemaoxfz/bi23/blob/main/202205_UCSD_Jun_course_homework_reaction_order_sampling.ipynb

Solution E: See the jupyter notebook at the following link. https://github.com/chemaoxfz/bi23/blob/main/202205_UCSD_Jun_course_homework_reaction_order_sampling_solution.ipynb

References

- [1] Stephanie K. Aoki, Gabriele Lillacci, Ankit Gupta, Armin Baumschlager, David Schweingruber, and Mustafa Khammash. A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature*, 570(7762):533–537, Jun 2019.
- [2] Elisabetta de Alteriis, Fabrizio Cartenì, Palma Parascandola, Jacinta Serpa, and Stefano Mazzoleni. Revisiting the crabtree/warburg effect in a dynamic perspective: a fitness advantage against sugar-induced cell death. *Cell Cycle*, 17(6):688–701, 2018. PMID: 29509056.
- [3] Noah Olsman, Fangzhou Xiao, and John C. Doyle. Architectural principles for characterizing the performance of antithetic integral feedback networks. *iScience*, 14:277–291, 2019.
- [4] Anjan Roy, Dotan Goberman, and Rami Pugatch. A unifying autocatalytic network-based framework for bacterial growth laws. *Proceedings of the National Academy of Sciences*, 118(33):e2107829118, 2021.